Watermarking services for medical database content security

Javier Franco Contreras

THÈSE / Télécom Bretagne
sous le sceau de l'Université européenne de Bretagne

pour obtenir le grade de Docteur de Télécom Bretagne
En accréditation conjointe avec l'Ecole doctorale Matisse
Mention : Traitement du signal et télécommunications

présentée par
Javier Franco Contreras

préparée dans le département Image et Traitement de l'Information
Laboratoire Latim

Watermarking services for medical database content security

Thèse soutenue le 9 décembre 2014
Devant le jury composé de :

Refik Molva
Professeur, Eurecom / président

Patrick Bas
Chargé de Recherche, École Centrale de Lille / rapporteur

Jacques Demongeot
Professeur, Université Joseph Fourier – Grenoble 1 / rapporteur

Stefan Darmoni
Professeur, CHU de Rouen / Examinateur

Gouenou Coatrieux
Professeur, Télécom Bretagne / Examinateur

Nora Cuppens
Directrice de Recherche, Télécom Bretagne / Examinateur

Christian Roux
Professeur, Mines Saint-Etienne / Co-Directeur de thèse

Fréderic Cuppens
Professeur, Télécom Bretagne / Directeur de thèse

N° d'ordre : 2014telb0347

Sous le sceau de l'Université européenne de Bretagne

Télécom Bretagne

En accréditation conjointe avec l'École Doctorale Matisse

École Doctorale – MATISSE

Watermarking services for medical database content security

Thèse de Doctorat

Mention : « Traitement du signal et télécommunications »

Présentée par **Javier Franco Contreras**

Département : ITI

Laboratoire : LaTIM (Laboratoire de Traitement de l'Information Médicale - U1101 Inserm)

Directeur de thèse : Frédéric Cuppens
Co-directeur de thèse : Christian Roux
Encadrant de thèse : Gouenou Coatrieux
Encadrante de thèse : Nora Cuppens-Boulahia

Soutenue le 9 Décembre 2014

Jury :

M. Refik Molva, Professeur, EURECOM (Président)
M. Patrick Bas, Chargé de Recherche CNRS, Ecole Centrale de Lille (Rapporteur)
M. Jacques Demongeot, Professeur, Université Joseph Fourier - Grenoble 1 (Rapporteur)
M. Stefan Darmoni, Professeur, CHU de Rouen (Examinateur)
M. Gouenou Coatrieux, Professeur, TELECOM Bretagne (Examinateur)
Mme. Nora Cuppens-Boulahia, Directeur de Recherche, TELECOM Bretagne (Examinateur)
M. Frédéric Cuppens, Professeur, TELECOM Bretagne (Directeur de thèse)
M. Christian Roux, Professeur. Directeur-adjoint chargé de la recherche et de l'innovation, Mines Saint-Etienne (Co-directeur de thèse)

A Eduardo y María del Carmen, mis padres,
que lo han dado todo para que yo llegara hasta aquí.
A Amparo, por su amor y su paciencia. Ahora empieza otra vida.

ACKNOWLEDGEMENTS

Without any doubt, this is one of the most difficult sections to write. A lot of people deserve to be mentioned in this limited space as their help and guidance has allowed me to get to this point. I really hope that the "forgotten" ones will not be annoyed.

First and foremost, I would like to thank my advisors, who trusted in me from the begging, when i was just an undergrad Erasmus student. Thanks for your help, your guidance and your patience. I would like to acknowledge Professor Gouenou Coatrieux for the time he spent helping, correcting and guiding me during these three years of hard work and learning.

I would like to thank Professor Patrick Bas and Professor Jacques Demongeot for accepting to evaluate my thesis. The comments and suggestions they provided helped me to improve the quality of my dissertation. I also thank Professor Stefan Darmoni and Professor Refik Molva for their participation in the evaluation jury.

Thanks to my colleagues in the ITI department of TELECOM Bretagne and in the LaTIM laboratory, who helped me to advance in my research and to improve the quality of my work. Thanks to Corinne Le Lann and Fabienne Guyader for all their help with the administrative procedures during these years.

I would like to thank my "French family", who prevented me from "going nuts" during these three years. A non-exhaustive list includes Isabel, Cristina, Lara, Mario, Angela, Daniela, Gabo, Dani y Daniel, Ronald, Juan David, Santiago, Luiz y Patricia, Andres, July, Sandra, Andrea, Oscar, Olivier E., Olivier B., Stephan, Pierre and an infinite *etcetera*. Also, a big thank you to my Spanish friends and to the link between these two worlds, Gonzalo (Pita), who has been there in the best and the worst since 1988. And to my parents, whose support made all this journey much easier.

Finally, thank you Amparo. For your patience and your love.

ABSTRACT

Watermarking services for medical database content security

Javier Franco Contreras

Department ITI, TELECOM - Bretagne

9 Décembre 2014

With the evolution of information and communication technologies, data gathering and management into data warehouses or simple databases represent today economical and strategic concerns for both enterprises and public institutions. Remote access and storage-resources pools are now a reality allowing different entities to cooperate and reduce expenses. In that context, data leaks, robbery as well as innocent or even hostile data degradation represent a real danger needing new protection solutions, more effective than the existing ones.

The work conducted during this Ph.D. thesis aims at the integration of watermarking in the family of existing database protection tools, considering the healthcare framework as case of study. Watermarking is based in the imperceptible embedding of a message or watermark into a document, watermark which allows us, for instance, to determine its origin as well as identifying its last user or verifying its integrity. A major advantage of watermarking in relation to other mechanisms is the fact that it enables to access the data while keeping them protected. Nevertheless, it is necessary to ensure that introduced distortion do not perturb the interpretation of the information contained in the database.

A first part of this work has focused on reversible and robust watermarking. The reversibility property ensures the recovery of the original data once the embedded sequence has been extracted. As defined, it allows us to relieve the constraints of watermark imperceptibility as well as to perform an update of the mark; watermark update which may be of interest in certain domains such as health or defense when tracing data. The results of this work stand in two lossless watermarking schemes. The first one is fragile and can be used for integrity control of databases. The second is robust, a property which makes it useful for traitor tracing and user identification purposes. Beyond the application framework, we theoretically evaluate the performance of our schemes in terms of capacity and robustness against two common database modifications: addition and removal of tuples. These theoretical results have been experimentally validated further demonstrating the suitability of our schemes for various security application frameworks (integrity control, traitor tracing ...).

In the continuity of this work, we studied traceability concerns of databases stored and merged into, for example, shared data warehouses. We have theoretically analyzed the problem so as to propose an optimized watermark detection strategy. This one allows us to identify a database representing a small piece of the total databases mixture (nearly 7%) with a detection rate close to 100%.

A second part of these research activities has dealt with the minimization of the semantic distortion produced by the watermark embedding process, even in the lossless case. Indeed, even in this latter case, there is an interest in minimizing the introduced distortion in order to keep the watermark into the database, ensuring thus a continuous protection. The solution we propose is based on the exploitation of ontologies the purpose of which is to allow the identification of existing semantic relations in-between attributes values of a tuple, relations that have indeed to be preserved. By doing so, our approach avoids the appearance of incoherent and unlikely values which could perturb

the interpretation of the database as well as indicate the presence of a watermark to a potential attacker. In this work, we demonstrate the convenience of this solution by means of an extension of the well known Quantization Index Modulation (QIM) to database watermarking. Again, theoretical performance of this scheme in terms of robustness has been established.

At last and in order to validate some of our hypothesis in relation with the distortion introduced by the watermarking process, we have developed and deployed an evaluation protocol in collaboration with the Biomedical Informatics Service of the Teaching Hospital (CHU) of Rouen. The objectives of this study were two: determine the distortion level considered as acceptable by health professionals and to clearly show the interest of a semantic distortion control.

Keywords— Health information, Databases, Security, Lossless watermarking, Semantic Distortion Control, Ontology.

RÉSUMÉ

Watermarking services for medical database content security

Javier Franco Contreras
Département ITI, TELECOM - Bretagne
9 Décembre 2014

Avec l'évolution des technologies de l'information et des communications, la collecte et la gestion de données au sein d'entrepôts voire tout simplement de bases de données sont aujourd'hui l'enjeu d'intérêts économiques et stratégiques tant pour les entreprises que pour les institutions publiques. L'accès à distance et la mutualisation des ressources de stockage de données sont une réalité, permettant à différentes entités de collaborer et réduire les coûts. Dans un tel contexte, les fuites, le vol ou encore la dégradation volontaire ou non de l'information représentent un danger réel qui nécessite des méthodes de protection nouvelles et plus appropriées que l'offre actuelle, par ailleurs limitée.

Les travaux réalisés dans cette thèse de doctorat ont pour objectif d'intégrer le tatouage ou "*watermarking*" à la famille d'outils existantes pour la protection de bases de données, avec pour cas d'étude le domaine de la santé. Dans son principe, le tatouage consiste à insérer un message ou une marque de façon imperceptible dans un document hôte de manière à pouvoir, par exemple, déterminer son origine, connaître le dernier utilisateur à y avoir accédé ou vérifier son intégrité. Un avantage majeur du tatouage par rapport aux autres outils est qu'il laisse l'accès à l'information tout en la gardant protégée. Cependant, il faut s'assurer que la distorsion introduite par le tatouage ne perturbe pas l'interprétation de l'information dans la base.

Une première partie de ces travaux a porté sur le tatouage réversible et robuste de bases de données. La propriété de réversibilité garantit la récupération des données originales après avoir retiré la marque insérée. Cette propriété permet de relâcher les contraintes d'imperceptibilité et aussi d'autoriser la mise à jour de la marque. Nous avons développé deux schémas de tatouage réversible : un schéma fragile, qui peut servir au contrôle d'intégrité, et un autre robuste, utile dans le cadre de la traçabilité et l'identification d'utilisateurs finaux. Au delà de ce contexte applicatif, nous avons évalué théoriquement la performance de ces schémas en termes de capacité et de robustesse face aux attaques classiques d'insertion et suppression de tuples. Ces résultats théoriques, validés expérimentalement, montrent l'adéquation de nos schémas aux applications envisagées.

Dans la continuité des ces travaux, nous nous sommes intéressés à la traçabilité de bases stockées et mélangées au sein d'entrepôts partagés, par exemple. Nous avons analysé théoriquement cette problématique pour ensuite proposer une méthode de détection optimisée qui permet d'identifier une base représentant une faible proportion des enregistrements du mélange de bases (environ 7%) avec une taux de détection proche de 100%.

Une deuxième partie de ces activités de recherche a focalisé sur la minimisation de la distorsion liée à l'insertion de la marque, y compris dans le cas réversible, où il y a en effet un intérêt à minimiser la distorsion afin de pouvoir garder la marque dans la base et assurer ainsi un protection continue. La solution proposée s'appuie sur une modélisation ontologique de la sémantique de la base qui permet d'identifier les relations clé à préserver entre les valeurs des attributs de la base. Ainsi guidée, notre approche évite l'apparition de valeurs incohérentes ou improbables qui pourraient perturber l'interprétation de la base comme aussi signaler la présence d'une marque à un

attaquant. Nous avons démontré l'intérêt de cette solution sur la base de l'extension de la *Quantization Index Modulation* (QIM) au tatouage de bases de données relationnelles. Ici aussi, la performance théorique de ce schéma en termes de robustesse a été établie.

Enfin, dans le but de valider certaines de nos hypothèses quant à la distorsion introduite par le marquage, nous avons mis en place un protocole de validation en collaboration avec le Service d'informatique Biomédicale du CHU de Rouen. Les objectifs de cette étude sont doubles : déterminer le niveau de distorsion considérés comme acceptable par les professionnels de la santé et montrer clairement l'intérêt d'un contrôle sémantique de la distorsion.

Mots clés— Information médicale, Bases de données, Sécurité, Tatouage réversible, Contrôle Sémantique de la distorsion, Ontologie.

Contents

List of Figures

List of Tables

Résumé en français

Avec l'évolution des technologies du multimédia et des communications, la collecte et la gestion de données au sein d'entrepôts ou simplement de bases de données sont aujourd'hui l'enjeu d'intérêts économiques et stratégiques tant pour les entreprises que pour les institutions publiques. L'intérêt pour le développement et la mise à disposition d'outils de « data-mining », c'est-à-dire d'outils d'exploration de données, sont des exemples qui soulignent la valeur grandissante de ces bases qui participent de plus en plus à la prise de décision. Dans un tel contexte, les fuites, le vol voire encore la dégradation volontaire ou non de l'information représentent un danger réel qui nécessite des nouvelles méthodes de protection, plus appropriées que l'offre actuelle, par ailleurs limitée.

Les établissements de santé n'échappent pas à cette problématique. Dans son fonctionnement quotidien, un hôpital génère une quantité énorme de données. Il s'agit tout aussi bien d'informations utiles à la prise en charge du patient (diagnostic, thérapeutiques, ...) qu'à l'évaluation de l'activité de l'établissement (e.g., évaluation médico-économique, traçabilité du médicament, ...) ou encore à la surveillance sanitaire. Plusieurs modèles pour les transferts d'informations, aujourd'hui standardisés comme DICOM pour l'imagerie radiologique ou HL7 pour les échanges entre les systèmes d'information hospitaliers, ont permis et servent encore à améliorer la qualité des soins et à rendre plus simple le partage d'information entre professionnels de santé. Aujourd'hui, nous en sommes à la constitution d'entrepôts de données de santé dont un intérêt est une meilleure compréhension des maladies.

Ce passage au numérique implique un accroissement des risques. Les besoins de sécurité à satisfaire sont cependant liés à la spécificité de ces données, au fait qu'elles touchent la santé d'un ou plusieurs individus, et au cadre législatif et déontologique que le citoyen à instauré. Par exemple, aux états Unis, la loi HIPAA (*Health Insurance Portability and Accountability Act*) établit les règles liées au respect à la vie privée et la sécurité des données médicales au format électronique. En France, l'Agence des Systèmes d'Information Partagés de santé (ASIP) affirme que la sécurisation des données de santé est une condition indispensable au développement du Dossier Médical Personnel (DMP) et de la télémédecine. Ainsi, il faut assurer l'intégrité et l'authenticité des données, garantir leur confidentialité et éviter les vols ou les fuites d'information comme aussi assurer leur disponibilité dans les conditions d'accès normalement prévues. On comprend pourquoi ... c'est la vie et l'honneur du patient qui sont en jeu ! Si l'on se réfère aux standards pour le déploiement de politiques de sécurité, comme ISO/CEI 27001 qui verse sur la sécurité de l'information et la protection des données sensibles, avec l'ISO 27799 dédiée aux données de santé en particulier, les risques pour l'information médicale peuvent être classés en trois catégories :

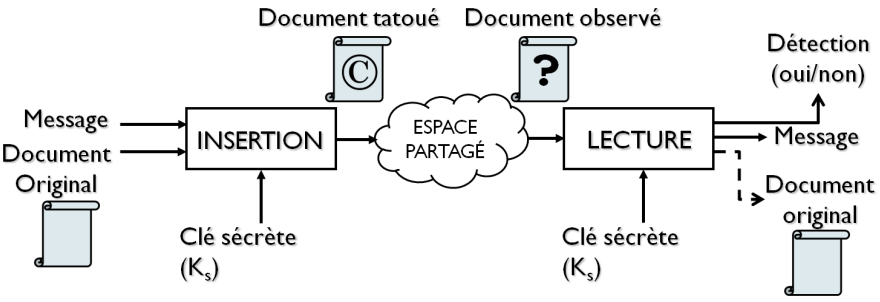


Figure 1: Étapes principales d'une chaine de tatouage classique. Le contenu tatoué est partagé (e.g. via l'Internet) et il peut être manipulé entre l'insertion et la lecture. À la lecture, dans le cas du tatouage réversible (*lossless*), le document original peut être complètement récupéré

- les accidents (e.g., les pannes matérielles, les phénomènes naturels, les négligences, etc.).
- les erreurs (lors de la saisie d'information, des transferts, des mauvaises utilisations, etc.).
- les malveillances ou autres malversations (fraudes, détournements d'information, chantage, intrusion dans les systèmes, etc.).

Pour répondre à ces menaces, différentes mesures de protection existent. Nous pouvons entre autres citer : le contrôle d'accès, la gestion des droits d'utilisation de l'information, l'enregistrement des accès, le chiffrement des données ou les signatures cryptographiques. Cependant, ces solutions de protection sont limitées et plutôt *a priori*, car une fois outrepassées les données ne sont plus protégées. C'est en particulier le cas des données déchiffrées. C'est ici que le tatouage s'impose, car il maintient une protection alors que les données sont manipulées. C'est une protection complémentaire pour l'information.

Par définition, le tatouage consiste à insérer un message dans un document multimédia hôte qui peut être une image, un signal audio voir, pour ce qui nous concerne, une base de données (voir figure 1). L'objectif de cette insertion peut varier selon le contexte. Elle dépend du lien entre le message et son hôte. Le message peut servir : à la protection des droits d'auteur, le contrôle d'intégrité, la traçabilité des données, l'ajout de méta-données, etc. Cette versatilité fait du tatouage une solution très intéressante dans le cadre de la protection des données de santé Coatrieux et al. [2006]. Cependant, les informations personnelles des patients étant particulièrement sensibles, des fortes contraintes existent qui font que la distorsion introduite par le tatouage doit être parfaitement maîtrisé car il existe un risque d'interférence avec les finalités de la base. Dans une telle situation, l'intérêt est porté par des méthodes à distorsion contrôlée, des méthodes de tatouage dites « libres de distorsion » (ou «*distortion free*») ou des méthodes réversibles.

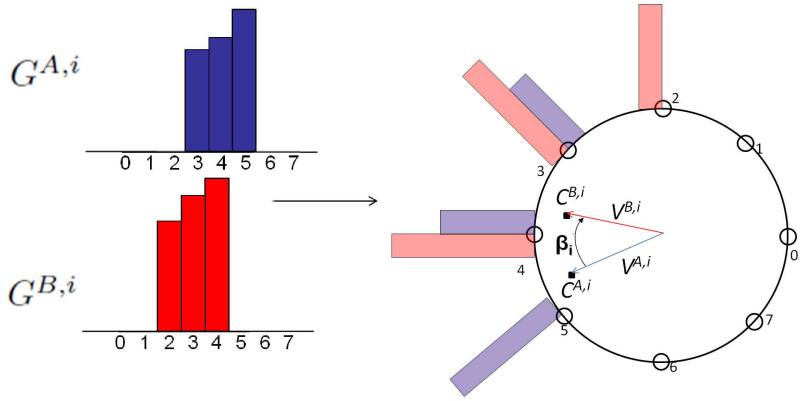
La protection par tatouage des contenus multimédia est largement étudiée et de nombreuses publications portent à ce sujet [Cox et al., 2008]. Par contre, peu de travaux se sont intéressés au cas des bases données relationnelles. En fait, moins d'une centaine d'articles ont été publiés sur ce sujet depuis la première contribution faite par Agrawal et Kiernan en 2002 Agrawal and Kiernan [2002] et les travaux sont très parcellaires dans le domaine du médical. C'est ici que se trouve le point de départ de mes travaux de thèse. Ils visent la mise en œuvre de méthodes de

tatouage adaptées aux contraintes de distorsion des bases de données avec comme cas d'étude les bases de données médicales.

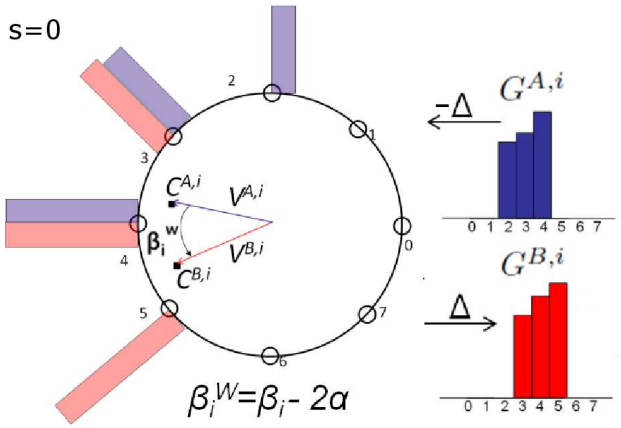
Dans un premier temps, nous nous sommes focalisés sur le tatouage réversible, qui permet d'enlever la marque et de récupérer la base de données originale. Cette caractéristique est intéressante dans le domaine médical, où les médecins et d'autres professionnels demandent parfois de pouvoir accéder aux données non-tatouées. La réversibilité permet aussi de mettre à jour la marque en cas de besoin sans avoir à introduire une distorsion additionnelle dans la base. Cependant, le nombre de travaux visant le tatouage réversible de bases de données relationnelles est assez faible, avec moins d'une dizaine de méthodes qui sont le résultat de techniques de tatouage réversible pour les images qui ont été adaptées. C'est la raison pour laquelle la plupart de celles-ci modifient des attributs numériques plutôt que des attributs catégoriels, avec l'exception de la méthode proposée par Coatrieux *et al.* [Coatrieux et al., 2011]. La majorité de schémas sont fragiles, avec deux exceptions remarquables : Gupta et Pieprzyk [Gupta and Pieprzyk, 2009] proposent une solution où un motif aléatoire qui identifie l'utilisateur est inséré dans des attributs numériques de tuples secrètement sélectionnés. Pour ce faire, un des bits de poids faible de la partie enti ere de la valeur d'un attribut est secr etement choisi et substituée par un bit généré de façon pseudo-aléatoire. inverser ce procédé d'insertion, la valeur originale du bit substitué est insérée dans la partie décimale de l'attribut, partie décimale dont la représentation binaire a été préalablement décalée à droite pour introduire un bit virtuel utilisé pour l'insertion. C'est la détection de la présence du motif aléatoire (i.e., la marque) qui permet au détecteur d'indiquer si la base a été tatouée ou non. Dans le but de réduire la distorsion introduite, Farfoura *et al.* [Farfoura et al., 2013] suggèrent de tatouer la partie fractionnelle d'un attribut numérique en utilisant la modulation par expansion de la prédiction d'erreur originellement proposée par Alattar Alattar [2004]. Bien que cette méthode soit dite robuste aux manipulations de la base (e.g., addition ou élimination de tuples), une simple opération d'arrondi peut effacer la marque. Plus généralement, la modulation par expansion de différences n'a pas été conçue pour être robuste a des attaques du document hôte. Une autre faiblesse de ces deux schémas est qu'ils dépendent de l'existence d'attributs numériques réels ou fractionnels.

Dans cette thèse, nous avons cherche à exploiter un domaine d'insertion plus pertinent que la modification directe des valeurs des attributs dans des tuples, dans le but d'obtenir une meilleure performance en termes de robustesse face aux attaques et de capacité. Pour ce faire, deux schémas ont été proposés qui étendent au tatouage de bases de données relationnelles la modulation robuste et réversible proposée à l'origine par De Vleeschouwer *et al.* De Vleeschouwer et al. [2003] pour les images. Pour insérer un message, les schémas proposés modulent la position angulaire relative du centre de masse de l'histogramme circulaire d'un attribut numérique. Plus concrètement, le processus d'insertion est le suivant. Comme dans la majorité de schémas de tatouage de bases de données, les tuples de la base sont premièrement organisés de manière sécrète en groupes par le biais d'une fonction de hachage cryptographique (e.g., SHA) qui prend en entre une clé secrète de tatouage K_S. Cette groupage à pour objectif de rendre la synchronisation entre l'insertion et la détection indépendante de la manière dans laquelle la base est stockée (e.g., de l'ordre des tuples dans une relation). Afin d'insérer un symbole dans un groupe G^i, celui-ci est divisé en deux sous-groupes de taille similaire ($G^{A,i}$ et $G^{B,i}$). Pour un attribut numérique à tatouer, l'histogramme dans un sous-groupe est calculé et projeté sur un cercle (voir figure 2(a)). Le centre de masses associé à l'histogramme circulaire

peut alors être obtenu ($C^{A,i}$ et $C^{B,i}$). L'information sur le symbole inséré dans le groupe va être portée par le signe de l'angle β_i entre les vecteurs centre de masses de chaque sous-groupe ($V^{A,i}$ et $V^{B,i}$), qui sera modifié par rotation de ces vecteurs (voir figure 2(b)).



(a)



(b)

Figure 2: a) Projection des histogrammes de chaque sous-groupe $G^{A,i}$ et $G^{B,i}$ sur un cercle. L'angle entre les vecteurs centre de masses est modulé afin d'insérer un symbole du message. b) Insertion d'un symbole s=0, correspondant à la rotation des histogrammes circulaires de $G^{A,i}$ et $G^{B,i}$ dans des directions opposées avec une modification angulaire α pour modifier le signe de β_i. C'est équivalent à l'addition de Δ aux valeurs de l'attribut dans $G^{B,i}$ et $-\Delta$ à celles de $G^{A,i}$.

Dans le premier schéma, un message binaire est inséré de manière fragile. Cette fragilité est due au fait que les attaques sur la base peuvent produire une perte de la synchronisation entre l'insertion et la détection. Il peut être utilisé, par exemple, pour le contrôle de l'intégrité de la base. Afin d'éviter ces problèmes liés à la synchronisation et rendre notre schéma robuste, on considère dans un deuxième temps l'insertion d'une séquence de symboles $s_i \in \{-1, 1\}$ de longueur fixe qui est détectée par corrélation. Ce deuxième schéma présente des applications dans la traçabilité et le contrôle de l'authenticité. Au-delà du contexte applicatif, les performances théoriques de ces schémas en termes de capacité et de robustesse face à

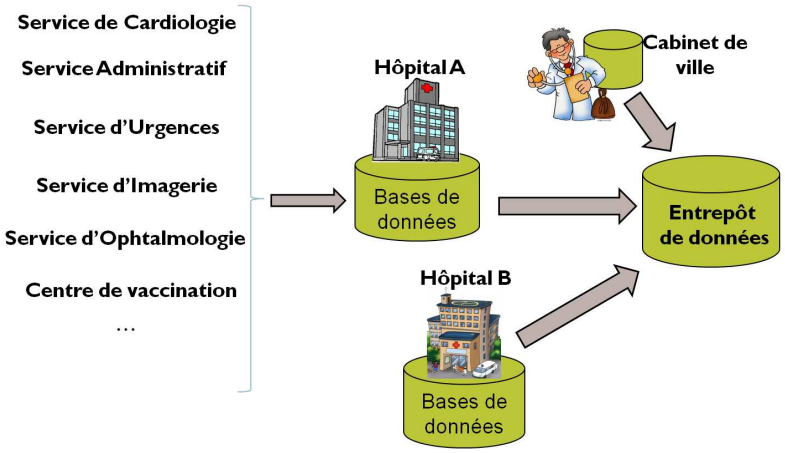


Figure 3: Exemple du partage de données au sein d'un hôpital (entrepôt de données interne) et entre différentes établissements de santé (entrepôt de données collaboratif), où différentes bases de données sont stockées dans un entrepôt de données partagé.

la suppression et l'insertion de tuples ont été aussi calculées à l'aide de résultats issus des statistiques circulaires, qui traitent des variables aléatoires représentant des angles ou des vecteurs. Ces calculs théoriques ont été verifiés de manière expérimentale.

Sur la base des résultats théoriques obtenus, nous nous sommes intéresés à l'optimisation du schéma robuste et réversible proposé pour la traçabilité de bases de données mélangées dans des entrepôts. Dans un tel scénario, l'objectif est de pouvoir détecter l'identifiant d'une base dont les tuples ont été agrégés avec ceux des autres bases (voir figure 3). Bien que l'utilisation de codes anti-collusion [Tardos, 2003] ait été écarte à cause des différences existantes entre le traçage de traitre et le mélange de bases, certains résultats ont été retenus et ont permis, conjointement avec la modélisation théorique du mélange, de proposer deux schémas de détection : le premier est basé sur le décodage souple, qui utilise des valeurs réels des angles au lieu de symboles; le deuxième est un décodeur informé qui exploit la connaissance que l'on a sur le impact du mélange de tuples dans la détection. Les deux schémas permettent d'obtenir de bonnes performances de détection pour une faible distorsion de la base, avec notamment une taux de détection de 100% pour une base ne représentant que 7% du mélange.

Dans un deuxième temps, notre travail a été focalisé sur la minimisation de la distorsion liée au marquage, y compris dans le cas réversible, où il y a en effet un intérêt à minimiser la distorsion afin de pouvoir garder la marque dans la base, assurant une protection continue, ainsi que d'éviter de faciliter l'accès aux paramètres de tatouage à un attaquant. Les solutions de contrôle de la distorsion proposées jusqu'à présent cherchent à préserver les statistiques des attributs, e.g., des contraintes définies sur la moyenne, l'écart-type ou la corrélation entre attributs [Kamran et al., 2013a] [Shehab et al., 2008]. Cependant, ces solutions ne tiennent pas compte de toutes les relations sémantiques qui existent dans la base et qui doivent aussi être préservées. La sémantique fait référence au sens d'une information. Par exemple, on peut considérer une base de données médicales qui contient deux attributs « genre » et « diagnostic ». Il existe une forte relation sémantique entre la valeur « femelle » de l'attribut « genre » et la valeur « grossesse » de l'attribut « diagnostic ». Il serait incohérent de trouver « genre » = « mâle ». Bien que les statistiques puissent fournir de pistes sur l'existence de ces liens

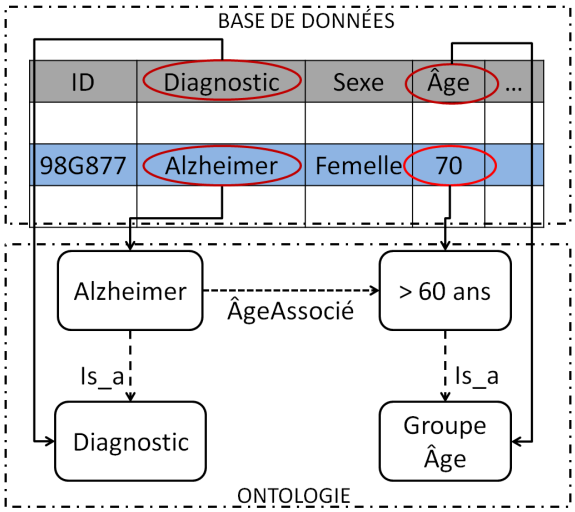


Figure 4: Connexions entre une base de donnés relationnelle et une ontologie. Les flèches en pointillés et en tirets représentent des relations ontologiques entre concepts dans l'ontologie. Les flèches en continue représentent des connexions entre des attributs ou des valeurs d'attributs et des concepts ontologiques.

sémantiques, évaluant les dépendances ou les co-occurrences des valeurs dans la base, elles ne permettent pas d'identifier directement une situation d'incohérence comme celle présentée. En effet, les tuples incohérents ou improbables peuvent être statistiquement insignifiants mais facilement détectables de manière sémantique.

Afin d'éviter l'apparition de ces enregistrements « problématiques », nous proposons une identification des contraintes de distorsion par le biais d'une ontologie modélisant la sémantique du contenu de la base de données. Comme il a été exposé par Gomez-Perez et Benjamins [Gomez-Perez and Benjamins, 1999], les ontologies fournissent un vocabulaire commun dans un domaine et définissent, avec différents niveaux de formalité, le sens des termes et des relations entre eux. Une ontologie est composée de concepts qui peuvent être instanciés sous la forme d'individus. Ces individus peuvent présenter des attributs qui servent à les décrire. Les concepts et individus dans l'ontologie sont sémantiquement liés par des relations. Bien que les ontologies aient été considérées pour des différentes applications, de l'extraction de connaissances à l'annotation et la recherche d'images, elles n'ont jamais été appliquées au contrôle de la distorsion produite par le tatouage. Pour ce faire, chaque concept de l'ontologie est associé à une valeur ou sous-ensemble de valeurs dans le domaine d'un attribut. Prenons par exemple le concept « Alzheimer » qui est associé à la valeur « Alzheimer » de l'attribut « diagnostic principal ». De la même façon, un concept « Plus de 60 ans » est associé à une plage de valeurs de l'attribut « âge ». Ces deux concepts sont liés par une relation « âge associée » qui modélise le lien sémantique qui existe entre les attributs. Dans un tel contexte, une requête à l'ontologie permet d'extraire les valeurs autorisées pour l'attribut « âge » dans le tuple, étant dans cet exemple celles supérieures à 60 (voir figure 4).

Les schémas réversibles présentés ne peuvent pas bien gérer ce type de contraintes de distorsion car la modification qu'ils introduisent sur chaque valeur de l'attribut à tatouer est de valeur fixe. C'est la raison pour laquelle, dans le but de montrer l'intérêt du contrôle

sémantique de la distorsion proposé, nous avons développé un schéma de tatouage robuste qui offre une dégrée de liberté dans la modification individuelle introduit dans chaque valeur d'attribut. En s'appuyant sur le domaine d'insertion utilisé pour nos schémas réversibles, nous avons proposé une extension de la modulation par quantification d'index (QIM) [Chen and Wornell, 1999] au tatouage de bases de données, où elle n'a jamais été considérée. Dans ce cas, le cercle est divisé en deux cellules qui correspondent aux valeurs possibles des symboles à insérer. Chaque cellule présente un centroïde défini à partir d'un paramètre qui permet de contrôler le niveau de distorsion (à différence du cas de la QIM traditionnelle où les centroïdes se trouvent au centre des cellules). Afin de coder un symbole, le centre de masses d'un groupe de tuples est quantifié au centroïde correspondant. La modification des valeurs de l'attribut à tatouer est faite de manière itérative, ce qui permet de considérer les contraintes de distorsion. Cette fois-ci aussi, les performances du schéma ont été évaluées de manière théorique et testées expérimentalement, montrant une robustesse importante, ainsi qu'une faible distorsion et une faible complexité tout en respectant les contraintes extraites grâce à l'ontologie.

Finalement, avec l'objectif de valider nos hypothèses en termes de distorsion, un protocole d'évaluation expérimentale a été développé en collaboration avec l'équipe CisMEF, basée au CHU de Rouen. Dans ce protocole, un expert évalue en termes de cohérence et plausibilité une série d'ensembles d'enregistrements tatoués avec des paramètres différents. Pour ce faire, deux tests sont proposés : 1) un test dit aveugle dans lequel l'évaluateur doit identifier des ensembles de données tatoués, l'objectif duquel est de montrer qu'il est possible de tatouer une base de données et d'établir un lien entre la visibilité de la marque et la gêne et les paramètres de tatouage; 2) un test non-aveugle ou informé dans lequel l'évaluateur doit identifier des enregistrements incohérents ou improbables afin de mettre en évidence que ces tuples peuvent être facilement identifiés par un expert. Le protocole d'évaluation est indépendant de la méthode de tatouage et de la base de données à analyser. Bien que notre étude se trouve dans sa phase initiale, les résultats préliminaires obtenus avec un évaluateur confirment nos hypothèses, en montrant notamment la nécessité de considérer un contrôle sémantique de la distorsion comme celui que nous avons proposé, et nous permettent d'extraire des autres conclusions par rapport à la sensibilité de certains attributs au tatouage due à leurs statistiques et liens sémantiques avec des autres attributs dans la base.

Introduction

The evolution of multimedia technologies and communications has resulted in a remarkable increase in the construction, transfer and sharing of databases. As a consequence, data gathering and management into data warehouses or more simply in databases have become important economical and strategic concerns for enterprises and public administrations. The expansion of data-mining and assisted analysis tools are just two examples that highlight the growing value of these databases, which are now of extreme importance in decision making. In that context, information leaks, thefts or even degradations, intentional or not, represent a real menace, that requires new and more adapted protection methods which complete the existing options, mostly based in encryption and access control mechanisms. This has recently been proved by the Wikileaks case [Rogers, 2010], where an enormous amount of data issued from the US intelligence agencies has been exposed due to an internal leak.

Health institutions are not indifferent to these issues [Orcutt, 2014]. In its daily operation, a hospital generates a huge amount of data about patient care information (diagnosis, therapy, ...), resource management or for public health questioning. This information is collected in relational databases that allow efficient storage and access to records. These databases are access by different users, shared and transferred in between hospitals. This open environment implies an increase of security risks. At the same time, some particular security needs arise due to the specificity of medical information, subject to strict ethics (e.g., Hippocratic Oath) and legislation, such as the Health Insurance Portability and Accountability Act (HIPAA) [U.S Government, 1996]. In France, the agency for shared health information systems (ASIP) states that the protection of medical data is an essential condition to the development and the use of electronic health records (e.g., *Dossier Médical Personnel* in France) and telemedicine.

In this scenario, it is necessary to ensure the integrity and authenticity of data, to protect their confidentiality avoiding data leaks or thefts, as well as to warrant their availability in the normally defined access conditions. It is also important to trace data from their origin until the end of their life. One can easily understand the reasons of these needs: any violation of this security principles endanger the patient, having also consequences for the helth care establishment or professional. Moreover, security is the base of a trustful relationship between patients and the health care system (practitioners, administration, etc.). As example, being sure his or her say are confidentially stored, a patient will talk about his or her health without any shadows.

As exposed by the standards for the deployment of security policies such as the ISO 27799 standard [ISO/IEC, 2008] (i.e., implementation of ISO/IEC 27002 [ISO/IEC, 2013] for health information), the existing risks for health information can be classified in three categories:

- accidents (e.g., material failures, natural phenomena, negligences, etc.).
- errors (e.g., mistyping, transmission errors, etc.).
- attacks and misappropriations (e.g., frauds, rerouting, blackmailing, etc.).

In order to counter-fight these threats, several protection mechanisms have been proposed. A non-exhaustive list includes access control, user rights management and encryption which are helpful for confidentiality while digital signatures will ensure data integrity and non-repudiation. However, these security solutions offer an *a priori* protection or in other words, once they are bypassed or more simply when the access is granted, data are vulnerable. In this context, watermarking complementary provides an *a posteriori* protection of data. Data can be accessed and manipulated while still being protected.

By definition, watermarking lies on the insertion of a message (some security attributes) or a watermark into a host document (e.g., an image, an audio signal or in our case, a database) by slightly perturbing host data based on the principle of a controlled distortion. The aim of this insertion may vary depending on the applicative context and it depends on the relation between the protected content and the message. This message may serve the protection of the owner rights, data integrity, data traceability or even the insertion of additional meta-data. This versatility makes of watermarking a really attractive solution in the health data protection framework. While there is a vast knowledge in the field of multimedia watermarking, even in the medical imaging [Coatrieux et al., 2008] [Pan et al., 2010] [Bouslimi et al., 2012], the interest in database watermarking has been limited to date, with about a hundred publications that have appeared since the seminal method by Agrawal and Kiernan which dates of 2002 [Agrawal and Kiernan, 2002]. In particular, relational database watermarking differs from multimedia contents watermarking in several points from which the following two are worth highlighting: i) records in a database can be reorganized without changing the meaning of the database, in opposition to highly correlated neighbor samples in a signal or pixels in images; ii) the existence of specific manipulations that a database may undergo, like tuple suppression and insertion that modify the intrinsic structure of the database. At the same time, the high sensitivity of medical information results in strong constraints which require a strict control of the distortion introduced by the watermark.

We fixed here the starting point of this Ph.D. thesis work, which focuses on the preservation of the information quality considering as security objectives integrity and authenticity control and traceability.

In a first moment, we focused on lossless watermarking, which allows removing the watermark with the exact recovery of the original database. This is an interesting feature in the medical domain, in which practitioners and other professionals are highly concerned by the quality of data and often demand an access to the original unmodified data. Moreover, the reversibility property allows updating the embedded message at any time without introducing new distortion into the database. However, lossless or reversible database watermarking is quite uncommon, with only a few existing approaches, from which no more than ten are conceived to be robust against common database manipulations (i.e., tuple insertion and suppression as well as attribute's values modification). Among them we can cite the methods by Gupta and Piepzryk [Gupta and Pieprzyk, 2009] and Farfoura *et al.* [Farfoura et al., 2013]. In this thesis,

we looked for an insertion domain more appropriate that working directly on tuples, expecting a better performance in terms of robustness against attacks and capacity. We propose two different schemes, one fragile and one robust [Franco-Contreras et al., 2013] [Franco-Contreras et al., 2014b] which work on the circular histograms, which as we will see, offers interesting properties.

As a second main axis of this thesis, we focused on watermark imperceptibility. Although reversible schemes provide the ability to recover the original database, there is an interest in keeping introduced distortion as low as possible so as to ensure the correct interpretation of the information and to remain invisible to an attacker. The allowable distortion that can be introduced in the database by the watermarking process strongly depends on the final use of data and it is not obvious to create a "psycho-perceptive model" as in the case of image, video or audio watermarking. Some authors have proposed to fix some statistical constraints on the attributes' values in order to preserve the result of prior or subsequent data-mining operations [Shehab et al., 2008] [Kamran et al., 2013a]. However, in our view these solutions do not preserve all the database semantics, that is to say the links that exist between attributes' values records. To go further, we propose to take advantage of the knowledge one can have about the database contents. Among the different approaches that can model such a knowledge, ontologies occupy an important place. They are also well developed in the health care domain. The open question is then how to use such a medical knowledge in order to determine the allowable modification in the values of records in the database, preserving its semantic coherence [Franco Contreras et al., 2014].

To deploy such a semantic distortion control process and prove its benefits, lossless watermarking schemes are not the most appropriate. Indeed, they impose a constant distortion over the whole database. This is the reason why we focus on how to adapt Quantization Index Modulation (QIM) [Chen and Wornell, 1999], not yet explored in relational databases, with the purpose of achieving a higher degree of freedom in the modification of individual attributes values and consequently, a better control of the robustness-distortion trade-off [Franco-Contreras et al., 2014a]. As for the previous schemes, the approach we propose works on the circular statistics domain.

It should be noticed that whatever the proposed watermarking solution, another important aspect we take care about is the theoretical and experimental evaluation of its performance.

Beyond developing new adapted relational database watermarking methods, we focus on how to use them in the particular case of interest in the medical domain which is the traceability of databases shared in collaborative data warehouses. One must know that there exist a growing interest in the constitution of collaborative projects between national or international institutions with the aim of offering researchers an access to more important amounts of data. Shared data warehouses are then disposed so as to store data records coming from different establishments. In order to ensure data traceability, we are interested in identifying the origins of records at any moment.This can be achieved by the watermarking each database with the identifier of its owner. Herein, one may ask on how to take into account the way data are merged in order to optimize the detection of such an identifier, that is to say to find a solution which allows detecting the presence of the message in particular attack conditions. The application of anti-collusion codes, originally proposed for traitor tracing applications, sounds a good approach but, as we will see, the mixture of watermarked tuples presents some

important differences from the classical collusion case, making anti-collusion codes inefficient in that context. We will show however that some optimizations can be found in order to improve the detection capabilities of our robust lossless scheme in this scenario.

In order to expose our work, this thesis is structured as follows. Chapter 1 contains general definitions on the main domains that will be addressed in the sequel: medical information, relational databases and database watermarking, and exposes the security needs of medical information. The focus on actual security tools, both general and specific for relational databases, underlaying their weaknesses and highlighting the interest of watermarking as a complement to these solutions. A second part of this chapter introduces the main principles of watermarking and the properties of watermarking systems as well as their applications. In particular, we explain the main differences between multimedia and database watermarking showing how to overcome some of them. An exhaustive state of the art of existing database watermarking methods concludes this chapter so as to the originality of our work.

Chapter 2 is devoted to lossless or reversible database watermarking. Specific features and restrictions of lossless watermarking are studied and existing applications are reviewed before focusing on the methods proposed for the protection of relational databases. The weaknesses of such methods lead us to focus on a more appropriate insertion domain which allows us to spread the watermark over several records. Circular histograms, as introduced by De Vleeschouwer *et al*. [De Vleeschouwer et al., 2003] for images, have interesting properties from which we derive two lossless database watermarking schemes. The performance of these schemes are theoretically analyzed and experimentally verified, showing their suitability for different security objectives like integrity control (fragile scheme), data authenticity and traceability applications (robust scheme), respectively. As we will show, they outperform existing methods in the literature.

In Chapter 3 we study the applicative scenario where our robust lossless scheme, proposed in the previous chapter, is used for database traceability in a shared data warehouse. The impact of the merging of records from different watermarked databases onto the detection process of our robust lossless scheme is first modeled. Then, we expose how the detection process can be optimized according to the *a priori* knowledge one can have on the databases mixture. Compared to a basic correlation based detector, the proposed approach offers an important gain of performance.

Chapter 4 addresses the control of the semantic distortion introduced by watermark embedding. We show that by using, medical knowledge through ontologies it is possible to identify the allowable values for the watermarked attributes. In particular, ontologies allow representing the semantic links that exist between attributes values in records in order to avoid incoherent or very rare record occurrences which may bias data interpretation or betray the presence of the watermark. In a second time, we present our robust watermarking scheme adapted from the well-known QIM to databases which is suitable to be used together with the proposed semantic distortion control method. As in chapter 2, the performance of this scheme is theoretically evaluated and empirically tested. It shows high performance in terms of robustness, introduced distortion and complexity.

In order to analyze the perception that medical users will have of the watermarked databases and verify our hypothesis that incoherent records can be easily identified by an expert, we developed an evaluation protocol in cooperation with the CISMEF team based at the Biomedical

Informatics Service of the Rouen University Hospital that is the main subject of Chapter 5. In our protocol, an expert evaluates, in terms of coherence and plausibility, differently watermarked sets of records. Two different tests are proposed: 1) a blind test in which the evaluator has to identify watermarked datasets, the purpose of which is to show that it is possible to watermark a database modifying its records as well as to establish the relationship in between the visibility and the watermark hindrance with the watermarking scheme parameterization; 2) an informed test in which the evaluator has to identify any incoherent or unlikely records, which proves that these kind of records can be easily detected by an expert and highlights the interest of a semantic distortion control process as the one we propose in the previous chapter.

Chapter 1

Health Databases Security and Watermarking fundamentals

Health information covers a wide range of data and contents that are employed in patient care, clinical research and health care administration. Some examples include analysis results, clinical images or patient demographic data. Due to this heterogeneity, different data representation models have been considered depending on the applicative context so as to allow the efficient storage and access to information. In the case of structured patients' records, drug inventories or more generally structured data, the most extended approach is the relational model. These databases are present in different layers of the health care information system, from the electronic patient record to the hospital management.

At the same time, the last few years has seen a remarkable increase of sharing and remote access of medical relational databases. This is caused by the reinforcement of their economical value and decisional interest, partially due to the progress of data mining and analysis tools, and also by the advances in the telecommunication field. Nevertheless, this ease of access intensifies security risks, as data records may be redistributed or modified without permission. Even in this highly sensitive domain, where information leaks and unauthorized modifications may represent a serious threat to patient's health and privacy, several security breaches are reported every year [McNickle, 2012]. Today, confidentiality of data is usually achieved by means of cryptographic mechanisms, user authentication and access control policies. However, once these mechanisms are bypassed or more simply when the access is granted, data are no longer protected. Herein, watermarking has been proposed as a complementary security mechanism that provides *a posteriori* protection, i.e., data can be accessed while still being protected.

This chapter hinges on three main parts. The first part presents the basic notions of what health information is, its security needs and its actual threats. In a second part, we introduce the database relational model, the most extended data model for structured information, and present existing mechanisms used nowadays to answer security requisites. As stated, these mechanisms do not completely protect data from leaks and malevolent manipulations and watermarking mechanisms, which are the matter of the third part and the core of our research work, can lead to appropriate solutions. In this chapter, we also come back on watermarking fundamentals, presenting the most important and recent methods in database watermarking.

1.1 HEALTH INFORMATION

1.1.1 HEALTH INFORMATION DEFINITION

Several definitions of what health information is can be found. According to the Council of Europe, "*health information refers to all data concerning the health of an identified or identifiable individual*" [Council of Europe, 1997]. A similar definition appears in the US Code of Federal Regulations, which defines medical information as "*any information related to the past, present, or future physical or mental health or condition of an individual, the provision of health care to an individual or the past, present, or future payment for the provision of health care to an individual*" [U.S Government, 2013]. From these definitions, pieces of health information correspond to data involved in the different patient care processes in relation, as example, with the diagnosis, the therapy, and so on. These data can also serve prevention as well as for public health purposes and economical evaluation. One can also underline the nominative aspect of a piece of health information, that allows the direct or indirect identification of an individual.

Another aspect one can identify is the heterogeneous character of health contents, the diversity of their sources and their finalities. They represent the medical history of one patient, including laboratory results, medical images or video sequences, medical reports, etc. Such an information can be stored, processed or communicated on electronic supports or over the network.

In this framework, the data of one patient are regrouped within his or her Patient Electronic Record. This aggregation is however somewhat virtual as patient records are usually distributed over several databases at the hospital. As example, the images of a patient are in practice stored within the Hospital PACS (i.e., Picture Archiving and Communication System) [Choplin et al., 1992].

Clinical data repositories (CDR) are another example. They contain information issued from different sources, providing the practitioners a complete informative view about patients for one or several specific health purposes (e.g., cancer registries), including notably administrative and demographic data, diagnoses, medical procedures, laboratory results, drug administrations, etc. Although these databases were initially conceived for patient cares, nowadays they also serve hospital administration, medical research or public health purposes [Collen, 2011].

These databases are manipulated in an open environment, where they are accessed by several users, shared by different establishments and transferred. Moreover, in this framework data can be extracted from hospitals' information systems in order to construct higher level data repositories serving, for instance, pharmacovigilance purposes or allowing the economical evaluation of the hospitals activities by governmental agencies (e.g., collection of anonymous inpatient stays summaries for the PMSI-MCO database [ATIH, 2013]).

Notice that these data are numerical or textual data in most cases encoded by means of medical terminologies. As we will see, this is of importance in our case due to fact the basic principle of watermarking consists in modifying host content. Some important examples of such terminologies are the following:

- Diagnoses encoding, by means of the International Classification of Diseases (ICD), which is nowadays on its 10^{th} revision (ICD10) [World Health Organization, 1992]. It contains around 18000 codes (notice that this value can differ from one country to another).
- Medical procedures encoding: in France the common classification of medical procedures (CCAM in French) is used. CCAM contains 8000 codes describing different medical procedures. Other examples are GOÄ codes (Medical Fee Schedule in English) used in Germany or the International Classification of Procedures in Medicine (ICPM) proposed by the World Health Organization (WHO).
- Laboratory result encoding: based on LOINC (Logical Observation Identifiers Names and Codes) composed of around 58000 codes.
- Drugs description: by means of the Anatomical Therapeutic Chemical Classification System (ATC) with around 5500 codes.

In this Ph.D. work, we focus on the protection of medical relational databases considering the framework of medical repositories supply. Due to the sensitive nature and the nominative aspect of pieces of medical information and the fact they are shared in open environments, different security needs have to be considered.

1.1.2 SECURITY NEEDS IN HEALTH CARE

The security of medical information is regulated by strict deontological ethics (e.g., Hippocratic Oath) and national and international legislative rules. In the US, one must take care of the Health Insurance Portability and Accountability Act (HIPAA) [U.S Government, 1996]. The goal of the Privacy and the Security Rules issued from HIPAA is to *assure that individuals' health information is properly protected while allowing the flow of health information needed to provide and promote high quality health care and to protect the public's health and well being*. To do so, they establish a set of commitments in terms of information security that medical entities must ensure.

Security commitments imposed by HIPAA, and more generally by legislative rules of different countries, include ensuring: data confidentiality, data integrity and availability. These three main security objectives are also completed by authenticity control and traceability of information, usually considered in order to secure the complete flow of information.

Beyond these legislative and deontological rules, one can also find national and international recommendations which provide implementation guidance, such as the rules stated by the ISO 27799 standard [ISO/IEC, 2008] (i.e., implementation of ISO/IEC 27002 [ISO/IEC, 2013] for health information). These recommendations add the health specific commitments to the security obligations covered by the original ISO/IEC 27002, devoted to information security. Some other specific standards like those proposed by IHE (Integrating the Healthcare Enterprise) [IHE, 2013], can also enrich such a framework.

We introduce now these security engagements before explaining how the can be assured in Sect. 1.2.3.

Confidentiality and Privacy: Confidentiality basically refers to the prevention of information disclosure to unauthorized individuals or systems. It is specially relevant in the case of nominative information. An example of a confidentiality violation, in which the medical records of a woman were posted on line without her knowledge or consent a few days after she was at the hospital, is exposed in [Hillig and Mannies, 2001].

This property represents an extension of the concept of privacy, which involves the protection of sensitive information of individuals. An example illustrating the difference between both concepts is an epidemiological study in which the access to a set of patients' records is granted to an external institution. Thus, data are no longer confidential but the re-identification of patients should not be possible so as to protect their privacy.

Integrity: Integrity control is defined, in the context of health information security, as the protection of the accuracy and consistency of data, avoiding non-authorized alterations or deletions. Data integrity can be compromised both by accidental or malevolent data manipulations or erasures. The worst scenario resulting from a modification in the medical record of a patient would affect its clinical background, diagnosis or prescribed therapy.

Malevolent examples are quite rare. In known cases, it was intended to erase evidence of a medical negligence or to take profit. For instance, it has been reported that some hospitals in the UK manipulated their death rates in order to put up their own reputation [Donnelly, 2014]. In another case, physicians and technicians were suspected to have altered medical records of patients affected with radiation overdoses [Epinal, 2013].

Availability: Technically, medical information must be accessible in any situation when needed. Availability is critical in case of emergency. This means that equipments, software and communication channels needed to store and access this information must be fully operational taking into account the supported workload as well as the security tools to implement. The correct behavior of an information system may be perturbed by different risks such as accidents, errors or malevolent actions. To counteract these security risks, different security mechanisms can be implemented as we will see in Sect. 1.2.3.

Authenticity: Authenticity represents the fact that information proceeds from the source it is supposed to come from. In health care, this consists for instance in asserting the origin of a piece of information and its link to a given patient.

The combined protection of data authenticity and integrity ensures the reliability of information. Reliable data can be used by a health care professional in total trust. The concept of reliability can be extended to traceability when it becomes possible to trace the information along its distribution.

Traceability: Traceability can be defined as the capacity of identifying all the elements that have accessed, transfered, modified or deleted an information from its origin to its final use or in a given period of time. When defining their security policies, medical establishments take special attention to this property as it serves to establish the responsible parts in case of negligence due to incorrect information manipulation [Coatrieux et al., 2012].

1.2 Relational Databases and Security

In this thesis work, our attention has been focused on relational databases. The relational model is still the most extended format of structured information storage and the security of relational databases is an issue that generates a growing concern in public and private entities.

1.2.1 Fundamental notions of the relational model

History shows that the efficiency in the information access is directly linked to the choice of the correct data representation model. This is the reason why database models have evolved in order to offer a more adapted solution to users' needs in terms of storage, update and querying. The first proposed approach was the hierarchical model, created by IBM for its Information Management System (IMS) in the 1960s (see [Blackman, 1998]). It is based on a tree structure where records are organized following parent-children relations. Due to its restrictions in the representation of information (a child can only be linked to one parent) and the complexity of records' access in the tree, its use is limited to some very specific contexts. The network model was proposed by Bachman in 1969 [Bachman, 1969] as an evolution of the hierarchical model, allowing many-to-many relationships. This model was the most popular until the late 1970s, when the relational model, introduced by Edgar. F. Codd [Codd, 1970] imposed itself due to the ease of data manipulation and query and the security it offers. It can now be found in most of the existing commercial database systems. The relational model is based on the idea of relations (tables) as the mathematic representation of data sets.

Relations, domains and attributes

A relation is based on the concepts of domain and attribute. A **domain** D is an indivisible set of values, which can be defined by a data type (e.g., string of characters of 10 elements, decimal number between 0 and 10000, ...) and a name that helps to interpret values in the domain. Domain examples are the list of units in a hospital, the set of telephone numbers in a country, the set of eye colors, etc. We can define **attribute** as the name of the role played in a relation by one domain. It is important to notice that categorical attributes differ from numerical attributes in the absence of order relationships in between the values of their domain. For example, taking the attribute *eye color*, no rule states *a priori* that 'blue" is greater or smaller than "green". It is then difficult to apply any mathematical operation in this context. We can not say what will be the result of "blue" plus "brown".

The **scheme** of a $n-ary$ relation R is the finite set of attributes $\{A_1, A_2, ..., A_n\}$. As a consequence, a **relation** is a subset of the Cartesian product of sets $r(R) \subseteq D_{A_1} \times D_{A_2} \times ... \times D_{A_n}$, where each of the sets D_{A_i} corresponds to the domain of an attribute [Elmasri and Navathe, 1989]. A relation has also a name used to identify it (e.g., Patient). To sum up, a relation represents a particular state of the real world.

Tuple

A **tuple** is an ordered set of n values in $D_{A_1}, D_{A_2}, ..., D_{A_n}$, which associate these values to the attributes' names. Consequently, a $n-ary$ relation is an unordered set of tuples.

From here on, we will call $t_u.A_t$ the value of the u^{th} tuple for the attribute A_t, with $t_u.A_t \in D_{A_t}$.

Primary key

By definition, all the tuples in a relation must be distinct, i.e., they must present different values in at least one attribute. We can define a **key** in a relation as a subset SK of minimal cardinality for which:

$$t_u.SK \neq t_v.SK \ \ if \ \ u \neq v \tag{1.1}$$

In general, a relation presents more than one key and it is necessary to design one among them as the primary key. The **primary key** PK is an attribute or set of attributes that allow the unique identification of each tuple in a relation. It must respect the temporal invariance property, meaning that any insertion or deletion of tuples should not modify its representativity. This is linked to the relational integrity constraints we define below.

Foreign key

Given two relations R_p and R_q in a database, a **foreign key** to R_q in R_p, if it exists, is an attribute or set of attributes in R_p having the same domains as the primary key of R_q. In that case, we can say that the attributes in the foreign key of R_p reference the relation R_q.

Integrity constraints

In order to ensure the accuracy and consistency of data stored in a relational database, two rules are imposed. First of all, the **entity integrity rule** states that every relation must have a primary key and for each tuple, it must be unique and non null. This allows to easily distinguish each tuple in a relation. Second, the **referential integrity rule** states that if an attribute in one relation is declared as foreign key to another relation, its value must either be part of a candidate or primary key from the parent relation or null, as it should make reference to an existing tuple.

Notice that these integrity constraints are not related to the notion of data integrity we introduced in Sect. 1.1.2

Notice that this model can evolve depending on applicative constraints. They are specified independently from the database structure but they should be respected if any modification over the values in the database is performed. For example, a patient can not leave the hospital before his arrival, so we can establish that the date of discharge from the hospital should be always subsequent to the date of arrival. This kind of rule can not be modeled in the scheme of the database.

1.2.2 RELATIONAL DATABASE OPERATIONS

Database management systems (DBMS) provide different functionalities that allow the user to manipulate the database. We can distinguish two main categories of operations: i) Update operations, which modify the structure or the content of the database; ii) Query operations, which allow to extract a certain part of the information according to some fixed criteria. These operations will help us to define the attacks that a watermarking system should cope with.

Example: Illustration of Relational model mathematical concepts.

Patient		
Patient_ID	**Name**	**Age**
1025786	Jérôme Legoff	25
75865	Michel Duberry	62
222158	Dominique Garcia	33
51485	Didier Leslandes	89
154838	David Batty	62

Diagnosis	
Patiend_ID	**Main_Diagnosis**
1025786	T424
75865	T435
222158	R51
51485	O800
154838	N10

Figure 1.1: Instances for a relational database composed of two tables. Primary keys for each relation are underlined.

We give in figure 1.1 a concrete example of the mathematical concepts of the relational database model, where an instance of a database is given. Different attributes are represented as columns in the tables while tuples, which represent elements in the real world, correspond to lines. The database scheme is $sch(I) = \{$*Patient, Diagnosis*$\}$. Relation *Patient* has a relation scheme $sch(Patient) = \{$*Patient_ID, Name, Age*$\}$, where each tuple is characterized by the patient ID, name and age. Relation *Diagnosis* has a relation scheme $sch(Diagnosis) = \{$*Patient_ID, Main_Diagnosis*$\}$, that patient diagnoses during his stays at the hospital.

In the relation *Patient*, the attribute *Patient_ID* correspond to the primary key, allowing uniquely identifying each patient. In relation *Diagnosis*, two attributes have this role: *Patient_ID* and *Main_Diagnosis*. At the same time, the attribute *Patient_ID* is a foreign key which allows to relate patients to diagnostics.

1.2.2.1 UPDATE OPERATIONS

There are three basic update operations over the relations:

- the insertion of new tuples into one relation.
- the deletion (or suppression) of tuples from one relation.
- the modification of attributes values for some tuples in the relation.

Any operations applied to the database should respect the integrity constraints and this is the reason why the DBMS should not allow manipulations that do not take such constraints into account. As an example, let us consider the database given in figure 1.1, where $Patient\_ID$ serves as primary key. If we decide to insert a new tuple with values $\{$*200866,Jean Marc LePont,58*$\}$, the **insertion operation** will be accepted as it respects the constraints. On the other side, the insertion of a new tuple with values $\{$*222158,Jean Marc LePont,58*$\}$ will be rejected as another tuple exists with the same primary key.

In the tuple **deletion operation**, the DBMS should be vigilant to the referential integrity constraint. Let us consider the same example as above. If a user wants to delete the tuple

{*154838,David Batty,62*} from the relation *Patient*, the DBMS will prevent it as it exists a reference to this tuple in the relation *Diagnosis*. The same stands for the **modification operation**. In the relation *Diagnosis*, it will not be permitted to give a value {*75238,O800*} to the last tuple in the figure, as it does not exist a tuple with the *Patient_ID* value 75238 in the relation *Patient*.

1.2.2.2 QUERY OPERATIONS

Database querying is derived from relational algebra operations. One basic principle to consider is that the operands as well as the result of an operation constitutes a relation. Four basic operations are : selection, projection, join and aggregation.

- **Selection:** A selection applied to a relation R with a condition $c1$ results in a new relation R' containing only those tuples for which $c1$ holds. If we consider the example in Fig.1.1 and we apply a selection from *Diagnosis* of the tuples for which the value of *Main_Diagnosis* starts with "T", we obtain the relation *Sel_diag* depicted in Fig.1.2(a).

Sel_diag	
Patiend_ID	**Main_Diagnosis**
1025786	T424
75865	T435

(a)

Proj_diag
Main_Diagnosis
T424
T435
R51
O800

(b)

Figure 1.2: a) Resulting relation from a selection operation applied to the relation *Diagnosis* under the condition *Main_Diagnosis* starts with 'T' . b) Resulting relation from a projection of the relation *Diagnosis* on the attribute *Main_Diagnosis*.

- **Projection:** A projection of a relation R on one or a set of attributes $\{A_i\}_{i=1,...,N}$ consists in the restriction of the set of values from the tuples in R for the set $\{A_i\}_{i=1,...,N}$, resulting in a new relation R' composed of the set $\{A_i\}_{i=1,...,N}$ where replicated values appear only once. If we consider the relation *Diagnosis*, the projection on the attribute *Main_Diagnosis* results in a new relation *Proj_diag* depicted in Fig. 1.2(b).
- **Join:** The join operation combines attributes of two relations into one. For sake of simplicity we present only the most common join operator, i.e., the natural join. For two relations R and S, sharing one or a set of attributes $\{A_i\}_{i=1,...,N}$, the natural join operation results in the set of all combinations of tuples for which the values of $\{A_i\}_{i=1,...,N}$ are equal in both relations. If we consider the natural join of the relations $Patient$ and $Diagnosis$, the new relation $Join\_Pat\_Diag$ is constituted of four attributes $Patient\_ID$, $Name$ Age and $Main\_Diagnosis$ and five tuples.
- **Aggregation:** An aggregation over a relation R allows gathering values from different tuples according to fixed conditions in order to form a single value of more significant meaning

or a measure, such as the mean, the sum, the maximal value, the number of elements, etc.

In the following sections of this work we will consider, without loss of generality, database attacks based only on update operations: suppression of tuples, insertion of new records and modification of attributes' values. Notice however that a selection query can be assimilated (in terms of an attack) to a suppression of tuples with a high number of removed records.

1.2.3 EXISTING SECURITY MECHANISMS AND THEIR LIMITATIONS

In practice, security relies on the definition and the deployment of a security policy. This one specifies security rules and requirements to be satisfied by an information system. These rules precise how information can and can not be accessed, the procedures of recovery management, new user registration and also how security services must be deployed/configured/parameterized, etc. Thus, the deployment of a security policy is a complex process. Nowadays, different standards exist to guide this process, the most recent like BS 7799-2 or ISO 27001 being based on the principle of continual improvement of the security, known as the model PDCA (Plan, Do, Check, Act).

The first step of the deployment of a security policy is a risk analysis, by means for example of the standard EBIOS [DCSSI, 2010] or MAHARI [CLUSIF, 2010]. This risk analysis is part of a process that allows the identification of the objectives and requirements of security that an information system must achieve. In order to counter/minimize the identified risks, existing protection tools and devices are then deployed. Protection means can be classified into physical and logical mechanisms. Physical protection mechanisms concern the materials and essentially aim at counteracting non-authorized physical access, robbery and natural risks such as fire, flooding, etc. In theory, an information system should be placed in a protected zone, isolated where the access is controlled. Anti-theft mechanisms should also be deployed, such as cable locks.

In addition to these mechanisms, maintenance safeguards the proper functioning of the information system in terms of hardware and software. Even though it depends of the security policy, as exposed above, it is usually provided by external service societies contractually linked to the health institution. These maintenance contracts take usually into account the constraints in terms of confidentiality, integrity and availability of information.

In practice, the efficiency of physical protection is arguable, as health establishments are open structures in which patients are mixed with health professionals and there exists an easy access to equipments. This is why logical (i.e., computer-based) security mechanisms are important. In order to have a more comprehensive view of the existing logical security mechanisms for databases, we propose to present them depending on the security objectives they answer to (see Sect. 1.1.2).

1.2.3.1 CONFIDENTIALITY

We can identify four different categories of risks related to the unauthorized access and disclosure of data: 1- illicit access to data and applications; 2- voluntary or involuntary errors in data

manipulation (copy of file, diffusion of information to a non-authorized third, ...); 3- transmission interception; 4- virus ("worms", "Troy horses", ...). In order to cope with these menaces, several counter-measures have been proposed. We suggest to distinguish them according to if they are at the level of the IS or they require some data manipulation.

INFORMATION SYSTEM SECURITY MECHANISMS

User authentication It is the first security barrier in order to protect data from unauthorized access. Several mechanisms have been proposed for user authentication. Nowadays, we must refer to strong authentication, the basic idea of which is the combination of two different criteria with the aim of : 1. Verify user's identity (e.g., password) 2. Provide a proof of the user's identity (e.g., smart card). This solution can be associated with a token, which ensures an unique connexion per user. Once the user connected, the token is assigned to the computer in which he is logged on. Then, no other connection will be permitted for the user somewhere else in the system.

Access control Once the user has accessed the system, it is mandatory to control the actions he/she can perform by defining user access rights. This problem has been largely treated in the literature and different models has been proposed: DAC model (Discretionary Access Control) [Lampson, 1971], MAC model (Mandatory Access Control) [Bell and Lapadula, 1976], RBAC (Role Based Access Control) [Sandhu et al., 1996], or more recently OrBAC (Organization Based Access Control) [Kalam et al., 2003]. Latter models like OrBAC have been extended beyond simply controlling the access and take into account usage control, right access propagation and so on.

Based on these models, a user's request is first validated before granting or rejecting the access to the solicited information. It is usually recommended to implement a reference monitor which verifies that each access is assured by an access right and refuses the access in case of absence of this right. Notice that the application of an access policy can not block all the possible attacks by itself, as a user can bypass its imposed restrictions. It is then necessary to adopt other security counter measures such as firewalls, encryption tools, data obfuscation, etc (see below).

Firewalls When the information system is connected to a network, it is necessary to protect it from intrusions. This is the main role of firewalls which allow surveying and restricting access from the outside (e.g., the Internet) to the inside (e.g., an equipment, local computer, intranet,...) and vice et versa [Zalenski, 2002].It works essentially as a filter: it permits the entrance (resp. transmission) of packages coming from (resp. destined for) an authorized address (IP + port number). A firewall is then one of the mechanisms of access control that can be implemented so as to deploy the security policy rules [Alfaro et al., 2007]. Nevertheless, it can also accomplish several other functions such as address translation (Network Address Translation or NAT) or act as an application proxy. Address translation allows to keep an internal address space, independent from the external network, i.e., addresses are neither known nor accessible from the outside. An application proxy acts as an intermediary which

interprets each interaction of an application (commands, requests and responses) in order to verify these exchanges follow the authorized protocol. However, this device does protect neither the confidentiality nor the integrity of data shared on the network.

Anti-viruses Viruses represent one of the most important threats. They can be introduced into a system in several manners, even if the system is not connected to an open network. Indeed, data transfered from a simple USB flash drive can be infected. An anti-virus policy should be deployed (prevention, detection and isolation of viruses and system restoration). The prevention is based on testing all memory units but also network connexions. Detection consists of checking with one or several virus detection tools the whole information system. According to the isolation principle, suspicious memory units should be disconnected. System restoration goes from virus removal performed by the anti-virus to the reformatting of the memory units and reinstallation of the information system.

DATA SECURITY MECHANISMS

Data Encryption Data encryption aims at preserving information confidentiality. It consists in transforming a plain message into an unreadable ciphertext by means of an encryption algorithm and a ciphering key K_c. Decryption involves the transformation of the ciphertext into a plain message identical to the original one through a decryption algorithm and a deciphering key K_d, as illustrated in figure 1.3. If $K_d = K_c$ we refer to this process as symmetric-key encryption or more simply, symmetric encryption. Until 1976, all the proposed ciphering algorithms were symmetric. Symmetric encryption is still highly used, mostly as a consequence of its rapidity. DES or Triple DES (Data Encryption Standard) and more recently AES (Advanced Encryption Standard) are the two most usual symmetric algorithms [Menezes et al., 1996]. Both algorithms apply a block ciphering process (i.e., they transform fixed-length strings from the plain message into cyphered strings of the same length). We can also find stream cipher algorithms that work directly with data flows, for instance the RC4 (Rivest Cipher 4) [Robshaw, 1995].

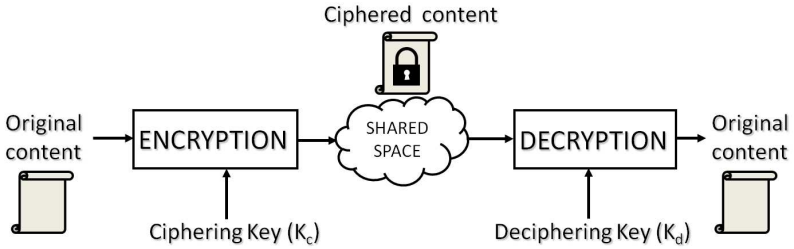


Figure 1.3: Main stages of a common encryption chain. The content is encrypted using a ciphering key K_c. We consider that the ciphered content is shared (e.g via the Internet) and then, decrypted by means of a deciphering key K_d.

On the other hand, asymmetric or public-key encryption is based on two different keys K_c and K_d. Then, if one key is used for encryption, the second key should be used for decryption.

These two keys are mathematically linked but theoretically, the knowledge of one of the keys does not allow obtaining the other one. In this pair of keys, which is associated to a user, one is called public key and is freely distributed. The second one, called private, must remain secret. Thus, in order to ensure data confidentiality, it is necessary to cipher the message with the receiver public key. Only him/her will be able to decipher by means of his/her private key. On the other side, if a user ciphers a message with his/her private key, everyone is able to decipher it by means of the published public key and the origin can not deny the emission of the message (Non-repudiation). The most widespread asymmetric algorithm is RSA (from its authors, Rivest, Shamir, Adleman) [Rivest et al., 1978]. Recently, some authors have dealt with the response to queries over encrypted databases with the help of homomorphic encryption [Hacıgümüş et al., 2004] [Tu et al., 2013], in order to avoid an intermediate decryption which could compromise data privacy.

Data obfuscation Another solution to be considered in order to avoid the unauthorized access to data consists in hiding the original values by replacing them. For example, a credit card number can be hidden by replacing some of its digits with "*". Several strategies have been proposed for the replacement of values: nulling out or deletion, data substitution, data swapping, etc. One important example of application of data obfuscation is the anonymization or sanitization, the purpose of which is avoiding the identification of an individual (i.e., identity disclosure) by means of some attributes called quasi-identifiers (e.g., ZIP code, age, date of birth) within a dataset containing sensitive information (e.g., salary, diseases, religion). To do so, the values of the quasi-identifiers are modified so as to make it impossible to distinguish an individual among a group of records (i.e., generalization). According to the final degree of generalization, we can distinguish *K-anonymity* techniques, in which information for each record contained in the release cannot be distinguished from at least $k-1$ individuals [Samarati and Sweeney, 1998][Samarati, 2001]; *L-diversity*, that creates groups of non distinguishable records, each of which contains at least l records [Machanavajjhala et al., 2007]; *t-closeness*, if the distance between the distribution of a sensitive attribute in this class and the distribution of the attribute in the whole table is no more than a threshold t [Li et al., 2007].

Data obfuscation processes require a strict distortion control, as the introduce modifications may perturb the subsequent uses of the database [Samarati and Sweeney, 1998] [Li et al., 2007], as for watermarking. Some of the proposed distortion measures will be analyzed in section 1.3.2.3.

Construction of database views Database views allow to limit the access to data contained in database tables by limiting the data presented to the end user. Essentially, a view is constructed as a result of a predefined query that populates the contents of an artificial database relation. They provide a granular view of data, allowing to define different sets of accessible attributes and tuples depending on the user access rights. For instance, in the case of the relations $Patient$ and $Diagnosis$ depicted in figure 1.1, it is possible to provide a view $VDoctor$ for emergency doctors where the patient name is hidden, giving only access to the patient IDs, age and main diagnosis.

This does not only protect the confidentiality of data but also limits the risk of accidental

or malevolent suppression or addition of information to the database, as deletion and insertion rights can be efficiently managed [Bertino and Haas, 1988].

1.2.3.2 INTEGRITY AND AUTHENTICITY

Content integrity and authenticity during its transmission or storage can be protected by means of another cryptographic mechanism: the digital signature. It proceeds as follows: 1) an imprint, i.e., a summary, a digest or a hash of the original document is computed. This is usually achieved by means of a cryptographic hash function such as SHA (Secure Hash Algorithm) [Gilbert and Handschuh, 2004] or MD5 (Message Digest version 5) [RFC1321, 1992]; 2) this imprint is then encrypted by means of an asymmetric crypto-system. This way, if a user ciphers an imprint with his/her private key, only the holders of his/her public key will be able to decipher the message. This allows the authentication of the signatory who can not repudiate the message as his/her unique private key has been used. Once the imprint has been deciphered, it is only necessary to compare it to the one recomputed from the received content so as to verify its integrity.

As exposed, the computation of the imprint is based on a cryptographic hash function. The principle of this function is that it provides a digest of fixed size whatever the size of the input document. For an ideal cryptographic hash function, it should be infeasible to generate the original document from its digest. It should also be impossible to modify a document without modifying its digest (diffusion property) and by extension, to have two different documents with the same digest. SHA-1 and MD5 are probably the most widespread cryptographic hash functions. However, after the detection of some security weaknesses, notably for MD5, they are being progressively replaced by new more secure functions such as SHA-3 [Bertoni et al., 2014].

In the case of relational databases, the DBsign^{TM}security suite by Gradkell Systems allows the implementation of a digital signature by means of the *de facto* standard format for digital signature storage PKCS#7 [RFC2315, 1998].

The usage of digital signatures is nowadays recognized by European and American laws [European Parliament, 1999] [U.S Government, 2000] if the employed system is certified. In the same way as for data cipher, a key management system should be deployed and keys should be kept and distributed by a certified trusted third party through a certificate.

In this context, a health-professional smart card can be helpful, as it provides support for both user authentication and encryption of data on the card and during transmission [Smart card alliance, 2003].

Notice that researchers are currently conducted with the objective of going beyond this "strict integrity control". Indeed, as exposed, one fundamental property of digital signatures is that two slightly different documents have completely distinct signatures (diffusion property). Thus, different studies seek to ensure a more relaxed integrity control based on perceptual hash functions [Monga, 2005], that permits the performance of innocent modifications (e.g., lossy compression) while allowing to detect hostile modifications (e.g., false data insertion).

1.2.3.3 AVAILABILITY

A violation of data availability corresponds to an "obstacle" that hinders or makes impossible an authorized access. Regarding material risks, they can be minimized by means of maintenance contracts, natural risks prevention, professional training, etc. Logical risks can result from a violation of data or software integrity or from a logical dysfunction at the access control level, the right attribution, the encryption, etc. This can only be overcome with a good security policy design, which will for example establish rules for: Back-up, crisis management (e.g., what to do in case of virus detection), security mechanisms deployment, etc. The basic idea is to limit the impact in patients' health while not perturbing the correct work of health care professionals.

1.2.3.4 TRACEABILITY

Classical traceability solutions are based in log-files (e.g., SYSLOG, ODBC ...) that record all interactions in the information system. They are later audited so as to verify that every security policy rules have been correctly applied and that all considered dispositions are coherent [Gerhards, 2009] and to raise and alarm if necessary.

Such an audit is a good instrument in order to identify users that may have rerouted or not respected the consent of the patient and their access right. This is important specially in case of emergency, when a physician accesses records of a patient he did not participate the cases of. The main interactions to be recorded are: users' connexions and logouts; creation, modification and destruction of security informations (e.g., access rights, passwords, etc.); rights changes, etc.

In this framework, we can highlight the contribution of the IHE (Integrating the Healthcare Enterprise) initiative, promoted by health care professionals and industry to improve the way computer systems in health care share information. To do so, they define several integration profiles that may be implemented in developed technical solutions. In particular, the Audit Trail and Node Authentication (ATNA) profile specifies when and how logs should be stored [IHE, 2013], both for successful (i.e., authorized) and failed (i.e., blocked by the access control mechanisms) interactions.

Beyond classic audit trail, one can found *a posteriori* access control which completes the *a priori* version (see section 1.2.3.1) and offers more functionalities than the audit itself. It aims at detecting the violations to the security policy in order to punish the culprit [Azkia et al., 2010].

1.2.3.5 LIMITATIONS

All above security mechanisms constitute a remarkable protection for sensitive documents. However, they do not offer a never-failing protection as accessed and consulted documents can escape these control measures. Indeed, once the access to the records in the database is granted and data decrypted and cut off from the ancillary data (header or encapsulation) that

protect them (e.g., digital signature, proof of origin and so on), they can be copied straightaway. Consequently, one can say that these mechanisms offer an "*a priori*" protection. In an "a posteriori" framework, threats one must consider concern data integrity and traceability.

Regarding data integrity, data may be degraded by different processes once the access is granted. One can exchange the ancillary data in between documents for fraud purpose or to hide a medical error for example. This can also happen when systems are not inter-operable.

Moreover, if ancillary data are erased, we can no longer ensure data traceability. All these reasons push us to consider the attachment of security information to data themselves in an inseparable way. This is the role of watermarking that is presented in the sequel.

1.3 WATERMARKING AS A COMPLEMENTARY SECURITY MECHANISM IN DATABASES

Watermarking interestingly complements the security mechanisms exposed in the previous section. In opposition to the "*a priori*" nature of the previously exposed solutions, watermarking provides an "*a posteriori*" protection, as it allows accessing a piece of information while keeping it protected by a watermark intrinsically linked to it. In the following, we come back on watermarking fundamentals and how it is applied to relational databases.

1.3.1 FUNDAMENTALS OF WATERMARKING

1.3.1.1 DEFINITION

Digital watermarking consists in the imperceptible insertion of a useful message (watermark) into a digital content, usually called host document, without disrupting its normal use or interpretation. While being based on the same principles as steganography, as old as cryptography, watermarking differs from it in terms of purpose and appears during the 90's. With steganography the host document has no importance. The objective of the user is to perform a secret communication using the host document as a covering channel. The embedded message must be completely imperceptible and undetectable. On it side, watermarking aims at protecting the host document by embedding a message.

Watermarking was originally proposed in the early 90s for copyright protection of multimedia contents [Zhao and Koch, 1995] [Boland et al., 1995]. Classically, one embed a message containing the owner identity as copyright information or the buyer identity for traceability purposes. The embedded message should be in this context resistant to attempts of a pirate to erase or modify the watermark. Since then, watermarking was extended to copy protection, integrity control and other applications we will come back with in section 1.3.1.3.

1.3.1.2 HOW DOES WATERMARKING WORK?

As depicted in figure 1.4, a common watermarking chain deeply resembles a communication system. In both cases, the objective is the transmission of a message. In the case of watermarking, the noisy communication channel corresponds to the host content and the available

bandwidth is the number of bits of message one can embed. The embedding process is performed by altering or modifying the host content under the principle of controlled distortion. For example, in the case of an image, a message can be inserted by slightly modifying the image pixels values or some transform coefficients of it (DCT, Wavelet, etc.). In databases, attributes' values of a relation or some characteristics of it can be modified (see section 1.3.2.4).

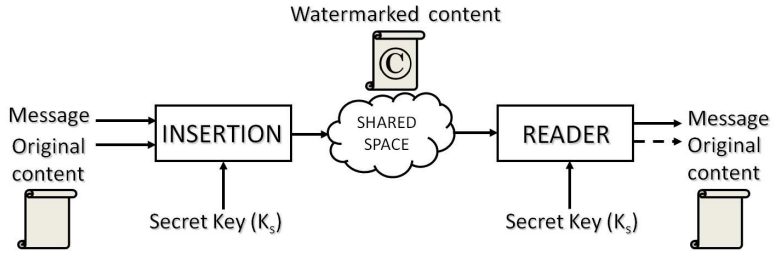


Figure 1.4: Main stages of a common watermarking chain. We consider that the watermarked content is shared (e.g via the Internet) and it can be manipulated between the insertion and the reading stages. At the reader stage, in the case of lossless or reversible watermarking (see Sect. 1.3.1.4), the original content can be fully recovered.

The security of such a watermarking system stands on one or several secret keys K_s. The idea is to follow the same principles as for cryptography [Kerckhoff, 1883a,b], that is to say that the system should be designed under the assumption that the enemy knows all its details except the key. These keys can be used in the selection of the embedding positions of the watermark into the content or in the construction of the watermark itself.

Depending on the application framework, the reader will perform a complete extraction of the message or it will simply detect it (see section 1.3.1.3). Additionally, in some cases the reader will be able to invert the introduced perturbations and to recover the original content. We will talk then of reversible or lossless watermarking, a property the schemes we conceived during this Ph. D. thesis have. Notice that the watermarked content is subject to authorized manipulations or attacks (innocent and malevolent) in between the insertion and the reader that could erase, weaken or modify the embedded message. The former are referred as innocent attacks, the latter being malevolent. The capability of a scheme to resist such an attack corresponds to the concept of robustness, while the length of the message one will be able to insert corresponds to the capacity. These properties are not the only ones, as it will exposed in Sect. 1.3.1.4.

1.3.1.3 WATERMARKING APPLICATIONS

COPYRIGHT PROTECTION

The first developed application of watermarking was copyright protection of multimedia documents. It is based on the insertion of an identifier associating the host document to its owner (creator or buyer) [Cox et al., 2008]. This identifier or the watermark should be imperceptible and resistant to any operations, specially those aiming at removing the watermark. Beyond

embedding the owner identification stands the problem of keeping a track of the link between identifiers and users. As example, the Digimarc company provides a tool for watermarking and maintains a server with owner contact information at the same time [Alattar, 2000]. Watermarking offers a more practical solution than the classical strategies based on the existence of a trusted third party who keeps a copy of the original content.

Moreover, this solution is hard to implement in the case of databases or software due to the storage capacity that will be requiered to keep a copy of this kind of contents, which can by themshelves have an important size.

Notice that the purpose of the first database watermarking technique introduced by Agrawal and Kiernan in 2002 [Agrawal and Kiernan, 2002] was also copyright protection, as for many other data.

TRAITOR TRACING

In some cases, the identification of the recipient of one content can be a priority so as to trace its possible illegal diffusion. Watermarking can be applied in that context being then referred as "fingerprinting" [Wagner, 1983]. Herein, each distributed copy of a content, like a video, is marked with an identifier or fingerprint which uniquely represents an individual. If one of the receivers decides to illegally distribute the document, it becomes possible to determine it [Li et al., 2005]. The way these fingerprints are built has received a lot of research effort in order to make them resistant to collusion attacks, when several users owning copies of a same content, each with a different watermark, cooperate in order to obtain an unwatermarked version. Proposed solutions, usually called anti-collusion codes [Boneh and Shaw, 1996] [Tardos, 2003], have as objective the detection of the attackers implied in such a coalition. We will discuss such anti-collusion codes in more detail in chapter 3 for the purpose of database tracing.

In the same vein, such a traitor tracing solutions can serve the identification of the origin of a data leak. As previously, a message identifying the user is embedded when he/she access the content. If the information is retrieved online, it will be possible to identify the responsible person by extracting the message. Contrarily to the previous problem, the collusion attack is out of concern. We present in chapter 2 a robust lossless database watermarking that allows to embed such a message.

INTEGRITY CONTROL

Integrity control represent the third main application of watermarking. Indeed, it is essential to ensure document integrity, especially when it acquires a legal value as it may be the casein health care, when a diagnosis decision is made. As we have seen, integrity is a component information reliability. A medical doctor will have trust into a piece of information for which he knows the content has not been altered and the origin is the expected one.

Two different integrity notions can nonetheless be identified. In the strict sense of the term, integrity guarantees that the received content is identical to the original one. Another notion

stands in "legal" integrity which is a bit more flexible and which warrants that information has not been modified without authorization.

As exposed in section 1.2.3, the use of cryptographic digital signatures (e.g., SHA-1) is appropriated to strict integrity, i.e., when no modification is tolerated. It is possible to embed such a signature into the content itself. The detection is then based on the comparison between the watermarked and the recomputed signatures. If they are not identical, an alarm is raised.

Fragile or semi-fragile watermarking constitutes an attractive alternative. In opposition to the robustness, the fragility of the mark to manipulations may serve integrity control. The absence or incorrect detection of the mark will indicate a loss of integrity. The basic idea is that any alteration of the content will partially or totally destroy the watermark. Then, depending on the applicative context, the mark can be designed to resist some specific manipulations and to be erased by others. In the case the method is fragile, the watermark should be erased or modified even after a slight manipulation [Li et al., 2004] [Kamel and Kamel, 2011]. On the contrary, we talk of semi-fragile watermarking, i.e., watermarking methods robust to innocent manipulations and fragile to hostile attacks [Zhang et al., 2006] [Coatrieux et al., 2011]. In chapter 2 we present a semi-fragile lossless watermarking scheme, based on the embedding of a digital signature so as to protect the integrity of the database.

Some have also proposed to embed some redundant information that can be used to reconstruct or recover some altered parts of a document. This idea was introduced by Fridrich and Goljan, in the case of images, by hiding a summary of certain zones of the image in other zones. This allows not only alerting a user in case of a non-authorized modification but also precisely identifying the modified areas and to partially restore them [Fridrich and Goljan, 1999] [Iwata et al., 2010]. Others have proposed to identify the nature of the manipulation suffered by the document [Huang et al., 2011].

INSERTION OF META-DATA

Watermarking can be used for meta-data transmission in order to provide an augmented content both in terms of description and functionalities. For instance, in the ARTUS project [Bailly et al., 2006], watermarking is applied so as to dissimulate into a television emission some informations about the movements of a virtual animated agent. On demand, these informations are extracted by an special decoder allowing to superimpose the agent which will provide the subtitle information in sign language, easing the access to the content for deaf people.

Contents can be enhanced by adding, for example, biographic information about their authors. Other semantic information may also be embedded allowing an easier indexation and retrieval of contents into an heterogeneous database. In this framework, the more information embedded, the more accurate the result of the possible queries. The number of referenced methods applying watermarking techniques in this context is not comparable to those developed for copyright protection. Some existing examples are [Chbeir and Gross-Amblard, 2006] or [Manasrah and Al-Haj, 2008].

1.3.1.4 WATERMARKING SYSTEM PROPERTIES

Each specific watermarking application is associated to several requirements, establishing a compromise between different properties. Among these properties, the **capacity** or data payload (i.e., the length of the message that can be embedded), the mark **imperceptibility** and its **robustness** to attacks are usually considered as the canonical features of a watermarking system. Figure 1.5 graphically represents the different trade-offs between these three properties that should be analyzed in order to provide a watermarking solution.

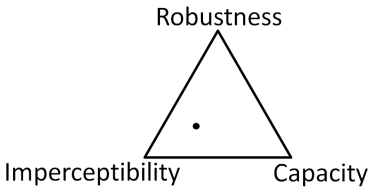


Figure 1.5: Graphical representation of the antagonism between the three canonical watermarking properties. A high performance in terms of two of them typically implies a very low performance in terms of the third one, as represented by the black dot in the figure.

Let us first consider the robustness/imperceptibility compromise. As noticed before, a watermarking system shares some basic notions with a communication system. Consequently, the ability to detect/extract the watermark directly depends of its energy in relation with host content (i.e., the signal to noise ratio). The more intense the watermark, the more robust it is. At the same time, this will introduce a higher distortion into the host content, making the watermark more perceptible (e.g visible for images, audible for audio). In order to increase watermark imperceptibility, some psycho-visual and psycho-auditive criteria are applied in image and audio watermarking, respectively, so as to select the most adequate positions to be altered and the watermark amplitude. The same question stands for database watermarking where, as we will see, statistical and semantic constraints are considered in order to identify modifiable attributes' values. The capacity requirements for copyright or copy control applications are typically small. For instance, the International Standard Audiovisual Number (ISAN), which allows to uniquely identify any audiovisual content, is composed of 64 bits [ISO/IEC, 2002]. Therefore, one can achieve high robustness and good watermarking imperceptibility.

For the same level of distortion, one can reduce robustness and increase capacity. Integrity checking and insertion of meta-data can then be considered. In the former scenario, a digital signature of the content itself can be embedded that allows the detection of non-authorized modifications. One can also embed an digest in order to identify the nature of these latter [Huang et al., 2011]. In general, the more information we embed about the original content, the more precise the identification of modifications will be. In applications based on the insertion of meta-data, capacity requirements are even higher, going to several kilobytes.

Additionally to this three properties and depending on the applicative context, some other important features can be considered, an interesting analysis of which can be found in [Cox

et al., 2008]: reversibility, watermark detection/extraction method, security of the watermark, low complexity, etc.

The **reversibility** property allows the recovery of the original host content by inverting back modifications induced by the watermarking process. It was introduced for image watermarking by Mintzer *et al.* in 1997 [Mintzer et al., 1997]. With such a property, watermark imperceptibility constraints can be relaxed as it is possible to retrieve the original data and embedded message update becomes possible avoiding watermark .In the case of integrity control, one can embed a digital signature of the whole host document. This is not possible with non-reversible schemes as the digital signature can only be computed on the document parts that will not be modified by the embedding process. In the case of database watermarking, it was first considered by Gupta and Pieprzyk [Gupta and Pieprzyk, 2009] in the case of numerical attributes and by Coatrieux *et al.* [Coatrieux et al., 2011] in the categorical case. We will come back on lossless watermarking for relational databases in chapter 2.

Regarding the **watermark detection/extraction**, the original unwatermarked content may or may not be necessary. The former case, we refer to as **non-blind detection/extraction**, will typically improve the system's performance as it allows to isolate the embedded watermark after subtracting the original document. Nonetheless, it is limited to some very specific copyright protection or traitor tracing applications, where the availability of the unwatermarked content is not impractical. On the other hand, **blind detection/extraction** works without any access to the original document. It is applied for instance in integrity control or copy control applications. Figure 1.4 is an example of such a blind detection.

Watermark **security** is directly related to the notion of robustness. While the latter refers to the resilience of the embedded watermark to content manipulations, security relates to the system's capability of resisting to malevolent attacks the purpose of which is to extract or remove the message or to embed a new one in order to hide the original message. The access to the embedded message must be then restricted, generally by means of a secret key K_s that allows only authorized users to extract inserted information. As exposed in Sect. 1.3.1.2, watermarking methods aim at respecting the Kerckhoffs's principle [Kerckhoff, 1883a,b]. This means that the security of the watermark system is based on K_s, being the watermarking method public itself. The reader must refer to [Bas and Furon, 2013] about how to estimate the optimal length of the secret key so as to ensure a secure embedding.

Embedding and reading **complexity** constraints can also be determinant in some applicative frameworks. For instance, in copy control and fingerprinting applications, a real-time detection is desired.

To sum up, the requirements of a watermarking system in terms of these properties are defined by the application framework it is intended to be used for. It is not possible to offer all of these features simultaneously and the final decision should be taken accordingly to the set of imposed constraints. Database watermarking schemes do not escape this rule.

1.3.2 DATABASE WATERMARKING

As previously exposed, a database DB is composed of a finite set of relations $\{R_i\}_{i=1,\ldots,N_R}$. From here on for sake of simplicity and without loss of generality, we will consider one database

based on one single relation constituted of N unordered tuples $\{t_u\}_{u=1,\ldots,N}$, each of M attributes $\{A_1, A_2, \ldots, A_M\}$. The attribute A_n takes its values within an attribute domain and $t_u.A_n$ refers to the value of the n^{th} attribute of the u^{th} tuple. Each tuple is uniquely identified by either one attribute or a set of attributes, we call its primary key $t_u.PK$.

1.3.2.1 DIFFERENCES BETWEEN MULTIMEDIA AND RELATIONAL CONTENTS

Although watermarking appears as a promising complementary tool for database security, the application of existing signal or image watermarking techniques is not a straightforward process. Relational databases differ from multimedia contents in several aspects that must be taken into account when developping a new watermarking scheme.

One of the main differences is that samples in a multimedia signal (e.g., pixels in an image) are sorted into a specific order, in a temporal (e.g., audio signal) and spacial domain (e.g., image or video), that gives a sense of the content itself to the user. Close samples are strongly correlated with usually an important information redundancy at a local level, i.e., between close samples or pixels. This is not the case of relational databases, the purpose of which is to provide an efficient storage of independent elements within a common structure. Thus, tuples in a relation are not stored in any specific order. At the same time, because tuples can be stored and reorganized in many ways in a relation without impacting the database content, the question arises of the synchronization between the insertion and the reading processes. Indeed, with signals or images, one can count on their intrinsic structure, working for instance on blocks or groups of consecutive samples. The same strategy is not so easy to apply in the case of relational databases where tuples can be displaced, added and removed. The identification of the watermarked elements of the database and consequently, the synchronization between the insertion and the detection stages, arise as an issue to be solved.

The frequency and nature of manipulations over the data are also different. In the multimedia case, filtering and compression operations are common. They modify the signal samples' values but don't change the signal structure (a filtered image will be close to its original version). In databases, insertion and deletion of tuples are frequent. They can be seen as subsampling and oversampling operations but with an irregular distribution, a quite rare situation in signal processing, especially if the process output should keep an image structure. Moreover, databases may be queried so as to extract a part of the information that presents an interest to the user.

Beyond the database structure and manipulation, one must also consider that the information contained in a database may come from different sources, such as different services in a hospital. Hence, values in the database are very heterogeneous while having a semantic logic. We may find numerical and categorical data, images, etc. This is not the case in multimedia signals, where all the samples are numerical with the same dynamic. As we will see in chapter 4, this results in a limitation of the watermark perturbations so as to preserve the meaningful value of the database.

Finally, multimedia watermarking is based on perceptual models, such as psychovisual phenomena in the case of images or video watermarking, that take advantage from human perception defects. These models allow watermark embedding into regions less perceivable

to the user. In the case of databases, the different options for querying information also complicate the construction of such a model and impose authors to search more appropriated approaches, based on data statistics or semantic links, as example.

1.3.2.2 INDEPENDENCE OF THE INSERTION/READER SYNCHRONIZATION FROM THE DATABASE STRUCTURE

As exposed in the previous section, contrarily to image or signal where pixels or samples, respectively, are ordered, the structure of a database, i.e., the way tuples and attributes are organized, is not so much constraint. This is one of the key issues to consider in database watermarking. Indeed, if for image and video, the insertion and watermark extraction processes take advantage of data organization, being synchronized through groups of samples or blocks of pixels, this question is largely open in the case of databases. A DBMS can reorganize the tuples in the way it wants, and this will not change the content or impact the value of the database. To overcome this issue, in order to make the watermark insertion/reading independent of the database structure or the way this one is stored, that is to say ensuring a correct synchronization between the embedding and the detection processes, several methods have been considered.

The first approach considered by Agrawal et Kiernan [Agrawal and Kiernan, 2002] consists in secretly constituting two groups of tuples based on a secret key. One of the groups contains the tuples to be watermarked. In order to obtain the group index of a tuple t_u in the relation R_i, they make use of a HASH function modulo a parameter $\gamma \in \mathbb{N}$ which controls the number of tuples to modify. If we define $t_u.PK$ as the primary key of a tuple, K_S the secret key, $\mod$ the modulo operation and $||$ the concatenation operation, the condition $HASH(K_S||HASH(t_u.PK||K_S)) \mod \gamma = 0$ indicates if the tuple must be watermarked or not. In [Agrawal et al., 2003a] and [Agrawal et al., 2003b], the HASH operation is replaced by a pseudo-random generator initialized with the primary key of the tuple concatenated with the secret key. Notice that this method allows embedding a message of one bit only. This consequently restricts the range of possible applications. In order to increase the capacity, Li *et al.* [Li et al., 2005] proposed an evolution of the previous method in which one bit of message is embedded per selected tuple. This allows the insertion of a multibit watermark, offering more applicative options.

A more advanced solution consists in a “tuple grouping operation" which outputs a set of N_g non-intersecting groups of tuples $\{G^i\}_{i=1,\ldots,N_g}$. This allows to spread each bit of the watermark over several tuples, increasing the robustness against tuple deletion or insertion. The first strategy proposed in [Sion et al., 2004] is based in the existence of special tuples called “markers” which serve as a boundary between groups. First of all, tuples are ordered according to the result of a cryptographic HASH operation applied to the most significant bits (MSB) of the tuples attributes concatenated to a secret key K_S as $HASH(K_S||MSB||K_S)$. Then, tuples for which $H(K_S||t_u.PK) \mod e = 0$ are chosen as group markers, where e is a parameter that fixes the size of groups. A group corresponds to the tuples between two group markers. This approach suffers from an important inconvenient as a deletion of some of the markers will induce a big loss of watermark bits. To overcome this issue, the most common strategy consists in calculating the group index number $n_u \in [0, N_g - 1]$ of t_u as in

eq.(1.2) [Shehab et al., 2008]. The use of a cryptographic hash function, such as the Secure Hash Algorithm (SHA), ensures the secure partitioning and the equal distribution of tuples into groups.

$$n_u = H(K_S||H(K_S||t_u.PK)) \mod N_g \tag{1.2}$$

One bit or symbol of the message is then embedded per group of tuples by modulating or modifying the values of one or several attributes according to the rules of the retained watermarking modulation (e.g., modifying the attribute's statistics as in [Sion et al., 2004] or the tuple order as in [Li et al., 2004]). Thus, with N_g groups, the inserted message corresponds to a sequence of N_g symbols $S = \{s_i\}_{i=1,...,N_g}$

Watermark reading works in a similar way. Tuples are first reorganized in N_g groups. From each group, one message symbol is detected or/and extracted depending on the exploited modulation. We come back on these aspects in Section 2.3.2. It is obvious that with such an approach, tuple primary keys should not be modified.

Some other approaches that do not make use of the primary key for group construction have also been proposed. Shehab *et al.* [Shehab et al., 2008] propose to base group creation on the most significant bits (MSB) of the attributes. The main disadvantage of this strategy stands on the fact groups can have very different sizes as MSB may not have a uniform distribution. A strategy based on k-means clustering has been proposed by Huang *et al.* [Huang et al., 2009]. In this case, the secret key K_S contains the initial position of cluster centers. However, this approach strongly depends on the database statistics which can be easily modified by attacks.

1.3.2.3 MEASURING DATABASE DISTORTION

As for any data protected by means of watermarking, one fundamental aspect is the preservation of the informative value of the database after the embedding process. The watermark should be invisible to the user, it should not interfere in the posterior uses of the database and should preserve the database integrity and coherence. Thus, it is necessary to be capable of evaluating and controlling the introduced distortion. Today, the distortion introduced in the database should be evaluated in two manners: statistically and semantically.

STATISTICAL DISTORTION

Most of the statistical distortion evaluation criteria are issued from database privacy protection, in which tuples are modified so as to make it impossible to distinguish an individual among a group of records (see Sect. 1.2.3). In this section we expose only those directly applicable to watermarking.

Usually, it is supposed that the database will undergo some *a posteriori* data mining operations. Some measures have been propose in the case the specific operation (e.g., clustering, classification, etc.) is *a priori* known. These measures are referred as "special purpose metrics", in opposition to general-purpose metrics.

For the latter, the simplest measure is the minimal distortion metric (MD) firstly introduced by Samarati [Samarati, 2001]. For each suppressed or modified record a penalty is added up.

For example, if ten records are altered, the total distortion is ten. This metric is not very useful as the penalty may not correspond to the real impact of the alteration. Other simple measures consist in the evaluation of the impact of an alteration on the attribute's values consist in measuring the attribute's mean and standard deviation variations [Farfoura and Horng, 2010] or the one of the median and the maximal value which may be more precise [Xiao et al., 2007]. Although this may procure a quick regard on the introduced distortion, it does not offer complete information about the variation in the attribute distribution. A better but more complex solution consists in the use of similarity measures between data distributions before and after the watermarking process, like the Kullback-Leibler and the Jensen-Shannon distances [Kamran and Farooq, 2012]. On the same line, Oliveira and Zaïane [Oliveira and Zaïane, 2002] evaluate the dissimilarity between the original and the modified attributes by means of the absolute difference between their histograms, h_{or} and h_{mod} respectively, defined as:

$$diff = \frac{\sum_{i=1}^{n} |h_{or}(i) - h_{mod}(i)|}{\sum_{i=1}^{n} h_{or}(i)} \tag{1.3}$$

Liu *et al.* [Liu et al., 2010] present a new metric called Semantic Information Loss Metric (SILM). This one takes into account the variation of the attribute distribution and of its statistical relation with other attributes. It is calculated as the variation of the ratio between the original and the modified joint probability distributions of the modified attribute C with respect to one statistically linked attribute Q_t (i.e., target and predictor attributes respectively). As it can be seen, despite its semantic denomination, the calculated metric only takes into account statistical parameters and not the real meaning of attributes' values.

On their side, Bertino *et al.* [Bertino et al., 2005] propose to measure the impact of a generalization, i.e., regrouping individual categories into bigger classes, by means of a defined magnitude called information loss (IL). This measure is proposed is the context of a specific generalization based watermarking scheme. For a categorical attribute whose domain corresponds to a set of hierarchically associated values forming a tree T (see figure 1.6), this is defined as:

$$IL_C = \frac{\sum_{i=1}^{M} (n_i \frac{|S_i|-1}{|S|})}{\sum_{i=1}^{M} n_i} \tag{1.4}$$

where S is the set of leaf nodes of T, S_i is the set containing the leaf nodes of a sub-tree rooted by a generalization node p_i and n_i is the number of entries in all the database containing values in S_i. For example, if we consider T the tree associated to the staff of a hospital depicted in figure 1.6, the value $Cardiologist \in S_i$ can be generalized into the generalization node $p_i = Doctor$ and $S_i = \{Radiologist, Cardiologist, Neurologist, Gynecologist\}$. This definition can also be adapted for numerical attributes. However, these measures fail to represent the perturbation in the attribute's distribution and its statistical links with other attributes.

Regarding special purpose metrics, Oliveira and Zaïane propose in the case of a clustering operation to evaluate the number of misclassified elements M_E after the embbeding [Oliveira and Zaïane, 2003].

$$M_e = \frac{1}{N} \sum_{i=1}^{k} (|Cluster_i(D)| - |Cluster_i(D')|) \tag{1.5}$$

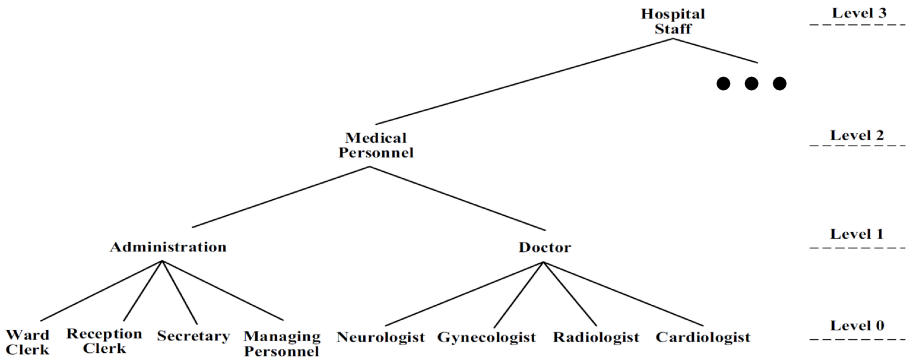


Figure 1.6: Example of domain hierarchy tree representing different roles for medical staff [Bertino et al., 2005].

Another example is given by Kamran and Farooq [Kamran and Farooq, 2012] who, for an previously classified dataset, propose to preserve the classification potential of attributes after watermarking. For a discrete attribute a, which takes values $\{v_j\}_{j=1,\ldots,N_v}$, this classification potential is defined as:

$$C_{P_a} = \left(\frac{IG_a}{\sum_{i=1}^{A} IG_i} \right) \times 100 \tag{1.6}$$

where A is the set of all attributes in the dataset and IG is the information gain is defined in terms of the class labels attribute R as:

$$IG_a = H(R) - \sum_j \mathbb{P}(a = v_j) \cdot H(R|a = v_j) \tag{1.7}$$

SEMANTIC DISTORTION

One important aspect of health databases is the existence of strong semantic links between attributes' values. Let us consider a tuple with attributes $diagnosis$, sex, ... The value $normal\ delivery$ in the domain of the attribute $diagnosis$ is directly linked to the value $female$ of the attribute sex. This link must be preserved in order to ensure the coherence and correctness of the database, keeping modifications invisible to attackers at the same time.

While statistical measures are of interest, they only partially capture the semantic meaning of the database. To our knowledge, semantic aspects of a database were first mentioned by Sion *et al.* [Sion et al., 2004]. However, they only consider higher-level semantics (e.g., end time being greater than begin time) that they finally translate into statistical constraints, without taking into account the existing links at a tuple level. As we will explain in chapter 4,it is possible to identify this semantic links by means of a knowledge base (ontology, terminology, etc.) and to control the introduced distortion in consequence.

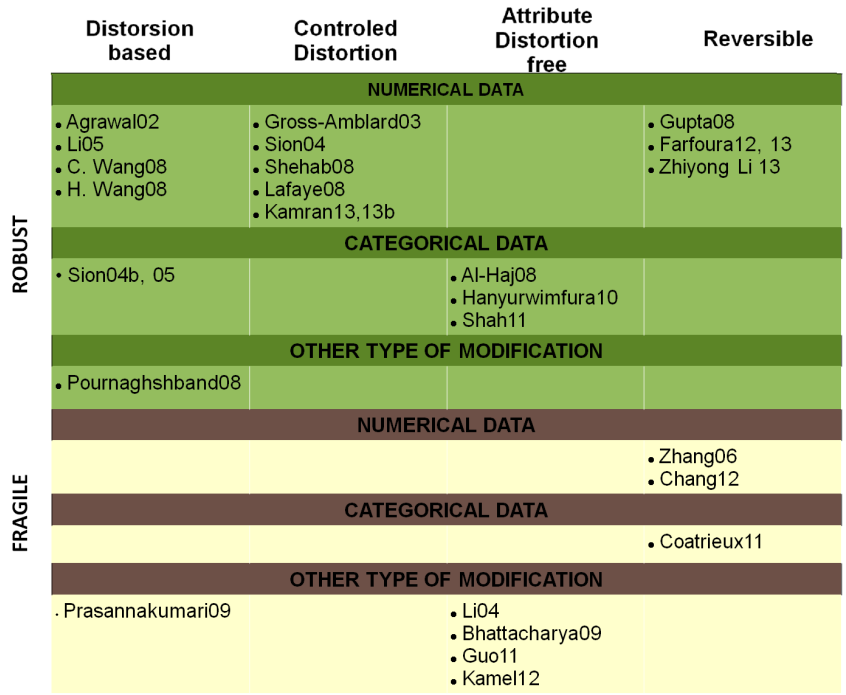


Figure 1.7: Relational database watermarking methods classified on the base of three criteria: approach to cope with introduced distortion, robustness against attacks and the type of modulated data.

1.3.2.4 OVERVIEW OF EXISTING METHODS

This section presents an overview of the state of the art in database watermarking. Marking modulations are classified according to three criteria. First, methods are classified based on their robustness against attacks, an important property in traitor tracing frameworks, or their fragility, a property at the basis of integrity control applications.

As aforementioned, imperceptibility is a fundamental issue in database watermarking. This is why in a second level, we distinguish methods on the way how they deal with data distortion (methods without or with distortion control, "distortion free" methods and lossless or reversible methods). Finally, a lower classification level is considered depending on the type of modulated data (categorical, numerical, order of tuples, etc.). The methods we analyze in this section are summarized according to the above in Fig. 1.7

ROBUST METHODS

As previously exposed, robustness is defined by the ability of the embedded watermark to survive database modifications, authorized or malevolent. This is an important property in copyright or fingerprinting/traitor tracing applications, where attackers will try to eliminate the proof of ownership or the user identification. Herein, we expose the

Distortion based methods

- Modification of numerical data

The first database watermarking method was proposed by Agrawal and Kiernan in 2002 [Agrawal and Kiernan, 2002]. In their method, insertion is carried out in the least significant bits (LSB) of attributes values by bit substitution. The tuples, attributes and bits to be modified are secretly selected by means of a hash operation (see Sect. 1.2.1). In their scheme, the embedded sequence depends of the database content and it is not known by the user, i.e., it corresponds to a database fingerprint. At the detection, if the database has been watermarked, the expected number of bit correspondences in secretly selected positions should be near to 100%, while this number logically falls down to 50% if has not.

Li *et al.* [Li et al., 2005] extended the previous method so as to allow the insertion of a sequence of bits. Considering a multi-bit message to embed m, the function $b \oplus m[q]$ (where b is a pseudo-random binary value and $m[q]$ is the q^{th} bit of the message m) is used in order to modify j^{th} bit of one attribute A_t in t_u. At the detection, for each selected tuple, the operation $b \oplus b_j(of\ \ t_u.A_t)$ is exploited to extract the binary value inserted in $t_u.A_t$, extraction followed by a majority vote strategy so as to determine the final value of the embedded bit. As each bit of the message is embedded several times, robustness is increased. Other approaches following this embedding strategy have been proposed with the aim of adding an additional complexity for an attacker to extract or tamper the watermark [Wang et al., 2008a] [Wang et al., 2008b].

In the previous methods, embedding is performed by modulation of the values of one or several numerical attributes. Nevertheless, some other database characteristics have been considered for watermark insertion: categorical attributes, tuple order, etc.

- Modification of categorical data

As exposed in section 1.2.1, categorical attributes differ from numerical data in the absence of order relationships in between the values of their attribute domain. Sion *et al.* [Sion, 2004] [Sion et al., 2005] proposed the first method allowing to watermark this kind of data. Let us consider an attribute A_t which takes its values in the finite set $\{a_1, a_2, a_3, \ldots, a_{N_a}\}$. These different values do not have a predefined order. However, a numerical value can be arbitrarily assigned to each categorical value creating thus a virtual attribute dynamic as for numerical attributes. By doing so, they can then apply a numerical attribute modulation, for instance LSB substitution. The main problem of this method is the strong distortion it can introduce when the meaning of the new value is considerably different from the original one.

- Other type of Modification

Another type of methods are based on the insertion of new pieces of information (e.g. tuples or attributes) into the database. In that case, even though the original information has not been modified, one can consider that a certain distortion results from the additional data, as they can bias the result of database queries. Pournaghshband [Pournaghshband, 2008] presents a method that inserts false tuples. As such a scheme may

impact database integrity, its algorithm constructs primary key values for each new tuple so as to respect the key integrity constraint (there should not be duplicated primary key values). At the detection, we seek for fake tuples. The presence of one of them indicates that the database has been watermarked, as they are only known of the database owner.

In a scenario not too different which concerns the watermarking of ontologies, we find the method of Suchanek et Gross-Amblard [Suchanek and Gross-Amblard, 2012] based on the same strategy of false information insertion in order to identify the owner of the ontology.

Distortion-control based methods

In order to increase the watermark imperceptibility and do not modify the normal use of the data, distortion control techniques have been considered, working all of them in numerical attributes. Gross-Amblard published in 2003 a theoretical work [Gross-Amblard, 2003] oriented to distortion minimization in the case of *a priori* known aggregation queries. Minimal distortion is considered obtained if the result of these queries is exactly the same as the one obtained with original data. In this framework, Gross-Amblard modulates pairs of tuples involved in the result of the same query with a distortion of identical amplitude but of opposite sign for each tuple in the couple so as to compensate introduced perturbation. This algorithm has been extended and implemented in the Watermill method proposed by Lafaye *et al.* [Lafaye et al., 2008]. A limitation of this approach is that queries should be known *a priori*. Moreover, the authors only considered aggregation queries. For other kind of queries, they propose to apply the method of Sion *et al.* [Sion et al., 2004] based on the modification of attribute's values statistics under information quality constraints, defined by means of the mean squared error (MSE). In this method, once groups of tuples are constructed, a reference value is calculated in each group according to the mean (avg) and the standard deviation (σ) of the attribute to watermark such as: $avg + c\sigma$, where $c \in (0, 1)$ is a user defined parameter. The embedded bit depends on the number of attribute's values in a group ν_c that are over this reference. More clearly, for a group of N_t tuples, insertion relies on two parameters $\nu_{true}, \nu_{false} \in (0, 1)$ in a way that a bit '0' is embedded if $\nu_c < N_t\nu_{false}$, a bit '1' if $\nu_c > N_t\nu_{true}$. At the same time, if the modification exceeds the quality constraints, a fixed threshold, a rollback operation is applied, i.e., all the operations performed on the tuples in a group are undone. However, in their article, authors do not explain how values of tuples should be modified.

Shehab *et al.* [Shehab et al., 2008] enhanced the method of Sion with a more efficient management of distortion constraints, solving at the same time some issues linked to the group creation strategy (see Sect. 1.2.1). Watermarking is presented as a constrained optimization problem, where a dissimulation function Θ is maximized or minimized depending on the bit value to embed. Optimization space is limited by the quality constraints set. In the example given by the authors, Θ represents the number elements which exceed a certain threshold (same function as in the method of Sion *et al.*). At the detection, the value of Θ is calculated and the detected bit is a 1 (resp. 0) if the obtained value is greater (resp. smaller) than a threshold T. The value of T is calculated so as to minimize the probability of decoding error.

Lately, Kamran *et al.* [Kamran et al., 2013b] have proposed the concept of "once-for-all"

155124	Inflammation of right knee

155124	Inflammation of right knee

Figure 1.8: Example of a tuple before and after the insertion with the methods of Al-Haj et Odeh [Al-Haj and Odeh, 2008] et Hanyurwimfura *et al.* [Hanyurwimfura et al., 2010]

usability constraints. Considering a database that should be transfered to several users, they proved that if detection threshold is fixed in order to ensure a correct detection for the most restrictive set of constraints, then detection reliability is independent of the constraints and these most restrictive set of constraints can be called "once-for-all". The problem of this method is that its robustness and low distortion stand on a very short mark embedded into a few number of tuples, with all the inconvenient that represents. If the good tuples are altered, the database is unprotected. The same authors in another work [Kamran et al., 2013a], propose a watermarking scheme that preserves classification results of an *a priori* known data-mining process. To do so, attributes are first grouped according to their importance in the mining process. Some local (i.e. for a set of attributes) and global constraints are then calculated according to some dataset statistical characteristics that are important in the mining process. Allowed perturbation for a set of attributes is obtained by means of optimization techniques.

As it can be seen, all these methods focus on preserving statistical properties of the database. However, another aspect to consider when modifying attributes values is the existence of strong semantic links between attributes values in a tuple, links that should be preserved. Indeed, tuples must remain semantically coherent in order to: i) assure the correct interpretation of the information without introducing incoherent or unlikely records; ii) keep the introduced perturbations invisible to the attacker. We will come back on this issue in more detail in chapter 4, where we propose an original solution to treat these aspects.

"Attribute Distortion-free" methods

In the above methods, it is supposed that a slight distortion can be carried out for message insertion without perturbing the interpretation or any *a posteriori* uses of data. However, if one may consider that no data perturbation can be introduced, "attribute distortion-free" methods, i.e., methods that do not modify attributes values, can represent an interesting alternative. Attribute Distortion-free robust embedding strategies play on the way textual or categorical attributes values are encoded.

Al-Haj et Odeh [Al-Haj and Odeh, 2008] embed a binary image by modifying the number of spaces between words. In the same vain, Hanyurwimfura *et al.* [Hanyurwimfura et al., 2010] take advantage of the Levenshtein distance between words in order to select the words between which the space can be modified, being those at smaller distance. Figure 1.8 illustrates the application of this modulation on a textual attribute. We consider such kind of modification does not induce any information quality loss.

Instead of modifying the spaces, Shah *et al.* [Shah et al., 2011] alter the encoding of attributes values and play with capital letters of secretly selected attributes. According to the bit to embed, the complete word (or phrase) or only the first letter are capitalized.

Notice that even if their authors present these methods as being robust, the watermark will be easily erased by means of simple manipulations on the way the attributes are

encoded (e.g., script giving a fixed value to all the spaces or changing words' capitalization). Reversible watermarking, which allows recovering the original database from its watermarked version are a more interesting option.

Lossless or reversible methods

In some cases, there is an interest or even a need of being able to recover the original database from its watermarked version, for example one may want to perform some operations on the original data or to update the watermark. The reversibility property is herein of great interest. Robust lossless watermarking has been recently considered in the context of relational databases. Most of the existing methods are an adaptation of techniques proposed for image watermarking. They are predominantly fragile with some exceptions.

Let us start by the latter. In [Gupta and Pieprzyk, 2009], Gupta and Pieprzyk propose a zero-bit watermarking method where a meaningless pattern is embedded into secretly chosen tuples. To do so, a secretly chosen LSB from the integer part of a numerical value is replaced by a pseudo-random generated bit. The original value is then inserted into the space left by right shifting the LSB representation of the fractional part. The presence of this pattern is checked by the detector, indicating if the database has been watermarked or not. In order to reduce data distortion, Farfoura *et al.* [Farfoura et al., 2012] [Farfoura et al., 2013] suggest watermarking the fractional part of one numerical attribute by means of prediction-error expansion modulation originally proposed by Alattar in [Alattar, 2004]. Although this method is said robust against common database manipulations (e.g. tuple addition or removal), a simple rounding integer operation will destroy the watermark. More generally, difference expansion modulation has not been designed for being robust to attributes' values modifications (this is the same for images). Notice also that the robustness of this method relies in the use of a majority vote strategy.

The method by Zhiyong Li *et al.* [Zhiyong Li and Tao, 2013] constructs groups of tuples according to a clustering technique. The maximal modification that can be introduced into a tuple should ensure that it will remain in the same group from the detector point of view. The watermarked value of an attribute is calculated from an expansion of the polar angle of the attributes to watermark. Unfortunately, the calculation details of this angle are not provided by the authors. Moreover, as they explain, the method is partially reversible as some little errors can be found in the recovered data.

FRAGILE METHODS

In contrast to robust methods, fragile methods have been designed so as to allow the disappearance of the watermark after database manipulations. This makes them of interest in integrity control applications, objective for which most of the following methods have been proposed to.

Distortion based methods

Prasannakumari [Prasannakumari, 2009] proposed the addition of virtual attributes into the relation, which will contain the watermark information. They propose the following steps. First of all, groups of tuples are constructed. An attribute of NULL value is

inserted in all the tuples of the relation. For each group, the value of the virtual attribute is replaced by an aggregate of the values of a chosen numerical attribute in the group. The aggregate can be the sum, the mean value, the median, etc. Then, for each tuple, the checksum of each attribute is calculated and concatenated to the virtual attribute value. At the verification stage, the same steps are followed. This method does not perturb the original data but introduces an additional information that we consider as a distortion to the database. Nevertheless, this method could be classified as distortion-free.

"Attribute Distortion-free" methods

Regarding "attribute distortion-free" methods, proposed strategies consist in the modulation of the tuples' (or attributes') organization in the relation. The first "attribute Distortion-free" method, presented by Li *et al.* in 2004 [Li et al., 2004], is constructed over this principle. It does not modify the values of attributes, hence the idea of a distortion-free watermarking. In order to embed the mark, tuples are grouped and ordered in the group according to the value of a hash function calculated on the attributes values concatenated with the primary key and the owner secret key. The mark to embed for a group i is a sequence W_i of length $l_i = \frac{N_i}{2}$ with N_i the number of tuples in the group. Insertion consist in altering the order of pairs of tuples in the group depending on the bit to embed. In the detection, if we do not obtain the same order of tuples, the database is considered as compromised. Other approaches have been later proposed in order to allow the identification of the manipulations the database underwent [Bhattacharya and Cortesi, 2009], [Kamel and Kamel, 2011] or to increase the embedding capacity [Guo, 2011].

Methods based on tuple reordering are extremely fragile as tuple order is easily manipulated by DBMS, a situation that does not correspond to a malevolent attack. As a consequence, their application context is limited.

Lossless or reversible methods

As explained in Sect. 1.3.1.4, lossless watermarking represents an interesting mechanism for integrity control, as it allows the embedding of a digital signature computed over the whole database. Being derived from image watermarking modulations, the majority work on numerical attributes, with only some exceptions we will see herein.

- Modification of numerical data

The histogram shifting modulation, a classical lossless modulation for images, has been applied by Zhang *et al.* [Zhang et al., 2006] to partial errors in a relation (i.e. the differences between values of consecutive tuples). For a unique digit of this difference (value comprise between 0 and 9), the histogram is calculated. Values that do not correspond to the maximum of the histogram (called "non-carriers") are shifted to the right. This creates a free class at the right side of the histogram maximum. The attributes that belong to this maximum, the "carriers", are then shifted to the right in order to code a '1' or let unchanged to code '0'. The problem of this approach is that the less significant digits of the calculated differences follow uniform distributions, which reduces embedding capacity to its minimum. In order to achieve a high capacity, it is necessary to introduce a high distortion by modifying the most significant digits.

Another approach proposed by Chang and Wu [Chang and Wu, 2012] consider the use of a support vector machine (SVM) classifier. One SVM is trained with a set of tuples selected so as to obtain a classification function $f(V)$ used by next to predict the values of one numerical attribute. Then, they apply difference expansion modulation for message embedding. Basically, they expand the differences between original and predicted values adding one virtual Least Significant Bit that is used for embedding message bits. The distortion magnitude is unpredictable and as underlined by its authors, it can be high in some cases.

- Modification of categorical data

Coatrieux *et al.* [Coatrieux et al., 2011] adapted the histogram shifting modulation to categorical data, being the first lossless watermarking method for this kind of attributes. For each group, the tuples are secretly divided in two sub-groups SG_1 et SG_2. The number of appearances of the values of the attribute in the sub-group SG_1 are used to construct a virtual dynamic, i.e an order relation between the different values that the attribute can take. The elements of the sub-group SG_2 serve to the insertion and the histogram shifting modulation is applied considering the virtual dynamic constructed from SG_1. The elements that belong to the class with the highest cardinality are considered as carriers. The others are shifted to the right so as to create a free bin. Carriers are then shifted or let unchanged depending on the bit to embed, '1' or '0' respectively. The embedded watermark can be a signature of the database used to verify its integrity.

1.4 CONCLUSION

As we have seen in this chapter, clinical data repositories, or more generally medical databases we want to protect, are very large databases involving a large number of very heterogeneous variables issued from different sources. They are manipulated in highly open environments in which different users access data at the same time and for different purposes, while being transfered and shared between different institutions. These specific characteristics together with the sensitive nature of medical data have leaded to the definition of strict legislative and ethical rules regarding the nominative aspect of medical records and the security of medical databases. Looking at databases and their content, security can be defined in terms of four essential points.

Data confidentiality: propriety that ensures that only the authorized users will be able to access the information;

Data availability: aptitude of an information system to be employed by intended users in the stipulated operation and access conditions;

Data reliability: which involves two aspects: 1- information integrity or the proof that it has not been modified by non-authorized users; 2- authenticity or the guarantee data is issued from the intended sources. The reliability notion ensures that a health professional can trustfully use the information.

Data traceability: Corresponds to an extension of data reliability when it is possible to verify the origin and modifications on data all over their existence.

We have also pointed out that the deployment of security solutions for health relational databases requires the definition of a security policy and the implementation of several security mechanisms. These mechanisms should be complementary and coherent in order to provide a high level of security while not impacting the daily medical practice. Consequently, a compromise has to be found in order to ensure an acceptable security level while not perturbing medical services. But at the same time such a flexibility weakens security which may intrinsically present some frailties due to the fact that security tools are mostly oriented to confidentiality and to the Information system protection. The classic access control based solutions are not enough as once these barriers are bypassed, it is difficult to ensure data integrity and authenticity.

In this context, watermarking provides a non-trivial contribution. It leaves access to the information while maintaining it protected. If many works have been conducted about medical signal watermarking, very few have been devoted to medical databases or to databases in general even for general public applications. This constitutes the main objective of this Ph. D. work, developing adapted watermarking mechanisms for medical relational databases. As we will see, our results are however not limited to the medical domain.

As discussed, the reversibility property is of special interest in the medical context as the original database can be completely recovered once the watermark extracted, leading thus to applications like integrity and authenticity control and traceability. In the next chapter, we detail two lossless watermarking schemes, one fragile and one robust, which can be applied to the protection of data integrity and authenticity. Moreover, in chapter 3 we illustrate how our robust lossless watermarking scheme can be considered so as to trace relational databases stored and mixed with some others in data-warehouses. The objective of this part being: is it possible to know if a database has been merged with other data?

Whatever the methods, we further theoretically demonstrate the limits of performance of our approaches in terms of capacity and robustness for a given level of distortion.

The control of introduced distortion is another important issue in database watermarking. We have exposed above different criteria that have been proposed to measure the statistical impact of the watermark insertion. However, much has to be done, especially in terms of semantic preservation of the data. In particular, how to preserve the semantic information that exists at a tuple level. As we will see in chapter 4, medical knowledge bases (e.g., ontologies) can help us to identify existing semantic relationships between attributes values in a tuple and to adapt the watermark amplitude in consequence.

Chapter 2
Lossless Database Watermarking

As exposed in the previous chapter, medical information and consequently medical databases present several security needs derived from strict ethics and legislative rules. Existing protection mechanisms for medical relational databases, such as encryption, data obfuscation or firewalls present some weaknesses especially in terms of protection continuity and traceability. They are in their majority designed to offer an "*a priori*" protection, i.e., they block non-authorized accesses to the database, but once the access is granted data are no longer protected. How to detect then that a database has been modified, partially erased or rerouted by users, either accidentally or on purpose is of main concern.

In that context, we have introduced watermarking as a complementary mechanism and presented how it has been adapted from multimedia data to relational databases. Whatever the method, it is usually assumed that database content can be altered without perturbing any of its *a posteriori* uses, being these automatic (e.g., data mining) or human-based. However, in some more restrictive scenarios (e.g., in medical or military data), one may have an interest to retrieve the original records so as to process them or to update the watermark content without watermark superposition, which will increase host data distortion. In that case, lossless or reversible watermarking appears as an attractive alternative.

The reversibility property (see section 1.3.1.4), allows the embedding of a message in a way that the host data perturbations or alterations can be "undone" or reversed.

In this chapter, we first introduce the applicative frameworks of lossless watermarking in the health care domain before analyzing the existing alternatives for relational database watermarking. This guides us to propose two reversible watermarking schemes, one robust and one fragile, based on the circular histogram modulation originally proposed by De Vleeschouwer *et al.* for images [De Vleeschouwer et al., 2003]. As another contribution of this chapter, a theoretical evaluation of the performance of our schemes in terms of capacity and robustness against common database modifications or attacks (tuple insertion and suppression) is given. These theoretical limits are verified by means of experiments conducted on one real medical database of patient stay records. In order to better evaluate the benefits and limitations of our robust scheme, we also compare it with two recent and efficient schemes ([Gupta and Pieprzyk, 2009] and [Farfoura et al., 2013]) in terms of robustness, distortion and complexity. Finally, we address some other important issues such as the watermark security and the false detection probability.

2.1 APPLICATIONS OF LOSSLESS WATERMARKING FOR MEDICAL DATABASES

Watermarking is a highly versatile tool which can serve different security purposes (see section 1.3.1.3). Its application to health relational databases is nonetheless limited to the work of Bertino *et al*. [Bertino et al., 2005], who aim at preserving simultaneously the privacy and the ownership rights of outsourced medical data, and Kamran *et al*. [Kamran and Farooq, 2012], who propose to apply optimization techniques in order to obtain a watermark that preserves the classification potential of attributes in a database within a subsequent data-mining process. Regarding lossless watermarking, the only approach is presented by Coatrieux *et al*. [Coatrieux et al., 2011], who reversibly watermark ICD-10 codes to ensure database integrity.

Due to this limited development, in this section we propose to go deeper into the foreseen potential use of lossless watermarking of medical relational databases, taking as reference solutions that have been proposed for medical images. Two main scenarios can be considered: database protection and meta-data insertion.

2.1.1 HEALTH DATABASES PROTECTION

In health care, watermarking has been first introduced as a mechanism providing an *a posteriori* protection complementary to the already existing mechanisms, so as to ensure data reliability control and traceability. Ensuring reliability of a database is based on the embedding into the data of proofs of their integrity, i.e., a proof that they have not been modified by a non-authorized user, and their authenticity, i.e., evidence of the database origin. In order to serve database traceability, the basic idea consists in inserting a message that allows us to identify at any moment the origin or the last database user, this latter being an operator or an information system.

2.1.1.1 RELIABILITY CONTROL

INTEGRITY CONTROL

As stated above, integrity control consists in proving that the information has not been altered by any unauthorized user. In order to ensure database integrity, the first possible solution consists in the insertion of a digital signature or a cryptographic digest of it. Notice that lossless or reversible watermarking may favorably contribute to this solution as it allows recovering the database original content once the watermark extracted from it. We depict in figure 2.1 one basic solution where lossless watermarking and digital signature are exploited so as to protect the integrity of a database [Zhang et al., 2006].

In this approach, a sequence of fixed length (a digest) is computed from the whole database by means of a cryptographic hash function, such as SHA-1 (Secure Hash Algorithm). This sequence can be asymmetrically encrypted using the private key of the data owner, obtaining then a digital signature. By next, this sequence is reversibly embedded into the database.

In order to verify the integrity of the received database, one has just to extract the embedded message, decipher it if needed and compare it to the recomputed digest from the restored database. If they match, the database has not been modified. Otherwise, we know that some modification has occurred. The interest of this strategy lies on the fact that lossless watermarking can be seen as more secure and practical than a system based on lossy watermarking in which the signature can only be computed on database elements not modified by the watermarking process [Coatrieux et al., 2011].

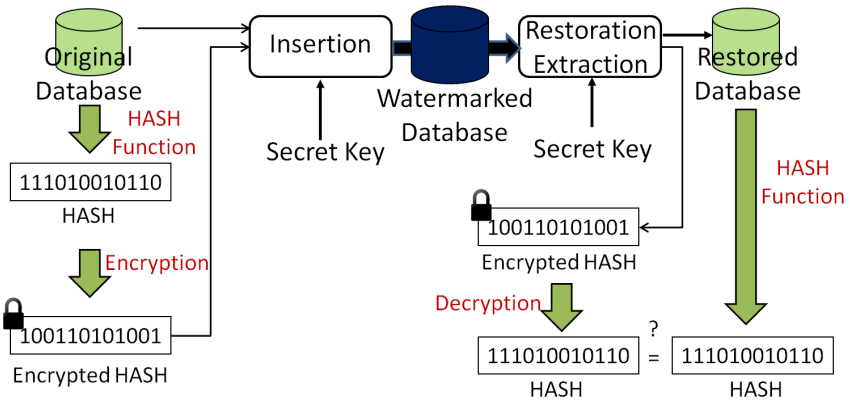


Figure 2.1: Database integrity control by means of lossless watermarking and the embedding of a digital signature.

Different signatures can be embedded so as to allow the localization of database modifications (e.g., which group of tuples has been modified). Proposed solutions are based on the non-reversible modulation of the order of tuples or attributes in the database [Li et al., 2004] [Guo, 2011]. In the case of images, some have proposed signatures based on geometrical moments extracted from the image [Huang et al., 2011]. Such a signature allows the user to detect and localize modifications but also to have an idea about the nature of the modification. One can expect to achieve the same objectives for relational databases but currently no solution has been proposed. This can be of interest informing, for instance, a health professional of the database parts that have not been compromised after an attack. This will avoid the retransmission of the database. From a medico-legal point of view, being able to identify altered tuples may also help to clarify the motivation of the attack.

Authenticity Control

In the case of authenticity control, the insertion of an authentication code will allow verifying the database origin or its expected destination. This authentication code may contain different informations relative to the owner of the database, the date of creation, the administrator, the health establishment, the final recipient and so on. For example, such a solution has been proposed in the case of images, where the standard DICOM ((Digital Imaging and Communications in Medicine) defines a UID (Unique Identifier) which is uniquely associated to the image and which can be embedded together with a patient identifier so as to prove the origin of the image [Woo et al., 2005]. Notice that this kind of test permits to prove direct assertions such

as "the establishment X is the sender of the database" and not to perform a more advanced inquiry about the identity of the owner. Nevertheless, this is usually enough in most of the information exchange frameworks.

2.1.1.2 DATABASE TRACEABILITY

Traceability refers to the possibility of identifying at any time the origin or the last user that accessed a database. As in the authenticity case, an identifier uniquely associated to a user or an establishment has to be inserted into the database. Lossless watermarking can be of interest while distinguishing two different applicative contexts.

In the case of a shared data warehouse, where different health establishments cooperate and upload their databases, each database may contain a watermark allowing to check its presence in the warehouse or to identify it in case of a leak. We address this case in more detail in Chapter 3.

In a second scenario, we may consider a chain of treatment which requires different users to access a database sequentially. In that case, lossless watermarking allows one authorized professional to remove the identifier from a previous user before the embedding of a new one corresponding to the current access, as depicted in figure 2.2. By doing this, if the database is outsourced by any user in the middle of the chain, it will contain his or her identifier, allowing an accusation. In such a framework, the question of watermarking access is of major concern, that is to say, who can read and modify the watermark and how to guarantee can not be counterfeit. In [Pan, 2012], the author proposes a secure lossless watermarking module, guided by a security policy which responds to this issues in health care. With the help of digital signatures and encryption, it is not possible for an attacker to illegally update the watermark content without being detected. Such an approach can be extended to databases.

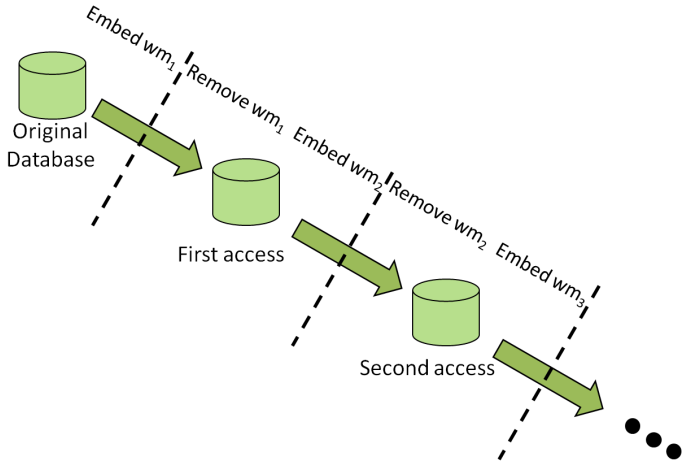


Figure 2.2: Traceability of a database in a chain of treatments. The reversibility of the embedding allows to replace the watermark after each step.

2.1.2 INSERTION OF META-DATA

A second application of lossless watermarking in health databases is the insertion of meta-data. The additional embedded information can help in the manipulation and query of the database as well as include indexing information for example.

Regarding databases, meta-data can include the attributes' data types and names, the total number of tuples, number of attributes and so on. We could also embed some statistical information about the values of the database, such as the correlation between attributes, the attributes means and standard deviations and so on. This could serve database repair in case of damage, as suggested in the context of an image database in [Coatrieux et al., 2006]. While some authors have proposed to embed meta-data into medical images by means of watermarking, with the purpose for instance of adding a knowledge digest to the image that eases the retrieval of similar images [Coatrieux et al., 2009], its extension to databases has not been still experimented.

2.2 OVERVIEW OF EXISTING LOSSLESS METHODS IN DATABASE WATERMARKING

Up to now, the few existing reversible approaches for database watermarking have been derived or adapted from lossless image watermarking. This is why they mostly work on numerical attributes rather than on categorical attributes. Notice that the first reversible scheme for categorical attributes has been presented by Coatrieux *et al.* in 2011 [Coatrieux et al., 2011]. We will come back on their algorithm at the end of this section.

Most database lossless schemes are fragile and devoted to database authentication. The approaches by Zhang *et al.* [Zhang et al., 2006] and Chang and Wu [Chang and Wu, 2012] are good illustrative examples of such schemes. The former applies histogram shifting modulation onto the differences between the values of a numerical attribute of two consecutive previously ordered tuples. Originally introduced by Ni *et al.* for images [Ni et al., 2006], the basic principle of Histogram Shifting modulation, illustrated in figure 2.3 in a general case, consists of shifting a range of the histogram with a fixed magnitude Δ, in order to create a 'gap' near the histogram maxima (C_1 in figure 2.3). Pixels, or more generally samples with values associated to the class of the histogram maxima (C_0 in figure 2.3), are then shifted to the gap or kept unchanged to encode one bit of the message, i.e., '0' or '1'. As stated previously, we name samples that belong to this class as "carriers". Other samples, i.e., "noncarriers", are simply shifted. At the extraction stage, the extractor just has to interpret the message from the samples of the classes and and invert watermark distortions (i.e., shifting back shifted values). Obviously, in order to restore exactly the original data, the watermark extractor needs to be informed of the positions of samples that have been shifted out of the dynamic range $[v_{min}, v_{max}]$, samples we refer as overflows or underflows (figure 2.3) only illustrates "overflows"). This requires the embedding of an overhead and reduces the watermark capacity. In [Zhang et al., 2006], this modulation is applied on one digit of the difference (in the integer value range $[1, 9]$). The capacity of this method directly depends on the probability distribution of the considered digit. In case of a flat histogram, i.e., a uniform distribution, the capacity is null (there is

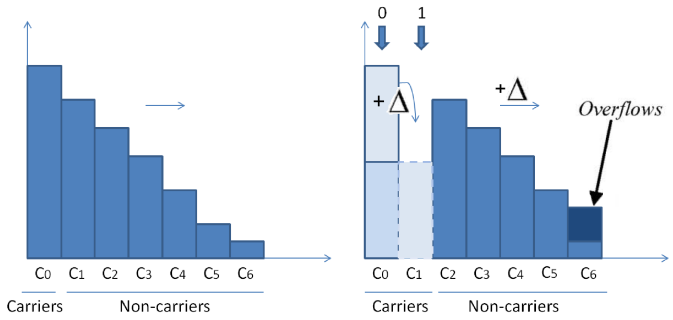


Figure 2.3: Embedding process for the histogram shifting method. As seen, non-carriers are shifted to the right so as to create a free class that will contain elements coding a '1'.

no maximum). As shown by the authors, except for the most significant digits, all the other have a uniform probability density. As a consequence, achieving an acceptable capacity may induce a high database distortion. Chang and Wu [Chang and Wu, 2012] consider the use of a support vector machine (SVM) classifier. One SVM is trained with a set of selected tuples so as to obtain a classification function $f(V)$ used by next to predict the values of one numerical attribute. Then, they apply difference expansion modulation to the difference between original and predicted values for message embedding. Basically, they "expand" these differences adding one virtual Least Significant Bit that is used for embedding the bit of a message. The distortion magnitude is unpredictable and as underlined by its authors, it can be very high in some cases.

Robust lossless watermarking has been experimented only recently. In [Gupta and Pieprzyk, 2009], Gupta and Pieprzyk propose a zero-bit watermarking method where a meaningless pattern is embedded into secretly chosen tuples. To do so, a secretly chosen LSB from the integer part of a numerical value is replaced by a pseudo-random generated bit. The original value is then inserted into the space left by right shifting the LSB of the fractional part. The presence of this pattern is checked by the detector, indicating if the database has been watermarked or not. In order to reduce introduced distortion, Farfoura *et al.* [Farfoura et al., 2012][Farfoura et al., 2013] suggest watermarking the fractional part of one numerical attribute by means of prediction-error expansion modulation proposed by Alattar in [Alattar, 2004]. To do so, this method computes the difference between the fractional part of the value of a numerical attribute $t_u.A_t$ and a value derived from the hash of the primary key of the tuple $t_u.PK$. Then, one bit of the message is concatenated to the binary representation of this difference (expanding it). The result is then converted into an integer value used as watermarked fractional part. Although this method is said robust against common database manipulations (e.g., tuple addition or removal), a rounding integer operation may destroy the watermark. More generally, difference expansion modulation has not been designed for being robust to attributes' values modifications (this is the same for images). Another possible disadvantage of this method is that it introduces new values in the attribute domain. More clearly, if the fractional part of an attribute is encoded on a fixed number of p bits, then its values will belong to the range $[0, \frac{1}{2^p}]$. By expanding the difference, we have no guarantee the resulting values belong to the attribute domain. This is the case for example for an

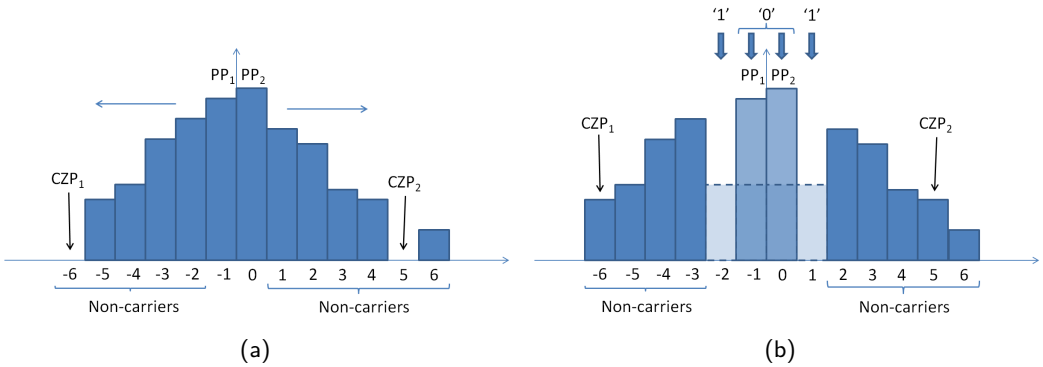


Figure 2.4: Histogram shifting of the differences between preordered attributes values. Two pairs [Peak point (PP), Closest Zero Point (CZP)] are considered so as to increase the embedding capacity. In order to embed a bit of message, carriers (i.e., values corresponding to the peaks) are shifted or left unchanged.

attribute "number of weeks" measured with a daily precision, its fractional part varies in discrete multiples of $\frac{1}{7}$ (encoded on $p = 3$ bits). Furthermore, adding new attribute values can be easily identified. However, the authors do not specify how to deal with such a situation.

To increase robustness, this last method considers a majority vote detection based on the redundant insertion of the message. More clearly, the message is embedded several times all along the available tuples and at the detection, the final value for one symbol of the message is calculated as the majority from the extracted values. This detection strategy is also applied by Chang *et al.* [Chin-Chen Chang, 2013]. For a set of secretly selected tuples, they propose to apply an adapted version of the histogram shifting modulation to the differences between the original attribute's values and the median value of the set. More clearly, they obtain two pairs [Peak point (PP), Closest Zero Point (CZP)] in the histogram and they shift the difference values in the range $[CZP1, PP1)$ to the left-hand side of histogram by 1 and the values in the range $(PP2, CZP2]$ to the right-hand side of the histogram by 1 in order to create two empty bins (see figure 2.4). The use of two peaks allows increasing the embedding capacity. Notice that some pieces of information concerning the position of the peaks and zeros of the histogram have to be transfered to the detector in order to allow the correct reconstruction of the original database.

Jawad and Khan [Jawad and Khan, 2013] present another approach seeking a better capacity-distortion trade-off. They extend a former scheme proposed by Gupta and Pieprzyk [Gupta and Pieprzyk, 2008] in which the difference expansion modulation is applied to pairs of integer attributes in tuples. More clearly, for two attributes to be modulated in secretly selected tuples, the difference between them is expanded allowing to insert one bit of message. However, if this expansion violates some pre-established distortion constraints (defined in terms of the number of modifiable LSBs), the tuple is not considered for watermarking. To reduce the amount of non-considered tuples, Jawad and Khan propose to apply a genetic optimization algorithm that selects the most suitable pairs of attributes to be considered in each tuple in order to respect fixed quality constraints. Nevertheless, in order to make their scheme fully reversible, the chromosome obtained in the optimization process should be transfered to the

reader.

The lossless watermarking of categorical attributes has not been yet fully addressed. One of the difficulties is that categorical attributes lack of order relationships in between values in their domain. To overcome this issue, Coatrieux *et al.* propose to build a virtual dynamic based on the statistical distribution of the attributes values [Coatrieux et al., 2011]. To do so, each group of tuples is secretly partitioned in two sub-groups SG_1 et SG_2, not necessarily of equal size. The number of occurrences of the values of an attribute in the sub-group SG_1 serves to establish an order relationship in between the different values of the attribute so as to build a virtual dynamic. The tuples of the sub-group SG_2 are modulated according to the histogram shifting modulation which is applied considering the virtual dynamic constructed from SG_1. The embedded watermark can be a signature of the database used to verify its integrity.

To sum up, the notion of lossless database watermarking is still uncommon, with only a few methods proposed. Robust lossless database watermarking has been only recently addressed. Most of the proposed lossless robust methods consider the application of a majority vote strategy in order to obtain robustness against common database attacks. That is to say, they are not intrinsically robust against these kind of manipulations. In the next section, we propose a lossless database watermarking scheme which naturally achieves a high robustness to common database manipulations.

2.3 PROPOSED LOSSLESS SCHEMES

In order to overcome some of the issues exposed above, we propose to exploit the robust lossless watermarking modulation originally proposed for images by De Vleeschouwer *et al.* [De Vleeschouwer et al., 2003] and integrate it within a common database watermarking scheme. This choice stands on the fact that, contrarily to histogram shifting or difference expansion modulation, this one is naturally robust. At the same time, it has never been considered in the case of databases. As we will see, this modulation works on circular histograms of numerical data. It does not depend on the existence of attributes with fractional parts, as it is the case in [Gupta and Pieprzyk, 2009] and [Farfoura et al., 2013]. The introduced distortion can be completely estimated before embedding, something not easy in some of the above presented methods.

This modulation allows us to derive two watermarking schemes: one fragile and one robust to tuple suppresion and insertion attacks as well as to attributes' values modification. Our schemes do not depend on the storing structure of the database, making them robust to tuple reordering in a relation, an important aspect considering that the DBMS can reorganize tuples in a relation in the way it wants.

2.3.1 CIRCULAR HISTOGRAM WATERMARKING

In [De Vleeschouwer et al., 2003], De Vleeschouwer *et al.* present two different modulations, one robust and one fragile, both based on a circular interpretation of bijective transformations.

In the robust case, we focus on here, they propose to divide a gray-scale image into N_b blocks of pixels. Each block is equally divided into two sub-blocks whose histograms are mapped onto a circle. In order to embed one bit in a block, the relative angle between both circular histograms' center of mass is modulated. Depending on the bit value to embed in a block, this operation consists in shifting of $\pm\Delta$ the pixel gray values of one pixel sub-block and of $\mp\Delta$ those of the other sub-block. In this work, we apply this robust modulation in order to embed one symbol s_i of the watermark (or equivalently of the message) in each group of tuples, i.e., $\{G^i\}_{i=1,...,N_g}$. Groups are constructed by means of a hash operation applied on the tuple primary key $t_u.PK$ concatenated with a secret key K_s as exposed in section 1.3.2.2:

$$n_u = H(K_S||H(K_S||t_u.PK)) \mod N_g \tag{2.1}$$

Let us consider one group of tuples G^i and A_n be the numerical attribute retained for embedding. The group is equally divided in two sub-groups of tuples $G^{A,i}$ and $G^{B,i}$, following the same principles of tuple grouping. The subgroup membership $n_{u_{sg}}$ of one tuple is given by:

$$n_{u_{sg}} = \begin{cases} G^{A,i} \ if \ \ H(K_S||t_u.PK)mod2 = 0 \\ G^{B,i} \ if \ \ H(K_S||t_u.PK)mod2 = 1 \end{cases} \tag{2.2}$$

Notice that the sub-group index is calculated in a slightly different manner from the group index in order to ensure no mutual correlation between these calculations. By next, the histograms of the attribute A_n in each sub-group $G^{A,i}$ and $G^{B,i}$ are calculated and mapped onto a circle. Then, and as illustrated in figure 2.5(a), the histogram center of mass $C^{A,i}$ (resp. $C^{B,i}$) of the sub-group $G^{A,i}$ (resp. $G^{B,i}$) and its associated vector $V^{A,i}$ (resp. $V^{B,i}$) are calculated. To do so, let us assume the attribute domain of A_n corresponds to the integer range [0,L-1]. Notice that if A_n is a numerical attribute encoded on a fixed number of bits b, then it can take 2^b distinct values. The module and phase of $V^{A,i}$ (resp. $V^{B,i}$) can be calculated from its Cartesian coordinates given by [De Vleeschouwer et al., 2003]:

$$\begin{array}{c} X = \frac{1}{Mass} \sum_{l=0}^{L-1} n_l \cos(\frac{2\pi l}{L}) \\ Y = \frac{1}{Mass} \sum_{l=0}^{L-1} n_l \sin(\frac{2\pi l}{L}) \\ Mass = \sum_{l=0}^{L-1} n_l \end{array} \tag{2.3}$$

where n_l is the cardinality of the circular histogram class l of $G^{A,i}$ (i.e., when A_n takes the value l). From that standpoint, the module of $V^{A,i}$ equals $R = \sqrt{X^2 + Y^2}$ and its phase, we also call mean direction μ, is given by:

$$\mu = \begin{cases} \arctan(Y/X) \ if \ \ X > 0 \\ \frac{\pi}{2} \ if \ \ X = 0, Y > 0 \\ -\frac{\pi}{2} \ if \ \ X = 0, Y < 0 \\ \pi + \arctan(Y/X) \ else \end{cases} \tag{2.4}$$

Let us now consider the embedding of a sequence of bits into the database, or more precisely the insertion of the symbol s=0/1 into G^i. As in [De Vleeschouwer et al., 2003], we modulate the relative angle $\beta_i = (\widehat{V^{A,i}, V^{B,i}}) \simeq 0$ between $V^{A,i}$ and $V^{B,i}$. Depending if we

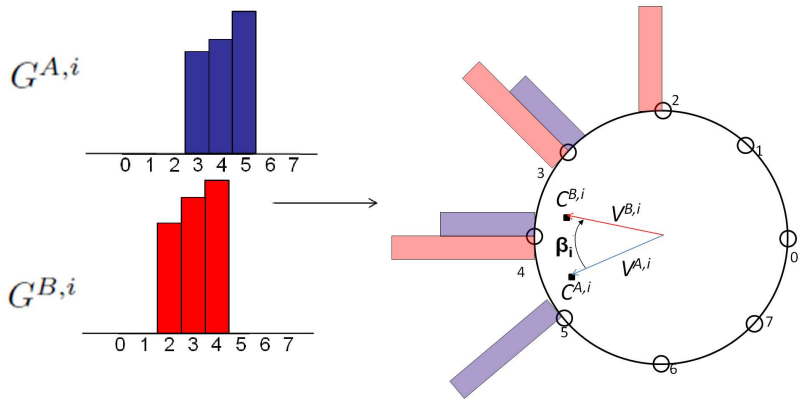
want to insert $s = 0$ or $s = 1$, we change it into its watermarked version β_i^W by rotating the circular histograms of $G^{A,i}$ and $G^{B,i}$ in opposite directions with an angle step α as follows (see figure 2.5(b)):

$$\begin{aligned} \beta_i^W &= \beta_i - 2\alpha \; if \; s = 0 \; (\beta_i^W < 0) \\ \beta_i^W &= \beta_i + 2\alpha \; if \; s = 1 \; (\beta_i^W > 0) \end{aligned} \tag{2.5}$$

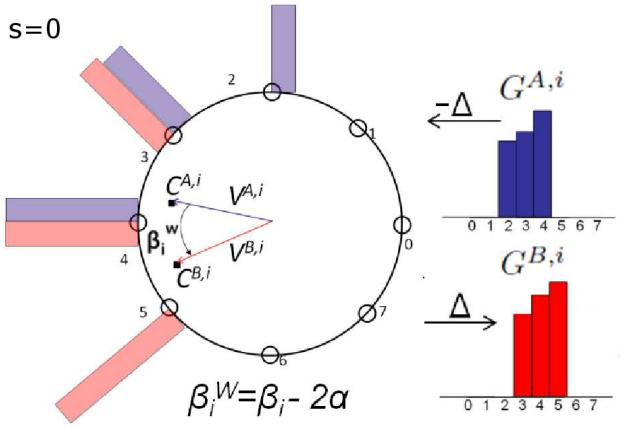
The angle step α is given by:

$$\alpha = \left| \frac{2\pi\Delta}{L} \right| \tag{2.6}$$

where Δ corresponds to the shift amplitude of the histogram (see figure 2.5(b)). More precisely, modifying the angle β_i of $2\alpha(2s-1)$ results in adding $(2s-1)\Delta$ to the attributes of $G^{A,i}$ and $(1-2s)\Delta$ to those of $G^{B,i}$.



(a)



(b)

Figure 2.5: a) 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. b) Embedding of a symbol s=0, corresponding to a rotation of the circular histograms of $G^{A,i}$ and $G^{B,i}$ in opposite directions with an angle step α in order to modify the sign of β_i. This is equivalent to the addition of Δ to the attribute values in $G^{B,i}$ and $-\Delta$ to those of $G^{A,i}$.

At the reading stage, the sign of β_i^W indicates the embedded symbol as well as the direction of rotation to follow so as to invert the insertion process and recover the original value of β_i.

However, at this point, not all of the groups of tuples can convey one symbol of message. In fact and from a more general point of view, we propose to distinguish three classes of groups. In the case $|\beta_i| < 2\alpha$ one can insert $s = 0$ or $s = 1$, as it is possible to swap the position of $V^{A,i}$ and $V^{B,i}$. We refer groups fulfilling this condition as "carrier-groups", due to the fact they can convey one symbol of the watermark. We identify two other kind of groups: "non-carrier groups" and "overflowed groups". They have to be considered separately and handled specifically so as to make the scheme fully reversible. Non-carrier groups are those for which the angle distortion α is not big enough to make change the sign of β_i (see figure 2.6(a)). In order not confusing such non-carriers with carriers at the reading stage, they are modified in the following way (see figure 2.6(a)):

$$\begin{array}{l} \beta_i^W = \beta_i + 2\alpha \; if \; \beta_i > 0 \\ \beta_i^W = \beta_i - 2\alpha \; if \; \beta_i < 0 \end{array} \tag{2.7}$$

In fact, this process results in increasing the angle $\widehat{V^{A,i}, V^{B,i}}$. At the reading stage, these watermarked non-carrier groups are those such as $|\beta_i^W| > 4\alpha$ and can consequently be easily retrieved and differentiated from the watermarked carriers, which belong to the range $[-4\alpha, 4\alpha]$. Thus the reader just has to add or subtract α based on eq. (2.7) so as to restore these watermarked non-carrier groups.

The last situation corresponds to groups of tuples we refer as "overflow-groups". This means groups for which an "angle overflow" may occur if modified. Basically and as exposed in figure 2.6(b), one overflow-group is a non-carrier group which angle $|\beta_i|$ exceeds $\pi - 2\alpha$. If modified according to rules given in eq. (2.7), signs of β_i and β_i^W will be different and the watermark reader will not restore properly the original angle β_i based on eq. (2.7). For instance, if $\beta_i > \pi - 2\alpha$ and $\beta_i > 0$ (see figure 2.6(b)) then adding 2α will lead to $\beta_i^W < 0$. On its side the reader will thus restore the group subtracting 2α instead of -2α. The solution we adopt so as to manage these problematic groups and to make the modulation fully reversible is the following one. At the embedding stage, these groups are left unchanged (i.e., not modified). We inform the reader about the existence of such groups by means of some extra data (a message overhead) inserted along with the message. By doing so, our scheme is blind. Basically, this message overhead avoids the reader confusing overflow groups with non-carriers. It corresponds to a vector O_v of bits stating that watermarked groups such as $\beta_i^W > \pi - 2\alpha$ or $\beta_i^W < -(\pi - 2\alpha)$ are overflow-groups (unmodified) or non-carrier groups (for which the angle has been shifted based on (2.7)). For instance, if $O_v(k) = 1$ then the k^{th} group such as $\beta_i^W > \pi - 2\alpha$ or $\beta_i^W < -(\pi - 2\alpha)$ is a non-carrier group; otherwise it is an overflow-group.

2.3.2 FRAGILE AND ROBUST DATABASE WATERMARKING SCHEMES

Database watermark robustness (resp. fragility) is defined as the ability (resp. inability) to extract/detect the embedded message after an attack such as tuple insertion, tuple deletion or attribute modification. In our approach, if we look at one watermarked group of tuples, this one will be said robust to an attack if it remains in the same class (carrier, non-carrier) while

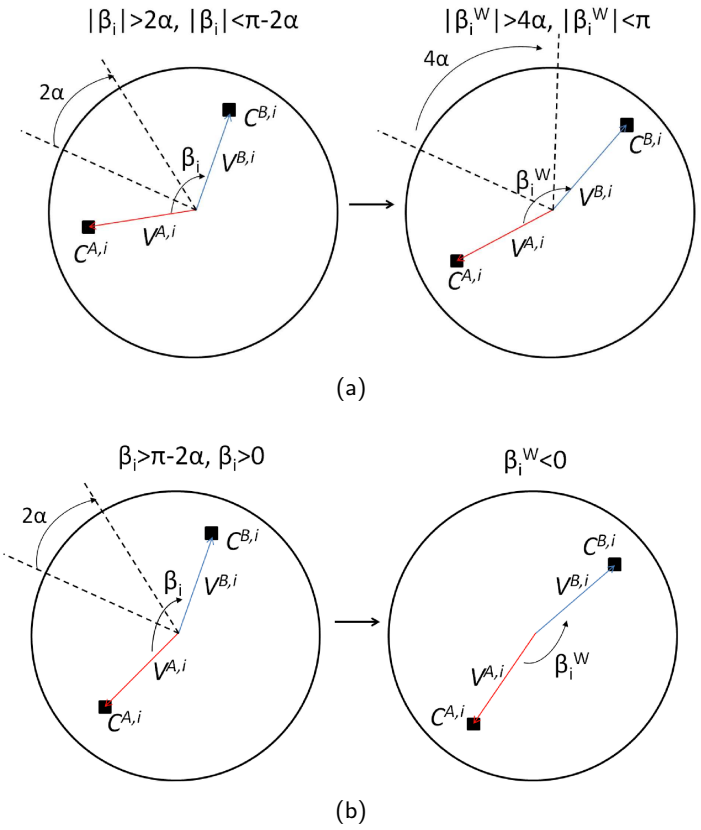


Figure 2.6: Problematic groups: Non-carrier groups and overflow groups (black squares represent circular histogram centers of mass). a) Non-carrier groups are such $|\beta_i| > 2\alpha$ (*on the left*); they are watermarked applying eq. (2.7) (*on the right*). b) Overflow groups are such as $|\beta_i^W| > \pi - 2\alpha$. In the given example $\beta_i^W > \pi - 2\alpha$ (*on the left*); if modified the reader will identify $\beta_i^W < 0$ and will not properly invert eq. (2.7); it will subtract 2α to β_i^W instead of -2α (*on the right*).

encoding the same symbol. In practice, this unlikely happens and three different situations can occur: i) a symbol error, when the symbol embedded into a carrier group changes of value; ii) a carrier loss, when a carrier group becomes a non-carrier one; iii) a carrier injection, when a non-carrier group becomes a carrier. The most harmful scenarios result from carrier injections and deletions. They lead to a loss of synchronization between the embedder and the reader. As example, in the case of a carrier injection, the reader will extract a longer sequence of symbols and, consequently, will face difficulties for interpreting the message. This is the reason why we propose two different reversible schemes. The first one is fragile while the second has been designed so as to be robust to different kinds of attacks.

Our fragile scheme consists in the embedding of a sequence of bits such as $S = \{s_i\}_{i=1,\ldots,N_c}$, $s_i \in \{0, 1\}$, where N_c is the number of available carriers. This sequence includes the message m_s the user wants to insert along with the overhead O_v if necessary (see Section 2.3.1). At the detection, the sequence of bits S is extracted directly from the carrier groups. In an applicative context, m_s may correspond to the digital signature of the database [Li et al., 2004]. At the reception, the recipient just has to compare the extracted signature to the one recomputed

from the restored database so as to decide about the database integrity.

The main problem to solve in building a robust reversible watermarking scheme is to counteract synchronization issues due to carrier injections or erasures. Notice that when De Vleeschouwer *et al.* introduce their robust modulation in [De Vleeschouwer et al., 2003], they do not specify how to manage such a situation. The solution we adopted to overcome these synchronization problems is based on the insertion of two watermark messages S^1 and S^2 of different nature. S^1 is made robust by means of a correlation based detection at the reading stage [Cox et al., 2008]. S^2 is fragile and contains at least the information required so as to ensure the reversibility of the scheme (i.e., it contains the overhead - see Section 2.3.1). To make more clear our proposal, let us describe in details how message embedding is conducted. As illustrated in figure 2.7, S^1 corresponds to a fixed length pseudo-random sequence of N_r symbols such as: $S^1 = \{s_j^1\}_{j=1,...,N_r}$, with $s_j^1 \in \{-1, +1\}$. S^1 is inserted into the N_r first groups of tuples (i.e., $\{G^j\}_{j=1...N_r}$) considering the previous modulation with s=-1/+1 in eq. (2.5). Notice that because all first N_r groups of tuples may not be carriers only (see Section 2.3.1), it is possible S^1 differs from its embedded version $\hat{S}^1$ (even without modifications of the database). Indeed, $\hat{s}_j^1$ will be equal to +1/-1 if the corresponding group G^j is a carrier-group or equal to "null" if it is a non-carrier or overflow group. As stated before, S^1 is detected by means of a correlation measure $C_{S_1} =< \hat{S}^1, S^1 >$, where $\hat{S}^1$ is the sequence of symbols extracted from the N_r first watermarked groups. If C_{S^1} is greater than a decision threshold Tr_{S^1}, S^1 is said present in the database. Extracted null values as well as their corresponding symbols in S^1 are not considered in the correlation measurement (i.e., in C_{S^1}). The second sequence S^2 is inserted like in the previous fragile scheme into the other $N_g - N_r$ groups of tuples. It contains a sequence of bits which encodes the overhead $O_{v_{info}}$ required for reconstructing the whole database ($O_{v_{info}}$ indicates also overflow-groups into the first N_r groups).

As exposed, the robustness of this scheme stands on the fixed length of S^1 which is detected by means of correlation. By doing so, any carrier injections or erasures can simply be considered as a symbol error. However, S^1 needs to be known for the detection process. This solution is rather simple and more complex ones based on error correction codes can be drawn [Sion et al., 2004].

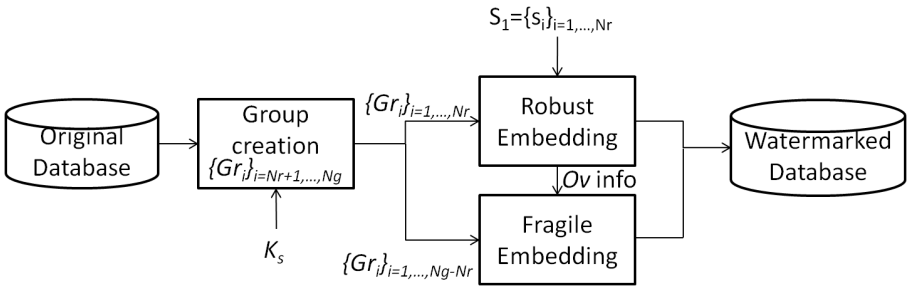


Figure 2.7: Proposed robust scheme. The sequence S^1 is robustly embedded in the first N_r groups while the reconstruction information fills the $N_g - N_r$ remaining groups.

From a more applicative point of view, such a robust-fragile scheme may help to identify the database origin or ownership as well as the recipient with traitor tracing objective. Beyond, if the fragile capacity is large enough, S^2 may convey not only the overhead information $O_{v_{info}}$

but also a message m_s, like a digital signature of the database. To summarize the way this robust-fragile system works, let us consider the process its watermark reader follows:

1. Based on the watermarking key K_s and primary keys, tuples are reorganized into N_g groups.
2. $\hat{S}^1$ is extracted from the N_r first groups. C_{S^1} is computed and if it is greater than Tr_{S^1} the database origins are confirmed.
3. $\hat{S}^2$ is extracted from the carriers of the other $N_g - N_r$ groups. If the database has not been modified, then m_s and O_v are error-free extracted, making it possible to restore the database and then verify its integrity in the case m_s contains the database digital signature. On the contrary, m_s and O_v will be extracted with errors and will inform the user about database integrity loss.

Performance in terms of capacity of the above schemes depend on the number of carrier-groups and overflow-groups. On the other hand, as previously exposed robustness is established based on the probability of symbol error, carrier injection and carrier deletion. We will see in Sect. 2.3.4 that these probabilities rely in part on the number of tuples per group and also on the properties of the numerical attribute retained for message embedding.

2.3.3 LINEAR HISTOGRAM MODIFICATION

β_i rotations can be performed in different ways in the linear domain, i.e on the attribute values. We propose two different strategies depending on the probability distribution of the numerical attribute A_n. Both are equivalent from the perspective of β_i but they allow us to minimize the database distortion.

In the case of numerical attributes of probability distribution centered on its domain range and concentrated around it, we propose to modify groups by adding Δ to the values in $G^{A,i}$ and $-\Delta$ to those in $G^{B,i}$ in order to modify the angle β_i of 2α (inversely for a modification of -2α). The idea is to distribute the distortion onto both groups instead of one and to limit the number of attributes' values jumps between attribute domain range extremities (as example a jump from the value 0 to 7 in figure 2.8). For an attribute range $[0, L-1]$, these jumps represent a modification of $|L - \Delta|$ to the corresponding attribute value.

If now the attribute has its probability density concentrated around one of its domain range extremities, let us say the lower one for example, one must avoid shifting to the left its histogram. Indeed, this will increase jump occurrences and maximize the database distortion. Thus, instead of modifying attribute's values in both $G^{A,i}$ and $G^{B,i}$, we propose to use one of them only, selected according to the sought final sign of β_i, and to shift its attributes values in the opposite direction of the lower domain range by adding them 2Δ. In this way, values in the lower extremity of the domain are never flipped, resulting in a significant reduction of introduced distortion compared to previous strategy. Nevertheless, this second strategy presents a disadvantage as the mean value of the attribute distribution is increased of Δ.

The selection of the most adequate strategy can be performed by the user according to his or her knowledge on the attribute distribution or it can be automatically done depending on the statistical moments of the attribute.

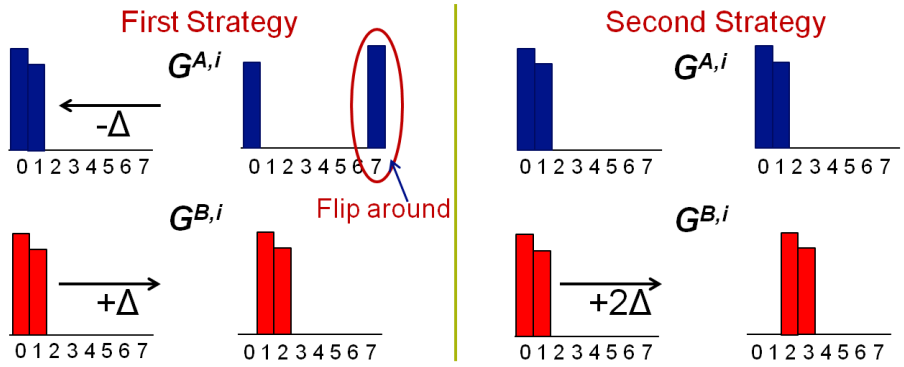


Figure 2.8: Different strategies for linear histogram modification according to the distribution of data. As depicted in the figure, in the case of attributes with probability density concentrated around their lower range extremities, the first proposed strategy results in an important number of value jumps, that can be avoided by means of the second strategy.

2.3.4 THEORETICAL PERFORMANCES

In this section we theoretically evaluate above schemes' performance in terms of capacity and robustness against two most common database attacks. Both depend on the statistical distribution of β_i and on the impacts of database modifications on this random variable.

2.3.4.1 CAPACITY PERFORMANCE

As stated, capacity directly depends on the number of carrier groups, i.e., those for which $|\beta_i| < 2\alpha$ (see section 2.3.1). Capacity can be established once the probability density function (p.d.f) of β_i over the database is known. To do so, let us first recall that β_i is associated to the group of tuples G^i and corresponds to the angle between the centers of mass of two circular histograms of the same attribute A_n considering two subgroups of tuples $G^{A,i}$ and $G^{B,i}$. Because each histogram represents the distribution of the attribute A_n, we can refer to some results issued from circular statistics, a sub-discipline of statistics that deals with data measured by angles or vectors, so as to get the p.d.f of β_i (see books [Mardia and Jupp, 1999] and [Fisher, 1993] as main references).

As a preliminary statement, let us consider the circular data distribution of one attribute θ (i.e., its histogram mapped onto a circle). This can be seen as the p.d.f $f(\theta)$ of a discrete random variable θ which takes L values around the circle, i.e., in the finite set $\{\frac{2\pi l}{L}\}_{l=0,\ldots,L-1}$. The mean direction μ of θ (or equivalently the phase of the vector associated to the center of mass of θ circular histogram) can be estimated based on a finite number of θ samples. Based on the Law of large numbers, it was shown by Fisher and Lewis [Fisher and Lewis, 1983] that for any circular data distribution $f(\theta)$ the difference between the real mean direction and its estimated value tends to zero as the number of samples used in the estimation tends to ∞. With the help of the central limit theorem, they also proved that the distribution of the mean direction estimator approaches a normal distribution centered on the real mean direction of the circular data distribution.

Let us come back now to our problem. By modulating β_i we in fact modulate the angle between two mean directions $\mu^{A,i}$ and $\mu^{B,i}$ of two circular histograms attached to the same attribute A_n within two sub-groups of tuples $G^{A,i}$ and $G^{B,i}$ respectively (see section 2.3.1). $\mu^{A,i}$ (resp. $\mu^{B,i}$) calculated on the sub-group $G^{A,i}$ (resp. $G^{B,i}$) can be seen as the estimator of the mean direction of the attribute A_n (i.e., $\theta = A_n$ in the above) using a number of samples or tuples $\frac{N}{2N_g}$, where (N and N_g are the number of tuples in the database and the number of groups respectively).

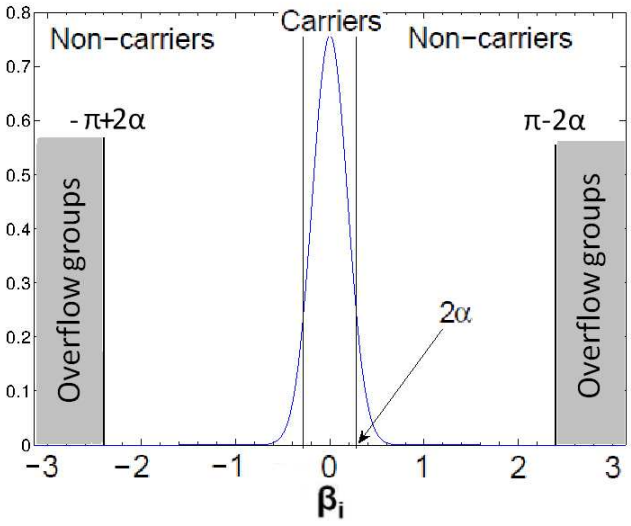


Figure 2.9: β_i distribution

Thus, from the above we can easily state that both $\mu^{A,i}$ and $\mu^{B,i}$ follow a normal distribution. Based on the fact the difference between two normally distributed random variables is also a normally distributed random variable, then $\beta_i = \mu^{A,i} - \mu^{B,i}$ follows a centered normal distribution $\mathcal{N}(0, \sigma^2_{\beta_i})$ where $\sigma^2_{\beta_i}$ corresponds to its variance.

From this standpoint, based on the p.d.f of β_i (see figure 2.9) and for a given angle shift α, the probability a group of tuples is a carrier-group (see section 2.3.1) and is defined as:

$$\mathbb{P}_{carrier} = \Phi(\frac{2\alpha}{\sigma_{\beta_i}}) - \Phi(-\frac{2\alpha}{\sigma_{\beta_i}}) \tag{2.8}$$

where Φ is the cumulative distribution function for a normal distribution, calculated as:

$$\Phi(\frac{2\alpha}{\sigma_{\beta_i}}) = \frac{1}{\sigma_{\beta_i}\sqrt{2\pi}} \int_{-\infty}^{2\alpha} e^{\frac{-t^2}{2\sigma^2_{\beta_i}}}\, dt \tag{2.9}$$

being t an auxiliary random variable. As common convention, we take $\Phi(-\infty) = 0$ and $\Phi(\infty) = 1$.

In practice, considering one numerical attribute, a database of N tuples and N_g groups, one just has to estimate $\sigma^2_{\beta_i}$ to find out the capacity limit of our fragile scheme. To do so,

let us first estimate the variance of the mean directions by means of the estimator proposed in [Quinn, 2010]

$$\sigma^2_{\mu^{A,i}} = \sigma^2_{\mu^{B,i}} = \frac{\sigma_s^2}{\frac{N}{2N_g}R^2} \tag{2.10}$$

where R corresponds to the module of center of mass vector (i.e., $V^{A,i}$, see Section 2.3.1) and σ_s^2 is defined as ([Quinn, 2010]):

$$\sigma_s^2 = \sum_{l=0}^{L-1} \sin^2(\frac{2\pi l}{L}) f(\frac{2\pi l}{L}) \tag{2.11}$$

Notice that the value of R provides an information about the uniformity and the dispersion of the values around the circle. Indeed, for $f(\theta)$ uniform, $R \approx 0$. On the other hand $R \approx 1$ indicates highly concentrated data. The values $\{\frac{2\pi l}{L}\}_{l=0,\ldots,L-1}$ are the bins of the circular histogram attached to the attribute A_n and $f(\frac{2\pi l}{L})$ their corresponding probabilities. Finally, due to the fact β_i results from the difference of two normally distributed random variables $\mu_{A,i}$ and $\mu_{B,i}$, its variance is:

$$\sigma^2_{\beta_i} = \frac{2\sigma_s^2}{\frac{N}{2N_g}R^2} \tag{2.12}$$

Notice that the above normal distribution assumption of β_i is verified in the cases when $\frac{N}{2N_g} \geq 30$ (see [Berenson et al., 2012] for further details). Due to the discrete nature of our circular distribution, the variance estimator can be slightly biased in some cases. In the case of attributes for which $R > 0.85$, i.e., highly concentrated data, we propose to calculate the variance of β_i taking into account the circular variance of the distribution defined by Schmidt [Schmidt, 2012] as $1 - R^2$, resulting in:

$$\sigma^2_{\beta_i} = \frac{2(1-R^2)}{\frac{N}{2N_g}} \tag{2.13}$$

The carrier probability can then be derived from eq. (2.8), and the total amount of bits C_T one may expect to insert into the database is given by

$$C_T = N_g \cdot \mathbb{P}_{carrier} \tag{2.14}$$

In order to get the real capacity, one must subtract to C_T the number of bits used for encoding of the overhead, i.e., $|O_v|$ bits. This number of bits is directly linked to the probability β_i belongs to the range $[-\pi, -\pi + 4\alpha] \bigcup [\pi - 4\alpha, \pi]$. We recall that the overhead is a vector which components indicate by '0' or '1' whether a watermarked angle β_i^w the reader sees in the range $[-\pi, -\pi + 2\alpha] \bigcup [\pi - 2\alpha, \pi]$ has been shifted or not (see end of Section 2.3.1).

$$|O_v| \leq N_g . \mathbb{P}_{[-\pi,-\pi+4\alpha] \bigcup [\pi-4\alpha,\pi]} \tag{2.15}$$

Where

$$\mathbb{P}_{[-\pi,-\pi+4\alpha] \bigcup [\pi-4\alpha,\pi]} = \left[\Phi(\frac{\pi}{\sigma_{\beta_i}}) - \Phi(\frac{\pi - 4\alpha}{\sigma_{\beta_i}})\right] + \left[\Phi(-\frac{(\pi - 4\alpha)}{\sigma_{\beta_i}}) - \Phi(-\frac{\pi}{\sigma_{\beta_i}})\right] \tag{2.16}$$

Finally, the length of the message one may expect to embed is upper bounded such as: $C \leq C_T - |O_v|$

From these results, we can conclude that, for a fixed value of α, the embedding capacity directly depends on the attribute's statistics. By extension, an uniformly distributed attribute will not be watermarkable as $\sigma^2_{\beta_i}$ will tend to ∞ (see eq. (2.12)) and the capacity to 0 (see eq. (2.8)).

2.3.4.2 ROBUSTNESS PERFORMANCE

Let us consider the watermarking of one database of β_i distribution given in figure 2.9, with a fixed angle shift amplitude α and a message constituted of a sequence S of uniformly distributed symbols $s_i \in \{-1, +1\}$ (i.e., like S^1 in section 2.3.2). The resulting distribution of the watermarked angles, i.e., of the random variable β_i^W, over the whole database is given in figure 2.10, where we retrieve the different classes of angles (or equivalently of groups of tuples): non-carriers and carriers (see the modulation rules in section 2.3.1). In such a framework, performance in terms of robustness depend on the probability a group of tuples changes of class (carrier or non-carrier) or of embedded symbol (in case the group is a carrier) after a database attack occurred. We propose to compute these probabilities considering two common database attacks or modifications: tuple addition or tuple removal. To do so, we need to express the impact of such an attack on the p.d.f of β_i^W; p.d.f we need to establish at first.

P.D.F OF β_i^W CLASSES

As depicted in figure 2.10, we propose to distinguish four classes depending if β_i^W is a carrier or a non-carrier and if it has been shifted by $+2\alpha$ or -2α. Notice that from here on, β_i overflow-angles (or equivalently overflow-groups of tuples, see section 2.3.1) are considered as non-carriers; they do not influence message robustness. The p.d.f of each class can be modeled by one truncated normal distribution functions (see Appendix A), where ϕ is the probability density function of the standard normal distribution:

- β_i^W Carriers - β_i shifted of 2α ($c+$ in figure 2.10).
  $f_{c+}(\beta_i^W) = \frac{\frac{1}{\sigma_{\beta_i}} \cdot \phi(\frac{\beta_i^W - 2\alpha}{\sigma_{\beta_i}})}{\Phi(\frac{2\alpha}{\sigma_{\beta_i}}) - \Phi(-\frac{2\alpha}{\sigma_{\beta_i}})}$, $\beta_i^W \in (0, 4\alpha)$
- β_i^W Carriers - β_i shifted of -2α ($c-$ in figure 2.10).
  $f_{c-}(\beta_i^W) = \frac{\frac{1}{\sigma_{\beta_i}} \cdot \phi(\frac{\beta_i^W + 2\alpha}{\sigma_{\beta_i}})}{\Phi(\frac{2\alpha}{\sigma_{\beta_i}}) - \Phi(-\frac{2\alpha}{\sigma_{\beta_i}})}$, $\beta_i^W \in (-4\alpha, 0)$
- β_i^W Non-carriers - β_i shifted of 2α ($nc+$ in figure 2.10).
  $f_{nc+}(\beta_i^W) = \frac{\frac{1}{\sigma_{\beta_i}} \cdot \phi(\frac{\beta_i^W - 2\alpha}{\sigma_{\beta_i}})}{1 - \Phi(\frac{2\alpha}{\sigma_{\beta_i}})}$, $\beta_i^W \in (4\alpha, \pi)$

- β_i^W Non-carriers - β_i shifted of -2α ($nc-$ in figure 2.10).
  $f_{nc-}(\beta_i^W) = \frac{\frac{1}{\sigma_{\beta_i}} \cdot \phi(\frac{\beta_i^W + 2\alpha}{\sigma_{\beta_i}})}{\Phi(\frac{-2\alpha}{\sigma_{\beta_i}})}$, $\beta_i^W \in (-\pi, -4\alpha)$

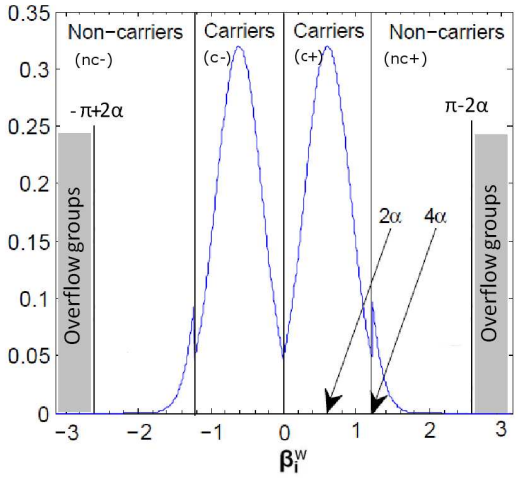


Figure 2.10: β_i^W distribution after the embedding process. We retrieve carrier and non-carrier classes.

DELETION ATTACK

Let us consider the attacker randomly eliminates N_d tuples in a way such that each group G_i loses in average $\frac{N_d}{N_g}$ tuples. In G^i, reducing the number of tuples influences the accuracy of $\mu^{A,i}$ and $\mu^{B,i}$ which are by definition estimators of the mean direction of $G^{A,i}$ and $G^{B,i}$ circular histograms respectively. Considering the whole database, this does not modify the original nature of the p.d.f of $\mu^{A,i}$ and $\mu^{B,i}$ but increase their variance as well as by extension the one of β_i^W. From our knowledge, such a variance increase can be modeled by adding to β_i^w a centered normally distributed random variable ϵ_i such as $\epsilon_i \sim \mathcal{N}(0, \sigma_{\epsilon_i})$. As a consequence the p.d.f of the random variable associated to the attacked watermarked angles β_i^{del}, i.e., $\beta_i^{del} = \beta_i^w + \epsilon$, is obtained after the convolution of each p.d.f of the previous classes with the p.d.f of ϵ_i (see [Turban, 2010]) leading to:

$$f_{c+}(\beta_i^{del}) = \frac{\frac{1}{\sigma_{\beta_i^{del}}} \phi(\frac{\beta_i^{del} - 2\alpha}{\sigma_{\beta_i^{del}}}) \left[\Phi(\frac{\beta_i^{del} - 4\alpha - \kappa_+}{\chi}) - \Phi(\frac{\beta_i^{del} - \kappa_+}{\chi}) \right]}{\Phi(\frac{2\alpha}{\sigma_{\beta_i}}) - \Phi(-\frac{2\alpha}{\sigma_{\beta_i}})}$$

$$f_{c-}(\beta_i^{del}) = \frac{\frac{1}{\sigma_{\beta_i^{del}}} \phi(\frac{\beta_i^{del} + 2\alpha}{\sigma_{\beta_i^{del}}}) \left[\Phi(\frac{\beta_i^{del} - \kappa_-}{\chi}) - \Phi(\frac{\beta_i^{del} + 4\alpha - \kappa_-}{\chi}) \right]}{\Phi(\frac{2\alpha}{\sigma_{\beta_i}}) - \Phi(-\frac{2\alpha}{\sigma_{\beta_i}})}$$

$$f_{nc+}(\beta_i^{del}) = \frac{\frac{1}{\sigma_{\beta_i^{del}}}\phi(\frac{\beta_i^{del}-2\alpha}{\sigma_{\beta_i^{del}}})\left[1-\Phi(\frac{\beta_i^{del}-4\alpha-\kappa_+}{\chi})\right]}{1-\Phi(\frac{2\alpha}{\sigma_{\beta_i}})}$$
$$f_{nc-}(\beta_i^{del}) = \frac{\frac{1}{\sigma_{\beta_i^{del}}}\phi(\frac{\beta_i^{del}+2\alpha}{\sigma_{\beta_i^{del}}})\left[\Phi(\frac{\beta_i^{del}+4\alpha-\kappa_-}{\chi})\right]}{\Phi(\frac{-2\alpha}{\sigma_{\beta_i}})} \quad (2.17)$$

with $\kappa_+ = \frac{\sigma^2_{\epsilon_i}(\beta_i^{del}-2\alpha)}{\sigma^2_{\epsilon_i}+\sigma^2_{\beta_i}}$,$\kappa_- = \frac{\sigma^2_{\epsilon_i}(\beta_i^{del}+2\alpha)}{\sigma^2_{\epsilon_i}+\sigma^2_{\beta_i}}$ and $\chi = \frac{\sigma_{\epsilon_i}\sigma_{\beta_i}}{\sqrt{\sigma^2_{\epsilon_i}+\sigma^2_{\beta_i}}}$.

Considering one real database, the value of $\sigma_{\beta_i^{del}}$ can be derived from eq. (2.12) substituting the sub-group number of tuples $\frac{N}{2N_g}$ by $\frac{(N-N_d)}{2N_g}$ which takes into account the reduction of tuples. Consequently, the value of $\sigma_{\beta_i^{del}}$ is:

$$\sigma^2_{\beta_i^{del}} = \frac{2\sigma_s^2}{R^2\frac{N-N_d}{2N_g}} \quad (2.18)$$

In order to compute the p.d.f $f(\beta_i^{del})$ and by next evaluate the robustness, we also need $\sigma^2_{\epsilon_i}$. Given that the variance of the sum of two independent random variables is the sum of their respective variances, we obtain:

$$\sigma^2_{\epsilon_i} = \sigma^2_{\beta_i^{del}} - \sigma^2_{\beta_i} = \frac{2\sigma_s^2}{R^2\frac{N-N_d}{2N_g}} - \sigma^2_{\beta_i} \quad (2.19)$$

INSERTION ATTACK

Let us consider the attacker inserts N_i tuples and assume the corresponding added attribute values follow the same distribution as the original un-watermarked attribute A_n. Because of the cryptographic hash function used for distributing tuples into groups G^i (see eq. (1.2) in Section 1.3.2.2), we can consider new tuples are uniformly distributed among the groups $\{G^i\}_{i=1,...,N_g}$ as well as in sub-groups $G^{A,i}$ and $G^{B,i}$. As described, such an attack can be modeled by a mixture of two populations: the watermarked tuples and the added un-watermarked tuples with mixture proportions parameters p_1 and p_2 such as $p_2 = 1 - p_1$ with $p_1 = \frac{N}{N+N_i}$.

Furthermore, under the central limit theorem conditions and from the work of Fisher and Lewis [Fisher and Lewis, 1983], the p.d.f of the resulting random variable β_i^{ins} (i.e., the p.d.f of β_i after watermarking and modification) remains normal. The variance and the mean of β_i^{ins} are however proportional to those of the angles β_i^W and angles β_{add} (angles related to the inserted tuples). The mean of β_i^{ins}, obtained with the expected value operator, is then such as

$$\mathrm{E}[\beta_i^{ins}] = p_1\,\mathrm{E}[\beta_i^W] + p_2\,\mathrm{E}[\beta_{add}] = p_1\,\mathrm{E}[\beta_i^W]\,(as\,\mathrm{E}[\beta_{add}] = 0) \quad (2.20)$$

while its variance is given by:

$$\sigma^2_{\beta_i^{ins}} = p_1^2\sigma^2_{\beta_i^W} + p_2^2\sigma^2_{\beta_{add}} \quad (2.21)$$

Values of $\mathrm{E}[\beta_i^W]$ and $\sigma^2_{\beta_i^W}$ depend on the previously defined classes (i.e., *c+*, *c-*, *nc+*, *nc-*) leading to four mean values $\mathrm{E}[\beta^{ins}_{i,c+}]$, $\mathrm{E}[\beta^{ins}_{i,c-}]$, $\mathrm{E}[\beta^{ins}_{i,nc+}]$ and $\mathrm{E}[\beta^{ins}_{i,nc-}]$ and four different variance values $\sigma^2_{\beta^{ins}_{i,c+}}$, $\sigma^2_{\beta^{ins}_{i,c-}}$, $\sigma^2_{\beta^{ins}_{i,nc+}}$ and $\sigma^2_{\beta^{ins}_{i,nc-}}$ which can be obtained with the help of appendix A. As a consequence, the p.d.f of β_{ins} is given per class as follow:

$$\begin{aligned} f_{c+}(\beta_i^{ins}) &= \phi(\frac{\beta_i^{ins} - \mathrm{E}[\beta^{ins}_{i,c+}]}{\sigma^2_{\beta^{ins}_{i,c+}}}) \\ f_{c-}(\beta_i^{ins}) &= \phi(\frac{\beta_i^{ins} - \mathrm{E}[\beta^{ins}_{i,c-}]}{\sigma^2_{\beta^{ins}_{i,c-}}}) \\ f_{nc+}(\beta_i^{ins}) &= \phi(\frac{\beta_i^{ins} - \mathrm{E}[\beta^{ins}_{i,nc+}]}{\sigma^2_{\beta^{ins}_{i,nc+}}}) \\ f_{nc-}(\beta_i^{ins}) &= \phi(\frac{\beta_i^{ins} - \mathrm{E}[\beta^{ins}_{i,nc-}]}{\sigma^2_{\beta^{ins}_{i,nc-}}}) \end{aligned} \tag{2.22}$$

In the case N_i tuples are added to the whole database, one can determine the standard deviation σ_{add} of β_{add} similarly as before:

$$\sigma^2_{\beta_{add}} = \frac{2\sigma_s^2}{(N_a/2N_g)R^2} \tag{2.23}$$

ROBUSTNESS PERFORMANCE - PROBABILITIES OF "ERROR"

The robustness of our scheme is characterized by three situations after an attack occurred:

- "Symbol error", of probability $\mathbb{P}_e$. It concerns carrier groups for which embedded symbols have been changed.
- "Carrier loss", of probability $\mathbb{P}_l$. Such a situation occurs when a carrier-group becomes a non-carrier group, it can be seen as a symbol erasure or deletion.
- "Carrier injection", of probability $\mathbb{P}_i$. This happens when a non-carrier-group turns into a carrier-group, it can also be viewed as a symbol injection.

$\mathbb{P}_e$, $\mathbb{P}_l$ and $\mathbb{P}_i$ can be derived from a hypothesis testing problem with the following set of four hypothesis:

- H_0 corresponds to the case $s_i = -1$, i.e., $\beta_i^W \in c-$.
- H_1 corresponds to the case $s_i = 1$, i.e., $\beta_i^W \in c+$.
- H_2 represents "negative" non-carriers, i.e., $\beta_i^W \in nc-$.
- H_3 represents "positive" non-carriers, i.e., $\beta_i^W \in nc+$.

The probability the watermark reader returns the wrong symbol value while the group remains a carrier-group, i.e., $\mathbb{P}_e$, corresponds to cases where only H_0 and H_1 hypothesis are considered with errors, i.e., with the acceptance of H_0 (resp. H_1) when the correct hypothesis is H_1 (resp. H_0). Thus $\mathbb{P}_e$ is expressed as:

$$\mathbb{P}_e = Pr(H_0)Pr(H_1|H_0) + Pr(H_1)Pr(H_0|H_1) \tag{2.24}$$

Depending on the attack, i.e tuple insertion or removal, P_e can be refined. As example, if the deletion attack is considered then

$$\begin{aligned}\mathbb{P}_e &= Pr(H_0)Pr(4\alpha > \beta_i^{del} > 0|H_0) + Pr(H_1)Pr(-4\alpha < \beta_i^{del} < 0|H_1) \\ &= \mathbb{P}_{carrier}\left(\frac{\int_0^{4\alpha} f_{c-}(\beta_i^{del})d\beta_i^{del} + \int_{-4\alpha}^{0} f_{c+}(\beta_i^{del})d\beta_i^{del}}{2}\right)\end{aligned} \tag{2.25}$$

The probability of carrier loss, i.e., $\mathbb{P}_l$, can be similarly derived and is calculated as:

$$\begin{aligned}\mathbb{P}_l &= Pr(H_0)Pr(H_2|H_0) + Pr(H_1)Pr(H_3|H_1) \\ &= Pr(H_0)Pr(\beta_i^{del} < -4\alpha|H_0) + Pr(H_1)Pr(\beta_i^{del} > 4\alpha|H_1) \\ &= \mathbb{P}_{carrier}\left(\frac{\int_{-\pi}^{-4\alpha} f_{c-}(\beta_i^{del})d\beta_i^{del} + \int_{4\alpha}^{\pi} f_{c+}(\beta_i^{del})d\beta_i^{del}}{2}\right)\end{aligned} \tag{2.26}$$

$\mathbb{P}_i$, i.e., the probability of a carrier injection, is obtained on its side as follows:

$$\begin{aligned}\mathbb{P}_i &= Pr(H_2)Pr(H_0|H_2) + Pr(H_3)Pr(H_1|H_3) \\ &= Pr(H_2)Pr(\beta_i^{del} > -4\alpha|H_2) + Pr(H_3)Pr(\beta_i^{del} < 4\alpha|H_3) \\ &= (1 - \mathbb{P}_{carrier})\left(\frac{\int_{-4\alpha}^{0} f_{nc-}(\beta_i^{del})d\beta_i^{del} + \int_{0}^{4\alpha} f_{nc+}(\beta_i^{del})d\beta_i^{del}}{2}\right)\end{aligned} \tag{2.27}$$

In order to get these probabilities in the case of the tuple insertion attack, one just has to use β_{ins} instead of β_{del} in eq. (2.25), (2.26), (2.27).

2.3.5 EXPERIMENTAL RESULTS

The purpose of this section is to verify the theoretical performance of our system in terms of capacity and robustness in the framework of one real database.

2.3.5.1 EXPERIMENTAL DATASET

The database we use is constituted of one relation of about ten million tuples issued from one real medical database containing pieces of information related to inpatient stays in French hospitals. Only approximately one million are watermarked (i.e., $N = 1048000$). The others will for example serve the tuple insertion attack. In this table, each tuple is represented by fifteen attributes like the hospital identifier (*id_hospital*), the patient stay identifier (*id_stay*),

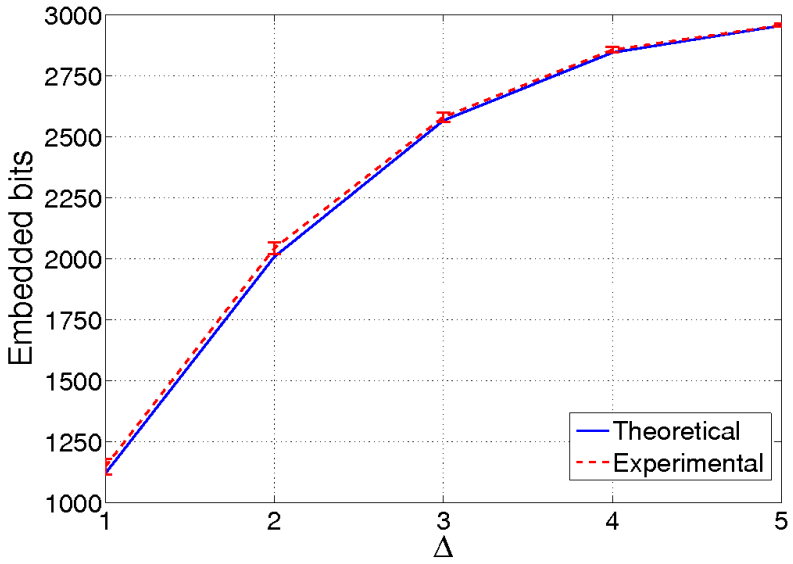


Figure 2.11: Capacity depending on the shift amplitude Δ for *Age* attribute taking 3000 groups.

the patient age (*age*), the stay duration (*dur_stay*) and several other data useful for statistical analysis of hospital activities. In the upcoming experiments, the attributes *id_hospital* and *id_stay* were concatenated and considered as the primary key used by next for tuple groups and subgroups constitution (see eq. (2.1)). Two numerical attributes were considered for the tests and watermarked independently: patient age (*age*) and stay duration (*dur_stay*). The domain of definition of the attribute *age* is the integer range $[0, 110]$ with a mean value of 49.5 and the one of the attribute *dur_stay* $[0, 248]$ with a mean value of 3.9. Notice that both attributes have distinct β_i variances. For instance, in the case of $N_g = 3000$ groups and using eq. (2.12), we have $\sigma^2_{\beta_i} = 0.0456$ for the *age* attribute whereas it is $\sigma^2_{\beta_i} = 3.1188E-5$ for *dur_stay*. For robustness experiments, the message we embed is a random sequence of two symbols '-1'/'+1' uniformly distributed. Furthermore, results are given in average after 30 random simulations with the same parameterization but with different tuples.

2.3.5.2 Capacity Results

In a first time, let us consider the attribute *age* (N occurrences) with a fixed number of groups $N_g = 3000$ and an attribute shift amplitude Δ varying in the range $[1, 5]$. We recall that the angle shift α of β_i depends on Δ (see eq. (2.6)). As it can be seen in figure 2.11, capacity increases along with Δ and verifies the theoretical limit we define in section 2.3.4.1. Obviously, one must also consider that the attribute distortion increases along with the capacity.

In a second experiment, *age* and *dur_stay* were watermarked with $\Delta = 3$ and $\Delta = 1$ respectively while considering varying number of groups such as $N_g \in 500, 700, 1500, 3000$. Notice that the more important the number of groups, the smaller is the number of tuples per groups. As depicted in figure 2.12 for both attributes, obtained capacities fit the theoretical limit we establish in section 2.3.4-A. Given results confirm that the capacity depends on the

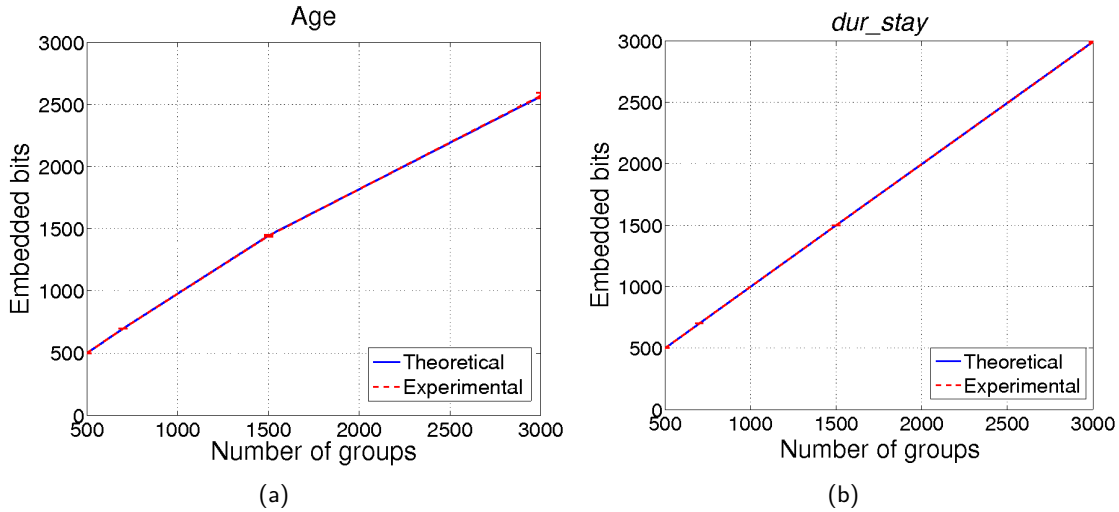


Figure 2.12: a) Capacity for the attribute age considering different number of groups and $\Delta = 3$. b) Capacity for the attribute dur_stay considering different number of groups and $\Delta = 1$.

properties of the attributes considered for embedding and especially of its standard deviation (see section 2.3.4.1). Indeed, we can insert more data within the attribute *dur_stay* which is of smaller variance.

2.3.5.3 Robustness Results

In a first experiment, we were interested in the validation of our theoretical results. Experimental and theoretical results are presented together in Tables 2.1 to 2.3. These tables give the probabilities of error of symbol $\mathbb{P}_e$ (Table 2.1), of carrier loss $\mathbb{P}_l$ (Table 2.2) and of carrier injection $\mathbb{P}_i$ (Table 2.3) considering two attacks, tuples insertion and removal, of various extent. More clearly, considering a database of $N = 10^6$ tuples, between 10% to 50% of tuples were removed or added. Indicated experimental results are given in average accompanied with their standard deviation.

Regarding the probability or error of symbol $\mathbb{P}_e$, it can be seen from Table 2.1, that experimental results are very closed to the theoretical ones we established in Section 2.3.4 whatever the attribute and the attack. This is also the same for $\mathbb{P}_l$, the probability of carrier loss, in the case of the tuple deletion attack (see Table 2.2). However if we look at the injection attack, obtained results for $\mathbb{P}_l$ are slightly different. This may be explained by the fact that, for small number of injected tuples, experiments do not verify the central limit theorem, an hypothesis we made in section 2.3.4.2 when establishing $\mathbb{P}_l$. However, when the number of added tuples per group increases, experimental results tend to fit theoretical ones whatever the attack and attribute. If now we look at the carrier injection probability (or equivalently of symbol injection), $\mathbb{P}_i$, experimental results are again very close to the theory.

From a more general point of view, whatever the attack $\mathbb{P}_l$, $\mathbb{P}_e$ and $\mathbb{P}_i$ increase along with the number of groups as well as with the standard deviation of the attribute. One can also

remark that in the case of the insertion attack, $\mathbb{P}_l$ and $\mathbb{P}_i$ decrease and increase respectively when the number of injected tuples rises up. Nevertheless, from all the above comments, it appears that once the statistical properties of the attribute we want to watermark are known, we are able to estimate the performance of our scheme in terms of capacity and also in terms of robustness for a given attack extent.

Table 2.1: Number of symbol errors (i.e $\mathbb{P}_e \cdot N_g$) for both attributes *age* and *dur_stay*. The table contains theoretical (*Th.*) and experimental results, the latters are given in average (*Avg.*) along with their corresponding standard deviation (*Std*).

		Symbol Errors								
Attack		Attribute *age* with $\Delta = 3$ Nb. groups								
		500			700			1500		
		Th.	Avg.	std	Th.	Avg.	std	Th.	Avg.	std
Deletion	$N_d = 10\%N$	0,1074	0,0667	0,2537	0,7052	0,6000	0,9322	13,5118	14,0333	3,056
	$N_d = 20\%N$	0,2724	0,3000	0,5350	1,5739	1,4333	1,0063	24,9977	25,9000	5,5606
	$N_d = 30\%N$	0,5870	0,3667	0,7184	2,9635	3,0000	1,9827	38,9857	38,9000	6,4077
	$N_d = 50\%N$	2,4157	2,4667	1,5025	9,0949	8,7667	2,9441	80,2505	82,7333	6,4001
Insertion	$N_i = 10\%N$	0,1458	0,0667	0,2537	1,0809	0,7000	0,8367	19,4604	13,8333	3,3330
	$N_i = 20\%N$	0,2432	0,2000	0,4068	1,6012	1,0667	1,0483	25,5356	22,9667	4,7596
	$N_i = 30\%N$	0,3770	0,3333	0,5467	2,2404	1,7333	1,5071	32,0431	29,9333	6,0168
	$N_i = 50\%N$	0,7679	0,5667	0,8172	3,8630	3,9000	1,7489	45,8714	47,5000	8,4761
Attack		Attribute *dur_stay* avec $\Delta = 1$								
Deletion	$N_d = 10\%N$	0	0	0	0	0	0	0,0197	0,0333	0,1826
	$N_d = 20\%N$	0	0	0	0	0	0	0,0636	0,1333	0,3457
	$N_d = 30\%N$	0	0	0	0	0	0	0,1771	0,1667	0,3790
	$N_d = 50\%N$	0	0	0	0,0021	0	0	1,2637	1,7667	1,4547
Insertion	$N_i = 10\%N$	0	0	0	0	0	0	0,0251	0	0
	$N_i = 20\%N$	0	0	0	0	0	0	0,0518	0,0333	0,1826
	$N_i = 30\%N$	0	0	0	0	0	0	0,0962	0,1333	0,3457
	$N_i = 50\%N$	0	0	0	0,0001	0	0	0,2624	0,2667	0,5208

In a second experiment, we evaluated the correlation rate obtained at the detection after different extents of the deletion and insertion attacks. Different values for the distortion amplitude Δ and the number of groups N_g were considered in order to show the existing trade-off between capacity, robustness and imperceptibility defined in our scheme by the values of N_g and Δ. Obtained results are illustrated in figure 2.13 and figure 2.14. As it can be seen, the watermark better resists the performed attacks for higher values of Δ and lower values of N_g. This was the expected result according to the aforementioned trade-off.

We can see that for $\Delta > 1$ and a low number of groups, e.g., $N_g = 300$, our scheme can resist a deletion attack removing 80% of the tuples in the database. The performance is better for the insertion attack, where for a lower value of Δ the obtained correlation is nearly 0.7 when 99% of new tuples were added to the database.

Table 2.2: Number of lost carriers (i.e $\mathbb{P}_l \cdot N_g$) for both attributes *age* and *dur_stay*. The table contains theoretical (*Th.*) and experimental results, the latters are given in average (*Avg.*) along with their corresponding standard deviation (*Std*).

		Lost Carriers								
Attack		Attribute *age* with $\Delta = 3$ Nb. groups								
		500			700			1500		
		Th.	Avg.	std	Th.	Avg.	std	Th.	Avg.	std
Deletion	$N_d = 10\%N$	0,1075	0	0	0,7061	0,8000	0,7611	13,5237	15,4000	3,4998
	$N_d = 20\%N$	0,2727	0,1333	0,3457	1,5755	1,4000	1,1017	25,0130	25,2667	4,9684
	$N_d = 30\%N$	0,5876	0,4667	0,6814	2,9659	3,0000	1,9298	39,0044	39,1333	6,1293
	$N_d = 50\%N$	2,4175	1,9333	1,1725	9,1001	10,1333	3,4614	80,2766	82,6667	9,5424
Insertion	$N_i = 10\%N$	0,0106	0	0	0,1496	0	0	6,0895	1,4667	1,2521
	$N_i = 20\%N$	0,0013	0	0	0,0315	0	0	2,6339	0,5000	0,8200
	$N_i = 30\%N$	0,0002	0	0	0,0064	0	0	1,1334	0,2333	0,5040
	$N_i = 50\%N$	0	0	0	0,0002	0	0	0,2078	0,0333	0,1826
Attack		Attribute *dur_stay* avec $\Delta = 1$								
Deletion	$N_d = 10\%N$	0	0	0	0	0	0	0,0232	0,0667	0,2537
	$N_d = 20\%N$	0	0	0	0	0	0	0,0723	0,1333	0,3457
	$N_d = 30\%N$	0	0	0	0,0001	0	0	0,1973	0,3000	0,6513
	$N_d = 50\%N$	0	0	0	0,0025	0	0	1,3674	1,7333	1,5071
Insertion	$N_i = 10\%N$	0	0	0	0	0	0	0,0007	0	0
	$N_i = 20\%N$	0	0	0	0	0	0	0	0	0
	$N_i = 30\%N$	0	0	0	0	0	0	0	0	0
	$N_i = 50\%N$	0	0	0	0	0	0	0	0	0

2.3.6 COMPARISON WITH RECENT ROBUST LOSSLESS WATERMARKING METHODS

Herein, we compare our approach with [Gupta and Pieprzyk, 2009] and [Farfoura et al., 2013], two recent and efficient robust lossless methods, in terms of distortion, robustness and complexity (see Sect. 2.2). For fair comparison, we have considered their experimental framework in which tuples with real-valued numerical attributes are randomly generated. However, for question of simplicity, only one attribute was considered for embedding. At the same time, [Gupta and Pieprzyk, 2009] and [Farfoura et al., 2013] were slightly modified, without changing intrinsically the strategy they follow, so as to adapt them to a correlation based watermark detection like our robust scheme does while considering a watermark S_1 of 64 bit long (see section 2.3.1). Both methods work at a tuple level, secretly selecting one out of γ tuples for watermarking.

In our experiments, 80000 tuples of one attribute with values following a normal distribution of mean 135 and standard deviation 28.787 were generated. As in the previous section, results are given in average after 30 random simulations. Both [Gupta and Pieprzyk, 2009] and [Farfoura et al., 2013] were parameterized with $\gamma = 7$, i.e., one out of seven tuples is watermarked, while our scheme makes use of all tuples with a shift amplitude $\Delta = 2$. This

Table 2.3: Number of injected carriers (i.e $\mathbb{P}_i \cdot N_g$) for both attributes *age* and *dur_stay*. The table contains theoretical (*Th.*) and experimental results, the latters are given in average (*Avg.*) along with their corresponding standard deviation (*Std*).

		Injected Carriers								
Attack		Attribute *age* with $\Delta = 3$ Nb. groups								
		500			700			1500		
		Th.	Avg.	std	Th.	Avg.	std	Th.	Avg.	std
Deletion	$N_d = 10\%N$	0,0491	0,0333	0	0,4073	0,5333	0,7611	11,7440	12,5333	3,4998
	$N_d = 20\%N$	0,0590	0,0667	0,3457	0,4974	0,7000	1,1017	14,8796	16,4000	4,9684
	$N_d = 30\%N$	0,0649	0,0333	0,6814	0,5525	0,7333	1,9298	16,9244	17,7667	6,1293
	$N_d = 50\%N$	0,0725	0	1,1725	0,6252	0,9667	3,4614	19,7847	20,5000	9,5424
Insertion	$N_i = 10\%N$	0,1429	0,3333	0,5467	1,1838	0,9667	0,9643	28,2523	31,4333	6,3933
	$N_i = 20\%N$	0,1621	0,3000	0,4661	1,4665	1,3667	1,1885	41,7108	45,1333	6,9219
	$N_i = 30\%N$	0,1638	0,3333	0,5467	1,5179	1,4000	1,1626	48,0390	50,2333	7,2048
	$N_i = 50\%N$	0,1640	0,3333	0,5467	1,5278	1,4000	1,1626	51,9391	54,7000	7,0572
Attack		Attribute *dur_stay* avec $\Delta = 1$								
Deletion	$N_d = 10\%N$	0	0	0	0	0	0	0,0061	0	0
	$N_d = 20\%N$	0	0	0	0	0	0	0,0073	0	0
	$N_d = 30\%N$	0	0	0	0	0	0	0,0081	0	0
	$N_d = 50\%N$	0	0	0	0	0	0	0,0090	0	0
Insertion	$N_i = 10\%N$	0	0	0	0	0	0	0,0210	0,0333	0,1826
	$N_i = 20\%N$	0	0	0	0	0	0	0,0221	0,0333	0,1826
	$N_i = 30\%N$	0	0	0	0	0	0	0,0222	0,0333	0,1826
	$N_i = 50\%N$	0	0	0	0	0	0	0,0222	0,0333	0,1826

parametrization was chosen so as to ensure a similar distortion for all algorithms; distortion we evaluate through the variations of the attribute's mean and variance after the insertion process like in [Gupta and Pieprzyk, 2009] and [Farfoura et al., 2013]. Distortion results are given in Table 4.1. It can be seen that these three methods provide closed performance and tend to preserve the attribute's mean and variance.

Table 2.4: Introduced distortion by compared methods in terms of the mean and the variance.

Method	Mean		Std. Deviation	
	Original	Modified	Original	Modified
Farfoura *et al.* [Farfoura et al., 2013]	134.9844	134.9969	28.792	28.798
Gupta and Pieprzyk [Gupta and Pieprzyk, 2009]	134.9838	134.9842	28.776	28.842
Proposed Method	135.0079	135.0048	28.785	28.853

By next, with the same parameterization, three attacks were considered in order to evaluate algorithms' robustness: insertion and suppression of tuples and attribute modification. Insertion and suppression attacks were performed with a percentage of suppressed/added tuples in the range $12.5\% - 87.5\%$. Attribute modification was conducted in two different manners: i) Gaussian noise addition; ii) rounding operation. As depicted in figure 2.15, the three methods

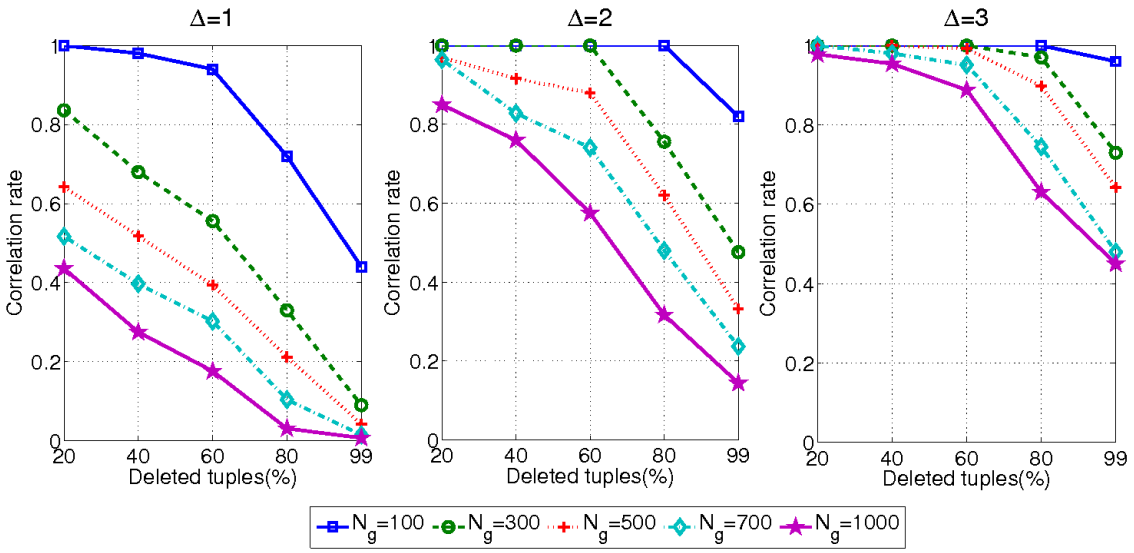


Figure 2.13: Correlation rate for different distortion amplitudes $\Delta = [1, 2, 3]$ and number of groups $N_g = [100, 300, 500, 700, 1000]$ for *Age* attribute considering a tuple deletion attack where $(20\%, 99\%)$ of the tuples in the database were removed.

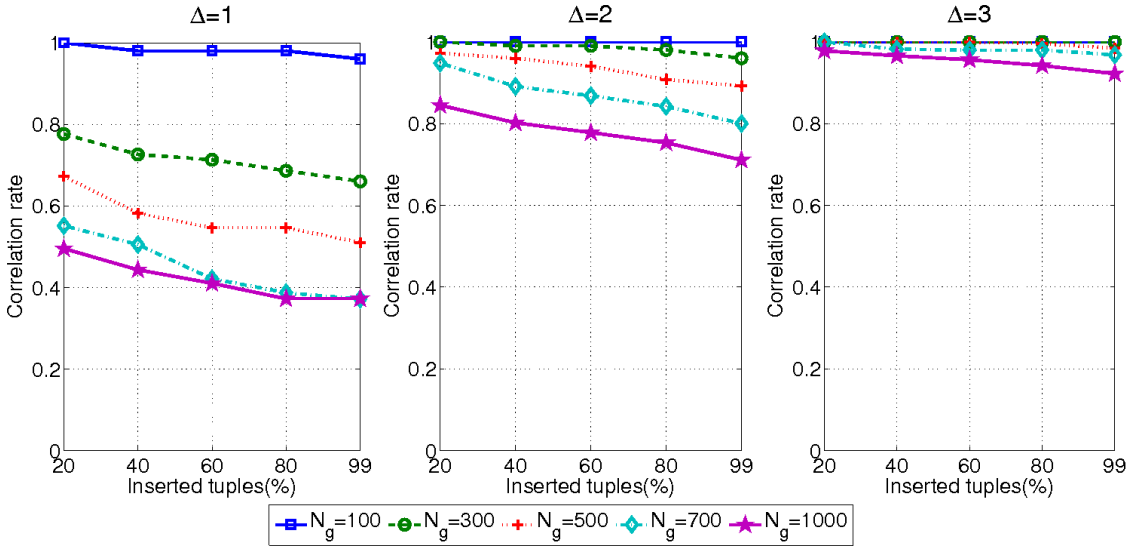


Figure 2.14: Correlation rate for different distortion amplitudes $\Delta = [1, 2, 3]$ and number of groups $N_g = [100, 300, 500, 700, 1000]$ for *Age* attribute considering a tuple insertion attack where $(20\%, 99\%)$ of new tuples were inserted in the database.

have a similar behavior under a tuple insertion attack. They perform well even when the percentage of new tuples is near 90%. In the case of a tuple deletion attack, our method performs worse under strong attack conditions, i.e., when more than 50% of tuples are removed. Nevertheless, it provides better performance than [Gupta and Pieprzyk, 2009] for smaller percentage of tuple removal.

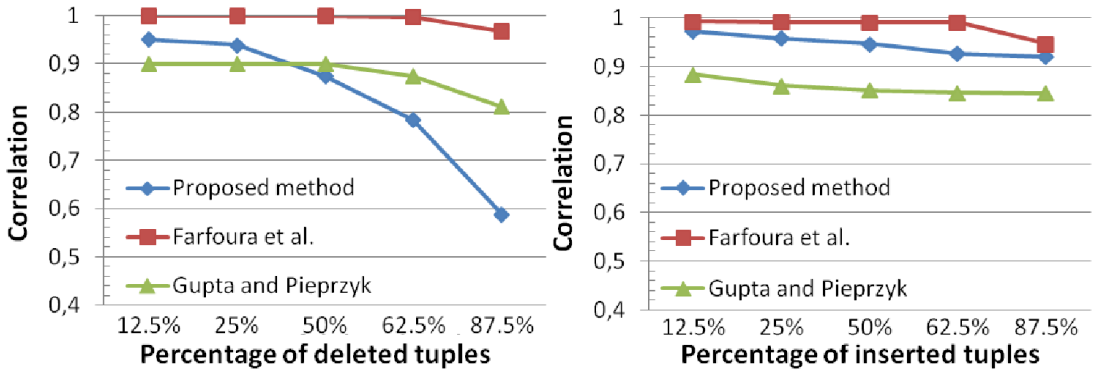


Figure 2.15: Methods' Correlation values in the case of the tuple deletion attack (Left) and the tuple insertion attack (Right) with various intensities.

The first attribute modification attack we applied consists in adding a centered Gaussian noise of standard deviation $\sigma = 0.1$ to all watermarked tuples. Herein, our method performs quite well with correlation values greater than 0.97, compared to [Gupta and Pieprzyk, 2009] which gets values above 0.87. On its side, [Farfoura et al., 2013] did not achieve a correlation greater than 0.53. Under the rounding attack, where values are rounded to the nearest integer, the method of [Farfoura et al., 2013] is inefficient due to the fact it embeds data into the attribute fractional part. That is not case of our scheme and of the one of Gupta and Pieprzyk [Gupta and Pieprzyk, 2009]. They show similar behaviors with correlation values greater than 0.96.

Computation time is used so as to evaluate the complexity of these approaches. [Gupta and Pieprzyk, 2009] and [Farfoura et al., 2013] show similar performance with an embedding/detection process conducted in about 60s. This is quite normal due to the fact they follow a similar strategy. Our method is nearly twice slower. The reason may stand in the histogram calculation for each sub-group of tuples.

To sum up, our approach provides better robustness performance than the scheme of Gupta and Pieprzyk [Gupta and Pieprzyk, 2009] (except for strong deletion attacks), but it is twice slower. The method of Farfoura *et al.* better resists to tuple addition and removal but not to attribute modification attacks. It also introduces new values in the attribute domain. This may limit its application. That is not the case of our scheme and of the method of Gupta and Pieprzyk [Gupta and Pieprzyk, 2009]. On the other hand, both the methods in [Gupta and Pieprzyk, 2009] and [Farfoura et al., 2013] depend on the existance of a fractional part to perform the embedding, which is not the case of our scheme.

2.3.7 Discussion

2.3.7.1 Security of the embedded watermark

In order to ensure the protection of the database, the content of the embedded watermark should be protected from hostile attacks which aim at extracting it. In our scheme, the access

to the embedded message is protected by the secret key K_s used in the construction of the groups and sub-groups of tuples. Indeed, due to the consideration of a cryptographic hash function in this process, any attempt of group and sub-group reconstruction with a different key $K_{s,attacker}$ will provide a completely different distribution of tuples into them. From a statistical point of view and based on the assumptions exposed in section 2.3.4, this results in a distribution of the angles β_i that is normal and centered. Then, for a detection threshold fixed at 0, the probability of retrieving the correct symbol in a group is $\mathbb{P}_{s_i} = \frac{1}{2}$. If we extend this analysis to the whole watermark sequence, we obtain a probability of correct extraction $\mathbb{P}_{extraction} = (\frac{1}{2})^{N_g}$.

In the case of our fragile scheme, where the embedded sequence can contain a meaningful message, another degree of protection can be considered based on encryption. We can consider both a symmetric or asymmetric encryption schemes in order to protect the content of the embedded sequence. In the first case, a secret key K_c unknown by an untrusted third is used for the encryption of the sequence to embed, avoiding the attacker to read the message. In the asymmetric case, the public key of the receiver is used to cypher the message, allowing only he or she to access its content.

Another possible malevolence consists in the embedding of a message by an unauthorized user in order to claim some rights on the database. This situation is harder to prevent once the database is shared. One possible solution we propose consists in the creation of a set of authorized secret keys $\{K_s\}_{s=1,...,N_{users}}$ that can be applied in the group construction process. Then, any mark embedding performed outside this limited set, using for example a secret key $K_{s,attacker}$, will be considered as an attack.

It is important to remark that we are assuming that only authorized users have access to the secret key K_s. As in the case of encryption based systems, the security of our scheme will be seriously compromised if an attacker gets to take over K_s.

2.3.7.2 FALSE POSITIVES

False positive refers to the detection of a random sequence S_{user1} in a non-watermarked database or in a database that contains another embedded sequence S_{user2}. The probability of false detection should be kept as low as possible in order to allow a real embedded watermark to be used as a proof of ownership or authenticity without any possible doubt. The fact that the embedding process is associated with a secret key K_s reduces the probability of detecting a non-embedded sequence. Indeed, as exposed in the previous section, the reconstruction of groups at the detection by means of any secret key that has not been applied in the embedding process leads to a centered normal distribution of the angles β_i. For a fixed detection threshold $Tr_\beta = 0$, this results in a probability of retrieving each symbol $\mathbb{P}_{s_i=1} = \mathbb{P}_{s_i=-1} = \frac{1}{2}$. Considering a correlation based detector, for a mark of length N_g the correlation value is a random variable that follows a centered normal distribution of variance $\sigma^2_{corr} = \frac{1}{N_g}$ (based on the central limit theorem) and consequently, for a given correlation threshold τ_{corr}, we have a probability of false positive $\mathbb{P}_{FP} = erfc(\sqrt{N_g} \cdot \tau_{corr})$. For example, for a mark of length $N_g = 100$ and $\tau_{corr} = 0.7$, the probability of false detection is $\mathbb{P}_{FP} = 4.1838 \cdot 10^{-23}$.

2.4 Conclusion

In this chapter we have focused on the lossless watermarking of relational databases, which can have interesting final security objectives in the context of medical databases: integrity protection, authenticity control or traceability. As exposed, database reversible watermarking presents some specific requirements derived from the fact that the detector should be able to inverse all perturbations introduced by the embedder. More clearly, the insertion of some additional information along with the message may be necessary to inform the detector about the reconstruction process.

Considering these requirements and the existing differences between multimedia and database watermarking (see section 1.3.2.1), we have derived two lossless watermarking schemes, one fragile and one robust, from the method proposed by De Vleeschouwer *et al.* for images [De Vleeschouwer et al., 2003]. They are based on the modulation of the relative angle between the centers of mass of circular histograms associated to secretly constituted groups of values of one numerical attribute of the relation. With the help of circular statistics, we have theoretically evaluated the performance of our method in terms of capacity and robustness against two common attacks: tuple deletion and tuple insertion. As it has been shown, both the capacity and robustness depend on the intrinsic statistical distribution of the selected attribute, on the number of groups and on the introduced distortion. Our theoretical results have been further validated by means of conducted experiments on a real medical database containing more that one million records of inpatient stays in French hospitals in 2011.

Being able to theoretically model the behavior of our scheme allows any user to correctly select our scheme parameters under constraints of capacity, robustness and also distortion, constraints that depend on the application framework (see Chapter 1). Our scheme has also been compared with two recent and efficient schemes, proposed by Gupta and Pieprzyk [Gupta and Pieprzyk, 2009] and Farfoura *et al.* [Farfoura et al., 2013], so as to prove the gain of performance our scheme provides, as well as to expose its disadvantages.

To complete the analysis of our schemes, we have also discussed its security. This latter fundamentally lies on the knowledge of a secret key K_s that parameterized the construction of groups and sub-groups of tuples. The lack of information about K_s avoids the attacker to access and modify the embedded message and the fact that the insertion of one message is linked to one key K_s makes our scheme ensure a low false detection probability.

As seen, being able to robustly and reversibly watermark a relational database is of major concern but such a solution can be improved, especially in terms of data distortion and also depending on the application framework. Two aspects we study in the next chapters.

Chapter 3

Traceability of medical databases in shared data warehouses

In health care, relational databases are commonly shared and remotely accessed for different purposes going from patient cares to medico-economic evaluations or clinical research. This explain in part why professionals and administrations are more and more interested in the storage of medical relational data in shared data warehouses or even in the cloud. This is for example the purpose of the MedRed BT Health Cloud (MBHC), a collaborative project between the United States and the United Kingdom [Brino, 2013].

In terms of security, traceability takes a particular importance in this context. Databases can be rerouted or redistributed without permission. The capacity of tracing data sets merged into one unique database presents interesting perspectives with the purpose of identifying the different contributors or data providers as well as localizing the origin of a data leak. Herein, one may want to insert an identifier into the database before uploading and merging it with some others. In this chapter, we expose how the lossless watermarking scheme we previously presented can be applied in the identification of one base in a data warehouse. Our idea being to prove someone has provided the data or that the recipient have rerouted them. In order to achieve this goal, in a first moment, we considered the use of anti-collusion codes, originally proposed for multimedia traceability applications. These codes have been specifically conceived to resist collusion attacks in which several users associate themselves by mixing their copies so as to make their identfiers undetectable [Tardos, 2003] [Nuida et al., 2009]. Although we pursue a similar objective, we will see that the differences between our database mixture scenario and the one of a collusion attack make anti-collusion codes ineffective.

Nevertheless, we will take this setback as the base to construct an optimized detection scheme. More specifically, we suggest a detection process based on scores rather than the classical correlation based detector as proposed in chapter 2. These scores are defined according to a theoretical analysis we made so as to model the impact of database mixtures on the embedded message. This analysis constitutes the first part of this chapter. Then, we explain the differences between usual anti-collusion codes applications and our particular scenario, exposing the reasons that make them ineffective. The proposed optimizations of our detection process are derived from this model. As in chapter 2, we experimentally verify our theoretical results in the case of a mixture of several real medical databases. Obtained detection rates by means of correlation and the detection scores are compared showing the gain of performance.

3.1 MEDICAL SHARED DATA WAREHOUSE SCENARIO

As stated, medical institutions constitute data warehouses so as to offer professionals an access to more and more data, opening new perspectives in patient care, resource management and research, and also with the objective to reduce information system costs.

In this scenario, different services of a hospital (e.g., cardiology, emergency, admissions, ...) or different hospitals share their databases. Such a framework is illustrated in figure 3.1. Notice that by constituting an inside data warehouse, a hospital will better manage its data and also facilitate data outsourcing with inter-operable systems. This shared access to mutualized data can serve different purposes, like epidemiological studies.

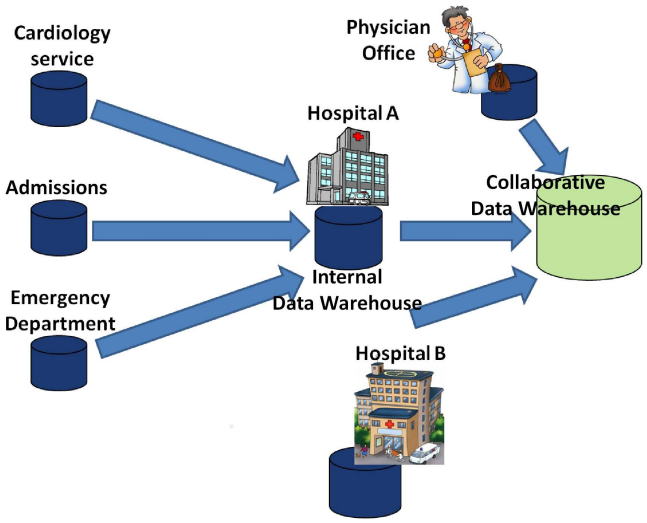


Figure 3.1: Example of shared storage inside a hospital (internal data warehouse) and between different health institutions (collaborative data warehouse) where several databases are stored into a shared data warehouse.

Several real examples of such a scenario can be found. One example is given by the Northwestern Medicine® enterprise data warehouse (NMEDW) based in Chicago, which integrates all clinical and research data from all patients receiving treatment through Northwestern healthcare affiliates [NUCTSI, 2011]. A second example corresponds to the collaborative project MedRed BT Health Cloud (MBHC) in-between the US and UK governments and deployed by MedRed and BT. The purpose of this project is to improve the quality of care, enhance patient safety and the speed of innovative drugs and medical technologies development [MedRed, 2013]. A last example concerns the US National Institute of Health (NIH) which supports several shared data repositories intended to make research data reusable [NIH, 2014].

To sum up, in this context, we are interested in the traceability of shared databases, that is to say, identifying the presence of a database, i.e, identifying the recipients who rerouted the data in case of an information leak.

3.2 THEORETICAL ANALYSIS

Let us consider two databases DB_1 and DB_2. For sake of simplicity, each of them is composed of one single relation containing N_1 and N_2 tuples respectively, with the same attribute set $\{A_1, ..., A_M\}$. Let us also consider that both of them are watermarked by means of the scheme studied in chapter 2, considering N_g groups of tuples and an embedding distortion Δ. The same attribute A_t for watermarking but different secret keys K_S^1 and K_S^2 for group and subgroup construction. Another hypothesis we consider is that A_t follows the same statistical distribution $f(l)$ in both databases. Watermarked versions of DB_1 and DB_2, i.e., DB_1^W and DB_2^W, are uploaded into a single relation of a shared data warehouse.

The objective we pursue in this section is to identify DB_1 when it is mixed with DB_2 (and *vice versa*). To perform this task and optimize the detection process, the question to answer is: Is it possible to model the impact of the aggregation of DB_1^W and DB_2^W over the detection of a watermark in DB_1^W with the only knowledge of the statistical parameters of A_t?

3.2.1 STATISTICAL DISTRIBUTION OF THE MIXTURE OF TWO DATABASES

From the "point of view" of DB_1, i.e., applying K_S^1 for group and subgroup construction at the reader, the tuples of DB_2 represent an added noise, the statistical distribution of which can be determined. This is due to the way groups and subgroups of tuples are constructed and which is based on a cryptographic hash function, parameterized with two distinct keys K_S^1 and K_S^2. So, based on the distribution of A_t and the one of its watermarked version, we propose to model the noise DB_2 will represent onto the watermark contained by DB_1.

By definition, under the previous assumption, the angle β_i between the mean directions of the circular histograms of A_t within the subgroups $G^{A,i}$ and $G^{B,i}$ in DB_2 follows a centered normal distribution $\mathcal{N}(0, \sigma^2_{\beta_{i,W}})$, the variance $\sigma^2_{\beta_{i,W}}$ of which can be calculated as:

$$\sigma^2_{\beta_{i,W}} = 2\frac{\sigma^2_{s,W}}{\frac{N_2}{2N_g}R^2_W} \tag{3.1}$$

We need then to calculate $\sigma^2_{s,W}$ and R^2_W. The value $\sigma^2_{s,W}$ is calculated as (see section 2.3.4.1):

$$\sigma^2_{s,W} = \sum_{l=0}^{L-1} \sin^2\left(\frac{2\pi l}{L}\right) f_W\left(\frac{2\pi l}{L}\right) \tag{3.2}$$

where $f_W\left(\frac{2\pi l}{L}\right)$ corresponds to the circular distribution of the watermarked attribute A_t in DB_2. As exposed in section 2.3.1, during the watermark embedding process, values of A_t in the tuples of DB_2 may be modulated in three different manners: addition of Δ, addition of $-\Delta$ (carriers and non-carriers) or no modification at all (overflow groups). The final distribution of the modulated attribute A_t can be modeled by a mixture of these three populations with mixture population parameters P_Δ, $P_{-\Delta}$ (probabilities of modifying of $\pm\Delta$) and P_{ov} with $P_\Delta = P_{-\Delta}$ and $P_{ov} = 1 - (P_\Delta + P_{-\Delta})$. Notice that P_{ov} can be calculated as exposed

in section 2.3.4.1. Assuming that A_t takes L values, i.e in the finite set $\{l\}_{l=0,...,L-1}$, the probability density function $f_W(l)$ of watermarked attribute's values for a given distortion Δ is given by:

$$f_W(l) = P_\Delta f((l+\Delta) \mod L) + P_{-\Delta} f(l-\Delta) \mod L) + P_{ov} f(l) \tag{3.3}$$

where the modulo function is considered so as to represent the circular shifting of the attribute's histogram.

In the circular domain, i.e., when this histogram is mapped onto a circle, the circular probability density function of A_t is:

$$f_W\left(\frac{2\pi l}{L}\right) = P_\Delta f\left(\frac{2\pi[(l+\Delta) \mod L]}{L}\right) + P_{-\Delta} f\left(\frac{2\pi[(l-\Delta) \mod L]}{L}\right) + P_{ov} f\left(\frac{2\pi l}{L}\right) \tag{3.4}$$

Due to the 2π-periodicity of the sin function, for any function $g : [0, L-1] \to [0, 1]$ and any integer $a \in [0, L-1]$, we have:

$$\sum_{l=0}^{L-1} sin^2\left(\frac{2\pi l}{L}\right)[g(l+a) mod L] = \sum_{l=0}^{L-1} \sin^2\left(\frac{2\pi(l-a)}{L}\right) g(l) \tag{3.5}$$

From (3.2), given the result in (3.5), we obtain after some trigonometric operations (see Appendix B.1):

$$\sigma^2_{s,W} = \left[P_{ov} + (1-P_{ov})\cos\left(\frac{4\pi\Delta}{L}\right)\right]\sigma_s^2 + (1-P_{ov})\sin^2\left(\frac{2\pi\Delta}{L}\right) \tag{3.6}$$

Once we have obtained $\sigma^2_{s,W}$, the value of R^2_W, i.e., the squared module of the center of mass vector of the watermarked attribute A_t in DB_2, can be established as follows. First, it is useful to realize that as DB_2 has been watermarked and A_t in DB_2^W corresponds to the mixture of three populations, we can consider that the center of mass vector $V_W^{A,i}$ (resp. $V_W^{B,i}$) is a linear combination of three vectors, each of which is associated to one of the populations. These vectors correspond to rotated versions of the original center of mass vector $V^{A,i}$ (resp. $V^{B,i}$). Then, considering the rotation matrix Ω (see Appendix B.2), the vector $V_W^{A,i}$ (resp. $V_W^{B,i}$) is calculated as:

$$V_W^{A,i} = V^{A,i}\left[P_\Delta \Omega\left(\frac{2\pi\Delta}{L}\right) + P_{-\Delta}\Omega\left(-\frac{2\pi\Delta}{L}\right) + P_{ov}\right] = V^{A,i}\left[P_{ov} + (1-P_{ov})\cos\left(\frac{2\pi\Delta}{L}\right)\right] \tag{3.7}$$

Consequently, the squared module R^2_W of the modified center of mass vector $V_W^{A,i}$ (resp. $V_W^{B,i}$) is:

$$R^2_W = ||V_W^{A,i}||^2 = ||V_W^{B,i}||^2 = \left[P_{ov} + (1-P_{ov})\cos\left(\frac{2\pi\Delta}{L}\right)\right]^2 \cdot R^2 \tag{3.8}$$

Applying (3.6) and (3.8) into (3.1), the resulting value for $\sigma^2_{\beta_{i,W}}$ is then such as:

$$\sigma^2_{\beta_{i,W}} = 2\frac{\left[P_{ov} + (1-P_{ov})\cos\left(\frac{4\pi\Delta}{L}\right)\right]\sigma_s^2 + (1-P_{ov})\sin^2\left(\frac{2\pi l}{L}\right)}{\frac{N_2}{2N_g}\left[P_{ov} + (1-P_{ov})\cos\left(\frac{2\pi\Delta}{L}\right)\right]^2 \cdot R^2} \tag{3.9}$$

Notice that, in practice, the values of σ_s^2 and R employed in this calculation can be estimated from the tuples in DB_1 as exposed in 2.3.4.1.

3.2.2 EXTENSION TO SEVERAL DATABASES

Let us extend our previous analysis to the case in which DB_1 is merged with a set of U different databases, constituting a new set U'. Each of these databases contains $\{N_i\}_{i=2,...,|\mathsf{U}|+1}$ tuples, where |U| represents the cardinality of U, with the same attribute set $\{A_1, ..., A_M\}$ and is watermarked with a different secret key. From the "point of view" of DB_1, we can consider the merge of these set of bases results in the mixture of tuple populations. The mixture proportions can be easily calculated as:

$$P_{DB_i} = \frac{N_i}{\sum_{k=2}^{|\mathsf{U}|+1} N_k}, \ \forall i \in [2, |\mathsf{U}| + 1] \tag{3.10}$$

The probability density function of attribute's values after this mixture can be defined by means of the previous proportions, corresponding to:

$$f_{mix}(l) = \sum_{k=2}^{|\mathsf{U}|+1} P_{DB_k} f_{W_k}(l) \tag{3.11}$$

where $f_{W_k}(l)$ is the probability distribution of the watermarked attribute A_t in each database k.

From this standpoint, we can linearly derive the parameters $\sigma^2_{s,mix}$ and R^2_{mix}:

$$\sigma^2_{s,mix} = \sum_{k=2}^{|\mathsf{U}|+1} P_{DB_k} \sigma^2_{s,W_k} \tag{3.12}$$

$$R^2_{mix} = \sum_{k=2}^{|\mathsf{U}|+1} P_{DB_k} R^2_{W_k} \tag{3.13}$$

where σ^2_{s,W_k} and $R^2_{W_k}$ are calculated for each database as exposed in the previous section.

Then, β_i in the groups of tuples created by means of K^1_S in the set of U databases follows a centered normal distribution $\mathcal{N}(0, \sigma^2_{\beta^{mix}_i})$ with:

$$\sigma^2_{\beta^{mix}_i} = 2 \frac{\sigma^2_{s,mix}}{\frac{\sum_{k=2}^{|\mathsf{U}|+1} N_k}{2N_g} R^2_{mix}} \tag{3.14}$$

In case the same parameters were used for the watermarking of all the databases in the mixture, it is not necessary to know *a priori* the number of databases in the warehouse. The mixture can be considered as a single database of $\sum_{k=2}^{|\mathsf{U}|+1} N_k$ tuples watermarked with a distortion amplitude Δ.

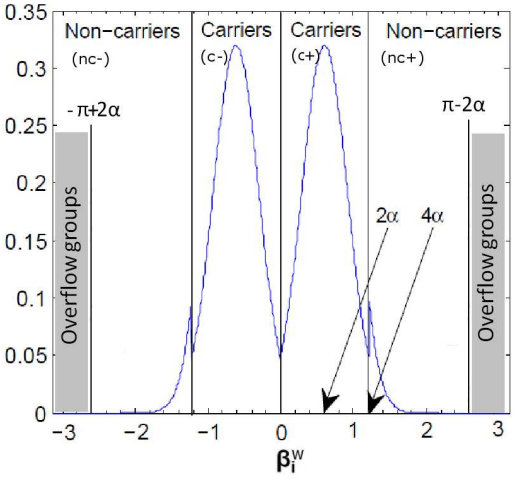


Figure 3.2: β_i^W distribution in DB_1 after the embedding process and before being merged with the other databases. We retrieve carrier, i.e., c+ and c-, and non-carrier classes, i.e., nc+ and nc-.

3.2.3 EFFECT ON THE DETECTION

Once the statistical distribution of the angles β_i calculated over the set U of mixed databases, the impact of this mixture on the detection of the watermark embedded in DB_1 can be modeled. To do so, we are going to consider a similar approach than the one defined in section 2.3.4.2 about the tuple insertion attack, attack where added tuples are non-watermarked tuples.

The distribution of β_i in the final database, or more precisely into the groups of tuples based on the secret key used to watermark DB_1, is closed to the one we have seen in chapter 2, section 2.3.4.2, except that is perturbed by the noise induced by the watermark tuples of the set U. As illustrated in figure 3.2, this density probability function can be seen as constituted of four classes: positive carriers (c+), negative carriers (c-), positive non-carriers (nc+) and negative non-carriers (nc-). For sake of simplicity, let us consider that all the constructed groups of DB_1 are carriers, that is to say that the p.d.f of the watermarked angle of DB_1^w, i.e., β_i^w, is only constituted of the two classes $c+$ and $c-$:

$$f_{\beta_i^w}(\beta_i^w) = \frac{1}{2} f_{\beta_i^w}(\beta_i^w | s_i = -1) + \frac{1}{2} f_{\beta_i^w}(\beta_i^w | s_i = +1) \tag{3.15}$$

where the angle β_i^w in the $c-$ and $c+$ is distributed according to the conditional density functions given the embedded symbol s_i. Under the hypothesis that symbols are uniformly distributed, these conditional distributions are:

$$\begin{array}{c} f_{\beta_i^w}(\beta_i^w | s_i = -1) \sim \mathcal{N}(-\frac{4\pi\Delta}{L}, \sigma^2_{\beta_i}) \\ f_{\beta_i^w}(\beta_i^w | s_i = +1) \sim \mathcal{N}(\frac{4\pi\Delta}{L}, \sigma^2_{\beta_i}) \end{array} \tag{3.16}$$

As in section 2.3.4.2, the conditional density functions of the resulting β_i^{merge} after the insertion of tuples remain normal. However, their mean and variance are modified proportionally

to those of β_i^w and β_i^{mix}. If we define the mixture proportions for DB_1, $P_{DB_1} = \frac{N_1}{N_1 + \sum_{k=2}^{|\mathsf{U}|+1} N_k}$ and for the databases in the set U, $p_{mix} = 1 - P_{DB_1}$, the mean and variance of β_i^{merge} are such as:

$$\begin{aligned} \mathrm{E}[\beta_i^{merge}|s_i = -1] &= P_{DB_1}\,\mathrm{E}[\beta_i^w|s_i = -1] + P_{mix}\,\mathrm{E}[\beta_i^{mix}] = P_{DB_1}\left(-\tfrac{4\pi\Delta}{L}\right) \\ \mathrm{E}[\beta_i^{merge}|s_i = +1] &= P_{DB_1}\,\mathrm{E}[\beta_i^w|s_i = +1] + P_{mix}\,\mathrm{E}[\beta_i^{mix}] = P_{DB_1}\left(\tfrac{4\pi\Delta}{L}\right) \end{aligned} \tag{3.17}$$

and

$$\sigma^2_{\beta_i^{merge}} = P^2_{DB_1}\sigma^2_{\beta_i^w} + P^2_{mix}\sigma^2_{\beta_i^{mix}} \tag{3.18}$$

As a consequence, the p.d.f of β_i^{merge} is given per class as follow:

$$\begin{aligned} f_{c-}(\beta_i^{merge}) &= f_{\beta_i^{merge}}(\beta_i^{merge}|s_i = -1) \sim \mathcal{N}(P_{DB_1}(-\frac{4\pi\Delta}{L}), \sigma^2_{\beta_i^{merge}}) \\ f_{c+}(\beta_i^{merge}) &= f_{\beta_i^{merge}}(\beta_i^{merge}|s_i = +1) \sim \mathcal{N}(P_{DB_1}(\frac{4\pi\Delta}{L}), \sigma^2_{\beta_i^{merge}}) \end{aligned} \tag{3.19}$$

Knowing the p.d.f of β_i^{merge} is very helpful so as to evaluate the performance of our system. Indeed, as we have seen, aiming at detecting a watermarked database within a mixture of watermarked databases is a problem very close to the tuple addition attack we have presented in chapter 2. For a given mixture hypothesis one can parameterize our algorithm so as to ensure watermark detection. To do so, he just has to use β_i^{merge} in error and erasure probabilities we computed in chapter 2, section 2.3.4.2. To go beyond, the question to answer now is how can we optimize the detection of one watermarked database within such a mixture. In order to achieve this goal, we studied different solutions, the first one being the use of fingerprinting codes, the second soft decoding.

3.3 IDENTIFICATION OF A DATABASE IN A MIXTURE

As exposed, we are interested in the traceability of databases merged with some others, as in collaborative data warehouses, for example. Our idea is to be able to identify that some parts of a database are issued from specific sources that were not allowed to communicate (i.e., case of an information leak). To do so, and as exposed in Figure 3.3, we propose to embed of a sequence of symbol or at least of bits that identifies each database by means of the robust lossless scheme presented in the previous chapter. More clearly, for a database j, a set of N_g groups of tuples is constructed by means of a secret key K_S^j. Then, each symbol of a uniformly distributed identification sequence $S^j = \{s_1^j, ..., s_{N_g}^j\}$, with $s_i^j \in \{-1, 1\}$, is embedded into a group of tuples considering an embedding distortion Δ. A graphical representation of the embedding stage for the proposed solution is illustrated in figure 3.3. Once these databases are uploaded to a warehouse, their tuples are mixed.

At the detection, in order to identify a database j, the groups of tuples are reconstructed over the whole mixed database by means of the secret key K_S^j. In order to determine if the

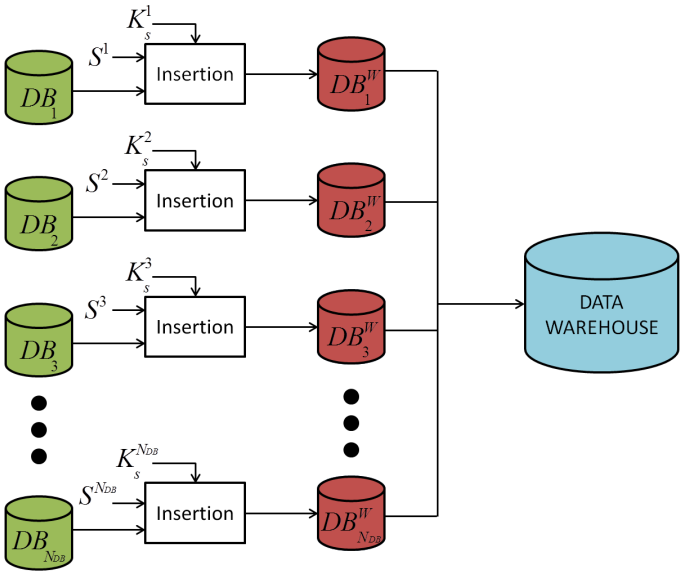


Figure 3.3: Embedding stage for the proposed solution. Each database is identified by means of a sequence S^j, the embedding of which is performed by means of a secret key K_s^j.

database j is present in the data warehouse, we seek to detect the sequence S^j. From here on, the question to answer is how to optimize the detection performance of system.

A largely analyzed related problem or similar scenario to our problem is traitor tracing, a well-known example of which corresponds to the video on demand framework, where one multimedia content is distributed to several users. One common traitor tracing solution stands on the use of fingerprinting codes, or more precisely anti-collusion codes inserted into the document to share. These codes are conceived and studied so as to resist to collusion attacks in which some dishonest users owning different copies of a same content cooperate by combining their copies with the aim of constructing a "pirated" copy, copy in which the embedded identifiers have been erased. In this section, we briefly introduce anti-collusion codes and we expose why they can not be directly applied in our case of interest. We then expose in the second part of this section how the detection of the watermark can however be optimized in order to improve identification results.

3.3.1 ANTI-COLLUSION CODES

Anti-collusion codes are a fundamental tool in fingerprinting applications. Let us come back on the classic example of video on demand scenario in which several buyers obtained individual copies of a same content. Each user is uniquely identified by means of sequence of m symbols embedded into the host content. This result in a set of different, but perceptually similar watermarked video copies. Following this strategy allows to identify a dishonest user who decides to redistribute a copy. In this framework, a collusion attacks consists of c users who cooperate, mixing their copies and consequently their identifiers. More clearly, at the detection or verification stage, one will not extract an individual identifier but a mixture of them.

To overcome this issue, different anti-collusion codes were conceived. All of them make a strong hypothesis in the construction of the pirate copy called the marking assumption. This assumption is intimately linked to the kind of data these codes have been devoted for, it means videos or images. Assuming that such a content is divided into blocks and that one symbol of the message, i.e., anti-collusion code, is embedded per block, the marking assumption states the following: *for one block b containing the same symbol S_b in all the merged copies, the resulting symbol in the pirate version of b is necessarily equal to S_b*. Figure 3.4 illustrates this hypothesis by means of an example where three copies of a document X is merged into a single document Y..

X_1 : **0** 0 0 0 1 1 1 **1**
X_2 : **0** 0 1 1 0 0 1 **1**
X_3 : **0** 1 0 1 0 1 0 **1**
Y : **0** 1 1 0 1 0 0 **1**

Figure 3.4: Example of a collusion between three users with identifiers X_1, X_2 and X_3 leading to the sequence Y at the detection stage. As stated by the marking assumption, positions 1 and 8 in the sequence are undetectable, as all the users share the same symbols.

Several approaches have been proposed that allow the identification of users having participated in a collusion. The construction of anti-collusion codes by concatenation of error correcting codes was initially considered [Boneh and Shaw, 1996] [Schaathun, 2008]). Even though these codes offered interesting properties, they were difficult to apply in practice because of the length of the codewords. Probabilistic fingerprinting codes were introduced in the seminal work by Tardos [Tardos, 2003]. Tardos codes, specifically constructed for fingerprinting, required a shorter length of the codewords with typically a length near the lower threshold $m = O(c^2 log(\frac{1}{\epsilon_1}))$, where c is the number of colluders that the code is prepared to detect and ϵ_1 is the false detection probability, i.e., probability of accusing an innocent user. Let us consider the binary case, a Tardos based fingerprinting scheme contains three main steps:

Initialization: A sequence of real numbers $p = (p_1, p_2, ..., p_m)$ (with $p_i \in [t, 1-t]$ and $0 < t < 1$) is obtained, where $p_i = \sin^2 r_i$ and r_i is a random value of uniform distribution in $[t', \pi/2 - t']$ with $0 < t' < \pi/4$. This sequence corresponds to the probabilities of symbol that will be used for the construction of the identification sequences as well as for the accusation process. These sequence p must be kept secret.

Construction: For a user j, the i^{th} symbol of the binary identification sequence is independently sorted with a probability $P(S_i^j = 1) = p_i$. Then, for a total of n users, we can construct a matrix of the form:

Finally, for a user j, the sequence S^j is embedded into the host document.

Accusation: At the detection or verification stage, the binary sequence $Y = \{y_1, ..., y_m\}$ is extracted from the pirate copy. The implication of each user in the attack is then measured by means of an accusation score Sc^j. Due to the central limit theorem, the distribution of an accusation score Sc^j can be approximated by a normal distribution $\mathcal{N}(0, m)$ when the user j is innocent and $\mathcal{N}(2m/\tilde{c}\pi, m(1 - 4/\tilde{c}^2\pi^2))$ when the user is

	p_1	p_2	p_3	...	p_m
S_1	1	0	1	...	0
S_2	0	1	0	...	1
S_3	1	1	0	...	1
⋮	⋮	⋮	⋮	⋮	⋮
S_n	0	0	0	...	1

Figure 3.5: Construction example of a Tardos code. Each line corresponds to the identifier for a user $j \in [1, n]$ while the columns are the different positions $i \in [1, m]$.

involved in the collusion, with $\tilde{c}$ the number of colluders (the reader must refer to [Furon, 2008] for more details). For a given false detection probability ϵ_1, this "gaussianity" of the scores allows defining a threshold $Z = \sqrt{2m} \cdot \mathrm{erfc}^{-1}(2\epsilon_1/n)$ which serves to determine if a user has participated in the collusion. In order to obtain the accusation scores, two accusation functions are employed which evaluate the degree of correlation between the identification sequence S^j and the extracted sequence Y:

$$Sc^j = \sum_{1}^{m} y_i U_{ji} \tag{3.20}$$

where the functions U_{ji} can be obtained from $g_{Y_i X_i}(pi)$ as:

$$U_{ji} = \begin{cases} g_{10}(p_i) = -\sqrt{\frac{p_i}{1-p_i}}, & \text{if } S_i^j = 0 \\ g_{11}(p_i) = \sqrt{\frac{1-p_i}{p_i}}, & \text{if } S_i^j = 1 \end{cases} \tag{3.21}$$

The idea behind the accusation process goes as follows. By construction, a majority of the sequences in a Tardos code present the same i^{th} symbol, for example '1'. Thus, the presence of the symbol in the extracted sequence Y does not provide a strong evidence about the colluders identity, as it was a highly probable symbol. On the other hand, if the symbol is unlikely, its presence in Y gives strong evidence of the identity of the attackers, as the symbol was contained only into a few of the embedded sequences. Thus, as depicted in (3.21), the increase in the detection score is pondered by the symbol probability p_i so as to obtain a higher increase for a less probable symbol.

The reason why this scheme is not really appropriated to our problem stands on the fact the marking assumption is not respected. Indeed, due to the use of different watermarking secret keys, the mixed databases do not share message symbols within the same groups of tuples. As a result, from the point of view of a database DB_1, i.e., groups created by means of a secret key K_s^1, the detected symbol in a group does not depend on the symbols embedded in the other databases $\{DB_n\}_{n=2,\ldots,|\mathsf{U}|}$. As shown in section, the identifier symbols of a database are merged with a noise. In that context, anti-collusion codes are not effective, resulting in a high probability of false detections, i.e., detecting the presence of a watermark in a content that contains a different watermark or has not been watermarked at all. We have been able to experimentally verify such a result. Another difference with our scenario, is that we want to identify one database at a time, i.e., identify one sequence S^j, contrarily to classical traitor tracing scenarios where several attackers can be identified at the detection.

If as defined fingerprinting codes seem to not being satisfactory, some of the improvements suggested in the literature can however be exploited so as to optimize our detection process.

3.3.2 DETECTION OPTIMIZATION BASED ON SOFT DECODING AND INFORMED DECODER

In the literature, a collusion is mathematically described as a vector of probabilities $\theta = (\theta(0), ..., \theta(c))$, where $\theta(\phi)$ is the probability of retrieving a symbol '1' in the sequence Y if ϕ colluders have participated in the forge of the pirate copy. That is:

$$\theta(\phi) = P[Y = 1| \sum_{j \in C} X_j = \phi] \tag{3.22}$$

with the marking assumption imposing that $\theta(0) = 1 - \theta(c) = 0$.

This vector depends on the strategy followed by the colluders in order to construct the pirate copy. Recently, authors have started to analyze a "relaxed" marking assumption hypothesis, considering that the actions of the colluders can erase the embedded bits. In this case, the detection is performed by means of a soft decoding strategy, i.e., considering the extracted sequence Y is composed of real values instead of binary values, and consequently that the collusions may occur over real values. This allows to analyze other kind of attacks, such as the addition of Gaussian noise or the averaging collusion [Kuribayashi, 2010] [Meerwald and Furon, 2012]. Notice that in the previous fingerprinting studies, users identifiers were assumed to be "merged" by means of an addition, a XOR operation and so on.

This scenario is more similar to our database traceability problem, as the watermarks' mixture can be followed by a noise addition, making each symbol of the extracted sequence Y be normally distributed. We have considered two detection strategies based on the calculation of detection scores and soft-decision decoding. The *a priori* knowledge we have on the distribution of the angles β_i^{merge}, i.e., watermarked angles after noise addition, leads us to consider soft-decision detection, as it provides higher detection performances than hard-decision based detectors in noise addition scenarios [Proakis, 2001].

3.3.2.1 SOFT-DECISION BASED DETECTION

The first strategy we studied is the application of soft-decision detection. This one was first considered in digital communications so as to minimize the transmission errors and later considered in watermarking by Baudry *et al.* [Baudry et al., 2001]. It improves the decision making process by supplying additional reliability information of the received symbol. In practice, each element of the sequence $\hat{S}$ extracted from a group of tuples i corresponds to the value of the angle β_i, i.e., a real value, instead of a symbol in $\{-1, 1\}$ and the reliability information is provided by the log-likelihood ratio (LLR) of the received symbols, which is defined for the i^{th} symbol of $\hat{S}$ as:

$$LLR = \log \frac{p(\hat{S}_i | s_i^j = 1)}{p(\hat{S}_i | s_i^j = -1)} \tag{3.23}$$

The absolute value of this function $|LLR|$ provides a measure of the reliability, i.e., the degree of trust we can have in the decision, while its sign indicates the most likely symbol (positive for '1', negative for '-1'). The nearer the LLR value is from zero, the less reliable is the detected

symbol. According to the Neayman-Pearson lemma, this score provides an optimal detection (see appendix C).

Thus, considering K_s^{emb} is the secret watermarking key used at the embedding stage, the detection score is of the form:

$$Sc^j = \sum_{i=1}^{N_g} s_i^j Sc_i^j = \sum_{i=1}^{N_g} s_i^j \log \frac{p(\hat{S}_i | K_s^{emb} = K_s^j, s_i^j = 1)}{p(\hat{S}_i | K_s^{emb} = K_s^j, s_i^j = -1)} \tag{3.24}$$

where $p(\hat{S}_i | K_s^{emb} = K_s^j, s_i^j)$ is the probability density function of $\hat{S}_i$, one symbol of the extracted sequence $\hat{S}$ conditionally to the embedded symbol s_i^j when the same secret key K_s^j is used for group construction in the embedding and the detection stages (see figure 3.6). This p.d.f can be calculated as exposed in section 3.2 such as:

$$p(\hat{S}_i | K_s^{emb} = K_s^j, s_i^j) \sim \begin{cases} \mathcal{N}(P_{DB_j}(-\frac{4\pi\Delta}{L}), \sigma^2_{\beta_i^{merge}}), & \text{if } s_i^j = -1 \\ \mathcal{N}(P_{DB_j}(\frac{4\pi\Delta}{L}), \sigma^2_{\beta_i^{merge}}), & \text{if } s_i^j = 1 \end{cases} \tag{3.25}$$

where L is the number of integer values in the attribute's domain, P_{DB_j} is the proportion of DB_j in regards to the total database of the data warehouse, $\sigma^2_{\beta_i^{merge}}$ is the variance of β_i conditionally to the embedded symbol and N_g is the number of groups, which is equivalent to the length of the sequence S^j. After some calculation, the details of which are given in appendix D.1, we obtain the complete expression:

$$Sc^j = \sum_{i=1}^{N_g} s_i^j 2P_{DB_j} \beta_i \frac{4\pi\Delta}{L\sigma^2_{\beta_i^{merge}}} \tag{3.26}$$

Based on this score, the decision about the presence of a database in a data warehouse can be made automatic by defining a decision threshold Z accordingly to a false detection probability, i.e., identifying the database while it is not present and vice et versa.

For a sufficiently long sequence so as to verify the central limit theorem (i.e., $N_g > 30$), the detection score of a database j that is not present is the data warehouse follows a centered Gaussian distribution $\mathcal{N}(0, N_g \sigma^2_{\beta_i^{mix}} (2P_{DB_j} \frac{4\pi\Delta}{L\sigma^2_{\beta_i^{merge}}})^2)$, where the value of $\sigma^2_{\beta_i^{mix}}$ is calculated as exposed in section 3.2. The knowledge of this distribution allows us to calculate the optimal detection threshold Z, which for a given false detection probability P_{FA}, Z is such as (see Apendix D.1):

$$Z = \sqrt{2N_g \sigma^2_{\beta_i^{mix}} (2P_{DB_j} \frac{4\pi\Delta}{L\sigma^2_{\beta_i^{merge}}})^2} \cdot \text{erfc}^{-1}(2P_{FA}/n) \tag{3.27}$$

where n is the number of possible identification sequences, with $n \geq |\mathsf{U}'|$. Then, the database j can be considered as present in the warehouse if $Sc^j > Z$.

3.3.2.2 INFORMED DECODER

The second strategy we propose is an informed decoder that takes into account that databases merged with the one we want to detect appear as a centered Gaussian noise added to the random variable β_i. Our detection problem can be then described by means of an hypothesis test between the following hypothesis:

- H0: Database j is not in the data warehouse, which means that K_s^j has not been used for group creation at the embedding stage and its identification sequence S^j is independent of $\hat{S}$.
- H1: Database j is in the data warehouse, which means that K_s^j has been used for group creation at the embedding stage and $\hat{S}$ has been created from its identification sequence S^j.

In that context, detection theory tells us that the score given by the log-likelihood ratio (LLR) is optimally discriminative in the Neyman-Pearson sense (see appendix C). Again, considering that K_s^{emb} the secret watermarking key used at the embedding stage, this score is calculated as:

$$Sc^j = \sum_{i=1}^{N_g} Sc_i^j = \sum_{i=1}^{N_g} \log \frac{p(\hat{S}_i | K_s^{emb} = K_s^j, s_i^j)}{p(\hat{S}_i | K_s^{emb} \neq K_s^j)} \tag{3.28}$$

where $p(\hat{S}_i | K_s^{emb} \neq K_s^j)$ is the probability density function of the i^{th} symbol $\hat{S}_i$ of the extracted sequence $\hat{S}$ when the secret key K_s^j has not been used for the construction of groups in the embedding stage. As exposed in section 3.2, due to the uniform distribution of tuples into groups, the p.d.f $p(\hat{S}_i | K_s^{emb} \neq K_s^j)$ is a centered normal $\mathcal{N}(0, \sigma^2_{\beta_i^{mix}})$, where $\sigma^2_{\beta_i^{mix}}$ corresponds to the variance of the angle β_i of a mixture of |U| databases. Figure 3.6 illustrates these different distributions.

Thus knowing both p.d.f. $p(\hat{S}_i | K_s^{emb} = K_s^j, s_i^j)$ and $p(\hat{S}_i | K_s^{emb} \neq K_s^j)$, the complete expression for the detection score is (see Appendix D.2):

$$Sc_i^j = \log\left(\frac{\sigma_{\beta_i^{mix}}}{\sigma_{\beta_i^{merge}}}\right) + \frac{\beta_i^2(\sigma^2_{\beta_i^{merge}} - \sigma^2_{\beta_i^{mix}}) + 2s_i^j\beta_i\sigma^2_{\beta_i^{mix}}P_{DB_j}\frac{4\pi\Delta}{L} - \sigma^2_{\beta_i^{merge}}(P_{DB_j}\frac{4\pi\Delta}{L})^2}{2\sigma^2_{\beta_i^{mix}}\sigma^2_{\beta_i^{merge}}} \tag{3.29}$$

Once again, due to the central limit theorem, the detection score of a database that has not been uploaded to the data warehouse is normally distributed. Thus, one can compute a detection threshold Z for a given false alarm probability such as:

$$Z = \sqrt{2\sigma_{SC}^2} \cdot \mathrm{erfc}^{-1}(2P_{FA}/n) + m_{SC} \tag{3.30}$$

Notice that in this case, additionally to the calculation of σ^2_{SC}, a term m_{SC} has to be calculated in order to obtain Z. This is due to the fact that the mean of the detection scores is not zero (see 3.29). However, m_{SC} can be calculated as (see Appendix D.2):

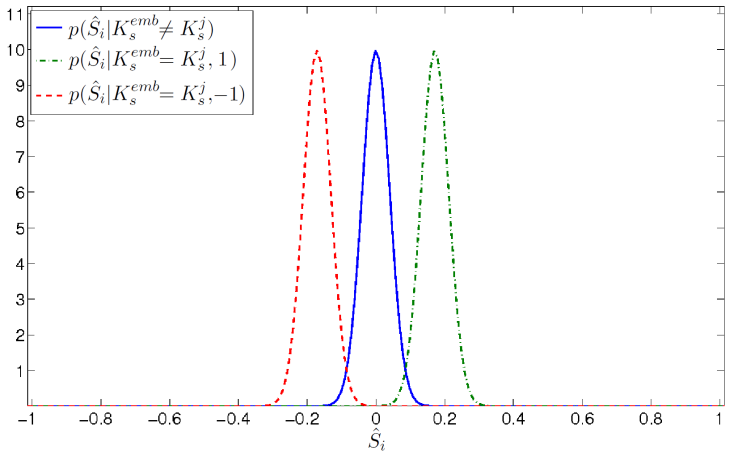


Figure 3.6: Probability density functions $p(\hat{S}_i|K_s^{emb} = K_s^j, s_i^j)$ and $p(\hat{S}_i|K_s^{emb} \neq K_s^j)$ considered in this section. In this particular example, the length of the attribute integer range is $L = 110$. This attribute was watermarked considering $N_g = 100$ groups and $\Delta = 3$. In the example, $\sigma^2_{\beta_i^{merge}} \approx \sigma^2_{\beta_i^{mix}} = 0.0016$.

$$m_{SC} = N_g \left[\log \left(\frac{\sigma_{\beta_i^{mix}}}{\sigma_{\beta_i^{merge}}} \right) - \frac{(P_{DB_j} \frac{4\pi\Delta}{L})^2}{2\sigma^2_{\beta_i^{mix}}} + \left(\frac{\sigma^2_{\beta_i^{merge}} - \sigma^2_{\beta_i^{mix}}}{2\sigma^2_{\beta_i^{merge}}} \right) \right] \tag{3.31}$$

The variance σ^2_{SC} is calculated as (see Appendix D.2):

$$\sigma^2_{SC} = N_g \sigma^2_{\beta_i^{mix}} \left[\left(\frac{P_{DB_j}}{\sigma^2_{\beta_i^{merge}}} \frac{4\pi\Delta}{L} \right)^2 + \left(\left(\frac{\sigma^2_{\beta_i^{merge}} - \sigma^2_{\beta_i^{mix}}}{2\sigma^2_{\beta_i^{mix}} \sigma^2_{\beta_i^{merge}}} \right)^2 2\sigma^2_{\beta_i^{mix}} \right) \right] \tag{3.32}$$

3.4 EXPERIMENTAL RESULTS

In this section we experimentally verify the above theoretical results in the following way. We recall that our basic idea is to identify one database merged with several other in a data warehouse. To do so, each database is watermarked by means of the robust lossless scheme presented in chapter 2 so as to insert its unique identifier, e.g., the binary sequence S^j, for the database j. Each database is watermarked using a distinct secret key K_s^j, key which is the base of tuple group construction.

Following experiments have been conducted on the same database described in chapter 2. It contains more than ten million tuples of 15 attributes each; attributes that describe inpatient stays in French hospitals. $N = 1048000$ tuples were extracted from this database for our test purposes. The attributes *age* and *dur_stay* were considered for watermark embedding while the attributes *id_hospital* and *id_stay* were used as the primary key. In order to compare the detection performance of the different approaches exposed above, the following experience plan was applied:

1. Creation of nine databases from the considered extract.
2. Watermarking of each database with its unique identifier S^j, based on a secret key K_s^j.
3. Databases merging under the constraint that DB_1 represent a $D\%$ of the total number of tuples.
4. Detection of the database of interest DB_1 considering its secret key K_s^1.

The plan of experimentation was conducted with various: watermark distortion amplitude (Δ); number of groups (N_g), which represents the length of the embedded identifier; percentage D of tuples DB_1 represents into the total database; false detection probability P_{FA}. Notice that for sake of simplicity, all nine databases were watermarked in the same way. The number of identification sequences was fixed to $n = 50$. Furthermore, the detection probability results are given after 40 random simulations with the same parameterization but with different tuple configuration.

3.4.1 CORRELATION-BASED DETECTION PERFORMANCE

In a first time, we were interested in evaluating the performance of the correlation detection process on which our scheme is originally based (see chapter 2) as a function of Δ, N_g, P_{FA} and D. In order to perform a fair comparison, the detection threshold was calculated based on the central limit theorem, from which we know that the correlation rate for a database not uploaded to the data warehouse follows a centered normal distribution $\mathcal{N}(0, \frac{1}{N_g})$. Then, for a fixed false alarm probability P_{FA}, the threshold is calculated as $Z = \sqrt{2/N_g} \cdot \mathrm{erfc}^{-1}(2P_{FA}/n)$.

To do so, the data warehouse was constructed so that DB_1 represents $D = [1\%, 10\%]$ of the total. The attribute *age* was watermarked considering two values of $\Delta = [2, 3]$ and four values of the number of groups $N_g = [70, 100, 130, 250]$, i.e., embedded sequence lengths. At the detection, two values of $P_{FA} = [10^{-3}, 10^{-4}]$ were also considered. Obtained results are depicted in figure 3.7 where we can observe that, in the best case, the database to detect should represent at least 9% of the total of tuples. As expected, the detection performance increases for higher values of Δ and for a lower number of groups N_g. However, for th, but in this case some exceptions appear. This is probably due to the fact that any error occurring in a shorter sequence has a harder impact on the final correlation score and consequently, on the final detection rate. Finally, a higher value of P_{FA} implies a more restrictive detection threshold, resulting in lower detection rates.

As it was exposed in chapter 2, the statistical distribution of the chosen attribute may have an important role in the detection performance. This is the reason of this second experiment. Attributes *age* and *dur_stay* were watermarked using the same parameterization: $\Delta = 2$, $N_g = 100$ and $P_{FA} = 10^{-3}$. As illustrated in figure 3.8, for the attribute *dur_stay*, which presents a more concentrated p.d.f, the detection rates are higher than those obtained for the attribute *age*, with a detection rate of 100% even if the DB_1 represents only 3% of the data warehouse. This result is logic based on the theoretical results we provide in chapter 2.

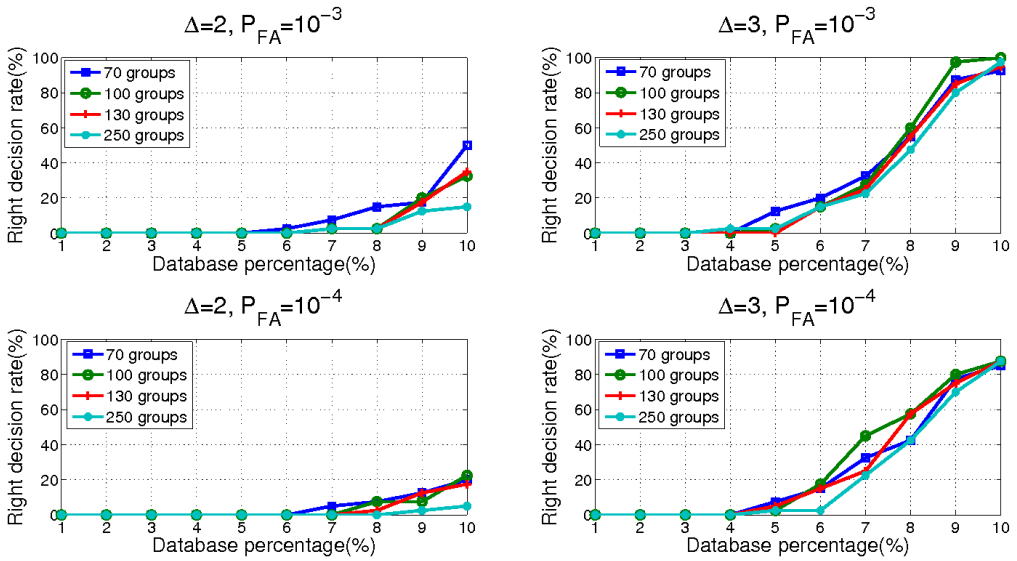


Figure 3.7: Detection performance obtained by the correlation based detector for the attributes *age* considering two values of $\Delta = [2, 3]$, four values of the number of groups $N_g = [70, 100, 130, 250]$ and two different false detection probabilities $P_{FA} = [10^{-3}, 10^{-4}]$.

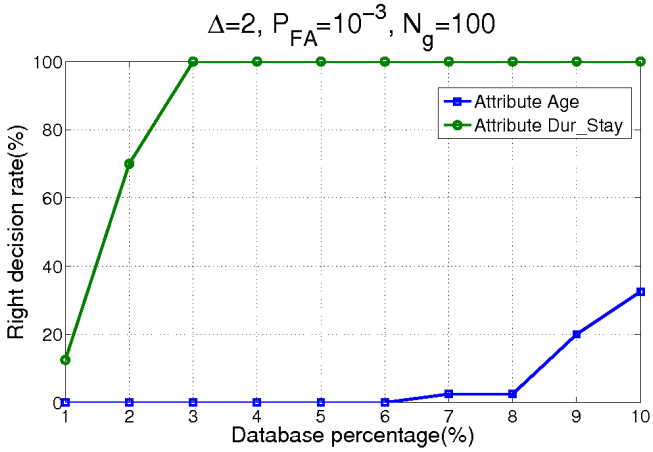


Figure 3.8: Detection performance obtained by the correlation based detector for two different attributes: *age* and *dur_stay* considering the following parameters: $\Delta = 2$, $N_g = 100$ and $P_{FA} = 10^{-3}$. As it can be seen, for an attribute with a smaller variance (*dur_stay*), we obtain a better performance.

3.4.2 PROPOSED DETECTORS PERFORMANCE

In a second time, we were interested in comparing the performance of the correlation-based detection with the two other detection strategies depicted above. As before, the attribute *age* is used for identifier embedding, with DB_1 representing $D = [1\%, 10\%]$ of the data warehouse. Two values of the distortion amplitude $\Delta = [2, 3]$, as well as four number of groups values $N_g = [70, 100, 130, 250]$, i.e., sequence lengths, and two values of $P_{FA} = [10^{-3}, 10^{-4}]$ were

considered. Results for the different parameters' combinations are illustrated in figures 3.9, 3.10, 3.11 and 3.12. Notice that the curves for the two proposed decoders are superposed in the figures, meaning that both detectors offer the same performance. They however outperform the results achieved with the correlation based detector in all the possible cases. In the best case, we obtain a detection rate of 100% even if DB_1 represents only 7% of the data warehouse.

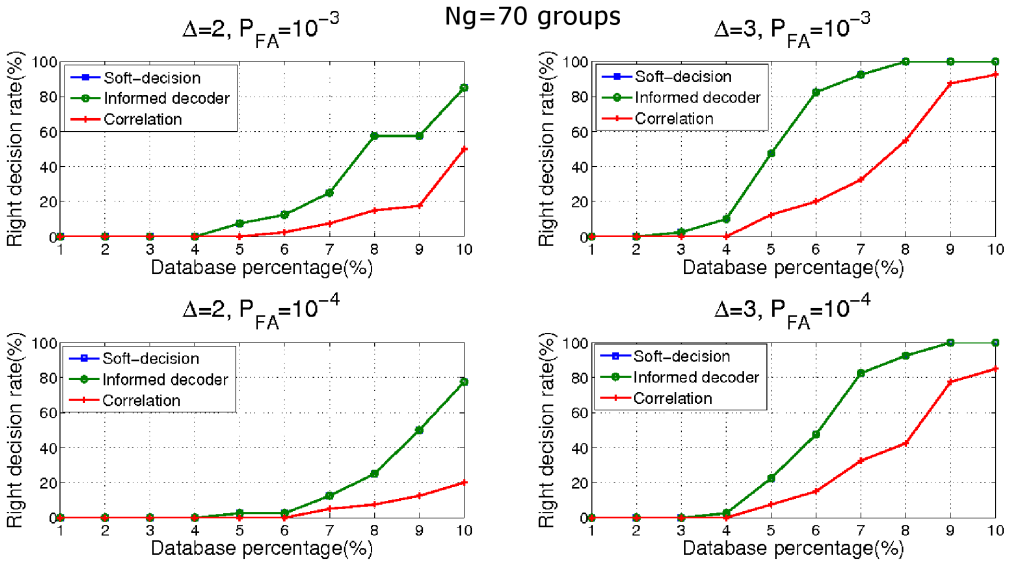


Figure 3.9: Detection performance obtained by the proposed detection approaches for $N_g = 70$.

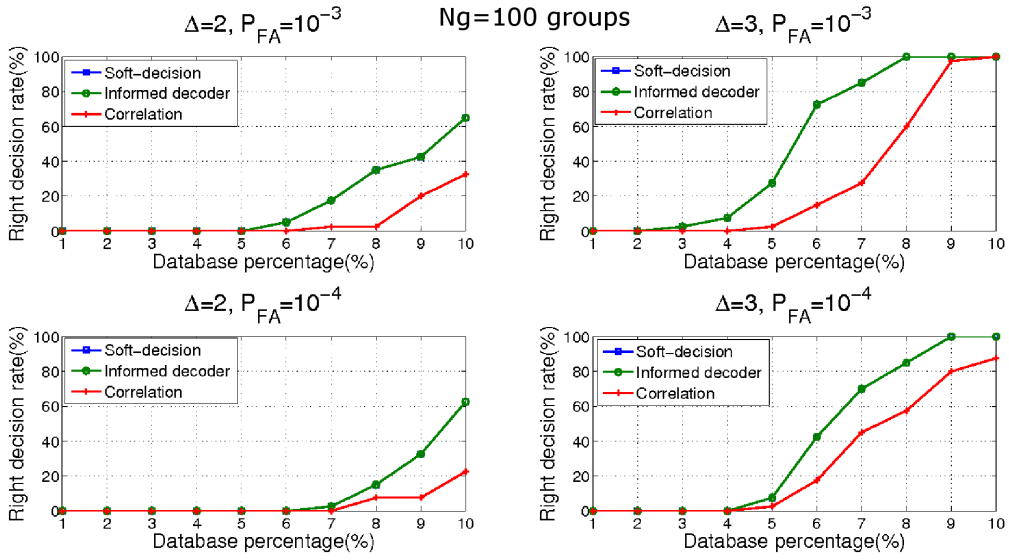


Figure 3.10: Detection performance obtained by the proposed detection approaches for $N_g = 100$.

In a second experiment we searched to compare these detectors in terms of the false detection rate when:

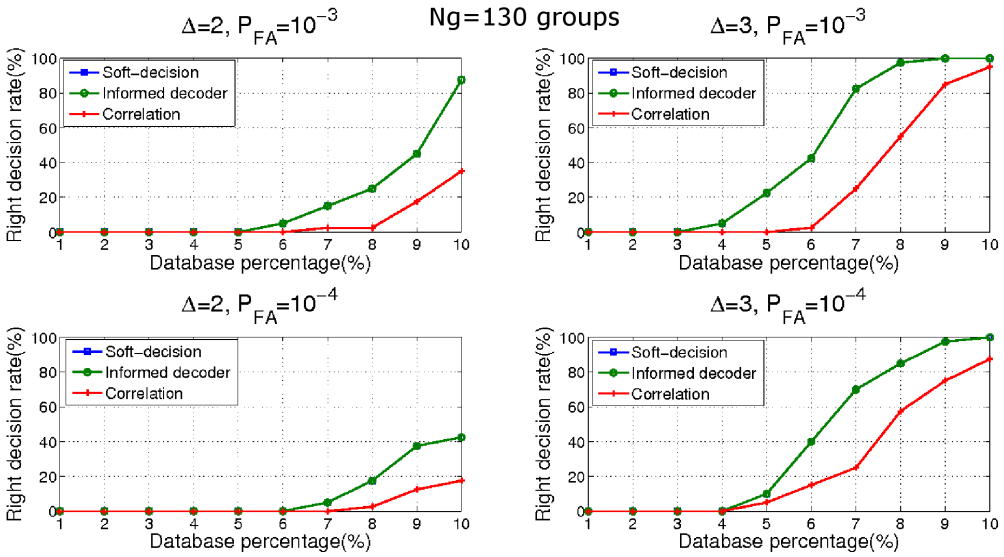


Figure 3.11: Detection performance obtained by the proposed detection approaches for $N_g = 130$.

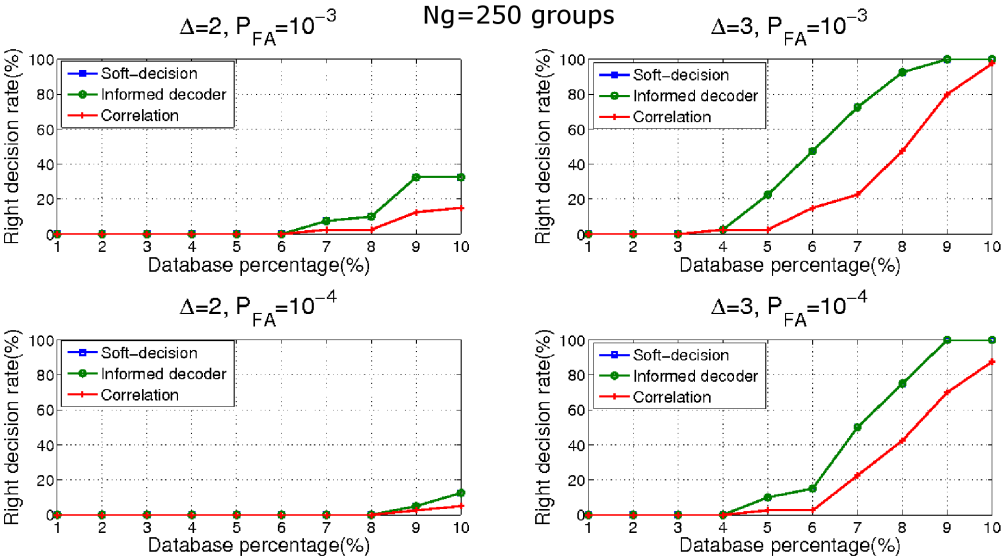


Figure 3.12: Detection performance obtained by the proposed detection approaches for $N_g = 250$.

- a different secret key is used at the detection
- the same secret key is used at the detection but with a different embedded identifier (e.g., sequence).

Obtained results show again that both detectors offer exactly the same performance, with a 0% of false detections in the first case. In the second case, i.e., the case where we try to detect a non-embedded sequence with a correct secret key, both detectors perform well, with only a few false detections when $N_g = 70$. Indeed, the length of the embedded sequence has

an impact in this case. A longer sequence will be more discriminative. It is then necessary to find an identifier length so as to obtain a good a trade-off between good detection (shorter sequences are more robust to the mixture) and false detection (longer sequences allow better discrimination).

As exposed, both detectors are based on the *a priori* knowledge we have about the impact of the data mixture on the statistical distribution of the watermarked angles. However, we made the hypothesis that all merged database attributes followed the same statistical distribution. This may not be necessarily the case and we have thus an interest in evaluating the impact of a database mixture where attributes are slightly different in their distributions. To do so, we performed a third experiment:

1. Construction of one database DB_1 which will represent $D\%$ of the data warehouse.
2. Estimation of the statistical distribution of the attribute to watermark in BD_1 by means of its histogram.
3. Generation of 8 "simulated" databases, the attribute of which to watermark has a distribution of the same nature than the original but of different parameters.
4. Insertion of one identifier S^j in each database, with different secret key K_s^j.
5. Merging of the databases.
6. Detection of the database of interest DB_1. To do so, its associated secret key (e.g., K_s^1) is considered for group construction.

The procedure we use in order to generate a database as described in step 3 is as follow. Let us consider the attribute to watermark takes its values in the integer range $[0, L-1]$. Then, the attribute distribution can be estimated from its normalized histogram by L probability values $\{P(l)\}_{l=0,\dots,L-1}$ with $P(l) \geq 0 \ \forall l$ and $\sum_0^{L-1} P(l) = 1$. For each of these values, a zero mean random value ϵ_l of fixed distribution is added in order to obtain the modified distribution ($P'(l) = P(l) + \epsilon_l$). Notice that in order to obtain the final modified distribution, all the values should be normalized so as to ensure $\sum_0^{L-1} P'(l) = 1$. The value ϵ_l follows a uniform distribution in $[-\frac{v \cdot P(l)}{100}, \frac{v \cdot P(l)}{100}]$, where $v \in (0, 100)$ is a parameter which allows controlling the intensity of the modification. Figure 3.13 illustrates the obtained distributions for the attribute *age*, which takes its values in $[0, 121]$, considering ten values of the parameter $v = [1, 2, 5, 10, 20, 30, 50, 75, 85, 99]$.

For the tests, we considered values of $D \in (7\%, 10\%)$, $N_g = [70, 100, 130, 250]$, $\Delta = 3$ and $P_{FA} = 10^{-3}$. As illustrated in figure 3.14, both proposed detectors offer the same high performance, with a correct detection rates close to the levels observed in our previous experiment and a false detection rate of 0% in the two previously presented false detection scenarios.

3.5 CONCLUSION

In this chapter, we addressed a practical application scenario of our lossless database watermarking scheme, which is the traceability of databases in the case when they are merged in a

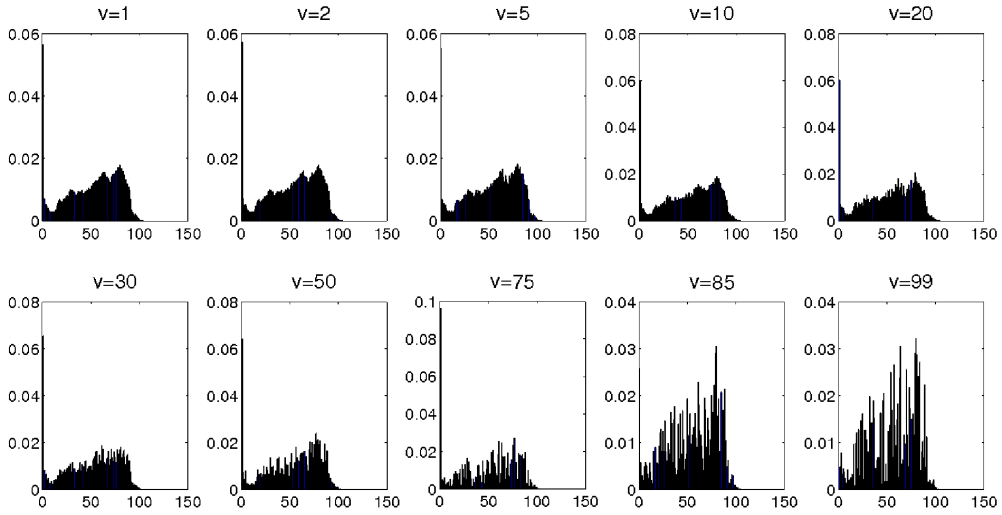


Figure 3.13: Examples of modified distributions when ϵ_l follows a uniform distribution in $[-\frac{v \cdot P(l)}{100}, \frac{v \cdot P(l)}{100}]$ and ten values of the parameter $v = [1, 2, 5, 10, 20, 30, 50, 75, 85, 99]$ were considered.

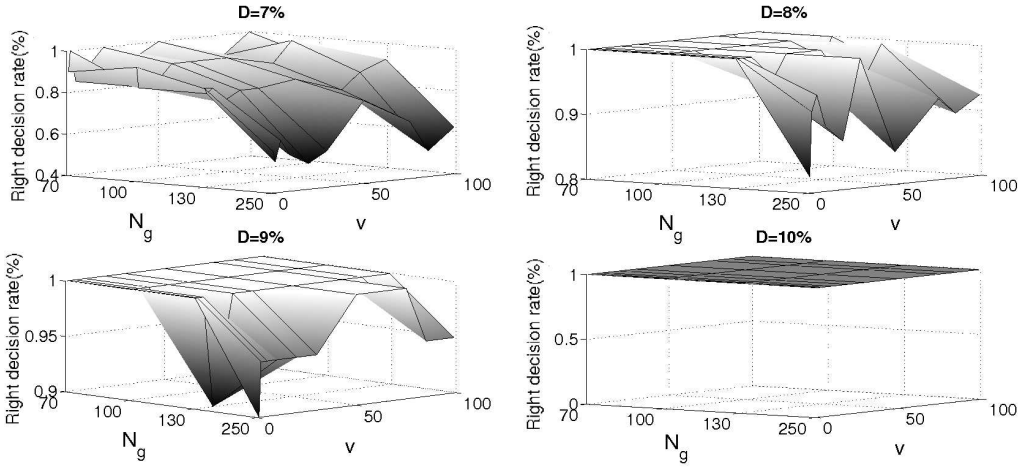


Figure 3.14: Detection performance for the proposed detectors with: $D \in (7\%, 10\%)$, $N_g = [70, 100, 130, 250]$, $\Delta = 3$ and $P_{FA} = 10^{-3}$ and $v = [1, 2, 5, 10, 20, 30, 50, 75, 85, 99]$.

data repository or even in a data warehouse. This is an issue of rising interest as there is a trend to the creation of inter-institutional shared data warehouses giving access to big amounts of data to researchers and administrators.

As we have exposed, the related problem of the traceability of distributed copies of a multimedia content has been largely analyzed. In that context, a class of optimal codes called anti-collusion codes has been proposed in order to offer an optimal detection of illegally shared copies in the case of a collusion attack. This attack is performed by several users who cooperate and mix their copies so as to erase the embedded identifiers and try to avoid being pointed as culprits of the content leakage. Although we can identify some similarities to our problem

we have also shown that several differences make anti-collusion codes ineffective in our case. Nevertheless, we proposed to take advantage of some improvements that have been considered for such traitor tracing solutions.

Our final database traceability system is based on two main aspects: first, the robust lossless watermarking scheme presented in chapter 2, which allows to embed a message that will resist an important degree of attacks to the database; second, the theoretical modeling of the impact of database aggregation onto the detection process of one single database. Both allow us to search the more optimal detector based on a soft-decision decoding.

In order to prove the benefits of the proposed detection improvements, our traceability solution has been tested on the mixture of several real medical databases, with different mixture proportions and watermarking parameterization. As exposed, the proposed detectors improve the performance obtained by the correlation based detection. Indeed, they allow to identify a database representing a low percentage of the total of tuples (around 7%) in nearly a 100% of the cases.

Chapter 4
Semantic distortion control

In the previous chapters, we have addressed the protection of relational databases by means of lossless watermarking. As exposed, lossless or reversible watermarking allows the removal of the embedded sequence, in order to access the original undistorted data. This makes that distortion constraints can be alleviated. Nevertheless, even in the lossless case, there exists an interest in controlling the introduced distortion. Indeed, keeping the watermark into the database ensures a continuous protection.

Until now, authors have focused on preserving the statistical distortion of the database in order to preserve the result of possible subsequent data-mining processes or of some aggregation queries as example [Shehab et al., 2008] [Kamran et al., 2013a] [Gross-Amblard, 2003]. However, minimizing statistic distorsion is not enough, one must also consider the strong semantic links that exist in-between attributes in the database. Indeed, watermarked tuples in the database must remain semantically coherent in order to ensure the correct interpretation of the protected database. Doing so will also enhance the invisibility of the embedded sequence to a potential attacker. As we will see in the first section of this chapter, previous methods fail to do so, as they only consider statistical criteria.

In order to go beyond this state of art, we propose to take advantage of the semantic knowledge one can have about the content of a database. In health care, this becomes possible based on the efforts on modeling medical knowledge and of the language of this domain. These efforts stand on the availability of medical ontologies, ontologies we propose to use in order to identify the allowable distortion that can be applied to each attribute value of a tuple. Basically, one ontology is attached to a specific area of knowledge and allows defining shared concepts with their relationships by means of a common vocabulary. As we will see, the domain of an attribute can be associated to a concept in an ontology.

In the second part of this chapter, we demonstrate the interest of such a semantic distortion control though the use of an extension of the Quantization Index Modulation (QIM) we propose. As for the previous schemes we developed until here, this one modulates the relative angle of the center of mass of circular histograms associated to groups of values of one numerical attribute of the relation. Our choice to use such non-reversible algorithm is based on the fact it provides us a degree of freedom in managing the distortion of individual attributes' values, allowing a more precise control of the robustness-imperceptibility trade-off. Nevertheless, this does not prevent us to establish a theoretical proof of its robustness against common attacks, i.e., tuple insertion and suppression. Again, we verify such theoretical results by means of experiments conducted on real medical databases.

4.1 OVERVIEW OF DISTORTION CONTROL METHODS

In Chapter 1 we briefly introduced database distortion constraints. They stand on quality criteria or rules that are established by both the owner and the final user of the database depending on its expected use. From our point of view, distortion constraints can be related to statistical properties of individual attributes or sets of attributes as well as to the semantic relationships between attributes. More clearly, the former can be used to guarantee the correctness of an aggregation query or a data-mining operation while the latter protect the coherence and plausibility of tuples.

Up to now, authors have focused in the minimization of the statistical distortion. Sion *et al.* [Sion et al., 2004] were the first to address this issue. They present a method in which the embedding process does not modify numerical attributes if some "data usability conditions", measured for example in terms of the mean squared error, are not respected. To do so, they developed an additional "plug-in" active during all the watermarking process in order to perform a "rollback" operation (i.e., restore the attributes original values) if static conditions are violated. In order to avoid the use of this plug-in, which introduces an additional complexity, Shehab *et al.* [Shehab et al., 2008] propose to adapt the watermark amplitude by means of optimization techniques computed before embedding. The optimization is constrained by the user-defined usability conditions. At the same time, this strategy allows them to optimize the detection process (i.e., the value of the decision threshold) resulting in a better robustness-imperceptibility trade-off. Optimization based embedding has been also proposed in [Meng et al., 2008].

These two approaches can be considered as of "general-purpose" distortion control methods, as they do not take into account any *a priori* knowledge about the subsequent operations data may undergo. However, one can also find some more "application-specific" methods. The preservation of some predefined aggregation queries results (e.g., sum, mean, etc.) has been addressed by Gross-Amblard [Gross-Amblard, 2003], who suggests modulating pairs of tuples involved in the query response. More clearly, for each of these pairs, the tuples attributes' values are modified with the same amplitude but with opposite sign in each tuple (e.g., $+1$ and -1). In their scheme, distortion minimization is ensured under some restrictions expressed in the Vapnik-Chervonenkis[1] dimension of the queries. More clearly, a query should not be able to separate tuples of the relation in any possible way. This avoids a query result that only contains positive alterations. An optimal method to identify the adequate pairs of tuples is presented in [Lafaye et al., 2008], where the authors propose a complete watermarking scheme based on Gross-Amblard's strategy.

Kamran *et al.* have focused on the watermarking of data sets intended for data-mining operations in particular those based in a classification process. In that case, they propose to group attributes in sets according to their classification importance, i.e., their influence in the classification result. Some local (i.e., for a set of attributes) and global constraints are then defined. The latter consist in keeping unchanged the classification potential of the attribute set produced by six different feature selector methods (information gain, information gain ratio, correlation based feature selection, consistency based feature subset evaluator and principal

[1] More information on the Vapnik-Chervonenkis dimension can be found in [Vapnik, 1995]

components analysis) [Kamran et al., 2013a]. Then, the allowed perturbation in tuples for this set of attributes is calculated with the help of optimization techniques. The embedding process is similar to the one they expose in [Kamran et al., 2013b], in which the distortion constraints are defined *a priori* by the data owner and recipient. They correspond to the mean and standard deviation of the attributes to watermark. In this work, authors introduce the concept of "once for all" usability constraints, which simply correspond to the more restrictive set of constraints that can be fixed for a database for a set of possible applications. According to the authors, this makes the detection performance independent of the constraints, as any alleviation of the modification restrictions will result in a better robustness.

As exposed, all the above methods focus on preserving statistical properties of the database. None consider that there exist strong semantic links in between attributes values in a tuple that should also be preserved. Indeed, tuples must remain semantically coherent in order to: i) ensure the correct interpretation of the information without introducing incoherent or unlikely records; ii) keep the introduced perturbations invisible to the attacker. As example, an "incoherent" tuple can be statistically insignificant but highly semantically detectable. Even though semantic distortion has been briefly evocated by Sion *et al.* in [Sion et al., 2004], to our knowledge no method has been yet proposed that allows respecting such semantic constraints. This is the objective we fixed ourselves and to reach it we propose to take advantage of medical ontologies so as to extract inter-attribute semantic constraints and adapt the watermark in consequence.

4.2 SEMANTIC KNOWLEDGE

4.2.1 INFORMATION REPRESENTATION MODELS IN HEALTHCARE

The heterogeneity of medical information results in the application of different information representation models that are used to encode data. The model is selected depending on the pursued objective (e.g., data storage and recovery, computer reasoning, information exchange and so on) that defines a set of needs in terms of semantic expressiveness, reasoning capability and information singularity (i.e., individuals vs general concepts). The compromise between these properties is depicted in figure 4.1(a) [Daconta et al., 2003].

As it can be seen, relational databases present a low semantic expressiveness. Indeed, they have been conceived so as to offer efficient storage and retrieval of individual records and to ensure data integrity (i.e., coherence), but their scheme does not provide a referential where existing semantic links between attributes would be explicitly defined. The relational database model considers a closed world where only the individuals in the database exist, making it possible to recover only previously stored records. In the medical context, relational databases are typically employed for the storage of medical records and the collection of administrative information (for economical evaluation – see PMSI database we use for experimentation, inventories or public health purposes).

On one upper level of semantic expressiveness we find taxonomies. A taxonomy is a hierarchical structure (i.e., a tree) containing a controlled vocabulary in some area of knowledge. It aims at providing a common terminology to be used for professionals in a domain, making

it easier to find the correct term from its hierarchical relations with others. In the health framework, the International Classification of Diseases (ICD) is constructed as a taxonomy [World Health Organization, 1992]. Each ICD code descends from only one upper class which can itself be a descendant of another category (see figure 4.1(b)).

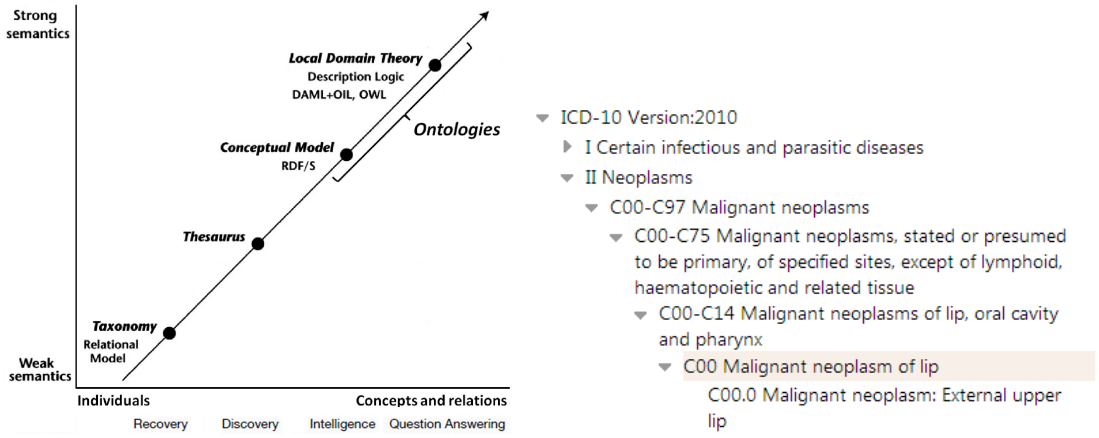


Figure 4.1: Classification of knowledge models according to their semantic expressiveness ("semantic spectrum") and the singularity of contained information (adapted from [Daconta et al., 2003]). b) Extract from the International Classification of Diseases (ICD) version 10. Terms are associated hierarchically from general categories to more specific terms.

Thesauri represent an extension of the taxonomy model. They offer some additional association (e.g., "related term") and equivalence (e.g., "synonym") semantic relations that may be useful when associating different concepts. Medical Subject Headings (MESH) is possibly the most important medical thesaurus [US-NLM, 1963]. It is organized in 16 trees, each of which is related to a general concept in medicine (e.g., Anatomy, Diseases, Chemicals and Drugs, ...). MESH is applied in the indexation and retrieval of medical documents and multimedia, more specifically for the indexation of bibliographic references in MEDLINE/PubMed.

The Systematized Nomenclature of Medicine-Clinical Terms (SNOMED-CT) [IHTSDO, 1999] is also a multilingual thesaurus, but it presents some features that approach it to the notion of ontology. More clearly, by means of a set of encoded rules and constraints, it allows the inference of new information from the *a priori* data introduced by the user. A simple example of a rule extracted from WikiPedia:

```
Viral upper respiratory tract infection equivalentTo
 Upper respiratory infection and Viral respiratory infection and
        Causative-agent some Virus and
        Finding-site some Upper respiratory tract structure and
        Pathological-process some Infectious process
```

Thus, an individual for which the *a priori* knowledge indicates an infectious process caused by a virus and located in the upper respiratory tract will be automatically classified as a viral upper respiratory tract infection.

The Unified Medical Language System (UMLS) [US-NLM, 1986] is at the same time a meta-thesaurus, containing more than 1 million medical concepts (more than 5 million concept names) coming from over 100 terminologies (e.g., ICD10, SNOMED-CT, ...) and an ontology, linking 135 semantic types (i.e., high level categories) by means of 54 different semantic relationships (e.g., is_a, co-occurs_with, interacts_with, ...). Each term in the meta-thesaurus is assigned to at least one semantic type. The purpose of UMLS is to facilitate the development of medical informatics systems, allowing them to "understand" the meaning of the language of biomedicine and health while providing human users a complete mapping between different terminologies.

It should be remarked that the biomedical knowledge representation literature refers to most of the existing medical terminologies (i.e., taxonomies, thesauri and ontologies) as ontologies, even though they lack is some cases of the principles of the ontological model. Nevertheless, they provide a common representation of entities in an area of knowledge, being enough in the majority of actual ontology applications. In the sections to come we review ontology principles in more detail before explaining the actual computer representation of these different knowledge models.)

4.2.2 DEFINITION OF ONTOLOGY

In philosophy, Ontology is the study of the nature of being, the existence as well as the basic categories of being. Although this definition can not be considered in a straightforward way in computer science and knowledge engineering communities, the idea of ontology as a representation of concepts or ideas and the relations between them is widely accepted. Nevertheless, authors do not come to an agreement when giving a standard definition of ontology. In the seminal work by Neches et al. [Neches et al., 1991], authors expose that an ontology "*defines the basic terms and relations encompassing the vocabulary of a topic area as well as the rules for combining terms and relations to define extensions to the vocabulary*". Another common and more general definition is given by Gruber who considers an ontology as an explicit specification of a conceptualization [Gruber, 1993], where a conceptualization is an abstract, simplified view of the world that we wish to represent for some purposes. In a more recent work, Gomez-Perez and Benjamins [Gomez-Perez and Benjamins, 1999] state that ontologies "*provide a common vocabulary of an area and define, with different levels of formality, the meaning of the terms and the relations between them*". From these definitions, we can deduce that an ontology allows the users in some specific area of knowledge to define shared concepts and how they are related by means of a shared vocabulary, helping them to overcome the intrinsic heterogeneity in the definition of real world entities. An important feature of ontologies is that they are interpretable by both human operators and by computer programs, representing a gateway between human and artificial knowledge.

4.2.3 COMPONENTS OF AN ONTOLOGY

Even though authors do not come to an agreement in terms of what components an ontology should have, most definitions contain the elements defined by Gruber. These elements are

classes, relations, functions, axioms and instances [Gruber, 1993]. We clarify these notions in the sequel.

Concepts or classes are abstract groups, sets, or collections of objects. Examples of concepts are Person, Car, Thing, etc. Depending on the ontology, classes can contain other classes and a universal class may contain every other classes. For instance, one can define the class "thing" that (in the abstract sense of the word) may contain anything one could imagine (e.g., Person, Car, Book, etc). An **individual** or instance corresponds to the ground level concept of the ontology; it is a concrete instantiation of an element or an object (e.g., a person named Peter or a car Renault Clio). Notice that the frontier between an individual and a class is quite blurred. It depends on the considered ontology.

Objects in the domain are associated by means of **relations**, which specify the interactions between them. We can have relations between classes, between an individual and a class, between individuals, etc. For example, we know that one person is-child-of another person or that Batman fights-against the Joker. Functions are a special case of relations in which the n^{th} element of the relationship is unique for the n-1 preceding elements, such as is-mother-of.

Attributes allow describing individuals in the ontology. Examples of attributes are has-name, has-age and so on. The value of an attributes is defined by a data type, e.g., integer, string.

Finally, **axioms** represent assumptions from which other statements are logically derived or are explicit rules in the relationships between concepts. For example, every person has a mother or every car brand should fabric at least one model of car. Axioms allow inferring new information.

4.2.4 EXISTING APPLICATIONS OF ONTOLOGIES

Ontologies have been considered for several applications in the health and biology domains. Herein, we expose the most common:

- **As a controlled vocabulary for indexing and annotation:** Concepts in a field can be described in many different manners. This heterogeneity can lead to misunderstandings situation and to increase information analysis complexity. In order to avoid this, users in a field of knowledge may agree on a set of terms that describe the concepts of their domain. Herein, the ontology is used as a reference which provides standard terms to describe real world elements, such as a disease, a gene, a zone of an image, etc.

- **Experiments methodology and results description application:** One important issue in health care is the reduction of "ambiguity" when sharing data between different users. For instance, the transfer of experiments or analysis results can derive in useless information if it is performed without care. In order to guarantee reproducibility and facilitate data analysis, ontologies allow associating standard labels that describe the experiments' methodology and its results, e.g., the Microarray Gene Expression Data (MGED) Ontology [Whetzel et al., 2006].

- **Definition of a data base scheme:** When constructing a data base, users are confronted to the definition and labeling of a set of attributes in order to represent information. Due to the different points of views of each user, different bases can overlap or give an incomplete or partial view of the domain. The use of non-standard attributes makes it hard to compare and merge these different bases in order to obtain a more complete and correct representation. An ontology constructed in agreement within a whole community provides a model from which the data base can be created so as to ensure interoperability within the domain.

- **Natural language processing applications:** Ontologies may be helpful in text-mining applications, allowing for instance to extract key words from long texts and associating them to a set of accurate sources of information in some specific situations. In a more simple application, they can provide a set of synonyms and related terms that can help a user not using the same words all the time.

- **Computer based reasoning application:** A classical example of such an application is given by aided decision making. An operator that has an *a priori* knowledge of some real world situation (e.g., patient symptoms) introduces the information into the system in order to obtain a set of possible responses that may help him to make a final choice. These responses are obtained after querying a "background" ontology.

- **Multiple information sources integration:** An ontology can be constructed so as to merge several knowledge resources into a single entity. This enables the access to all of the knowledge by means of a unique user interface and provides an homogeneous information representation.

4.2.5 How are relational database and ontology models represented?

As exposed in chapter 1, a relational database consists of one or several tables, each of which contains a set of records (tuples) that associate different attributes. These tables can be encoded in a native database format (e.g., Oracle, Microsoft Access, Lotus Notes, ...) but also in other formats such as spreadsheets (e.g., Microsoft Excel), or even text files where tables are constructed by means of specific delimiters (e.g., tabulations, commas, etc.). While this format is efficient for data storage and retrieval, it is not optimal in other contexts, such as artificial intelligence, where the focus is fixed in the meaning and the semantic relationships between concepts.

Usually, in ontologies and ontology-like resources (e.g., Thesauri, taxonomies, etc.) knowledge is represented in the form of triplets, i.e., relations of the type "subject" + "verb" + "object" (e.g., *fish* + *is_an* + *animal*), which can be modeled in different manners. For instance, one can consider semantic networks in which triplets are displayed in the form of oriented graphs where the concepts are represented by nodes while arcs define relations (see figure 4.2). Although semantic networks may allow representing taxonomies or thesauri, they present limitations for the handling of non-taxonomic knowledge and of concepts' attributes or properties.

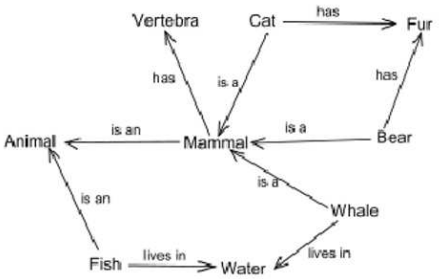


Figure 4.2: Knowledge representation by means of a semantic network. Arcs indicate different types of relations while nodes represent concepts.

In order to overcome these issues, other representation models have been proposed, among which it is worth highlighting the frame language, description logics, conceptual graphs or the semantic web languages (RDF, OWL, etc.) [Minsky, 1974] [Brachman and Schmolze, 1985] [Chein and Mugnier, 2009].

Frames represent classes or objects which are hierarchically organized [Minsky, 1974]. Each frame comprises a set of slots that define the class attributes. These attributes may consist of a value or set of values, a relation to another frame or even procedural rules that indicate to an external agent which actions to perform after some of the slots have been filled in. A simple example of knowledge representation by means of frames is depicted in figure 4.3, where a frame "Lecture" is defined by a set of slots, among which we can identify fixed values ("Context"), values to be filled up by an agent ("Course", "Level"), relations to other frames ("Lecturer") and rules defining actions according to other slots ("If difficult, then pay attention"). The frame model also allows objects to inherit properties from ancestors.

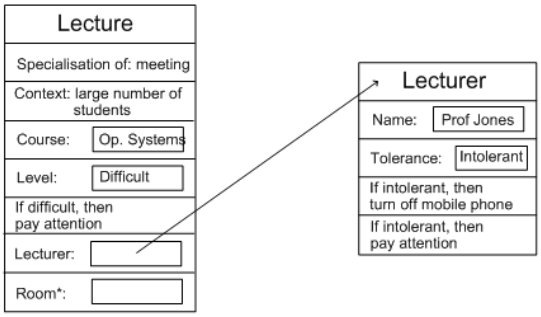


Figure 4.3: Knowledge representation by means of frames [Colton, 2005].

Description Logics were introduced more specifically so as to overcome the lack of formal semantics suffered by the frame language [Brachman and Schmolze, 1985]. They rely in three main notions: concepts, roles and individuals, and the relationships between them. In order to represent these elements, two different structures are considered: the Terminological box (T-Box) contains the description of the concepts and the hierarchical relations between them. On the other hand, the Assertional box (A-Box) is constituted of the description of individuals, of the rules to which they are associated and the definition of their relations to concepts. Description logics are more flexible than frames and they lie on strict semantics and syntax

[Baader and Hollunder, 1991].

Conceptual graphs were introduced in the seminal paper by Sowa [Sowa, 1976] but their application to knowledge representation is based in the work by Chein and Mugnier [Chein and Mugnier, 2009]. They are conceived to be an intermediate formalism between natural language and first order logic. Predicates and arguments in a phrase are modeled as a labeled graph with two kind of different nodes: concept nodes, which represent entities, events or states, and relation nodes which show how concepts are interconnected [Sowa, 1976]. The particular organization of these nodes in the graph represents a logic formula, or more generally, knowledge.

During the last few years, the concept of Semantic web has firmly settled as the expected evolution of the classical World Wide Web that pretends to make all the resources in the web directly understandable by computers. The objective is to provide more semantically meaningful and accurate responses to human requests. This evolution involves the use of a new set of languages that allow describing at the same time documents and their links (as in HTML) but also other concepts such as people, places, meetings, etc. For instance, the Resource Description Framework (RDF) is a model that allows encoding, exchanging and the reuse of structured metadata in a simple way, that is to say, by means of statements of the type subject + predicate + object, as depicted in figure 4.4. The RDF-Schema (RDFS) is a language that enables the creation of RDF models, i.e., sets of items specific to a particular domain. For example, the medical community can be interested in defining items such as disease, drug, etc. and how these items are associated. RDFS is composed of a set of classes and properties based on the RDF model that describes concepts and the relations between them.

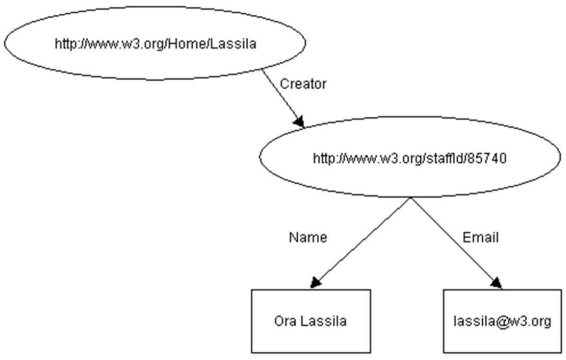


Figure 4.4: Representation of the sentences "The individual referred to by employee id 85740 is named Ora Lassila and has the email address lassila@w3.org. The resource http://www.w3.org/Home/Lassila was created by this individual." by means of RDF [Swick, 1999].

Ontology Web Language (OWL) was born as a more expressive and flexible evolution of RDF/RFDS. Although both of them share the same basic concept, OWL offers a larger vocabulary and a set of constraints (disjunction, cardinality, equivalence, etc.) that allow for instance to describe concepts in terms of set operations (e.g., the concept Mother is the union of the concepts Parent and Woman). These features make OWL a more powerful language for

ontology creation. However, it should be noticed that RDF/RDFS can be completely suitable for lite ontologies, where we require lower needs in terms of inference and reasoning.

In this thesis, we only have considered ontological resources developed in RDF/RDFS and OWL due to the fact they can be considered as the *de facto* standards and that there exist a large number of tools that allow creating, handling and querying such a kind of entities. For example, we use the ontology editor Protégé [Protégé, 2004] so as to manipulate an ontology in our experiments in section 4.4.4.

4.3 ONTOLOGY GUIDED DISTORTION CONTROL

4.3.1 RELATIONAL DATABASES AND ONTOLOGIES

As exposed in section 1.2.1, a relational database consists in a finite set of relations $\{R_i\}_{i=1,...,N_R}$ where one relation R_i contains a set of N unordered tuples $\{t_u\}_{u=1,...,N}$, each of which having M attributes $\{A_1, A_2, ..., A_M\}$. As defined, this data structure lacks of semantic information about the attributes meaning and links between different attributes' values in a tuple, pieces of information about the database content an ontology can thus offer. The question now is how to make interact these two structures.

As previously exposed, concepts in an ontology are linked by means of relationships 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 figure 4.5 such a mapping considering the example of one database containing pieces of information related to inpatient stay records (e.g., Diagnosis, Gender, Age and so on) which is linked to an ontology describing the semantic relations between diagnosis and age ranges.

In this example, 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 "≥ 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 relationships informs us 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". This can not be easily identified by means of a simple statistical analysis of the database. 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., $t_u.\{A_1, ..., A_{t-1}, A_{t+1}, ..., A_M\}$, or a subset of them.

More generally, concepts and relations in an ontology can be exploited in two manners: i) determinate the tolerated distortion for the attributes' values in each tuple. In the case of numerical attributes', the result of this process is a set of ranges of possible values the watermarked attribute can take. For categorical attributes, it gives the codes or words that can replace the original one; ii) identify the more suitable attributes to be watermarked according to their relationship degree with others. In this Ph.D. work we only have focused on the first approach, as exposed in the next sections.

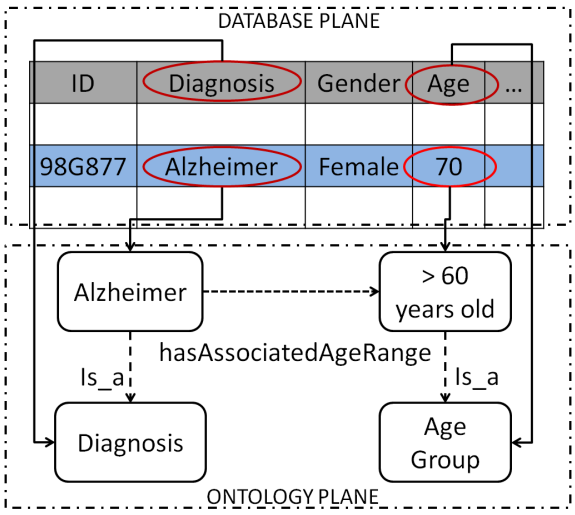


Figure 4.5: Existing connections between a relational database and an ontology. Dotted and dashed arrows represent ontological relations between concepts in the ontology. Solid arrows represent connections between attributes or attributes values and ontological concepts.

4.3.2 IDENTIFICATION OF THE LIMIT OF NUMERICAL ATTRIBUTE DISTORTION

As exposed above, in order to identify the maximum allowable distortion of the attribute values, we propose to use the concepts and relations of an ontology. In the case of a numerical attribute A_t, this limit in the tuple t_u corresponds to the range of possible values $t_u.A_t$ can take: $Rg_{t_u.A_t}$, under the semantic constraints of $S_{t_u.A_t}$ (i.e., the values of the other attributes of t_u). If we come back to the example in section 4.3.1, where $A_t =$"age", the value this one can take in a tuple t_u, $t_u.age$, belongs to an integer range $Rg_{t_u.age}$ imposed by the set $S_{t_u.age} =$"Alzheimer" (in our example $S_{t_u.age}$ is constituted of only one value, the one of the attribute Diagnosis in t_u). 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 intervals such as: $Rg_{t_u.A_t} = [A_{tmin_1}, A_{tmax_1}] \bigcup ... \bigcup [A_{tmin_{N_{rg}}}, A_{tmax_{N_{rg}}}]$ and $Rg_{t_u.A_t} \subseteq [A_{t,min}, A_{t,max}]$. This set of intervals can be identified by querying the ontology considering the other attributes' values in t_u, i.e., $S_{t_u.A_t}$. We illustrate such a situation in Fig. 4.6, where .... Our semantic distortion control approach is based on $Rg_{t_u.A_t}$ that will be used as reference so as to constraint the values watermarked attributes can take.

It is important to notice that the semantic distortion control we propose is complementary to any other statistical distortion control method. For instance, additionally to the ontology constraints one may also aim at preserving the correlation or the mutual information between attributes. In a more advanced construction, a global relational database watermarking scheme should associate semantic distortion control and statistics distortion control, combining for example our approach with optimization based technique suggested by Shehab *et al.* Shehab et al. [2008].

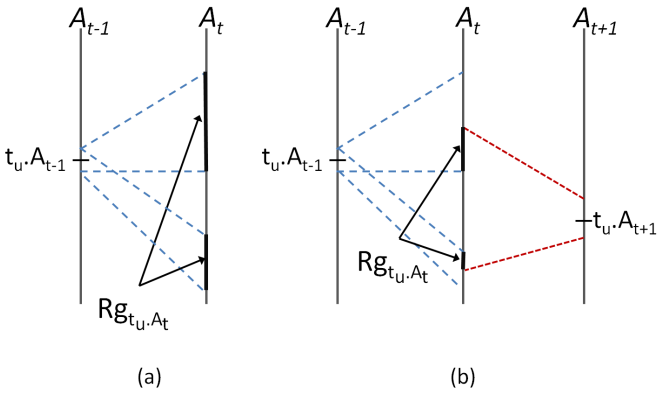


Figure 4.6: Identification of the possible range values $Rg_{t_u.A_t}$ of an attribute value $t_u.A_t$ depending on its relation with: a) a value $t_u.A_{t-1}$ in $S_{t_u.A_t}$; b) two values $t_u.A_{t-1}$ and $t_u.A_{t+1}$ in $S_{t_u.A_t}$. In the first case, $Rg_{t_u.A_t}$ corresponds to the union of different intervals. In the second case, the additional constraints imposed by the second value $t_u.A_{t+1}$ are represented as the intersection of the allowable ranges.

Minimization of the number of required queries In practice, the above process requires querying the ontology for each tuple in the database. In order to reduce such a complexity, we propose an ontology preprocessing stage which takes advantage of the fact that in general two numerical attributes have relationship in terms of range of values as illustrated in the previous paragraph where $Rg_{t_u.A_t}$ is the range of possible values of A_t in t_u under the constraint $S_{t_u.A_t}$ in t_u.

Let us generalize and look from the point the point of A_t. From the above, it appears that one range of its values is associated to a range or set of values of the numerical attributes and categorical attributes, respectively, in S_{A_t} (as illustrated in figure 4.6b). Returning to our example with A_t ="age" and as illustrated in figure 4.8, the range of ages $[60, 110]$ can be associated to a range of values $[117, 145]$ of the attribute A_{t+1} ="Systolic blood pressure" and to a set of values {Alzheimer, atherosclerosis,...} of the categorical attribute A_{t+2} ="Diagnosis".

In this context, the preprocessing stage we propose then to perform before database watermarking process consists in the construction of a correspondence table or mapping between ranges of A_t and ranges of attributes' values in S_{A_t}. Notice that a range of A_t is not necessarily associated to all the attributes in S_{A_t}. The construction of this table is based in the execution of inverse queries going from each value $\{Val_l\}_{l=1,...,L}$ of the domain of A_t to those of S_{A_t}. An example of such a query is given in figure 4.7, in the case of an age value 60. We assume that each "$diagnosis$" concept is associated to an "$age\ range$" concept (as depicted in figure 4.5); a concept that presents two attributes "*hasUpperLimit*" and "*hasLowerLimit*". The query returns the set of main diagnoses associated to numerical ranges the value 60 belongs to and the limits of these ranges. This results in a set of ranges Val_l may belong to under the constraints of the attributes' values in S_{A_t}. Figure 4.8 illustrates such a correspondence table or mapping for the attribute "age" (to be watermarked) in regard with the attributes "Systolic blood pressure" and "Diagnosis". Once the table constructed, it is used during the watermarking process and for one tuple t_u one just has to look for the values of $S_{t_u.A_t}$ in

```
SELECT ?max ?min ?diag
WHERE {
?vals hasUpperLimit ?max .
?vals hasLowerLimit ?min .
?vals hasDiagnosisAssociatedToAgeRange ?diag.
FILTER (?min< 60 && ?max > 60 )
}
```

Figure 4.7: Example of query taking an age value 60 as input. We are assuming that each "diagnosis" concept is associated to an "age range" concept; a concept that presents two attributes "hasUpperLimit" and "hasLowerLimit". The query returns the set of main diagnoses associated to numerical ranges the "age" value 60 belongs to and for each of these diagnoses, the limits of the associated range.

S_{Age} / R_{Age}	...	Systolic blood pressure	Diagnosis	...
[2,12]		-	Reye's syndrome, Infantile eczema, ...	
[0,10]		[110,124]	-	
[60,110]		[117,145]	Alzheimer, atherosclerosis, ...	

Figure 4.8: Example of a correspondence table that maps different value ranges of the attribute "age" depending on the ranges or sets of values of the other attributes in the relation, in particular the attributes "Systolic blood pressure" and "Diagnosis" in this example.

the columns so as to get all the possible ranges of values for the watermarked version of $t_u.A_t$. For instance, as seen in figure 4.8, for the values $t_u.Systolic\,blood\,pressure = 113$ and $t_u.Diagnosis =$"Reye's syndrome", we have $Rg_{t_u.age} = [2, 10]$.

4.3.3 EXTENSION OF THE PROPOSED APPROACH TO CATEGORICAL ATTRIBUTES

The above approach also works for categorical attributes. Let us thus consider a categorical attribute A_c (see section 1.2.1). As for numerical attributes, the values $t_u.A_c$ can take are semantically linked to the set $S_{t_u.A_c}$ of values of the other attributes of t_u, i.e., $t_u.\{A_1, ..., A_{c-1}, A_{c+1}, ..., A_M\}$ or a subset of them. However, due to the fact A_c is a categorical attribute, $Rg_{t_u.A_c}$ is no longer a range of values but a set of categorical values $Rg_{t_u.A_c} = Val_1, ..., Val_{N_{vals}}$. Again, $Rg_{t_u.A_c}$ can be identified from the ontology by querying it for the values in $S_{t_u.A_c}$. For example, if we consider $A_c =$"diagnosis" in regard to the attribute $A_{c+1} =$"gender", we know that for a tuple where $t_u.A_{c+1} =$"Male", $t_u.A_{c+1}$ can not be equal to "Multiple gestation".

From a practical point of view, in order to do not introduce new values in the database, we may be interested in letting values of A_c to be replaced only by other values already present in DB. To do so, each tuple is assigned to a category depending on the result of the ontology query. Then, values being part of the same category are interchangeable. This process is depicted in figure 4.9.

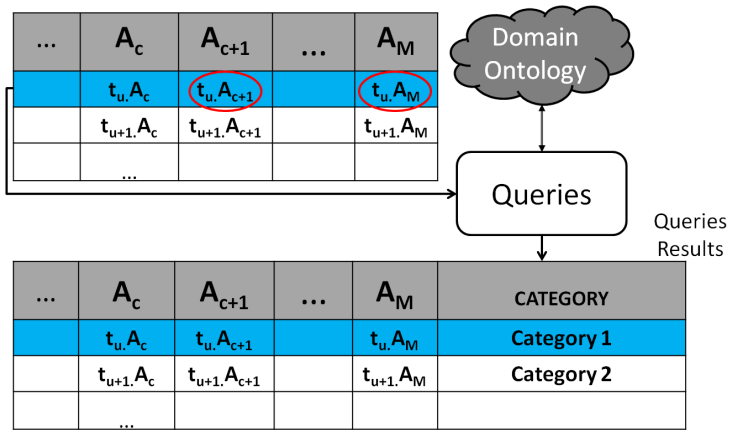


Figure 4.9: Identification of the possible values a categorical attribute can take in each tuple of DB. Tuples are assigned to categories depending on the result of the ontology queries. Then, values in the same category can be interchanged.

4.4 APPLICATION TO ROBUST WATERMARKING: QIM-BASED CIRCULAR HISTOGRAM MODULATION

In order to demonstrate the interest of the previous solution in the definition of a better robustness-imperceptibility trade-off, we propose an extension of the Quantization Index Modulation (QIM) to database watermarking. To our knowledge, this modulation, originally proposed by Chen and Wornell [Chen and Wornell, 1999] for images, has never been considered in database watermarking. Herein, we propose to use and adapt it so as to modulate the relative angle of the center of mass of circular histograms associated to groups of values of one numerical attribute of the relation. The use of this modulation offers us an important degree of freedom in the alteration of each individual value, contrarily to the previous schemes we proposed which alter attribute values with the same distortion.

4.4.1 QIM WATERMARKING AND SIGNALS

Quantization Index Modulation (QIM) is based on the quantization of the elements (samples, group of samples or transform coefficients) of a host signal according to a set of quantizers based on codebooks in order to embed the symbols of a message [Chen and Wornell, 1999]. More clearly, to each symbol s^i issued from a finite set $S = \{s_u^i\}_{u=0,\ldots,U}$ the QIM associates a codebook $\{C_{s_u^i}\}_{u=0,\ldots,U}$ such that:

$$C_{s_u^i} \bigcap C_{s_v^i} = \emptyset \text{ if } u \neq v \tag{4.1}$$

In order to embed the symbol s_u^i into one element X of the signal, this one is replaced by X_W which corresponds to the nearest element of X in the codebook $C_{s_u^i}$. This process can be seen as:

$$X_W = Q(X, s_u^i) \tag{4.2}$$

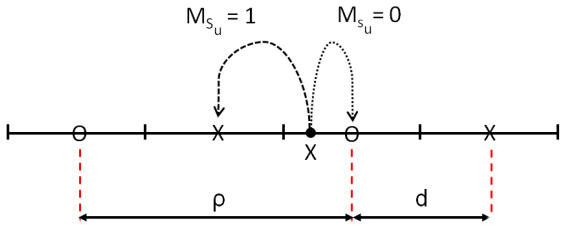


Figure 4.10: Example of QIM in the case where X is a scalar value for the embedding of a binary sequence. Codebooks are based on an uniform quantization of quantization step ρ. Cells centered on crosses represent $C_0(s_u^i = 0)$ whereas cells centered on circles represent $C_1(s_u^i = 1)$. $d = \rho/2$ establishes the measure of robustness to signal perturbations.

where the function Q returns the nearest element to X in $C_{s_u^i}$. Notice that the watermarking distortion corresponds to the distance between X and X_W. To exemplify this process, let us consider one pixel X of an image, which may take its values from a one-dimensional space $[0, 255]$. This scalar space is divided into non overlapping cells or intervals of equal size. Each cell is then related to only one codebook $\{C_{s_u^i}\}_{u=0,\ldots,U}$ so as to satisfy (4.1). Consequently, a symbol s_u^i has several representations in $[0, 255]$ and Q corresponds to a scalar quantizer. In the insertion process, if X belongs to a cell that encodes the desired symbol s_u^i, its watermarked version X_W corresponds to the centroid of this cell. Otherwise, X is replaced by the centroid of the nearest cell encoding s_u^i. In the extraction, the knowledge of the cell to which X_W belongs is enough to identify the embedded symbol. This process is illustrated in figure 4.10 in the case of a binary message, i.e., $s_u^i \in 0, 1$ and two codebooks C_0 and C_1 for which the cells are defined according to a uniform scalar quantization of quantization step ρ. In this example, X will be quantized to the nearest square or circle in order to encode s_u^i.

An extension of this approach whose purpose is to reduce the distortion is the Compensated QIM [Chen and Wornell, 1999], where a fraction of the quantization error is added back to the quantized value so as to better manage the watermark robustness/imperceptibility tradeoff.

Herein, we propose to modulate the angle of the vector associated to the center of mass of the circular histogram of an attribute in a group of tuples. The calculation of the center of mass was addressed in section 2.3.1 so we will not detail it in this section. In the sequel, in order to embed a symbol s^i into a group G^i, we modulate the value the angular position μ_i of the center of mass, which is generally called mean direction. In the next sections, we explain the QIM codebook construction and present our complete scheme.

4.4.2 Modified Circular QIM watermarking

4.4.2.1 Construction of the codebooks

For sake of simplicity, the considered message is a sequence of bits $S = \{0, 1\}$. Thus, two codebooks C_0 and C_1 are necessary. Another simplification we make in this work is that only one cell is associated to each codebook as illustrated in figure 4.11(b). Two questions

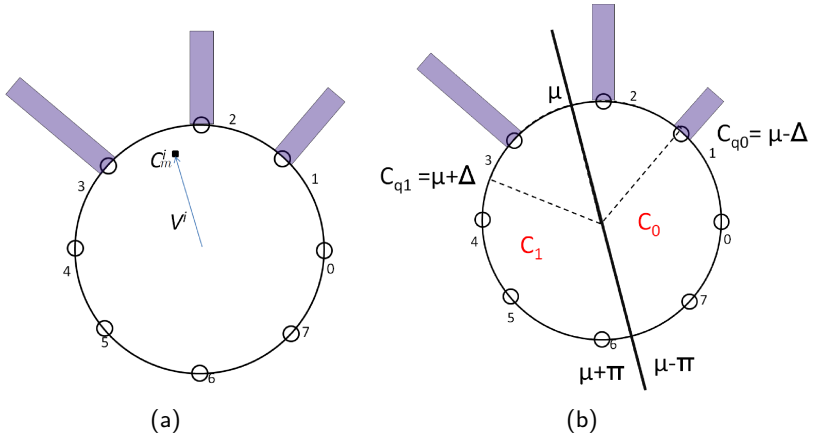


Figure 4.11: a) Histogram mapping of one group G^i onto a circle. The angle of the vector pointing its center of mass is modulated in order to embed one message symbol $s^i = \{0, 1\}$. b) C_{q0} and C_{q1} are the centroids of the cells of each codebook unique cell C_0 and C_1.

need then to be answered: the determination of the cell's boundaries and the position of their centroids.

Let us define μ as the mean direction of A_t calculated over all the tuples of the database. Based on the fact attribute circular histograms of tuple groups are all positioned around this mean direction, we decided to define the cells' frontiers as the intersection between μ and the unit circle as illustrated in figure 4.11(b). So in order to encode 0 or 1 the histogram will be rotated to the left or to the right of this frontier.

Unlike the previously presented QIM based on uniform scalar quantization, the centroids C_{q0} and C_{q1} of our cells C_0 and C_1 respectively do not correspond to the cell's center. This allows us to refine the imperceptibility/robustness trade-off. They are defined as:

$$\begin{aligned} C_{q0} &= \mu - \Delta \\ C_{q1} &= \mu + \Delta \end{aligned} \tag{4.3}$$

where Δ corresponds to the rotation angle shift, a user defined parameter that allows controlling the compromise robustness/distortion. As defined the maximum robustness is achieved when $\Delta = \frac{\pi}{2}$ while the maximum distortion is achieved when $\Delta = \pi$. The main difference with compensated QIM exposed above stands in two facts: 1) the cell center position is no longer at the cell's centroid and 2) the quantization error is not added back (see Section 4.4.1).

To sum up, each codebook is then associated to a one cell (see figure 4.11(b)) defined as:

$$C_0 = (\mu - \pi, \mu),\ C_1 = (\mu, \mu + \pi) \tag{4.4}$$

4.4.2.2 MESSAGE EMBEDDING AND DETECTION

Let us now consider the embedding of the binary sequence symbol $\{s^i\}_{i=0,\ldots,N_g-1} = \{0/1\}$ into the tuple group of tuples $\{G^i\}_{i=0,\ldots,N_g-1}$. As stated above, the mean direction value μ_i

of G^i is replaced by the centroid of the cell that encodes the value of s^i. This embedding process can be synthesized as:

$$\mu_i^w(\mu, s^i) = \mu + (2s^i - 1)\Delta \tag{4.5}$$

where μ_w is the watermarked mean direction, Δ is the rotation angle shift that allows the rotation of V^i so as to align it onto the cell centroid. This rotation is performed in the linear domain, i.e., on the attribute histogram, by modifying the attribute's values of certain tuples of G^i. We come back on this manipulation with more detail in Section 4.4.2.3. Regarding the message extraction stage, groups of tuples are reconstructed and angles μ_i^{det} calculated from each group. It is important to notice that in order to extract the message, the value of μ should be known of the detector so as to make it possible to reconstruct the dictionaries C_0 and C_1. μ can be sent to the reader as part of the watermarking key, as example.

The value of μ_i^{det} can differ from μ_i^w in case the watermarked database has been attacked. Whatever the situation, the cell to which μ_i^{det} belongs allows us to extract one bit s^i:

$$s^{i,det} = \begin{cases} 0 \text{ if } \mu_i^{det} \in C_0 \\ 1 \text{ if } \mu_i^{det} \in C_1 \end{cases} \tag{4.6}$$

4.4.2.3 LINEAR HISTOGRAM MODIFICATION

As previously exposed, a rotation of the center of mass vector V_i can be performed by changing the values of the attribute A_t in a certain number of tuples of the group G^i. For instance, if we call α the elementary angle between two consecutive bins of the circular histogram of A_t, $\alpha = \frac{2\pi}{L}$, modifying μ_i of α in the clockwise direction results in adding -1 to the attribute's value for every tuple in the group. Notice that $+1$ (resp. -1) is the minimal perturbation of an integer value.

In practice, we execute an iterative process so as to modify the attribute's values in G^i to rotate V^i onto C_{q0} or C_{q1}. In an iteration, the attribute values in the group are increased of $+1$ (resp. -1), under the distortion constraints established as exposed in section 4.3.2, so as to rotate positively (resp. negatively) μ_i^w to make it converge to C_{q1} (resp. C_{q0}). One can compute the number the number n_{mods} of tuples to be modified of the minimal perturbation $+1$ (resp. -1) in an iteration. n_{mods} depends on the elementary angle α and the number of tuples N_{G^i} in the group:

$$n_{mods} = round(\frac{|\mu_i^w - C_{q0}|}{\alpha} N_{G^i}) \tag{4.7}$$

As exposed, after each iteration the distance between μ^w and C_{q1} (resp. C_{q0}) decreases. However because A_t is an integer attribute, being at least modifiable of ± 1, μ_i^w may not reach the codebook cell centroid after an infinite number of iterations. This is why we introduced a user defined parameter ε, such as our algorithm stops when $|\mu_i^w - C_{q0}| < \epsilon$. Notice that the lowest value ϵ can take depends on the attribute. Indeed, due to the fact A_t is an integer its circular histogram center of mass can be rotated of a minimal angle $\frac{2\pi}{L}\frac{N_g}{N} = \alpha\frac{N_g}{N}$ (i.e., a

modification of ± 1 of one individual attribute's value). This results in a minimum value of ϵ for this attribute of a half of this rotation, $\epsilon_{min} = \frac{\alpha}{2}\frac{N_g}{N}$.

Notice that this iterative process could be replaced by an optimization based scheme, considering the semantic distortion constraints and the value of ϵ in order to obtain the perturbation to be introduced to each tuple in the group.

4.4.3 THEORETICAL ROBUSTNESS PERFORMANCE

In this section we theoretically evaluate the performance of our scheme in terms of robustness against two most common database attacks: tuple deletion and tuple insertion. As we will show, robustness depends on the number of groups N_g, the rotation angle shift Δ (on which depends the codebook cell centroids), the statistical distribution of the watermarked mean directions μ_i^w as well as on the strength of the database modifications, i.e percentage of deleted/inserted tuples. The calculations we provide in the following make use of some of the results exposed in section 2.3.4 of chapter 2.

Let us thus consider the watermarking of one numerical attribute A_t in a database by means of the above scheme, where two unique cell codebooks C_0 and C_1 with centroids $C_{q0} = \mu - \Delta$ and $C_{q1} = \mu + \Delta$ are used so as to embed a sequence S of uniformly distributed symbols $s^i \in \{0, 1\}$, respectively. The result of such an insertion process on the normal distribution of the original mean direction μ^i (see figure 4.12 and section 4.4.2) is illustrated in figure 4.13 which gives the p.d.f of the watermarked angles μ_i^w. One can easily identify the centroids of the codebook cells as well as the frontier between the two cells (or codebooks) established by μ.

As exposed in section 4.4.2, the modulation of μ_i is performed by introducing a controlled distortion of the values of A_t. This modification is carried out by means of an iterative process that stops when $|\mu_i^w - C_{q0}| < \epsilon$ (resp. $|\mu_i^w - C_{q1}| < \epsilon$) with the error ϵ fixed by the user. Contrarily to the QIM, the p.d.f distribution of μ_i^w does not present only two peaks on the cells' centroids, but two Gaussians centered in C_{q0} and C_{q1}; Gaussians with a variance that depends on the error ϵ, as seen in figure 4.13.

Performance in terms of robustness of our scheme depends on the probability that a group of tuples changes of embedded symbol after an attack. We propose to compute these probabilities considering two common database attacks or modifications: tuple addition or tuple removal. To do so, we need to express their impact on the p.d.f of the watermarked angles, i.e., of the random variable μ_i^w given in figure 4.13.

Notice that for the sake of simplicity, we consider in the sequel that the error ϵ, with $|\mu_i^w - C_{q0}| < \epsilon$ (resp. $|\mu_i^w - C_{q1}| < \epsilon$), equals zero. We thus make the hypothesis that the p.d.f $f_{\mu_i^w}(\mu_i^w)$ of μ_i^w is such as:

$$f_{\mu_i^w}(\mu_i^w) = \begin{cases} \frac{1}{2} \ if \ \ \mu_i^w = \mu - \Delta \\ \frac{1}{2} \ if \ \ \mu_i^w = \mu + \Delta \\ \ \ 0 \ otherwise \end{cases} \tag{4.8}$$

Under this hypothesis, the probability of symbol error $\mathbb{P}_e$ is reduced increasing thus the theoretical robustness of our scheme. As we will see in the experimental section, ϵ can be

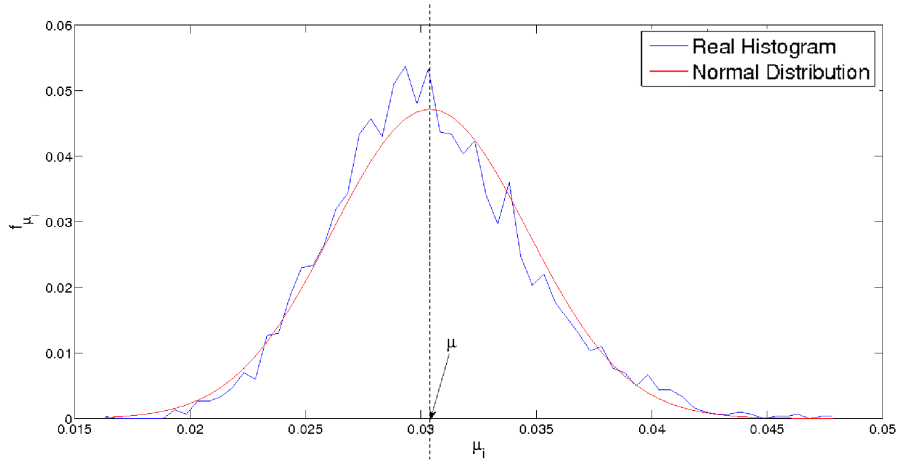


Figure 4.12: Distribution of the mean direction μ_i for an exponentially distributed numerical attribute taking its values in $[0, 707]$ ($L = 708$) with a number of groups $N_g = 500$. As shown, the real distribution obtained by means of the normalized histogram perfectly fits a normal distribution with the theoretically calculated statistical moments.

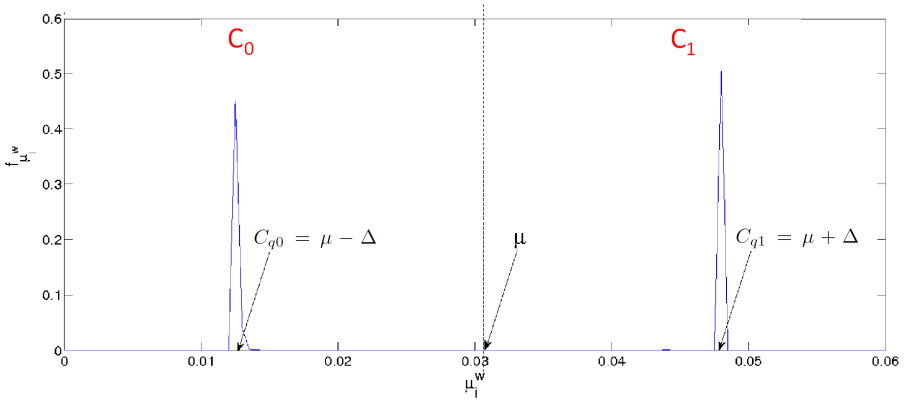


Figure 4.13: μ_i^w distribution after the embedding process for an exponentially distributed numerical attribute taking its values in $[0, 707]$ ($L = 708$) with $N_g = 500$ and $\Delta = 2\frac{2\pi}{L} = 0.0177$.

made small enough to reach such performance at the price however of time computation which depends of the number of iteration of our algorithm (see section 4.4.2).

The p.d.f $f_{\mu_i^w}(\mu_i^w)$ of the watermarked angles can be expressed in terms of the conditional p.d.f of μ_i^w given the quantization cell C_0 or C_1 they belong to. As the symbols of the sequence S are uniformly distributed, we can suppose that angles belong to each cell with equal probability $\mathbb{P}_{C_0} = \mathbb{P}_{C_1} = \frac{1}{2}$. Then, we have:

$$f_{\mu_i^w}(\mu_i^w) = \frac{1}{2} f_{\mu_i^w}(\mu_i^w | \mu_i^w \in C_0) + \frac{1}{2} f_{\mu_i^w}(\mu_i^w | \mu_i^w \in C_1) \tag{4.9}$$

The conditional p.d.f of μ_i^w given cell correspond to:

$$f_{\mu_i^w}(\mu_i^w | \mu_i^w \in C_0) = \begin{cases} 1 \ if \ \mu_i^w = C_{q0} \\ 0 \ otherwise \end{cases} \tag{4.10}$$

$$f_{\mu_i^w}(\mu_i^w|\mu_i^w \in C_1) = \begin{cases} 1 \ if \ \mu_i^w = C_{q1} \\ 0 \ otherwise \end{cases} \tag{4.11}$$

Notice that under the assumption of $\epsilon = 0$, each conditional p.d.f can be seen as a normal distribution centered in the cell centroid with a variance equal to zero.

4.4.3.1 DELETION ATTACK

In this attack, N_d tuples are randomly eliminated. Based on the fact that tuples are uniformly distributed into N_g groups (see section 1.3.2.2), we can assume that each group G^i loses in average $\frac{N_d}{N_g}$ tuples. This reduction impacts the accuracy of μ_i^w which by definition is an estimator of the attribute mean direction in the group G^i (see section 4.4.2). If we consider the previously exposed conditional distributions of μ_i^w given the cell they belong to, this attack only modifies their variances, leaving their means unchanged. We can model this variance modification as the addition of a centered normally distributed random variable ψ to the value of μ_i^w, such as $\mu_i^{del} = \mu_i^w + \psi$.

Due to the fact the tuples in a group have not been watermarked with the same amplitude or distortion, the impact of the deletion attack variably depends on each tuple. This makes incoherent to theoretically calculate the value of the variance σ_ψ^2 of ψ and consequently the probability of symbol error $\mathbb{P}_e$. However, we can obtain an upper bound of P_e considering the case where all the tuples in a group are modified with the same maximum distortion. The variance of ψ is then obtained from eq. 4.12, with σ_s^2 (see section 2.3.4) calculated over all the tuples in the database.

$$\sigma_\psi^2 = \begin{cases} \frac{\sigma_s^2}{\frac{N-N_d}{N_g}R^2}, & \text{if } R \leq 0.85 \\ \frac{1-R^2}{\frac{N-N_d}{N_g}}, & \text{otherwise} \end{cases} \tag{4.12}$$

The resulting conditional p.d.f $f_{\mu_i^{del}}(\mu_i^{del}|\mu_i^w \in C_0)$ (resp. C_1), i.e., the p.d.f of μ_i after watermarking and deletion attack, is a normal density function given by:

$$\begin{array}{l} f_{\mu_i^{del}}(\mu_i^{del}|\mu_i^w \in C_0) \sim \mathcal{N}(C_{q0}, \sigma_\psi^2) \\ f_{\mu_i^{del}}(\mu_i^{del}|\mu_i^w \in C_1) \sim \mathcal{N}(C_{q1}, \sigma_\psi^2) \end{array} \tag{4.13}$$

4.4.3.2 INSERTION ATTACK

In this situation the attacker inserts N_i tuples. Herein, we assume that added attribute values follow the same distribution as the original un-watermarked attribute A_t. As previously, the fact a cryptographic hash function is used to construct groups of tuples (see section 1.3.2.2) allows us to consider that new tuples are uniformly distributed among these groups $\{G^i\}_{i=1,...,N_g}$. As a consequence, such an attack can be modeled by a mixture of two populations: the watermarked tuples and the added un-watermarked tuples with mixture proportion parameters

p_1 and p_2 such as $p_2 = 1 - p_1$ with $p_1 = \frac{N}{N+Ni}$, where N is the number of tuples in the original database.

The conditional p.d.f $f_{\mu_i^{ins}}(\mu_i^{ins}|\mu_i^w \in C_0)$ (resp. C_1), i.e., the p.d.f of μ_i after watermarking and tuple insertion, is a normal density function. Its mean $\bar{\mu_i}^{ins,0}$ (resp. $\bar{\mu_i}^{ins,1}$), which corresponds to the conditional mean given that ${\mu_i}^w \in C_0$ (resp. C_1), and its variance $\sigma_{\mu_i^{ins}}$ can be calculated as:

$$\begin{aligned} \bar{\mu_i}^{ins,0} &= \bar{E}[{\mu_i}^{ins}|\mu_i^w \in C_0] = p_1 C_{q0} + p_2\mu \\ \bar{\mu_i}^{ins,1} &= \bar{E}[{\mu_i}^{ins}|\mu_i^w \in C_1] = p_1 C_{q1} + p_2\mu \end{aligned} \tag{4.14}$$

$$\sigma^2_{\mu_i^{ins}} = \begin{cases} p_2^2 \frac{\sigma_s^2}{\frac{N_i}{N_g} R^2}, & \text{if } R \leq 0.85 \\ p_2^2 \frac{1-R^2}{\frac{N_i}{N_g}}, & \text{otherwise} \end{cases} \tag{4.15}$$

The cell conditional p.d.f are obtained as:

$$\begin{aligned} f(\mu_i^{ins}|\mu_i^w \in C_0) &\sim \mathcal{N}(\bar{\mu_i}^{ins,0}, \sigma^2_{\mu_i^{ins}}) \\ f(\mu_i^{ins}|\mu_i^w \in C_1) &\sim \mathcal{N}(\bar{\mu_i}^{ins,1}, \sigma^2_{\mu_i^{ins}}) \end{aligned} \tag{4.16}$$

PROBABILITIES OF ERROR

The robustness of our scheme is characterized by the symbol error probability $\mathbb{P}_e$, the probability the symbol of a group changes after an attack. $\mathbb{P}_e$ can be determined through an hypothesis testing problem with the following set of hypothesis:

- H_0 corresponds to the case $s_i = 0$, i.e., $\mu_i^w \in C_0$.
- H_1 corresponds to the case $s_i = 1$, i.e., $\mu_i^w \in C_1$.

The probability the watermark reader returns the wrong symbol value, i.e., $\mathbb{P}_e$, results from the acceptance of H_0 (resp. H_1) when the correct hypothesis is H_1 (resp. H_0). Thus, $\mathbb{P}_e$ is calculated as:

$$\mathbb{P}_e = \frac{1}{2} Pr(H_1|H_0) + \frac{1}{2} Pr(H_0|H_1) \tag{4.17}$$

$\mathbb{P}_e$ can be refined depending on the database attack::

$$\begin{aligned} \mathbb{P}_{e,del} &= \frac{\int_{C_1} f(\mu_i^{del}|H_0) d\mu_i^{del} + \int_{C_0} f(\mu_i^{del}|H_1) d\mu_i^{del}}{2} \\ \mathbb{P}_{e,ins} &= \frac{\int_{C_1} f(\mu_i^{ins}|H_0) d\mu_i^{ins} + \int_{C_0} f(\mu_i^{ins}|H1) d\mu_i^{ins}}{2} \end{aligned} \tag{4.18}$$

where $\mathbb{P}_{e,del}$ and $\mathbb{P}_{e,ins}$ correspond to the probability of symbol error under a deletion and an insertion attack respectively

4.4.4 EXPERIMENTAL RESULTS

4.4.4.1 EXPERIMENTAL DATASET

The following experiments have been conducted on a test database constituted of one relation of 508000 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 stay duration (*dur_stay*), the attribute GHM (patient homogeneous group), the attribute ICD10 principal diagnosis and several other data useful for statistical analysis of hospital activities. If *age* and *dur_stay* are numerical attributes, GHM and ICD10 are categorical attributes. 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 diagnostics, the stay duration, and several other elements.

In order to constitute the groups of tuples (see section 1.3.2.2), the attributes *id_stay* and *id_hospital* were considered as the primary key. Two numerical attributes were considered for message embedding and watermarked independently: patient age (*age*) and stay duration (*dur_stay*). These attributes were chosen because of the specificity of their distributions. The attribute *age* is slightly uniformly distributed, while the attribute *dur_stay* presents an exponential distribution concentrated over the lower values of its domain.

4.4.4.2 DOMAIN ONTOLOGY

In this experiment, for sake of simplicity, we summed up the domain ontology to the two sets of relations that exist between the attributes GHM, age and $stay\ duration$ and between the attributes $ICD10\ principal\ diagnosis$ and age. These two sets have been selected due their different natures. In the first, GHM, age and $stay\ duration$ are related by means of a formula, while in the second set, relations between attributes are issued from semantic rules established by DATIM (e.g. a diagnosis code is highly unlikely for of range of age).

If we go further into the details, our ontology represents thus a subset of the rules associated to the calculation of the GHM codes. For instance, as depicted in figure 4.16, the code "$25Z033$: *VIH related disease, age lower than 13 years old, level 3*" is related to the group of ages "*Less than 13 years old*" which corresponds to a numerical range of values $(0, 12)$. To integrate this rule in our ontology, we build a hierarchy of classes with a top class named "GHM" which as a parent class. From this top class, a set of subclasses representing each group of major diagnosis (e.g., nervous system diseases (01), eye diseases (02)) was constructed. Each of these classes may present from one to three children classes representing the nature of GHM codes (e.g., surgical (C), medical (M) and so on). iIndividuals or instances of each of these classes were created in order to represent one GHM code, (e.g., $25Z033$). Another class "age" has been created. Its instances represent all possible age ranges (e.g., "*Less than 13 years old*"). Age and GHM code instances were associated by means of the

relation "*hasAssociatedAgeRange*". Notice that these "age ranges" are known *a priori* and one instance presents two integer attributes: "*hasUpperLimit*" and "*hasLowerLimit*". For example, the instance "*Less than 13 years old*" has as attribute "*hasUpperLimit*" and "*hasLowerLimit*" values of 12 and 0, respectively. The same procedure has been followed to represent the rules associating the GHM codes and the hospital stay durations. Notice that the example depicted in 4.14 has been simplified, in order not to represent all the "GHM" hierarchy.

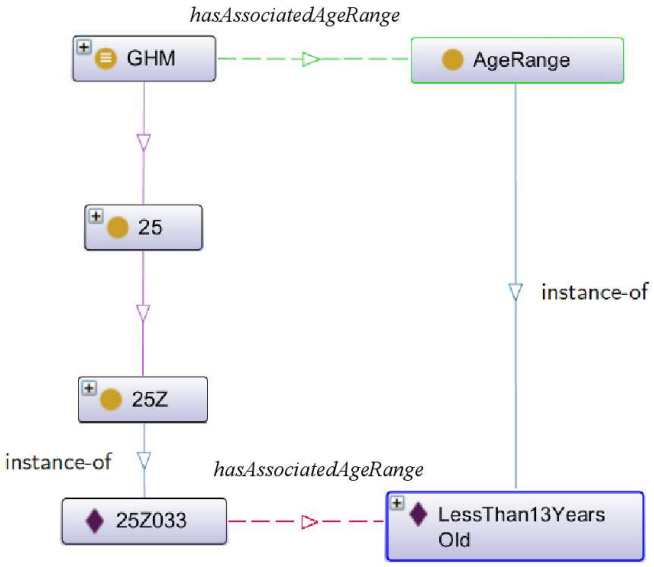


Figure 4.14: Extract of the hierarchy "GHM" in the domain ontology. Hierarchical and "instance-of" relations are represented by vertical lines. Relation "*hasAssociatedAgeRange*" is represented by an horizontal dashed line.

Regarding the second set of relations of our ontology, they correspond to semantic relationships between ICD10 codes and age ranges extracted from DATIM. In fact, among the different information DATIM provides, it establishes association links between ICD10 codes and Age ranges. As for the GHM codes, a hierarchy of classes was first constructed and instances were created at the lower class level. Lower level which represents each ICD10 code (e.g., "*A00.9 - cholera, unspecified*").These codes were associated to instances of the class "*age range*", described previously, through a relationship: "*ForbiddenAgeRange*". As example, the ICD10 code "*C58 - Malignant neoplasm of placenta*" is forbidden for patients under 9 years old and over 65 years old. Thus, the instance "*C58*" is associated to two age instances "*Less than 9 years old*" and "*More than 65 years old*". This example is illustrated in figure 4.15.

In our implementation, the domain ontology was constructed in Protégé [Protégé, 2004], an open source ontology editor, and queried by means of the SPARQL query language. The choice on the query language was made for a practical reason, due to its similarity to the database query language SQL. Moreover, SPARQL queries can be easily constructed in Java by means of the Apache Jena[2] framework and subsequently implemented in Matlab. An

[2] `https://jena.apache.org/`

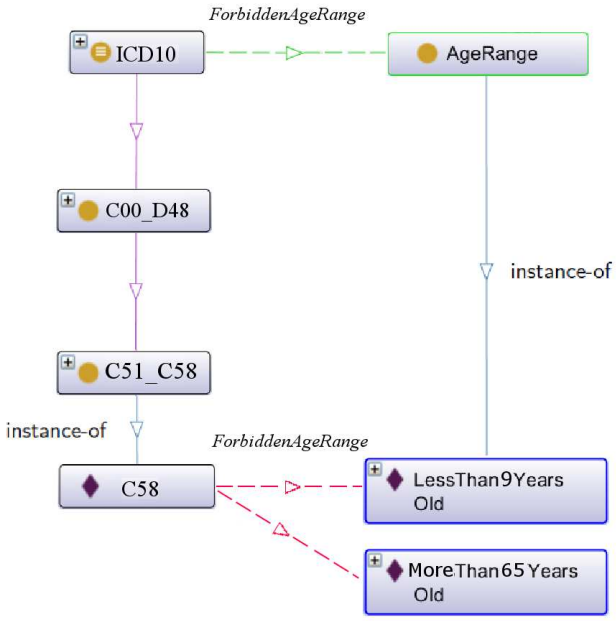


Figure 4.15: Extract of the hierarchy "$ICD10$" in the domain ontology. Hierarchical and "instance-of" relations are represented by vertical lines. Relation "*ForbiddenAgeRange*" is represented by a dashed lines.

example of SPARQL query is illustrated in figure 4.7 (see section 4.3.2). In this query, we request for all the ICD10 codes associated to age ranges to which an age value 60 belongs to and for each of these codes, the upper and lower limits of the associated range.

Illustrative example of the ontology interest An example presenting the advantage of controlling semantically the database distortion by means of an ontology is given in figure 4.16. 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 from table a) and b), 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.

4.4.4.3 PERFORMANCE CRITERIA

The performance of our scheme is evaluated in terms of statistical distortion, robustness against tuple suppression and insertion attacks and complexity. In order to get a global vision of the variation of the attribute's distribution, we quantify its statistical distortion through the variations of the attribute's mean and standard deviation, the Kullback-Leibler divergence

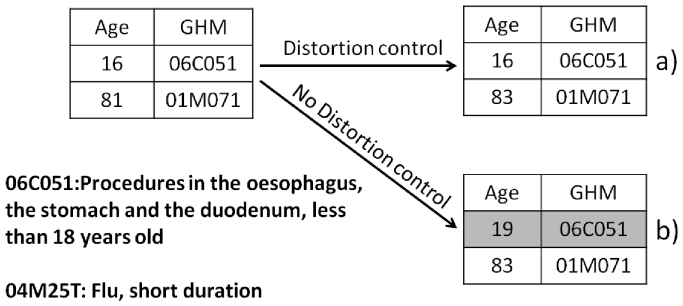


Figure 4.16: Example of modification of two tuples taking and not into account semantic distortion limits. Semantically incorrect tuples are highlighted.

(see eq. 4.19) and the mean absolute error (MAE) (see eq. 4.20) between histograms of the attribute before and after watermark embedding. If we call h_{A_t} and $h_{A_t^{wat}}$ the histograms of the original attribute A_t and of its watermarked version A_t^{wat} respectively, we have:

$$D_{\mathrm{KL}}(h_{A_t} \| h_{A_t}^{wat}) = \sum_{l=0}^{L-1} \ln\left(\frac{h_{A_t}(l)}{h_{A_t^{wat}}(l)}\right) h_{A_t}(l) \tag{4.19}$$

$$\mathrm{MAE} = \frac{1}{L \cdot N} \sum_{l=0}^{L-1} \left| h_{A_t}(l) - h_{A_t^{wat}}(l) \right| \tag{4.20}$$

We recall that the attribute's domain of A_t corresponds to the integer range $[0, L-1]$ and N is the total number of tuples in the database. Robustness is evaluated by means of the bit error rate (BER), i.e., the probability the value of an extracted symbol is incorrect after attacks, we compute as:

$$\mathrm{BER} = \frac{\sum_{i=1}^{N_g} (s^i \oplus s^{i,det})}{N_g}; \tag{4.21}$$

Complexity is established as the computation time or more clearly, the amount of time taken by the execution of the insertion and the extraction processes. It is important to notice that all of the following results are given in average after 30 random simulations with the same parameterization but different group configuration.

4.4.4.4 STATISTICAL DISTORTION RESULTS

As stated above, we evaluate the statistical database distortion through the variations of the the mean, the standard deviation, the Kullback-Leibler divergence (D_{KL}) and the histogram Mean Absolute Error (MAE) of the attribute. These variations mainly depend on the rotation angle shift Δ of the center of mass and of the number of tuples per group. Table 4.1 provides the results we obtained for the attribute *Age* for a different number of groups $N_g = 100, 500, 1000$ and different values of Δ which are multiples of the elementary angle α (see Section 4.4.2). We recall that our test database contains $N = 508000$ tuples. The original mean and standard deviation values of *Age* are 50.078 and 25.236 respectively. As it can be seen they tend to differ from their original value with the value of Δ but the variation remains below 1%. It

is the same for the Kullback-Leibler divergence and the histograms MAE which quantify the distortion of the attribute's distribution. These measures increase with the number of groups but the augmentation is not significant. Thus, if our scheme minimizes semantic distortion, it also induces low statistical distortions and may not bias most data-mining operations.

In order to evaluate the gain of performance in terms of distortion when considering our semantic distortion control, the same experiments were conducted applying our scheme with the same parameterization but without the ontology. Obtained results are given in Table 4.2. As we can see, without semantic constraints the distortion is at least 4 times greater whatever the criteria.

Table 4.1: Introduced statistical distortion in terms of mean, standard deviation, D_{KL} and histograms MAE for the attribute Age, considering a test database of N= 508000 tuples for different number of groups and various rotation angle shifts Δ. α is the elementary angle (see section 4.4.2). Moments' variations are indicated in parenthesis.

Nb. groups		$\Delta = \alpha$	$\Delta = 2\alpha$	$\Delta = 3\alpha$
Mean	100	50.113 (0.06%)	50.159 (0.1%)	50.153 (0.15%)
	500	50.138 (0.11%)	50.164 (0.17%)	50.204 (0.25%)
	1000	50.158 (0.16%)	50.194 (0.23%)	50.227 (0.29%)
Std. dev.	100	25.24 (0.01%)	25.266(0.11%)	25.306 (0.27%)
	500	25.233 (0.01%)	25.258 (0.08%)	25.304 (0.26%)
	1000	25.222 (0.05%)	25.25 (0.05%)	25.295 (0.23%)
D_{KL}	100	0.001	0.002	0.004
	500	0.001	0.002	0.004
	1000	0.001	0.002	0.004
MAE	100	2.36 10^{-4}	2.73 10^{-4}	4.54 10^{-4}
	500	2.54 10^{-4}	3.55 10^{-4}	4.82 10^{-4}
	1000	3 10^{-4}	3.73 10^{-4}	4.82 10^{-4}

Table 4.2: Introduced statistical distortion in terms of mean, standard deviation, D_{KL} and histograms MAE with no semantic constraints

Nb. groups		$\Delta = \alpha$	$\Delta = 2\alpha$	$\Delta = 3\alpha$
Mean	100	50.201 (0.24%)	50.326 (0.49%)	50.4 (0.64%)
	500	50.238 (0.32%)	50.324 (0.49%)	50.519 (0.88%)
	1000	50.307 (0.45%)	50.394 (0.63%)	50.514 (0.87%)
Std. dev.	100	25.2 (0.14%)	25.186 (0.19%)	25.187 (0.19%)
	500	25.179 (0.22%)	25.175 (0.24%)	25.189 (0.18%)
	1000	25.156 (0.31%)	25.159 (0.3%)	25.175 (0.24%)
D_{KL}	100	0.018	0.039	0.072
	500	0.019	0.034	0.063
	1000	0.024	0.034	0.055
MAE	100	6.11 10^{-4}	9.07 10^{-4}	11 10^{-4}
	500	6.41 10^{-4}	8.62 10^{-4}	11.23 10^{-4}
	1000	7.27 10^{-4}	8.75 10^{-4}	11.56 10^{-4}

4.4.4.5 ROBUSTNESS RESULTS

Robustness or the symbol error rate of our scheme against tuple deletion and insertion attacks mainly stands on the rotation angle shift Δ and on the number of tuples per group, established by the number of groups N_g as the number of tuples in the database N is fixed. In this experiment, the attribute *Age* of our test database was watermarked with an uniformly distributed binary message S considering different values of Δ (as multiples of the elementary angle α) and N_g. These watermarked databases were then attacked by tuple insertion or deletion. The degree of the attack is measured in percentage from 20% to 99%, i.e., the percentage of tuples added to or deleted from the protected database. Also a fixed value of $\epsilon = 0.0001$ was considered. Herein, we confront experimental to theoretical performance we established in section 4.4.4.5.

In figure 4.17, we show the bit error rate (BER) we achieved in the case of a deletion attack for two values of N_g and the lowest rotation angle shift value, i.e., $\Delta = \alpha$. As it can be seen, experimental curves are lower than the theoretical BER upper limit we defined in Section 4.4.3.1. However, they tend to this limit along with the increase of the degree of the deletion attack. Figure 4.18 provides more tuple deletion attack robustness results making varying Δ and N_g. Obviously, the BER increases along with the degree of the attack but also with the number of groups. This is due to the limited size of the database, the more the number of group increases, the more the number of tuples per group decreases. In general, decreasing the number of tuples per group by mean of a deletion attack or a high number of groups directly impacts the mean direction estimation and consequently the robustness of the scheme. At the same time, the higher the value of Δ, the further the codebook cell centers are (see section 4.4.2) and the greater the robustness is.

Similar experiments were conducted regarding the tuple addition attack. As in section 4.4.3.2, where we theoretically established the BER, new added tuple attributes' values follow the original distribution of the attribute. Results are provided in figure 4.19. First, it can be seen that experimental results fit theoretical ones. A little error can be seen, which is related to the hypothesis we made in section 4.4.4.5 with ϵ equals 0. Beyond, as in the previous attack, the BER decreases with the increase of Δ, and increases along with the number of groups. Notice also, that our scheme better resists to the addition attack than the deletion attack. This is due to the fact that the addition attack only increases the variance of the p.d.f. of μ_i^w while the deletion attack impacts also its mean.

A third experiment was conducted so as to evaluate the influence of the attribute distribution itself. To do so attributes *age* and *dur_stay* were watermarked using the same number of groups $N_g = 1000$ and rotation angle shift $\Delta = \alpha$. As depicted in figure 4.20, the BER obtained for the attribute *dur_stay* is the lowest one due to its low dispersion around its mean values, which makes its mean direction more stable faced to the addition or suppression of tuples.

Finally, we have also evaluated the robustness of our scheme against attribute's values modifications, modifications performed in two different manners: i) Gaussian noise addition of standard deviation $\sigma = 2$; ii) uniform noise addition of amplitude in [-4; 4]. Considering the attribute age with a number of groups $N_g = 1000$ and $\Delta = \alpha$, we obtained a BER≈ 0.04 in the first case and BER≈ 0.09 in the second, even when 99% of tuples were modified. The

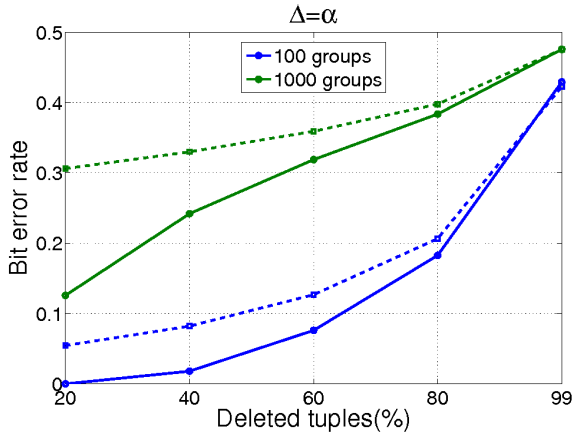


Figure 4.17: Tuple deletion attack - Bit error rate obtained with for the attribute *Age* considering $\Delta = \alpha$ and $N_g = 100$ and 1000 groups. Theoretical and experimental results are indicated by a dashed and solid lines, respectively.

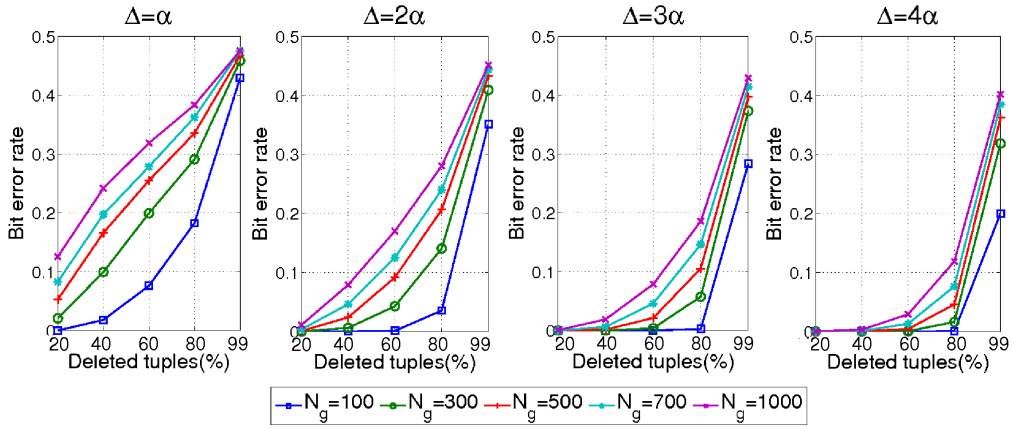


Figure 4.18: Bit error rate for the attribute *Age* with different rotation angle shifts Δ taking $N_g = 100, 300, 500, 700$ and 1000 groups for a tuple deletion attack.

BER decreases for a lower N_g and for a higher Δ. We can then conclude that our scheme is highly robust against attribute value modifications.

4.4.4.6 COMPUTATION TIME

The computation time of our scheme depends on the construction of groups of tuples, on the identification of the allowable distortion and the watermark embedding/extraction processes. The complexity increases along with the number of tuples in the relation, i.e., N. However, the complexity of the allowable distortion identification task is in addition dependent on the number of attributes considered for watermarking and of the number of attributes semantically connected with them (complexity of the ontology queries).

In this experiment, the attribute *Age* was considered for embedding. It is important to

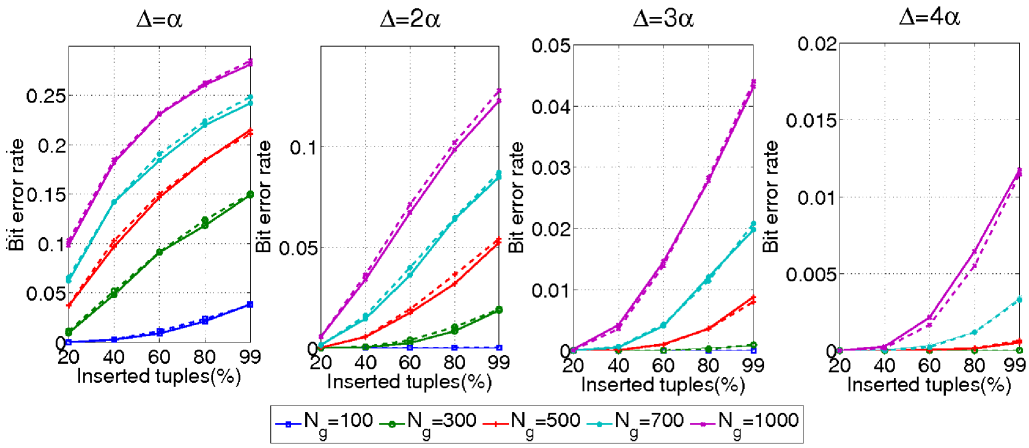


Figure 4.19: Bit error rate for the attribute *Age* with different rotation angle shifts Δ taking $N_g = 100, 300, 500, 700$ and 1000 groups for a tuple insertion attack. Theoretical and experimental results are indicated by a dashed and solid lines, respectively.

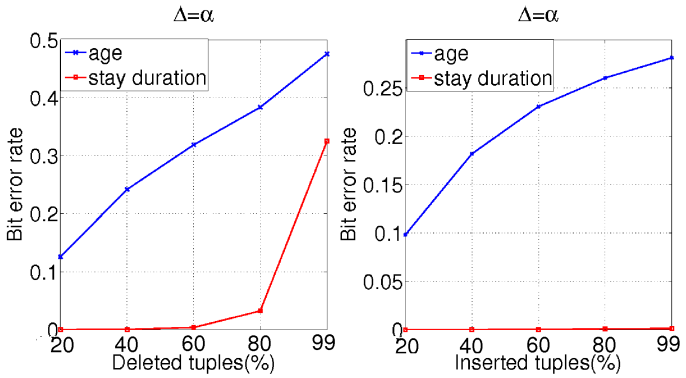


Figure 4.20: Bit error rate for the attributes *age* and *dur_stay* with $N_g = 1000$ and $\Delta = \alpha$ considering the tuple deletion attack (left) and the tuple insertion attack (right).

notice that all the following computation times are those of an implementation of our scheme made with Matlab® running on a Intel® Xeon® E5504 at 2Ghz with 3GB of physical memory and four cores. Table 4.3 evaluates the elapsed time for the first two tasks for several values of N. It can be seen that the time increases linearly with N and with the number of attributes semantically connected with the one to be watermarked.

The complexity of the insertion process depends on the number of groups N_g, on the value of Δ and the value of the error ϵ manually fixed by the user. The smaller ϵ, the more iterations our scheme will have to make to reach this value. On its side, the extraction stage complexity is essentially related with the number of groups. Indeed, once the groups reconstituted one just has to interpret their center of mass values to decode the message. Results for both stages processes are given in Table 4.4. It can be observed that the extraction complexity increases with the number of groups only. The insertion computation time increases also with Delta. Indeed, our scheme iteratively modified the tuples of one group so as to reach the codebook cell centroid which encodes the desired bit under an epsilon value constraint.

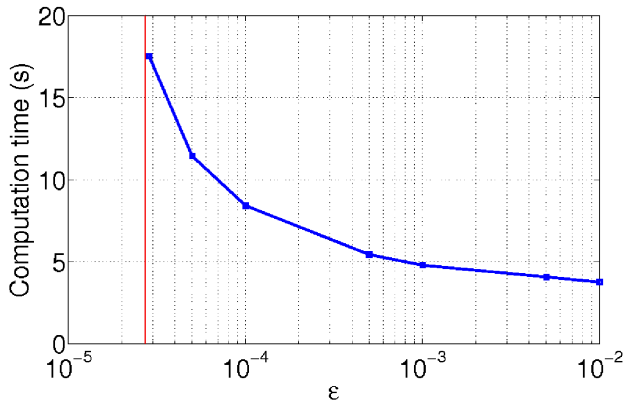


Figure 4.21: Computation time for the attribute *Age* with $\Delta = \alpha$ taking $N_g = 500$ and several values of ϵ. The vertical solid line represents the asymptotic value of ϵ for this attribute..

In order to evaluate the influence of the value of ϵ on the computation time, we insert a watermark into the attribute *age* considering $N_g = 500$ and $\Delta = \alpha$ and several values of ϵ. Results are depicted in figure 4.21. As expected, the computation time inversely grows with ϵ until reaching an asymptote in $\epsilon = 2.79 \cdot 10^{-5}$. This later value is directly related to the definition of the attribute *age*, an integer variable (see section 4.4.2.3).

Table 4.3: Computation time for the identification of semantic distortion limits and the construction of groups using Matlab®.

Identification	$N = 200000$	$N = 400000$	$N = 500000$
One attribute	3.54s	4.63s	4.98s
Two attributes	6.97s	9.15s	10.72s
Group creation	$N = 200000$	$N = 400000$	$N = 500000$
	67s	134s	170s

Table 4.4: Computation time for the insertion and the detection stages for the attribute age with $\epsilon = 0.0001$ using Matlab®.

Insertion	$N_g = 100$	$N_g = 500$	$N_g = 700$
$\Delta = \alpha$	2s	8.5s	11.3s
$\Delta = 2\alpha$	2.7s	10.4s	14s
$\Delta = 3\alpha$	3.5s	12.35s	17s
$\Delta = 4\alpha$	4.05s	14.2s	19.6s
Detection	$N_g = 100$	$N_g = 500$	$N_g = 700$
	0.3s	1s	1.3s

4.4.5 PERFORMANCE COMPARISON RESULTS WITH STATE OF ART METHODS

In this section, we compare our approach with existing distortion control methods. As exposed in section 4.1, distortion control methods one can find in the literature are based on statistical criteria. As example, most recent schemes take the attributes' mean and standard deviation variations [Shehab et al., 2008], [Kamran et al., 2013b], the mean squared error [Sion et al., 2004] as well as attributes values co-occurrences (evaluated in terms of the correlation, the information gain and so on) [Kamran et al., 2013a] as constraints. Even though these statistical aspects are linked to the database semantics, they represent only a part of the knowledge attached to the database content. For instance, the modification of the age of a newborn can be statistically insignificant but it will be easily pointed out by means of a semantic analysis. Our semantic distortion control allows avoiding such a situation. In a more general way, it completes the statistical distortion control.

Based on this statement and for fair comparison, we decided to compare our scheme with the ones of Sion *et al.* [Sion et al., 2004] and Shehab *et al.* [Shehab et al., 2008] (two efficient robust methods) under statistical distortion constraints only. By doing so, we only compare our scheme based on the adapted QIM (see Sect. 4.4.2) with these two approaches but without considering ontology based semantic distortion control.

The method by Sion *et al.* is based on the modification of the attribute value statistics in a group of tuples G^i so as to embed one bit $s^i = \{0, 1\}$ of the message S (see Sect. 1.3.2.2). To do so, a threshold value is first derived from G^i: $Tr = avg + c\sigma$, where avg and σ are the mean and standard deviation values of A_t in G^i and $c \in (0, 1)$ is a user defined parameter. The embedded bit value is encoded depending on the number ν_c of watermarked attributes values over or under this threshold. More clearly, for a group of N_t tuples, a bit value 0 is embedded if $\nu_c < N_t\nu_{false}$ and a bit value is embedded if 1 if $\nu_c > N_t\nu_{true}$ where $\nu_{true}, \nu_{false} \in (0, 1)$ are user defined parameters exploited so as to control watermark robustness. At the reading stage, the message is extracted simply by verifying if ν_c is greater or smaller than Tr. At the same time, [Sion et al., 2004] was slightly modified, without changing intrinsically the strategy it follows in order to adapt the construction of groups of tuples to the one the other schemes use.

In the method of Shehab *et al.* [Shehab et al., 2008], watermarking is presented as a constrained optimization problem, where a dissimulation function Θ is maximized or minimized depending on the bit value to embed. Optimization space is limited by the quality constraints set. In the example given by the authors, Θ represents the number elements which exceed a certain threshold defined as in [Sion et al., 2004] (see above). At the detection, the value of Θ is calculated and the detected bit is a 1 (resp. 0) if the obtained value is greater (resp. smaller) than a threshold T. The value of T is calculated so as to minimize the probability of decoding error.

In the following experiments, 506800 tuples of the attribute *age* of mean 50.078 and standard deviation 25.236 were watermarked and the same statistical constraints, allowing a change in data values within ± 10 percent were considered for all methods. We considered a watermark S of 500 uniformly distributed bits and consequently, a number of groups $N_g = 500$. All schemes were parameterized so as to ensure a similar distortion in terms of mean and

standard deviation, that is to say: for [Sion et al., 2004], $c = 0.85$, $\nu_{false} = 0.05$ and $\nu_{true} = 0.09$; for [Shehab et al., 2008], $c = 85$; for our scheme, a rotation angle shift $\Delta = 1$. As in the previous section, results are given in average after 30 random simulations.

4.4.5.1 ATTRIBUTE P.D.F PRESERVATION

Although all the methods preserve the attribute's mean and standard deviation, the do not have the same behavior in preserving the attribute's p.d.f, as shown by the D_{KL} and the MAE criteria in Table 4.5. [Shehab et al., 2008] provides the best results but at the price of a very high complexity due to the use of an optimization process (see below).

Table 4.5: Distance between distributions in terms of the D_{KL} and the MAE for our scheme and the methods proposed by Sion *et al.* [Sion et al., 2004] and Shehab *et al.* [Shehab et al., 2008].

Method	D_{KL}	MAE
Sion *et al.* [Sion et al., 2004]	0.0805	0.00218
Shehab *et al.* [Shehab et al., 2008]	1.916 10^{-4}	5.619 10^{-5}
Proposed Method	0.0024	2.8288 10^{-4}

4.4.5.2 ROBUSTNESS

With the same parameterization, three attacks were considered in order to evaluate algorithms' robustness: insertion and deletion of tuples and attributes values modification. All the attacks were performed impacting a percentage of tuples in the range $20\% - 99\%$. The attribute alteration attack consisted in the addition of a centered Gaussian noise of standard deviation $\sigma = 2$.

As depicted in Table 4.6, our method performs in general better than [Sion et al., 2004] and [Shehab et al., 2008], being [Sion et al., 2004] the worst solution. [Shehab et al., 2008] provides a better robustness than our method in the case of a suppression attack only. In the case of an attribute alteration attack, our scheme provides a BER 100 times smaller than [Shehab et al., 2008]. Here is the interest of working with the angular position of the center of mass. It also achieves better performance regarding the tuple insertion attack.

4.4.5.3 COMPLEXITY

Computation time is used so as to evaluate the complexity of these approaches. Regarding the embedding process, it is conducted in about 3s with our method and the one by Sion *et al.*. It takes about 4 hours to [Shehab et al., 2008], due to its optimization process. The detection stage duration is approximately the same in all cases, being of less than 3s.

To sum up, our approach provides better robustness performance than the scheme of Sion *et al.* [Sion et al., 2004] and it introduces less statistical distortion. Regarding the scheme by Shehab *et al.* [Shehab et al., 2008], our scheme is more robust against tuple insertion and

Table 4.6: Bit error rate for our scheme and the methods proposed by Sion *et al.* [Sion et al., 2004] and Shehab *et al.* [Shehab et al., 2008] for various attacks.

Deletion	**20%**	**40%**	**60%**	**80%**	**99%**
Sion *et al.* [Sion et al., 2004]	0.2643	0.3183	0.3453	0.3875	0.4387
Shehab *et al.* [Shehab et al., 2008]	0.0652	0.0776	0.0944	0.1548	0.464
Proposed Method	0.0434	0.1208	0.2041	0.3053	0.4624
Insertion	**20%**	**40%**	**60%**	**80%**	**99%**
Sion *et al.* [Sion et al., 2004]	0.487	0.4932	0.5	0.5	0.5
Shehab *et al.* [Shehab et al., 2008]	0.074	0.0956	0.1268	0.1776	0.218
Proposed Method	0.0263	0.0721	0.1104	0.1449	0.1656
Modification	**20%**	**40%**	**60%**	**80%**	**99%**
Sion *et al.* [Sion et al., 2004]	0.5	0.5	0.5	0.5	0.5
Shehab *et al.* [Shehab et al., 2008]	0.1192	0.134	0.2392	0.4386	0.4804
Proposed Method	0.0028	0.0031	0.0039	0.0043	0.0037

attribute's values modification attacks. It introduces more distortion in terms of the distance between distributions but is of a much lower complexity.

4.5 CONCLUSION

In this chapter, we have addressed the control of the distortion resulting from the watermark embedding process on database semantics. This is an important issue in database watermarking, as any careless modifications of the data may result in incoherent or unlikely records while betraying the presence of the watermark. As shown, current research mainly focus on statistical distortion control, in order to ensure the correct result of *a posteriori* data mining operations the database may undergo. Until now, no solution has been yet proposed to protect the semantic links in between attributes values in tuples. This is the main originality and contribution of the work we exposed in this chapter.

We proposed a new semantic distortion control method which relies in the identification of existing semantic links in between values of attributes in a tuple by means of an ontology. These constraints can be extracted both in the case of numerical and categorical attributes, allowing to define ranges of allowable values in the first case and a set of equivalent values in the second.

We have also experimentally shown that this distortion control can be easily integrated within a general watermarking approach and that it can complete statistical distortion means. In particular we combined it with a new extension of the Quantization Index Modulation (QIM) for database watermarking. Watermark embedding is conducted by modulating the relative angular position of the circular histogram center of mass of one numerical attribute. As we have explained, this watermarking scheme offers us a degree of freedom in the individual distortion to introduce into each value of an attribute. We theoretically demonstrate the robustness performance of our scheme against most common attacks (e.g., tuple insertion and deletion). This makes it suitable for copyright protection, owner identification or traitor

tracing purposes. Our theoretical results were experimentally verified within the framework of a medical database of more than one half million of inpatient hospital stay records. Finally, our approach was also compared with an efficient scheme so as to prove its benefits. It improves the robustness-imperceptibility trade-off.

Chapter 5

Expert validation of watermarked data quality

As mentioned earlier, the choice of a watermarking scheme depends on the application framework and the pursued security purposes. This framework basically constraints the scheme trade-off needs in terms of capacity, robustness and imperceptibility. For instance, if data integrity control is of major concern, the embedding capacity (i.e., the length of the signature that can be hidden into the database) will define the ability of the reader to detect and locate the alterations the database may have undergone. On the contrary, robustness is capital in authenticity control and traceability, where the considered watermarking modulation should ensure that embedded identifiers remain detectable after innocent or hostile manipulations of the database tuples.

If the robustness and capacity needs of relational database watermarking schemes can be easily defined, that is not the case of imperceptibility. As underlined in the previous Chapter, imperceptibility can be analyzed in terms of two different criteria: i) the preservation of the results of statistical data-mining operations applied to the dataset; ii) the preservation of semantic coherence of the database attributes' values. Even though one can preserve the statistical properties of the database and take advantage of ontologies so as to represent the database semantic, these ontologies may not be available all the time, in health care as well as in general public domain. As a consequence, existing watermarking techniques will be applied without considering these semantic associations when embedding a message into a relation. In that case, watermark imperceptibility, can only be judged by the database owner and final users, as it is strongly related to the use to which database records are intended for.

In this Chapter, we are interested in measuring the impact of the watermarking process onto the interpretation of a medical database. To do so, we developed a data quality evaluation protocol the objective of which is twofold: first, we are interested in the analysis of the perception of the practitioner towards the watermark for different levels of distortion of different attributes. The answer to this question will provide useful information on the watermarking parameterization but may also put in evidence that the perception of the practitioner of the data alteration strongly depends on his *a priori* knowledge about the database content. Second, we want to prove that a careless embedding may result in the appearance of incoherent (impossible) or unlikely records.

In the protocol we propose, a physician is invited to thoroughly analyze several sets of tuples in order to point out possible issues caused by the watermark embedding in their interpretation of the information. This test has been performed in two different stages: a blind test in which

he is asked to indicate if a dataset has been watermarked or not; second, an informed test in which he is asked to identify problematic records in watermarked sets. Today, for constraints of time and cost, these tests were only performed by one physician expert attached to the Biomedical Informatics Service of the Rouen University Hospital. Although much more tests must have to be conducted, these preliminary results give us important information about the sensitivity of certain attributes and the influence of the watermarking process, both being the keys of the trade-off between capacity, robustness and imperceptibility. Moreover, it confirms the importance of the semantic distortion control in order to avoid the occurrence of incoherent or unlikely tuples.

In the following, after having identified the perception issues related to attributes values in databases and having explained the different parameters considered of the watermarking scheme used, we detail our evaluation protocol. Obtained experimental results in the study are then analyzed in terms of imperceptibility, capacity and robustness.

5.1 SUBJECTIVE PERCEPTION OF THE WATERMARK

In order to subjectively evaluate the imperceptibility of the embedded watermark, we propose to develop a protocol the objective of which is to measure at first the coherence and likelihood of watermarked data according to the observations of health professionals, the only ones able to judge if the level of disturbance modify or not their interpretation of the data. It should be remarked that the perception of a practitioner may differ from the ones of his or her colleagues for various reasons, making it difficult to identify some specific values of a watermarking algorithm parameters, as example. The main exigencies we identified for the deployment of such an evaluation protocol in database watermarking are:

- Selection of the dataset: The nature of medical databases is very diverse. While the selection of a representative dataset is a difficult task in other domains such as image or audio watermarking, it becomes a real challenge in database watermarking, where there exist an infinity of possible combinations of attributes with different domains and types. The choice of the dataset will be then linked to a limited and specific domain.

- Variability of the evaluators: the analysis of data can be performed by professionals with different roles inside the health system: doctors, nurses, operators, etc. who may prioritize different attributes. There is also an important influence of the competences of each professional and of the conditions in which the test will be deployed: *a priori* knowledge, repeated datasets and so on.

- Final use of data: The required quality of the information in the database will not be exactly the same for different contexts. For instance, if the database corresponds to patient records intended for diagnosis, we should manipulate with care all the pieces of information allowing a practitioner to take a decision. On the other hand, in medico-economical databases, some personal information can be slightly modified without altering the subsequent analysis of the records. Notice that these latter constraints are "semantic" type.

In light of these requirements, we can conclude that building a completely representative test is of very high complexity. The evaluation protocol we propose is not an exception and only represents a small fraction of all the possible datasets and experimental conditions that one may encounter. However, it can be helpful to provide some interesting clues about the constraints to consider in the parameterization of a watermarking scheme, or more clearly about the trade-off robustness-imperceptibility-capacity and in the future development of the proposed watermarking schemes.

5.1.1 PERCEPTION VS ACCEPTABILITY

The search for a higher capacity or a greater robustness, in order to meet the requirements of an application, can lead us to a more important database distortion and consequently, to have a more perceptible watermark. As exposed above, this may perturb or interfere with **the final interpretation** of the database records. However, the **perception** of the presence of a watermark, i.e., the evidence that some pieces of information have been altered, is not directly linked to such interferences. Indeed, the evaluation of the perturbation should take into account the final use of the database.

We can define the **acceptability** of a watermarking system as the boundary distortion that should not be overpassed so as to ensure that professionals will have the same analysis conclusions with watermarked database as with the original one. In some cases, acceptability can directly correspond to the notion of imperceptibility, as for example, when the modified information directly impacts the analysis. In some other contexts, it will also depend on the practitioners, as they may be able to correctly analyze the data even if they know that some values have been modified in order to embed a watermark.

Even though the watermark perception and acceptability may not be equivalent, one should also take into account that an attacker may look for modified values in order to erase or alter the embedded sequence. From this point of view, an interest exists in linking watermark imperceptibility, acceptability and security, the latter being the capability of an algorithm in keeping modulated values invisible to a hostile user.

5.1.2 INCOHERENT AND UNLIKELY INFORMATION

The uncontrolled modification of attributes values can lead to undesired results that may interfere with the interpretation of the database. More precisely, **incoherent** and **unlikely** records may occur. As defined in Chapter 4, the first ones correspond to tuples in which the values of different attributes are illogically combined. For instance, the uncontrolled modification of the patient's gender in a patient in-stay record can result in a record where a male patient presents a principal diagnosis "Single delivery by cesarean section". This record will be easily detected by a final user and it can be problematic as it indicates the presence of the watermark to an attacker.

The second case, i.e., unlikely records, appears when the modified values cause a combination of attributes values that occur with a very low probability. For instance, if a patient

presents a main diagnosis "Acne", we can hardly expect an associated diagnosis "Sudden cardiac death, so described". The association of these two diagnoses is not impossible but it is highly unlikely to find them together.

The objective of our study is to evaluate the influence of the distortion amplitude and the attribute choice in the apparition of these impossible or unlikely tuples that may distort the interpretation the physician makes of records in the database.

5.2 WATERMARKING SCHEME

In this study we essentially focused on the distortion induced by a watermarking process. Even though distortion is closely related to watermark robustness and capacity, this is the main parameter we played within the following.

So, in order to evaluate the impact of the distortion amplitude onto the watermark perception and to demonstrate the need of a distortion control process, we propose to use the scheme we developed in Chapter 2 and which has the advantage of introducing an equivalent distortion in each tuple of the database.

In the objective to embed a sequence of uniformly distributed bits into the values of an integer attribute A_t of dynamic range $[0, L]$, this algorithm introduces a distortion of amplitude Δ in approximately a half of the values of A_t and $-\Delta$ in the other half. There are two exceptions to this rule: first, values in the extremities of the attribute range will be modified of $\Delta - L$ and $L - \Delta$ respectively; second, tuples belonging to an overflow group for which the values remain unmodified.

With this scheme, it is the number of groups of tuples which conditions the capacity. We recall that one bit of message is inserted per group of tuples; groups secretly constructed based on the secret watermarking keys. On the contrary, watermark robustness to tuple suppression and addition stands on the number of tuples per group and on the watermark amplitude. Once the capacity fixed, robustness only depends on Δ. Based on this statement and on the fact that the distortion is uniform over the whole database (in absolute value), we decided to only study the variation of the watermark amplitude over the database interpretation in our evaluation protocol.

5.3 PROTOCOL

Our protocol is composed of two different tests:

- a blind test the purpose of which is to evaluate the perception towards the watermark for different degrees of distortion,
- a non-blind or “informed” test, where the evaluator is informed the database is watermarked so as to know if he or she identifies “problematic” or improperly watermarked records, consequently proving the need for a suited control of the database distortion.

In order to find a response to these interrogations, we defined different sets of questions; questions the evaluator will have to answer during the test. In this section, we present the tests questions as well as their respective experimental plans.

Notice that, even though our protocol is quite general and can be adapted to other databases and application contexts, the subjective appreciation one can have about the database distortion is intrinsically linked to the database content. Indeed, a same attribute may present different distortion constraints depending on the application and the database it belongs to. The database we consider in the following has been provided directly by the evaluators so as to stay in their area of expertise and their daily activities. It contains nearly 5000 tuples associating attributes referring to inpatient stays in the cardiology service at the hospital. Although it presents some similarities with the database we exploited in the previous Chapters, it contains a different set of attributes. In particular, these attributes are:

- *patient ID*, an integer value.
- *gender*, a value "M", i.e., male, or "F", i.e., female.
- *stay duration*, an integer value.
- *age*, an integer value.
- *main diagnosis*, a string containing the ICD10 code and description, e.g., "*G90.0A - Syncope*".
- *associated diagnosis 1 – associated diagnosis 5*, a string containing the ICD10 code and description, e.g., "*E86 - Volume depletion*"
- *medical act 1 – medical act 4*, a string containing the CCAM code and description, e.g., "*ZBQK002 - thoracic radiography*".

For sake of simplicity, only the attributes *patient age* and *stay duration* have been considered for embedding. The two first order statistical moments of these attributes are given in table 5.1.

Table 5.1: Statistical moments for the attributes *age* and *Stay duration* considered for watermarking.

	Moments	
Attribute	**Mean**	**Standard Deviation**
age	70.698	20.147
Stay duration	4.802	6.849

5.3.1 BLIND TEST

The purpose of this first test is to determine the acceptable level of distortion one can introduce without perturbing the interpretation for a given database and a given watermarking scheme that introduces the same distortion into the attributes (in the sequel we use the robust lossless scheme presented in Chapter 2). To do so, in this test, the evaluator has no *a priori* knowledge about the database he or she is facing has been watermarked or not.

5.3.1.1 PROPOSED QUESTIONNAIRE

This one should allow us to evaluate the perception one practitioner has of the watermark. From our point of view, this perception stands on two aspects: the visibility of the watermark and the produced disturbance. This is the reason why our questionnaire contains two questions:

- **"In your opinion, has this database been watermarked (do you perceive a watermark)?"**. Expected answer is **"yes/no"** so as to avoid any ambiguities.
- **"Does the watermark perturb your interpretation of the database?"**. This question is conditional to the previous one and aims at evaluating a degree of perturbation perceived by the physician waiting an answer within the set of values: **No perturbation/medium/strong hindrance**.

As in any subjective experiments, we recall that the answers to these questions strongly depend on the evaluator skills and of his or her degree of expertise inducing more or less variability from one evaluator to another. Anyway, it gives us some clues about the impact of the watermark.

5.3.1.2 EXPERIMENTAL PLAN

It consists in presenting to the evaluator a sequence of extracts of the original test database (i.e., not watermarked) and of different watermarked versions of it (i.e., under different watermarking parameterization). For each extract, the evaluator has to answer to the questions exposed above.

The test experiment was conceived for a duration of less than one hour per person as time has an influence in the capacity of the evaluator to stay concentrated and also in their availability.

TEST SEQUENCE CONSTRUCTION

As stated above, one test sequence is obtained as a random selection of n watermarked extracts alternated with a number y of original extracts.

We considered that our original database can be divided into a set of non-overlapping extracts of records, each of which contains 30 records. This number of records has been chosen in agreement with the evaluators in order to make their analysis easier and respect the test duration. From the complete set of extracts the database is divided into, 20 were randomly selected for test purposes. This number of extracts is sufficient for our experimental purposes and low enough to be handled with ease.

We decided to consider two different attributes for watermark embedding: *age* and *Stay duration*. This will allow us to validate our hypothesis that some attributes are more sensitive than others to watermark distortion.

Notice however that watermarking several attributes of different nature with various watermark parameters will lead to a highly complex analysis. Indeed, this is also out of the

scope of our objective which is to show we need to control the watermark distortion. So, in order to be able to establish a simple relationship between the database distortion perceived by one practitioner with the watermark parameterization, only the values of one attribute are watermarked in an extract.

In order to have a reproducibility indicator on the analysis of the answers of one evaluator, a number v of randomly selected extracts is repeated k times in one sequence. An alternative to this approach is to request the same evaluator to re-conduct the test at a later time. However, this approach is time consuming. Notice that whatever the approach, the second evaluation (or the k^{th} evaluation) is biased by the practitioner memory. Considering the first solution, high values of the repetition number k should be avoided, especially if the sequence of extracts is small.

It is also recommended not using the same original database extract for different watermarking parameterizations. More clearly, extracts of the test sequence should correspond to different set of records of the database, these ones being watermarked or not. This will avoid two extracts having similar attributes values except those that have been differently watermarked, fact that may help to identify watermarked extracts.

As stated above, the watermarking scheme we consider is the robust lossless one exposed in Chapter 2. The distortion it introduces is completely parameterized by the watermark amplitude Δ (see section 5.2). More clearly, the attribute age or duration stay will be modified by adding or subtracting Δ. In the first test we made, we considered four different distortion amplitude values, i.e., $\Delta = [1, 2, 3, 4]$. As seen in Chapter 2, these values offer high robustness against common database attacks.

To sum up, one extract of a sequence corresponds to a triplet (original database extract, Watermarked attribute, Watermark amplitude), knowing that no extract in the sequence corresponds to a same original database set of records watermarked differently. v extracts can be however repeated k times in the sequence. Notice also that each parameterization is represented at least once among the n watermarked extracts.

We give an example of such a sequence in Table 5.2. Shadowed lines are extracts which have been repeated so as to measure the reproducibility of the evaluator judgment, using a number of repetitions $k = 1$.

TEST DURATION

The total duration of this test is a function of: the number n of watermarked extracts; the number y of original extracts; v the number of repeated extracts and k the repetition factor. Then, if an evaluator takes x_{blind} minutes to analyze an extract, we have:

- Duration of a test for one evaluator: $T_{test} = (y + n + v \cdot k)x_{blind}$

As envisioned by the evaluator before the experiment, x_{blind} was estimated to 3 minutes. Thus the analysis of a sequence should take 48 minutes. As we will see in section 5.4.2, this was far from the reality.

Table 5.2: Example of a sequence of extracts one evaluator will have to analyze. Shadowed lines indicate an extract that is repeated.

Extract number in the sequence	Watermarked Attribute	Watermark amplitude Δ	original database extract
1	*age*	1	3
2	*Stay duration*	4	20
3	*age*	3	12
4	*Stay duration*	2	2
5	-	-	13
6	*age*	2	10
7	*Stay duration*	1	15
8	*Stay duration*	1	6
9	*Stay duration*	3	5
10	*age*	1	9
11	*age*	4	17
12	-	-	14
13	*age*	2	10
14	*Stay duration*	3	5
15	-	-	13
16	*Stay duration*	1	15

5.3.2 INFORMED TEST

The informed test works in the same way as the previous one except that the presence of a watermark is not hidden to the evaluator. Its purpose is to see if the evaluator is able to detect the watermark through the presence of "problematic" tuples containing incoherent or unlikely information.

5.3.2.1 PROPOSED QUESTIONNAIRE

Its questions are oriented so as to know if the evaluator founds some tuples he or she estimates as abnormal in an extract. As exposed in section **??**, such tuples may correspond to impossible or very rare combinations of attributes' values.

We planned three questions to obtain an idea of the practitioner's perception:

- **"Knowing that the database has been watermarked and from the values of the attributes in the sequence, could you identify if the database has been watermarked?"**. Expected answer to this question is simply **"yes/no"**.
- **"Do you identify impossible tuples in this database extract? (indicate the tuple line)"**. The purpose of this question is to see if incoherent tuples identified by the evaluator have been watermarked and if yes with which watermark parameterization.
- **"Do you identify tuples containing unlikely information? (indicate the tuple line)"**.

5.3.2.2 EXPERIMENTAL PLAN

As previously stated, all the extracts of the sequence presented to an evaluator are watermarked. Knowing that the evaluator may pay more attention to records' details with the objective of founding out which attribute values have been modulated, an increase in the amount of time needed to analyze one extract is expected. We take this into account in order not having too long experiments.

TEST SEQUENCE CONSTRUCTION

One sequence consists of distinct randomly selected sets of records (30 records) of the original database differently watermarked. In agreement, with the evaluator team, a sequence of 8 extracts sounded to be a good compromise. Contrarily to the previous test, we did not consider test reproducibility in this experiment in order not increasing the test duration, knowing also that the evaluator will easily recognize an extract.

To sum up, an extract of the test sequence is associated to three parameters: Original database set of records, watermarked attribute, watermark amplitude Δ.

An example of such a sequence is given in Table 5.3

However, as we explained, the length of this sequence has to be conveniently chosen so as to do not increase the duration of the test in excess. After discussion with the evaluator, we decided to consider a sequence of 8 datasets, which were randomly selected among all the possible datasets. In our case, the considered sequence is exposed in Table 5.3

Table 5.3: Example of selected extracts for the informed test. For each of these datasets the watermarked attribute, the value of Δ and the sequence number of the extract in the database are given.

Extract number in the sequence	Watermarked Attribute	Watermark amplitude Δ	original database extract
1	*age*	3	5
2	*Stay duration*	2	4
3	*Stay duration*	4	8
4	*age*	4	7
5	*Stay duration*	3	6
6	*age*	2	3
7	*Stay duration*	1	2
8	*age*	1	1

TEST DURATION

The test duration is a function of the number m of selected extracts. If an evaluator takes $x_{informed}$ minutes to analyze one extract, it is given by:

$$T_{test} = m \cdot x_{informed}$$

Before the experiment, the evaluator team estimated $x_{informed} \approx 7$ minutes for the analysis of one extract. Thus, for a complete sequence $T_{test} \approx 56$ minutes.

5.3.3 BIAS IN THE STUDY

Several factors may harm the quality of the result analysis of our protocol. Herein, we expose and discuss separately the most evident ones even though they are related.

5.3.3.1 TEST REPRESENTATIVENESS

The test database was chosen in agreement with the evaluator so as to stay in their area of expertise and facilitate distortion database analysis. Such a database content and use are specific and mainly devoted to the activity evaluation of a hospital. As seen in Chapter 1, there exist many other medical databases the acceptable distortion of which is obviously dependent of their content and purpose. Thus, if our protocol remains general enough for these ones to, results analysis will lead to different results. For example, a same attribute may have limits of distortion different from one database to another. As a consequence, the results we provided thereafter may not be easily generalized to other databases and domains.

Beyond, a sequence of extracts covers only a few samples of all the possible combinations of attributes values. This is the consequence of the time limitation we have and limits the generalization of our results.

The number of evaluators is also a limiting factor. Such a study should be made with several physicians of various degrees of experience. As example, a watermark may not be seen by a young evaluator while a highly experienced one will easily detect it. Notice that we were not able to verify such an hypothesis due to the fact that only one expert participated to the experiment we present in section 5.3. At the same time, obtained results will not be relevant enough to establish a clear relationship in between watermark invisibility and our algorithm parameterization but we expected they would serve at least to validate our hypothesis regarding the interest of a distortion control process.

5.3.3.2 EVALUATOR UNDERSTANDING OF THE PROBLEMATIC

Before starting the tests, the protocol was presented to the evaluator by means of a presentation exposing the context, the watermarking process as well as the details and objectives of each test. Although we assumed that every detail had been well understood by the evaluator team we were not able to verify our assumption.

5.3.3.3 A PRIORI KNOWLEDGE OF THE POSSIBLE PRESENCE OF A WATERMARK

For both tests, this knowledge induces the following problem: when the watermark appears as an unlikely record, the practitioner will be prone to state “the watermark is present” whereas in normal situations he or she will have interpreted it just as an unlikely record. This remark essentially concerns the visibility interpretation of the watermark. Another formulation of the test question may help us to avoid such an ambiguity. Anyway, this issue underlines the difficulty to build up subjective protocols.

Notice, that this remark is also valid for incoherent records as databases do not escape input errors (e.g., a health professional errors in completing the patient records).

5.3.3.4 QUEST FOR A WATERMARK VS DATABASE CONTENT INTERPRETATION

It would have been interesting to conduct the first experiment in a framework where the evaluator interprets and manipulates the data as he or she does every day and to see if anomalies or data misuses due to the watermark presence can be identified. Deploying such a framework is however extremely complex especially in terms of data representativeness as well as in terms of expert investments. Additionally, knowing the existence of potentially watermarked data, it is highly probable that practitioners focus on the watermark detection rather than on their daily tasks (i.e., data analysis and management). This is the reason of the existence of our second test. By asking evaluators to analyze some database extracts watermarked differently, we importantly reduce our protocol complexity at the price however of rerouting the evaluator attention from the real meaning of the records to the detection of the watermark

5.4 EXPERIMENTAL RESULTS

The results we provide in this section are those of the first experiment we conducted with only one expert for the purpose of validating our protocol. This experiment is in fact the preliminary step of a larger study we expect to conduct before the end of the year with more experts from the CISMeF team, Rouen University Hospital. Just to give an idea, we started to build up our protocol in June. It then took 3 months to interact with CISMeF team so as to expose the issue we pursue, to define a test database, to decide about the data encoding format, to define the evaluation framework and finally to conduct the first evaluation and to analyze its results. As a consequence, results we give thereafter are far from being generalized; they however provide us clues on the correctness of our hypothesis about the higher sensitivity of some attributes to modifications and the need of a controlled insertion.

5.4.1 PRESENTATION OF THE EVALUATOR

Both tests have been carried out by one expert evaluator of the CISMeF team, Rouen University Hospital with whom we also interacted before the experiment so as to decide about the choice and appropriateness of the test database.

He is a highly experienced physician with more than 20 years of clinical practice (cardiology and intensive care) and several years of experience in the medical coding domain (coding evaluation, co-occurrences analysis, terminological mapping, ...).

Our choice for soliciting him stands on the fact that he will have a more reliable opinion and enough knowledge about the unlikely or incoherent character of a record. In the following, we analyze the responses we obtained from this expert after he conducted the two previously exposed tests.

5.4.2 BLIND TEST

We recall that the objectives of this test are threefold:

1. identify the level of distortion or equivalently the watermark amplitude values at which the watermark becomes perceptible;
2. if the watermark is visible, evaluate its degree of hindrance;
3. identify if an attribute is more sensitive than another to watermarking.

For this experiment the test sequence we used is given in Table 5.4. As stated in section 5.3.1, some of database extracts are repeated so as to verify the experiment reproducibility.

Table 5.4: Sequence of extracts analyzed by the evaluator. Shadowed lines indicate an extract that is repeated.

Extract number in the sequence	Watermarked Attribute	Watermark amplitude Δ	original database extract
1	*age*	1	8
2	*Stay duration*	4	7
3	*age*	3	1
4	*Stay duration*	2	20
5	-	-	12
6	*age*	2	15
7	*Stay duration*	1	5
8	*Stay duration*	1	11
9	*Stay duration*	3	7
10	*age*	1	4
11	*age*	4	12
12	-	-	3
13	*age*	2	15
14	*Stay duration*	3	7
15	*age*	3	1
16	*Stay duration*	1	11

5.4.2.1 EXPERIMENT DURATION ISSUE

One issue we faced in this experiment is related to its duration. Even though we estimated it with the evaluator team, it took to our expert about 15 minutes per extract instead of 3 minutes. This is due to the fact that the evaluator analyzed each record in the extract more exhaustively than it was initially expected.

This is obviously not a good news for the larger study we are preparing with a new emerging question: how much time will this study take? Solutions on how to reduce experiment time have thus to be explored.

5.4.2.2 RESPONSES ANALYSIS

Responses provided by the evaluator are recorded in Table 5.5. We can verify the matching in the responses provided by the evaluator for the four repeated extracts, indicating the test reproducibility. However, this reproducibility evaluation is relative. Indeed, the sequence of extracts is quite short, as it takes 15 minutes to the expert to analyze each extract, and he surely had recognized repeated database extracts.

Table 5.5: Results to the blind test provided by the evaluator. The second column indicates if the dataset was watermarked (yes) or not (no). The third column indicates if the evaluator was able to detect the presence of the watermark in the dataset. The last column corresponds to the hindrance degree perceived by the evaluator. Highlighted lines represent watermarked extracts for which the evaluator did not perceive the watermark.

Dataset number	Watermarked	Identified by the Evaluator	Hindrance degree
1	yes	no	-
2	yes	yes	strong
3	yes	no	-
4	yes	yes	strong
5	no	no	-
6	yes	no	-
7	yes	yes	medium
8	yes	yes	medium
9	yes	yes	strong
10	yes	no	-
11	yes	no	-
12	no	no	-
13	yes	no	-
14	yes	yes	strong
15	yes	no	-
16	yes	yes	medium

As we can see, the evaluator did not identify as watermarked any of the non-watermarked extracts. On the contrary, none of the extracts in which age attribute has been watermarked have been detected as watermarked whatever the watermark amplitude value (i.e., $\Delta = [1, 4]$). This is not the case with extracts where the attribute *stay_duration* has been used with the same watermark amplitude. This allows us drawing a first conclusion: the watermark is more easily concealed in some attributes than in others. Going deeper in the analysis, we found a relationship between the watermark imperceptibility with the attribute statistical distribution and on its semantic links with other attributes. In this experiment, $Stay\ duration$ has a more concentrated distribution (see Sect. 5.1) than the attribute age. As a consequence the watermark amplitude at which it becomes visible is smaller than for the age. Nowadays, no statistic controlled based watermarking algorithms take this constraint into account. Notice that some database extracts where $Stay\ duration$ was watermarked were more easily identifiable due to value jumps at the extremities of the attribute domain range (e.g., 0 becoming 135 after the watermarking process). Such situation can simply be avoided by taking care of the attribute

domain definition during the insertion (see Chapter 2, section 4.4.2.3).

At the same time, $Stay\ duration$ presents stronger semantic links with the patients' diagnosis than age does. Thus, in a tuple, the values the watermarked version of $Stay\ duration$ can take are constrained by the value $patient\ diagnosis$ has. These remarks confirm our hypothesis and the needs for an insertion process where distortion is controlled based on the knowledge one has about the database content, this one being statistic or semantic, in order to ensure watermark imperceptibility.

Regarding the watermark degree of hindrance, which can only be evaluated on 7 extracts where the attribute $Stay\ duration$ has been watermarked (see Table 5.4 and Table 5.5), it increases along with the value of the watermark amplitude Δ, as expected. However, it can be noticed from Table 5.5 that the evaluator indicates directly a somehow high degree of hindrance, i.e., from medium to strong, when he declares the watermark visible. That is to say that for some attributes, a hindrance occurs immediately when the watermark becomes visible. If we look at the corresponding values of Δ, the watermark is visible for any Δ. These results are however not statistically representative in order to establish a clear relationship between Δ and the watermark imperceptibility. They should be completed with a larger study which will allow the estimation of an empirical distortion threshold. Nevertheless, based on this experiment one comes easily to the conclusion that the maximum watermark amplitude for the attribute $stay\ duration$ to avoid a strong hindrance equals 1. However, as we know that $stay\ duration$ value strongly depends on the attribute diagnosis value, it should be possible to increase the distortion for some $Stay\ duration$ values in the database. Here comes the interest for a semantic distortion control.

Notice also that the age attribute can be more strongly modified than $stay\ duration$, leading thus to a greater watermark capacity or more robustness. This suggests that the attribute statistics can help to chose the most suitable attributes for watermarking.

5.4.3 INFORMED TEST

The objective of this test is, in a first time to reinforce the conclusions of the previous analysis and secondly, to demonstrate the importance of taking into account the semantic links between values in a tuple. The same expert conducted this experiment in the continuity of the previous one.

The sequence he has to analyzed is given in Table 5.6. As it can be seen, it is 2 times smaller than the sequence of the first test. Indeed, we recall that for this test, the evaluator knows the extracts he has to analyze have been watermarked and we want him to focus on the detection on attributes values he founds suspicious. However, he does not know which attribute in between age and $stay\ duration$ has been watermarked. He does not know what is the strength of the embedding either.

As in the previous test and for the same reason, the real analysis duration of an extract is much greater than the one expected. Indeed, it took about 160 minutes to the evaluator.

Table 5.7 contains the responses of the evaluator to the questions of this second test; questions we recall: "Even though you know this set of records is extracted from a watermarked

Table 5.6: Extracts for the informed test. For each of these datasets the watermarked attribute, the value of Δ and the sequence number of the extract in the database are given.

Extract number in the sequence	Watermarked Attribute	Watermark amplitude Δ	original database extract
1	*age*	3	5
2	*Stay duration*	2	4
3	*Stay duration*	4	8
4	*age*	4	7
5	*Stay duration*	3	6
6	*age*	2	3
7	*Stay duration*	1	2
8	*age*	1	1

Table 5.7: Results of the expert evaluator for the informed test with the sequence given in Table 5.6. For each extract the evaluator had to if he considered the extract as watermarked (Column 2) and if he did, he was requested to give the lines of the records he find as incoherent or unlikely (column 3 and 4 respectively)

Dataset number	Identified by the Evaluator	Incoherent records	Unlikely records
1	?	-	12, 14
2	yes	2, 6, 9,10, 12,16, 23, 24, 26, 31	-
3	yes	5, 6, 8,14, 16, 19, 20	-
4	no	-	-
5	yes	3, 4, 5, 6, 11, 13, 14, 19, 20, 25, 29	-
6	no	-	-
7	yes	5, 8, 9, 19, 24, 30	4, 22, 29
8	no	-	-

database, would you be able to say that this extract is or not watermarked? "; "if yes, please indicate the records' lines you found incoherent or unlikely". Due to the fact the experiment is a prelude to a larger experiment, we took time to discuss with the evaluator about the incoherent and unlikely records he identified. As we will see in the sequel, his comments are of great interest.

As it can be seen through the answers to the first question, the evaluator was able to identify several watermarked extracts. These "detection" results are similar to those of the previous test. Indeed, none of the extracts where the attribute age was watermarked were identified by the evaluator, whatever the watermark amplitude. Regarding the first extract, the evaluator emitted some doubts due to some $stay\ duration$ values, which he identified as too low regarding the values of $main\ diagnosis$ in the record. However, as exposed in table 5.7, this extract was watermarked through the attribute age. This case can be assimilated to a false positive detection, i.e., saying the extract is watermark through $stay\ duration$ while it is not. This is in agreement with the bias we commented in section 5.3.3. The evaluator focuses on the watermark and considers it corresponds to the occurrence of unlikely records.

If we now focus on the extracts where the attribute $stay\ duration$ has been used, all of them have been identified by the evaluator through several tuples he identified as incoherent.

Most indicated incoherent records correspond to jumps in-between values at in the extremities of the $stay\ duration$ attribute domain range after the watermarking process. Even though one may think such modifications are easily detectable, it underlines the idea that an attribute value cannot be modified without care.

The case of the dataset number 7 is more interesting. Incoherent and unlikely records are not the consequence of value jumps. Indeed, the feedback we had from the evaluator is that, for both incoherent and unlikely tuples, their $Stay\ duration$ values are too short regarding the patient's diagnosis. This highlights the importance of the semantic links that exist between attributes' values in a tuple and proves the need of a distortion control method that takes them into account. This evidences the interest of the semantic distortion control we developed in Chapter 4.

5.5 CONCLUSION

In this Chapter, we presented a protocol we developed in order to study the subjective perception of a watermarking process so as to validate different of our hypothesis, that is to say: i) establish that a relationship exist in-between the watermark visibility and the hindrance it may occurs with the watermark distortion or equivalently with the watermarking algorithm parameterization; ii) put in evidence that if attribute values are modified without taking care of the attribute statistical properties and of its semantic meaning, the watermark can be easily identifiable by the user and consequently have an impact on database post-processing operations.

Our protocol is constituted of two independent tests: one blind and one informed, based on various questions. If in the former the evaluator does not know if a database extract he has to analyze has been watermarked or not, that is not the case in the latter where the evaluator is informed.

Even though the experiment we conducted is based on one evaluator and a watermarking algorithm simply parameterized by a distortion parameter (a constant watermark amplitude for all the attribute values), it points out the existence of a relationship between the watermark amplitude and the perceived perturbation in the case of two different attributes. It is however not possible to generalize the threshold values above which the watermark becomes visible in a attribute due to the fact this experiment is not statistically representative (only one evaluator). These results have to be refined by means of a larger study. Beyond, the results analysis also shows that some attributes, due to their statistical properties and semantic links to other attributes in the database, are more appropriate for watermark embedding than others. As seen in section 5.4.2, the attribute age can be more distorted than the attribute $stay\ duration$, leading thus to a greater watermark capacity or more robustness, under the constraint however that watermarked tuples remain semantically correct. Indeed, attribute visibility thresholds are closely linked to the whole database content and use. For instance, for a same attribute, a small variation can be insignificant in some databases while having a high impact in some others. Anyway, both tests can be used to provide some guidance in the search of the more adapted watermarking algorithm parameterization in order to ensure an imperceptible watermark or at least, a perceptible watermark which does not introduce hindrance or discomfort for the user.

This sole experimentation of our protocol, and especially of our second test, also helps us to clearly put in evidence that incoherent or unlikely records can be easily identified by users, especially those who work daily with the same databases. It also supports the need for a semantic distortion control method that takes into account the semantic links between attributes values. The solution we proposed in Chapter 4 is fully adapted. If used, no impossible records would have been added.

Notice that these tests are also general enough to be exploited with different databases and some other watermarking algorithms, these ones being or not equipped of a semantic and statistic distortion control. Our questionnaires can be used in the case some modifications are applied in the experimental plan so as to broaden our study, e.g., increase of the number of evaluators, different kinds of datasets, different parameterizations and so on. However, as exposed above, the main obstacle that we may find when planning a larger study is the test duration. The reduction of the number of extracts in the test sequences or the number of records in each extract would imply an important loss of representativeness. Thus, there is not an evident solution to this issue.

An aspect we did not consider in this evaluation is the hindrance a watermarking process may induce in database post-processing, e.g., clustering, classification. Thus, in a more advanced study, some statistical analysis tools will be provided to the evaluator in order to measure the impact of watermarking on their results.

Chapter 6
Conclusion and Perspectives

The protection of relational databases by means of watermarking within the health care domain as case study, is the focus of this Ph.D. thesis work. As exposed, medical databases are important elements of the health care system supporting or serving various objectives from the patient care, the hospital management as well as public health (e.g. repositories). They are also of sensitive nature and submitted to strict legislative and ethical rules especially in regards with the nominative aspect of medical records. Nevertheless, they are manipulated in highly open environments in which different users access information at the same time. Databases can also be transferred and shared between different health institutions, the most common scenario being the one where several hospitals feed some mutualized repositories. Indeed, as in any other domains, databases are of high economical and decisional value. In this context, where data are outsourced, common security mechanisms such as access control or encryption mechanisms remain limited. Once these barriers bypassed, it is difficult to protect data reliability, which involves the outcomes of their integrity and authenticity, as well as to ensure data traceability. Here comes the interest of watermarking. It leaves access to the information while maintaining it protected. It provides an "*a posteriori*" data protection which by definition is independent of the way date are encoded and stored. Despite this interest only a few database watermarking algorithms have been proposed until now.

In a first moment, we have studied relational database reversible or lossless watermarking for different purposes: integrity control and traceability. The reversibility property is indeed of special interest in the medical context as it allows recovering the original database content. Such schemes are also very rare in the literature. We have derived two lossless schemes, one being fragile and the other one robust [Franco-Contreras et al., 2013] [Franco-Contreras et al., 2014b]. Both schemes modulate the circular statistics of numerical attributes. Especially, the center of mass of the circular histogram of one attribute is a characteristic with very interesting "watermarking" properties. Modulating the difference angle in between the vectors associated to the centers of mass of group of tuples, our first scheme is fragile and allows the embedding and extraction of a message. This message, a sequence of symbols, can be used for integrity control applications, allowing for example the embedding of a digital signature of the database within itself. Any differences between the extracted and the recomputed signatures will indicate the database integrity loss. The second algorithm is lossless and robust. It takes advantage of the fact that the center of mass of the circular histogram does not vary too much after tuple insertion or insertion attacks as well as after attribute modification attack. This is not the case when working directly at the attribute value level. It is like conducting the embedding into the Fourier Transformed of an image, the effect of which is to spread

the message over the whole image pixels in the spatial domain. In our case, we spread our message over several attributes values in the database. To succeed, an attack will have to modify a lot of tuples. Nevertheless, even though this database characteristic is robust, we had to face synchronization issues. Indeed, after the database has been attacked, symbols embedded in independent groups of tuples may be erased (i.e., impossibility to detect the value of the symbol in a group) or injected (i.e., a symbol of message is detected into a group where no symbol was inserted). In order to counteract such synchronization issues, we propose to embed two different messages: one is robust and of fixed length detected by means of correlation; the second is fragile and contains all pieces of information the watermark reader will need so as to reconstruct the original database. This second scheme can be used for database traceability purposes, identifying its origins as well as its recipients even under database attacks. In addition to the reconstruction information, the fragile message can also convey a digital signature of the database so as to allow verifying if the database is the original one or not.

One of the constraints of the above schemes is that attribute values are modified with a fixed distortion. This is imposed by the reversibility property which cannot be achieved if this rule is not satisfied. In order to get a degree of freedom in the modification of each value, we proposed a non-reversible robust scheme. The originality of the latter stands on the fact that it extends the well-known Quantization Index Modulation, never considered in database watermarking, to the center of mass of circular histograms for a numerical attribute. The main difference of our adapted QIM with its original version stands on the definition of the QIM dictionary cell centroids (see Chapter 4) we no longer design as the cell center so as to minimize database distortion. Beyond, we also propose an iterative process that allows us modifying attribute values faster than other optimization based approaches while taking into account any semantic and statistical constraints [Franco-Contreras et al., 2014a].

A second contribution of this work corresponds to the theoretical demonstration of the performance of the above schemes in terms of capacity and robustness. This theoretical analysis is possible due to the behavior of the circular histogram center of mass which distribution can be modeled by means of Gaussian distributions. We thus showed that both the capacity and the robustness of our schemes depend on the intrinsic statistical distribution of the selected attribute for embedding, on the number of groups and obviously on the amplitude of the watermarking. These theoretical results have been validated experimentally on a real medical database of inpatient stays in French hospitals in 2011 and allow us to precisely predict the performance of our schemes. Moreover, they prove the suitability of the proposed schemes to the applications they are conceived for.

Another contribution of our work corresponds to the optimization of watermark detection in the context of databases traceability when they are merged into a data repository or a data warehouse. This is an issue of rising interest with the nowadays trend towards big-data. After having established that, from the point of view of one database we want to detect, tuples of the other mixed databases can be seen as a noise, the statistical properties of which can be calculated, we optimized the detection process of our robust lossless watermarking scheme by means of a soft-decision decoding. Experimental results we obtain are such that a database can be detected even if it only represents a 7% of the total number of records of the data warehouse, under high constraint of distortion. A fourth contribution of this work stands on

a semantic distortion control process the purpose of which is to take care not introducing incoherent or unlikely records. Such records may interfere with the post processing of the database as well as betray the presence of the watermark. Our approach is complementary to actual distortion control solutions which only focus on preserving database statistics (statistics of attributes or in-between attributes, such as the correlation, the mutual information and so on). If statistics provide hints about the existence of semantic links, identifying the co-occurrences of values in the database, they only give access to a part of the knowledge one can have about the database content. Indeed, a statistically insignificant modification may be highly semantically detectable. The semantic distortion control we propose takes advantage of an ontology defined over the relational database scheme. This ontology identifies the semantic relationships in between values of attributes in a tuple and gives access to the set of values a watermarked attribute can take in order not introducing incoherent records [Franco Contreras et al., 2014].

Finally, we have also developed an experimental validation protocol in cooperation with the CISMEF team based at the Biomedical Informatics Service of the Teaching Hospital (CHU) of Rouen so as to validate our distortion constraint hypothesis. Even though we were only able to make one test with one expert, the result analysis of this test shows that: i) depending on their statistical properties, certain attributes are more appropriate for watermark embedding than others; ii) it is important to preserve the database semantic so as not introducing incoherent and unlikely tuples, highly detectable by the user, requiring thus a specific distortion control approach; iii) there exist a relation between the watermark amplitude and the database distortion perceived by the practitioner. However, these preliminary results have to be validated by a larger study in order, for example, to clearly establish for a watermarking method the relationship between the watermark parameterization and its degree of hindrance once it becomes visible.

Notice that some parts of these work are actually subject to a technology transfer. This underlines the interest and the needs for an *a posteriori* protection of shared databases.

Even though this Ph.D. thesis provides some contributions to database watermarking, there are still several open issues. These are both general of database watermarking and specific of our approach.

In our watermarking schemes, we considered the modulation of the center of mass of circular histograms in order to obtain a domain of insertion robust against common database attacks (i.e., tuple deletion and insertion and attribute values modification). This allows spreading each symbol of the message or of the watermark over several tuples, like for images when working in the frequency domain rather than in the spatial domain. The choice of the insertion domain has to be explored in database watermarking.

In this work only numerical attributes were considered for embedding. An extension to categorical data has to be considered so as to broaden the possible applications of our schemes.

With respect to our theoretical performance analysis, it can be enlarged in order to consider other kind of "attacks". The attribute value modification attack has to be studied as well as the tuple addition attack where new tuples do not follow the original attribute distribution. Moreover, an ontology based attack, in which the attacker uses the semantic knowledge contained in the ontology in order to modify values in the database has to be explored.

Regarding our database traceability scheme, it requires to be completed by a security protocol in order to manage the secret watermarking key delivery and user identifiers. Such a security protocol has not been yet explored in the literature. A further analysis has to be carried out so as to construct it and allow the deployment of our traceability scheme.

Beyond, in order to make the watermark insertion and extraction independent of the way the database is stored and robust to tuple deletion and addition, most watermarking systems use a preliminary group construction stage. This latter assigns one tuple into a group through a cryptographic function which takes as input its primary-key. This key is of crucial importance and the weakness of these schemes. Indeed, if it is altered or simply removed from the database, one will not be able to reconstruct groups of tuples, desynchronizing the detection from the insertion, making thus the watermark detection impossible. Existing solutions, such as the use of attributes MSB or clustering techniques, do not represent a real alternative as they present important drawbacks, notably the non-uniform distribution of tuples in groups and the sensibility to attribute value modifications. New and more adapted strategies have to be found.

Moreover, with the current "Big-Data" trend, data processing and analysis are becoming more and more difficult due to the number of dimensions, i.e., attributes, number of tuples, etc. This difficulties will be also encountered in database watermarking. It is necessary to find techniques adapted to this increasing amount of data.

Additionally, the database may contain attributes of highly sensitive nature that are usually encrypted in order to protect patients' privacy. An analysis of the joint use of watermarking and encryption in databases should be carried out in order to offer a solution ensuring the protection of privacy and allowing the use of watermarking at the same time.

Regarding the control of the database distortion, the joint application of statistical and semantic constraints should be analyzed. Indeed, in order to minimize the risk of incorrect data interpretation by a physician and to ensure the correct result of data-mining operations both semantic and statistical criteria must be considered. In this framework, our experimental validation protocol can be useful. If actually, it already has to be enlarged, considering a greater number of evaluators , it can be be modified so as to include automatic post-processing. This will provide a more complete view of the impacts of watermarking on the interpretation of physicians and also extend on the results of data-mining operations, such as classification or clustering.

Bibliography

Agence technique de l'information sur l'hospitalisation (2013). Programme de médicalisation des systèmes d'information en médecine, chirurgie, obstétrique et odontologie (PMSI MCO). `http://www.atih.sante.fr/mco/presentation`, Accessed: 09/03/2014.

Agrawal, R., Haas, P. J., and Kiernan, J. (2003a). A system for watermarking relational databases. In *Proceedings of the 2003 ACM SIGMOD international conference on Management of data*, SIGMOD '03, page 674?674, New York, NY, USA. ACM.

Agrawal, R., Haas, P. J., and Kiernan, J. (2003b). Watermarking relational data: framework, algorithms and analysis. *The VLDB Journal*, 12(2):157?169.

Agrawal, R. and Kiernan, J. (2002). Watermarking relational databases. In *Proc. of the 28th International Conference on Very Large Databases, VLDB '02:*, pages 155 – 166. Morgan Kaufmann, San Francisco.

Al-Haj, A. and Odeh, A. (2008). Robust and blind watermarking of relational database systems. *Journal of Computer Science*, 4(12):1024–1029.

Alattar, A. (2004). Reversible watermark using the difference expansion of a generalized integer transform. *IEEE Trans. on Image Processing*, 13(8):1147 –1156.

Alattar, A. M. (2000). Smart images using digimarc's watermarking technology. In *Proc. SPIE, Security and Watermarking of Multimedia Contents II*, volume 3971, pages 264–273.

Alfaro, J., Cuppens, F., and Cuppens-Boulahia, N. (2007). Aggregating and deploying network access control policies. In *Availability, Reliability and Security, 2007. ARES 2007. The Second International Conference on*, pages 532–542.

Azkia, H., Cuppens-Boulahia, N., Cuppens, F., and Coatrieux, G. (2010). Reconciling ihe-atna profile with a posteriori contextual access and usage control policy in healthcare environment. In *IAS*, pages 197–203. IEEE.

Baader, F. and Hollunder, B. (1991). A terminological knowledge representation system with complete inference algorithms. In Boley, H. and Richter, M., editors, *Processing Declarative Knowledge*, volume 567 of *Lecture Notes in Computer Science*, pages 67–86. Springer Berlin Heidelberg.

Bachman, C. W. (1969). Data structure diagrams. *SIGMIS Database*, 1(2):4–10.

Bailly, G., Attina, V., Baras, C., Bas, P., Baudry, S., Beautemps, D., Brun, R., Chassery, J.-M., Davoine, F., Elisei, F., Gibert, G., Girin, L., Grison, D., Léoni, J.-P., Liénard, J., Moreau, N., and Nguyen, P. (2006). ARTUS: synthesis and audiovisual watermarking of the movements of a virtual agent interpreting subtitling using cued speech for deaf televiewers. *Modelling, Measurement and Control - C*, 67SH(2, supplement : handicap):177–187. AMSE - ISSN: 1259-5977.

Bas, P. and Furon, T. (2013). A new measure of watermarking security: The effective key length. *Information Forensics and Security, IEEE Transactions on*, 8(8):1306–1317.

Baudry, S., Delaigle, J.-F., Sankur, B., Macq, B., and Maître, H. (2001). Analyses of error correction strategies for typical communication channels in watermarking. *Signal Processing*, 81(6):1239–1250.

Bell, D. E. and Lapadula, L. J. (1976). Secure Computer System: Unified Exposition and MULTICS Interpretation. Technical Report ESD-TR-75-306, The MITRE Corporation.

Berenson, M., Krehbiel, T., and Levine, D. (2012). *Basic Business Statistics: Concepts and Applications*. Prentice-Hall.

Bertino, E. and Haas, L. (1988). Views and security in distributed database management systems. In Schmidt, J., Ceri, S., and Missikoff, M., editors, *Advances in Database Technology*, volume 303 of *Lecture Notes in Computer Science*, pages 155–169. Springer Berlin Heidelberg.

Bertino, E., Ooi, B. C., Yang, Y., and Deng, R. H. (2005). Privacy and ownership preserving of outsourced medical data. In *Proceedings of the 21st International Conference on Data Engineering*, ICDE '05, pages 521–532, Washington, DC, USA. IEEE Computer Society.

Bertoni, G., Daemen, J., Peeters, M., and Van Assche, G. (2014). SHA-3 standard: permutation-based hash and extendable output functions. Technical report, National Institute of Standards and Technology (NIST).

Bhattacharya, S. and Cortesi, A. (2009). A Distortion Free Watermark Framework for Relational Databases. In *International Conference on Software and Data Technologies*, pages 229–234.

Blackman, K. R. (1998). Technical note: IMS celebrates thirty years as an IBM product. *IBM Syst. J.*, 37(4):596–603.

Boland, F., Ruanaidh, J., and Dautzenberg, C. (1995). Watermarking digital images for copyright protection. In *Image Processing and its Applications, 1995., Fifth International Conference on*, pages 326–330.

Boneh, D. and Shaw, J. (1996). Collusion-secure fingerprinting for digital data. In *IEEE Transactions on Information Theory*, pages 452–465. Springer-Verlag.

Bouslimi, D., Coatrieux, G., Cozic, M., and Roux, C. (2012). A joint encryption/watermarking system for verifying the reliability of medical images. *IEEE Transactions on Information Technology in Biomedicine*, 16(5):891–899.

Brachman, R. J. and Schmolze, J. G. (1985). An overview of the kl-one knowledge representation system*. *Cognitive Science*, 9(2):171–216.

Brino, A. (2013). US and UK share health data via cloud. In Healthcare IT News `http://www.healthcareitnews.com/news/us-uk-share-health-data-cloud`. Accessed: 04/23/2014.

Chang, J.-N. and Wu, H.-C. (2012). Reversible fragile database watermarking technology using difference expansion based on svr prediction. In *Computer, Consumer and Control (IS3C), 2012 International Symposium on*, pages 690–693.

Chbeir, R. and Gross-Amblard, D. (2006). Multimedia and metadata watermarking driven by application constraints. In *Multi-Media Modelling Conference Proceedings, 2006 12th International*, page 8 pp.

Chein, M. and Mugnier, M. (2009). *Graph-based Knowledge Representation*. Advanced Information and Knowledge Processing. Springer-Verlag New York, Inc., New York, NY, USA.

Chen, B. and Wornell, G. W. (1999). Quantization index modulation: A class of provably good methods for digital watermarking and information embedding. *IEEE Trans. on Information Theory*, 47(4):1423–1443.

Chin-Chen Chang, Thai-Son Nguyen, C.-C. L. (2013). A blind reversible robust watermarking scheme for relational databases. Article ID 717165, 12 pages.

Choplin, R. H., Boehme, J. M., and Maynard, C. D. (1992). Picture archiving and communication systems: an overview. *RadioGraphics*, 12(1):127–129.

Coatrieux, G., Chazard, E., Beuscart, R., and Roux, C. (2011). Lossless watermarking of categorical attributes for verifying medical data base integrity. In *2011 Annual International Conference of the IEEE EMBS, EMBC*, pages 8195–8198. IEEE.

Coatrieux, G., Le Guillou, C., Cauvin, J.-M., and Roux, C. (2009). Reversible watermarking for knowledge digest embedding and reliability control in medical images. *Information Technology in Biomedicine, IEEE Transactions on*, 13(2):158–165.

Coatrieux, G., Lecornu, L., Sankur, B., and Roux, C. (2006). A review of image watermarking applications in healthcare. In *Engineering in Medicine and Biology Society, 2006. EMBS '06. 28th Annual International Conference of the IEEE*, pages 4691–4694.

Coatrieux, G., Quantin, C., and Allaërt, F. A. (2012). Watermarking as a traceability standard. *Studies in Health Technology and Informatics*, (180):761–765.

Coatrieux, G., Quantin, C., Montagner, J., Fassa, M., Allaert, F.-A., and Roux, C. (2008). Watermarking medical images with anonymous patient identification to verify authenticity. In Andersen, S. K., Klein, G. O., Schulz, S., and Aarts, J., editors, *MIE*, volume 136 of *Studies in Health Technology and Informatics*, pages 667–672. IOS Press.

Codd, E. F. (1970). A relational model of data for large shared data banks. *Communications of the ACM*, 13(6):377–387.

Collen, M. (2011). *Computer Medical Databases: The First Six Decades (1950–2010)*. Health Informatics. Springer-Verlag London Limited.

Colton, S. (2005). Lecture 4: Knowledge representation. `http://www.doc.ic.ac.uk/~sgc/teaching/pre2012/v231/lecture4.html`. Accessed: 07/11/2014.

Council of Europe (1997). Recommendation no. R(97)-5 on the protection of medical data.

Cox, I., Miller, M., Bloom, J., Fridrich, J., and Kalker, T. (2008). *Digital Watermarking and Steganography*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2 edition.

Daconta, M. C., Obrst, L. J., and Smith, K. T. (2003). *The Semantic Web: A guide to the future of XML, Web Services and Knowledge Management*. Wiley Publishing, Inc., Indianapolis, Indiana, USA, 1 edition.

de la Sécurité de l'Information Français (CLUSIF), C. (2010). Mehari. Version 2010 in: `http://www.clusif.asso.fr/fr/production/mehari/index.asp`, Accessed: 09/05/2014.

de la Sécurité des Systèmes d'Information (DCSSI), D. C. (2010). Expression des besoins et identification des objectifs de sécurité : la méthodologie. Version 2010 in: `http://www.ssi.gouv.fr/IMG/pdf/EBIOS-1-GuideMethodologique-2010-01-25.pdf`, Accessed: 09/05/2014.

De Vleeschouwer, C., Delaigle, J.-F., and Macq, B. (2003). Circular interpretation of bijective transformations in lossless watermarking for media asset management. *Multimedia, IEEE Trans. on*, 5(1):97 – 105.

Donnelly, L. (2014). Fears that hospitals are covering up death rates. The Telegraph `http://www.telegraph.co.uk/health/healthnews/10728189/Fears-that-hospitals-are-covering-up-death-rates.html`, Accessed: 04/11/2014.

Elmasri, R. and Navathe, S. B. (1989). *Fundamentals of Database Systems*. Addison-Wesley Longman.

Epinal (2013). French doctors and radiologist jailed for radiation overdoses. The Telegraph, Accessed: 04/11/2014.

European Parliament (1999). Directive 1999/93/EC of the european parliament and of the council.

Farfoura, M. and Horng, S.-J. (2010). A novel blind reversible method for watermarking relational databases. In *Parallel and Distributed Processing with Applications (ISPA), 2010 International Symposium on*, pages 563 –569.

Farfoura, M. E., Horng, S.-J., Lai, J.-L., Run, R.-S., Chen, R.-J., and Khan, M. K. (2012). A blind reversible method for watermarking relational databases based on a time-stamping protocol. *Expert Systems with Applications*, 39(3):3185–3196.

Farfoura, M. E., Horng, S.-J., and Wang, X. (2013). A novel blind reversible method for watermarking relational databases. *Journal of the Chinese Institute of Engineers*, 36(1):87–97.

Fisher, N. I. (1993). *Statistical Analysis of Circular Data*. Cambridge University Press.

Fisher, N. I. and Lewis, T. (1983). Estimating the common mean direction of several circular or spherical distributions with differing dispersions. *Biometrika*, 70(2):333–341.

Franco-Contreras, J., Coatrieux, G., Cuppens-Boulahia, N., Cuppens, F., and Roux, C. (2013). Authenticity control of relational databases by means of lossless watermarking based on circular histogram modulation. In *Security and Trust Management*, volume 8203 of *Lecture Notes in Computer Science*, pages 207–222.

Franco-Contreras, J., Coatrieux, G., Cuppens-Boulahia, N., Cuppens, F., and Roux, C. (2014a). Adapted quantization index modulation for database watermarking. In *Digital-Forensics and Watermarking*, Lecture Notes in Computer Science. (In press).

Franco-Contreras, J., Coatrieux, G., Cuppens-Boulahia, N., Cuppens, F., and Roux, F. (2014b). Robust lossless watermarking of relational databases based on circular histogram modulation. *IEEE Trans. on Information Forensics and Security*, 9(3):397–410.

Franco Contreras, J., Coatrieux, G., Cuppens-Boulahia, N., Frédéric, C., and Roux, C. (2014). Ontology-guided distortion control for robust-lossless database watermarking: Application to inpatient hospital stay records. In *2014 Annual International Conference of the IEEE EMBS, EMBC*, pages 4491–4494. IEEE.

Fridrich, J. and Goljan, M. (1999). Images with self-correcting capabilities. In *Image Processing, 1999. ICIP 99. Proceedings. 1999 International Conference on*, volume 3, pages 792–796.

Furon, T., G. A. C. F. (2008). On the design and optimization of tardos probabilistic fingerprinting codes. In Solanki, K., S. K. M. U., editor, *Information Hiding*, volume 5284 of *Lecture Notes in Computer Science*, pages 341–356. Springer Berlin Heidelberg.

Gerhards, R. (2009). The syslog protocol. Request for Comments: 5424, Accessed: 04/16/2014.

Gilbert, H. and Handschuh, H. (2004). Security analysis of sha-256 and sisters. In Matsui, M. and Zuccherato, R., editors, *Selected Areas in Cryptography*, volume 3006 of *Lecture Notes in Computer Science*, pages 175–193. Springer Berlin Heidelberg.

Gomez-Perez, A. and Benjamins, V. R. (1999). Applications of ontologies and problem-solving methods. *AI Magazine*, 20(1):119–123.

Gross-Amblard, D. (2003). Query-preserving watermarking of relational databases and xml documents. In *Proceedings of the Twenty-Second ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems*, pages 191–201.

Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowl. Acquis.*, 5(2):199–220.

Guo, J. (2011). Fragile watermarking scheme for tamper detection of relational database. In *Computer and Management (CAMAN), 2011 International Conference on*, pages 1 –4.

Gupta, G. and Pieprzyk, J. (2008). Reversible and blind database watermarking using difference expansion. In *Proceedings of the 1st International Conference on Forensic Applications and Techniques in Telecommunications, Information, and Multimedia and Workshop*, e-Forensics '08, pages 24:1–24:6, ICST, Brussels, Belgium, Belgium. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering).

Gupta, G. and Pieprzyk, J. (2009). Database relation watermarking resilient against secondary watermarking attacks. In *ICISS*, pages 222–236.

Hacıgümüş, H., Iyer, B., and Mehrotra, S. (2004). Efficient execution of aggregation queries over encrypted relational databases. In Lee, Y., Li, J., Whang, K.-Y., and Lee, D., editors, *Database Systems for Advanced Applications*, volume 2973 of *Lecture Notes in Computer Science*, pages 125–136. Springer Berlin Heidelberg.

Hanyurwimfura, D., Liu, Y., and Liu, Z. (2010). Text format based relational database watermarking for non-numeric data. In *Proceedings of the 2010 International Conference on Computer Design and Applications (ICCDA)*, pages V4–312–V4–316. IEEE.

Hillig, T. and Mannies, J. (2001). Woman sues over posting of abortion details: her records from granite city hospital were put on web site. hospital, protesters are defendants. *St Louis Post-Dispatch*.

Huang, H., Coatrieux, G., Shu, H., Luo, L., and Roux, C. (2011). Medical image integrity control and forensics based on watermarking - approximating local modifications and identifying global image alterations. In *Engineering in Medicine and Biology Society,EMBC, 2011 Annual International Conference of the IEEE*, pages 8062–8065.

Huang, K., Yue, M., Chen, P., He, Y., and Chen, X. (2009). A Cluster-Based watermarking technique for relational database. In *First International Workshop on Database Technology and Applications*, pages 107–110. IEEE.

Integrating the Healthcare Enterprise (2013). IHE IT infrastructure (ITI) technical framework – volume 1 integration profiles. Revision 10.1.

International Health Terminology Standards Development Organisation (1999). Systematized nomenclature of medicine-clinical terms (SNOMED-CT).

ISO/IEC (2002). International standard audiovisual number (ISAN) – part 1: Audiovisual work identifier.

ISO/IEC (2008). Health informatics – information security management in health using ISO/IEC 27002 (ISO 27799:2008).

ISO/IEC (2013). Information technology – security techniques – code of practice for information security controls (ISO/IEC 27002:2013).

Iwata, M., Hori, T., Shiozaki, A., and Ogihara, A. (2010). Digital watermarking method for tamper detection and recovery of jpeg images. In *Information Theory and its Applications (ISITA), 2010 International Symposium on*, pages 309–314.

Jawad, K. and Khan, A. (2013). Genetic algorithm and difference expansion based reversible watermarking for relational databases. *J. Syst. Softw.*, 86(11):2742–2753.

Johnson, N. L., Kotz, S., and Balakrishnan, N. (1995). *Continuous univariate distributions*. Wiley & Sons.

Kalam, A., Baida, R., Balbiani, P., Benferhat, S., Cuppens, F., Deswarte, Y., Miege, A., Saurel, C., and Trouessin, G. (2003). Organization based access control. In *Policies for Distributed Systems and Networks, 2003. Proceedings. POLICY 2003. IEEE 4th International Workshop on*, pages 120–131.

Kaliski, B. (1998). PKCS#7: Cryptographic message syntax. Technical report, IETF.

Kamel, I. and Kamel, K. (2011). Toward protecting the integrity of relational databases. In *2011 World Congress on Internet Security (WorldCIS)*, pages 258–261. IEEE.

Kamran, M. and Farooq, M. (2012). An information-preserving watermarking scheme for right protection of EMR systems. *IEEE Transactions on Knowledge and Data Engineering*, 24(11):1950–1962.

Kamran, M., Suhail, S., and Farooq, M. (2013a). A formal usability constraints model for watermarking of outsourced datasets. *IEEE Transactions on Information Forensics and Security*, 8(6):1061–1072.

Kamran, M., Suhail, S., and Farooq, M. (2013b). A robust, distortion minimizing technique for watermarking relational databases using once-for-all usability constraints. *IEEE Transactions on Knowledge and Data Engineering*, 25(12):2694 – 2707.

Kerckhoff, A. (1883a). La cryptographie militaire. *Journal des sciences militaires*, IX:5–83.

Kerckhoff, A. (1883b). La cryptographie militaire. *Journal des sciences militaires*, IX:161–191.

Kuribayashi, M. (2010). Experimental assessment of probabilistic fingerprinting codes over awgn channel. In Echizen, I., Kunihiro, N., and Sasaki, R., editors, *Advances in Information and Computer Security*, volume 6434 of *Lecture Notes in Computer Science*, pages 117–132. Springer Berlin Heidelberg.

Lafaye, J., Gross-Amblard, D., Constantin, C., and Guerrouani, M. (2008). Watermill: An optimized fingerprinting system for databases under constraints. *IEEE Transactions on Knowledge and Data Engineering*, 20:532–546.

Lampson, B. (1971). Protection. In *Proc. Fifth Princeton Symposium on Information Science and Systems*, pages 437–443.

Li, N., Li, T., and Venkatasubramanian, S. (2007). t-closeness: Privacy beyond k-anonymity and l-diversity. In *Data Engineering, 2007. ICDE 2007. IEEE 23rd International Conference on*, pages 106–115.

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

Li, Y., Swarup, V., and Jajodia, S. (2005). Fingerprinting relational databases: schemes and specialties. *Dependable and Secure Computing, IEEE Transactions on*, 2(1):34 –45.

Liu, Y., Wang, T., and Feng, J. (2010). A semantic information loss metric for privacy preserving publication. In *Proceedings of the 15th International Conference on Database Systems for Advanced Applications - Volume Part II*, DASFAA'10, pages 138–152, Berlin, Heidelberg. Springer-Verlag.

Machanavajjhala, A., Kifer, D., Gehrke, J., and Venkitasubramaniam, M. (2007). L-diversity: Privacy beyond k-anonymity. *ACM Trans. Knowl. Discov. Data*, 1(1).

Manasrah, T. and Al-Haj, A. (2008). Management of medical images using wavelets-based multi-watermarking algorithm. In *Innovations in Information Technology, 2008. IIT 2008. International Conference on*, pages 697–701.

Mardia, K. V. and Jupp, P. E. (1999). *Directional statistics*. Wiley Series in Probability and Statistics. Wiley, Chichester.

McNickle, M. (2012). Top 10 data security breaches in 2012. In Healthcare Finance News `http://www.healthcarefinancenews.com/news/top-10-data-security-breaches-2012`. Accessed: 11/21/2013.

MedRed (2013). Medred BT health cloud. `http://www.medred.com/content.cfm?m=14&id=14&startRow=1`, Accessed: 07/23/2014.

Meerwald, P. and Furon, T. (2012). Toward practical joint decoding of binary tardos fingerprinting codes. *Information Forensics and Security, IEEE Transactions on*, 7(4):1168–1180.

Menezes, A. J., van Oorschot, P. C., and Vanstone, S. A. (1996). *Handbook of Applied Cryptography*. Discrete Mathematics and Its Applications. CRC Press, Inc., Boca Raton, FL, USA.

Meng, M., Cui, X., and Cui, X. (2008). The approach for optimization in watermark signal of relational databases by using genetic algorithms. In *Computer Science and Information Technology, 2008. ICCSIT '08. International Conference on*, pages 448–452.

Minsky, M. (1974). A framework for representing knowledge. Technical report, Cambridge, MA, USA.

Mintzer, F., Lotspiech, J. B., and Morimoto, N. (1997). Safeguarding digital library contents and users: Digital watermarking. *D-Lib Magazine*, 3(12).

Monga, V. (2005). *Perceptually Based Methods for Robust Image Hashing*. PhD thesis, The University of Texas at Austin, USA.

Neches, R., Fikes, R., Finin, T., Gruber, T., Patil, R., Senator, T., and Swartout, W. R. (1991). Enabling technology for knowledge sharing. *AI Mag.*, 12(3):36–56.

Neyman, J. and Pearson, E. S. (1933). On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 231:pp. 289–337.

Ni, Z., Shi, Y.-Q., Ansari, N., and Su, W. (2006). Reversible data hiding. *Circuits and Systems for Video Technology, IEEE Transactions on*, 16(3):354–362.

NIH (2014). NIH data sharing repositories. `http://www.nlm.nih.gov/NIHbmic/nih_data_sharing_repositories.html`, Accessed: 08/13/2014.

Northwestern University Clinical and Translational Sciences Institute (2011). Northwestern medicine enterprise data warehouse. `http://www.nucats.northwestern.edu/resources-services/data-and-informatics-services/enterprise-data-warehouse/`, Accessed: 08/13/2014.

Nuida, K., Fujitsu, S., Hagiwara, M., Kitagawa, T., Watanabe, H., Ogawa, K., and Imai, H. (2009). An improvement of discrete tardos fingerprinting codes. *Des. Codes Cryptography*, 52(3):339–362.

Oliveira, S. R. M. and Zaïane, O. R. (2002). Privacy preserving frequent itemset mining. In *Proceedings of the IEEE International Conference on Privacy, Security and Data Mining - Volume 14*, CRPIT '14, pages 43–54, Darlinghurst, Australia, Australia. Australian Computer Society, Inc.

Oliveira, S. R. M. and Zaïane, O. R. (2003). Privacy preserving clustering by data transformation. In *XVIII Brazilian Symposium on Databases (SBBD 2003)*, pages 304–318.

Orcutt, M. (2014). Hackers are homing in on hospitals. In MIT Technology Review, `http://www.technologyreview.com/news/530411/hackers-are-homing-in-on-hospitals/`, Accessed: 09/03/2014.

Pan, W. (2012). *Protection des images médicales - tatouage réversible pour le contrôle d'accès et d'usage*. PhD thesis, TELECOM Bretagne and University of Rennes I, France.

Pan, W., Coatrieux, G., Cuppens-Boulahia, N., Cuppens, F., and Roux, C. (2010). Watermarking to enforce medical image access and usage control policy. In *Signal-Image Technology and Internet-Based Systems (SITIS), 2010 Sixth International Conference on*, pages 251–260.

Pournaghshband, V. (2008). A new watermarking approach for relational data. In *Proceedings of the 46th Annual Southeast Regional Conference*, ACM-SE 46, pages 127–131, New York, NY, USA. ACM.

Prasannakumari, V. (2009). A robust tamperproof watermarking for data integrity in relational databases. *Research Journal of Information Technology*, 1(3):115–121.

Proakis, J. G. (2001). *Digital Communications*. McGraw-Hill, 4 edition.

Quinn, B. (2010). Phase-only information loss. In *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*, pages 3982–3985.

Rivest, R. (1992). The MD5 message-digest algorithm. Technical report, IETF.

Rivest, R. L., Shamir, A., and Adleman, L. (1978). A method for obtaining digital signatures and public-key cryptosystems. *Commun. ACM*, 21(2):120–126.

Robshaw, M. (1995). Stream ciphers. Technical Report TR-701, RSA Laboratories.

Rogers, S. (2010). Wikileaks embassy cables: download the key data and see how it breaks down. In The Gurdian Online `http://www.theguardian.com/news/datablog/2010/nov/29/wikileaks-cables-data`. Accessed: 11/21/2013.

Samarati, P. (2001). Protecting respondents' identities in microdata release. *IEEE Trans. on Knowl. and Data Eng.*, 13(6):1010–1027.

Samarati, P. and Sweeney, L. (1998). Protecting privacy when disclosing information: k-anonymity and its enforcement through generalization and suppression. Technical report.

Sandhu, R. S., Coyne, E. J., Feinstein, H. L., and Youman, C. E. (1996). Role-based access control models. *Computer*, 29(2):38–47.

Schaathun, H. (2008). On error-correcting fingerprinting codes for use with watermarking. *Multimedia Systems*, 13(5-6):331–344.

Schmidt, J. M. (2012). Transforming between discrete and continuous angle distribution models: application to protein χ_1 torsions. *Journal of Biomolecular NMR*, 54(1):97–114.

Shah, S., Xingming, S., Ali, H., and Abdul, M. (2011). Query preserving relational database watermarking. *Informatica*, 35(3):391–397.

Shehab, M., Bertino, E., and Ghafoor, A. (2008). Watermarking relational databases using optimization-based techniques. *IEEE Trans. on Knowledge and Data Engineering*, 20:116–129.

Sion, R. (2004). Proving ownership over categorical data. In *Data Engineering, 2004. Proceedings. 20th International Conference on*, pages 584 – 595.

Sion, R., Atallah, M., and Prabhakar, S. (2004). Rights protection for relational data. *Knowledge and Data Engineering, IEEE Trans. on*, 16(12):1509 – 1525.

Sion, R., Atallah, M., and Prabhakar, S. (2005). Rights protection for categorical data. *IEEE Transactions on Knowledge and Data Engineering*, 17:912–926.

Smart card alliance (2003). Hipaa compliance and smart cards: Solutions to privacy and security requirements. Technical report.

Sowa, J. F. (1976). Conceptual graphs for a data base interface. *IBM Journal of Research and Development*, 20(4):336–357.

Stanford Center for Biomedical Informatics Research (BMIR). Protégé ontology editor. `http://protege.stanford.edu/` Accessed: 12/09/2013.

Suchanek, F. M. and Gross-Amblard, D. (2012). Adding fake facts to ontologies. In *Proceedings of the 21st international conference companion on World Wide Web*, WWW '12 Companion, pages 421–424, New York, NY, USA. ACM.

Swick, O. L. . R. R. (1999). Resource description framework (RDF) model and syntax specification, W3C recommendation. `http://www.w3.org/TR/1999/REC-rdf-syntax-19990222/`.

Tardos, G. (2003). Optimal probabilistic fingerprint codes. In *Proceedings of the Thirty-fifth Annual ACM Symposium on Theory of Computing*, STOC '03, pages 116–125.

Tu, S., Kaashoek, M. F., Madden, S., and Zeldovich, N. (2013). Processing analytical queries over encrypted data. In *Proceedings of the 39th international conference on Very Large Data Bases*, PVLDB'13, pages 289–300. VLDB Endowment.

Turban, S. (2010). Convolution of a truncated normal and a centered normal variable. Online in `http://www.columbia.edu/~st2511/notes.html`, Accessed: 04/17/2013.

U.S Government (1996). Health insurance portability and accountability act of 1996, public law 104-191.

U.S Government (2000). Electronic signatures in global and national commerce act, public law 106-229.

U.S Government (2013). Code of federal regulations - title 45: Public welfare.

US National Library of Medicine (1963). Medical subject headings (MESH).

US National Library of Medicine (1986). Unified medical language system (UMLS).

Vapnik, V. N. (1995). *The Nature of Statistical Learning Theory*. Springer-Verlag New York, Inc., New York, NY, USA.

Wagner, N. (1983). Fingerprinting. In *Proceedings of the 1983 Symposium on Security and Privacy*, pages 18–22. IEEE Computer Society.

Wang, C., Wang, J., Zhou, M., Chen, G., and Li, D. (2008a). ATBaM: an arnold transform based method on watermarking relational data. In *International Conference on Multimedia and Ubiquitous Engineering, 2008. MUE 2008*, pages 263–270. IEEE.

Wang, H., Cui, X., and Cao, Z. (2008b). A speech based algorithm for watermarking relational databases. In *Information Processing (ISIP), 2008 International Symposiums on*, pages 603–606. IEEE.

Whetzel, P. L., Parkinson, H., Causton, H. C., Fan, L., Fostel, J., Fragoso, G., Game, L., Heiskanen, M., Morrison, N., Rocca-Serra, P., Sansone, S.-A., Taylor, C., White, J., and Stoeckert, C. J. (2006). The mged ontology: a resource for semantics-based description of microarray experiments. *Bioinformatics*, 22(7):866–873.

Woo, C., Du, J., and Pham, B. (2005). Multiple watermark method for privacy control and tamper detection. In *Proc. APRS Workshop on Digital Image Computing Pattern Recognition and Imaging for Medical Applications*, pages 43–48.

World Health Organization (1992). International statistical classification of diseases and related health problems, 10th revision.

Xiao, X., Sun, X., and Chen, M. (2007). Second-LSB-Dependent robust watermarking for relational database. pages 292–300. IEEE.

Zalenski, R. (2002). Firewall technologies. *Potentials, IEEE*, 21(1):24–29.

Zhang, Y., Niu, X., and Yang, B. (2006). Reversible watermarking for relational database authentication. *Journal of Computers*, Vol.17(2).

Zhao, J. and Koch, E. (1995). Embedding robust labels into images for copyright protection. In *Proceedings of the International Congress on Intellectual Property Rights for Specialized Information, Knowledge and New Technologies*, pages 242–251.

Zhiyong Li, J. L. and Tao, W. (2013). A novel relational database watermarking algorithm based on clustering and polar angle expansion. *International Journal of Security and Its Applications*, 7(2):1–14.

Chapter A
Truncated Normal distribution

The p.d.f of a normally distributed random variable whose values are bounded can be represented by a truncated normal distribution. Let us consider $\gamma \sim \mathcal{N}(\mu, \sigma)$ which lies in the interval $\gamma \in [a, b]$, the truncated density function is:

$$f(\gamma; \mu, \sigma, a, b) = \begin{cases} \frac{\frac{1}{\sigma}\phi(\frac{\gamma-\mu}{\sigma})}{\Phi(\frac{b-\mu}{\sigma})-\Phi(\frac{a-\mu}{\sigma})}, & \text{if } a \leq \gamma \leq b \\ 0, & \text{elsewhere} \end{cases}$$

The moments of this distribution are given by Johnson et al. [1995]:

$$\mathrm{E}[\gamma \mid a < \gamma < b] = \mu + \frac{\phi(\frac{a-\mu}{\sigma}) - \phi(\frac{b-\mu}{\sigma})}{\Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma})}\sigma$$

$$\sigma^2_{(\gamma|a<\gamma<b)} = \sigma^2 \left[1 + \frac{\frac{a-\mu}{\sigma}\phi(\frac{a-\mu}{\sigma}) - \frac{b-\mu}{\sigma}\phi(\frac{b-\mu}{\sigma})}{\Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma})} - \left(\frac{\phi(\frac{a-\mu}{\sigma}) - \phi(\frac{b-\mu}{\sigma})}{\Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma})}\right)^2\right]$$

It is important to remark that a truncated normal distribution defined by a symmetric truncation interval that extends five standard deviations to either side of the mean value would look very similar to a general normal p.d.f, and the scaling would be so mild that the function values, the mean and the variance would also be close.

Chapter B

Impact of database mixtures on the watermark: Additional calculation

In this chapter we present some additional calculus that complete the results on the impact of database mixtures over the embedded message presented in section 3.2 of chapter 3.

B.1 CALCULATION OF $\sigma^2_{s,W}$

First, let us recall some important trigonometric equations that have been employed to obtain $\sigma^2_{s,W}$ in (3.6) from equations (3.2), (3.4) and (3.5).

$$\begin{aligned}
\sin(a+b) &= \cos(b)\sin(a) + \cos(a)\sin(b) \\
\sin(a-b) &= \cos(b)\sin(a) - \cos(a)\sin(b) \\
\cos(a+b) &= \cos(a)\cos(b) - \sin(a)\sin(b) \\
\cos(a-b) &= \cos(a)\cos(b) + \sin(a)\sin(b)
\end{aligned}$$

Then, the calculation of $\sigma^2_{s,W}$ goes as follows:

$$\sigma^2_{s,W} = \sum_{l=0}^{L-1} \sin^2(\frac{2\pi l}{L}) \left[P_\Delta f\left(\frac{2\pi[(l+\Delta) \mod L]}{L}\right) + P_{-\Delta} f\left(\frac{2\pi[(l-\Delta) \mod L]}{L}\right) + P_{ov} f\left(\frac{2\pi l}{L}\right)\right]$$

Then, given that $P_\Delta + P_{-\Delta} = 1 - P_{ov}$ and $P_\Delta = P_{-\Delta}$

$$\sigma^2_{s,W} = \sigma^2_s P_{ov} + \sum_{l=0}^{L-1} \frac{1-P_{ov}}{2} f\left(\frac{2\pi l}{L}\right) \left[\sin^2\left(\frac{2\pi[(l-\Delta)]}{L}\right) + \sin^2\left(\frac{2\pi[(l+\Delta)]}{L}\right)\right]$$

From here, given that $\sin^2(a) = \frac{1-cos(2a)}{2}$

$$\sigma^2_{s,W} = \sigma^2_s P_{ov} + \sum_{l=0}^{L-1} \frac{1-P_{ov}}{2} f\left(\frac{2\pi l}{L}\right) \frac{1}{2} \left[2 - \cos\left(\frac{4\pi[(l-\Delta)]}{L}\right) - \cos\left(\frac{4\pi[(l+\Delta)]}{L}\right)\right]$$

Given that $\cos(a+b) = \cos(a)\cos(b) - \sin(a)\sin(b)$, we have

$$\sigma^2_{s,W} = \sigma^2_s P_{ov} + \sum_{l=0}^{L-1} \frac{1-P_{ov}}{2} f\left(\frac{2\pi l}{L}\right) \left[1 - \cos\left(\frac{4\pi l}{L}\right)\cos\left(\frac{4\pi\Delta}{L}\right)\right]$$

again, given that $1 - 2\sin^2(a) = \cos(2a)$

$$\sigma_{s,W}^2 = \sigma_s^2 P_{ov} + \sum_{l=0}^{L-1} \frac{1 - P_{ov}}{2} f\left(\frac{2\pi l}{L}\right) \left[1 - \cos\left(\frac{4\pi\Delta}{L}\right)\left[1 - 2\sin^2\left(\frac{2\pi l}{L}\right)\right]\right]$$

from (3.2), we obtain

$$\sigma_{s,W}^2 = \sigma_s^2 \left(P_{ov} + (1 - P_{ov})\cos\left(\frac{4\pi\Delta}{L}\right)\right) + \sum_{l=0}^{L-1} \frac{1 - P_{ov}}{2} f\left(\frac{2\pi l}{L}\right)\left[1 - \cos\left(\frac{4\pi\Delta}{L}\right)\right]$$

Finally, given that $\sum_{l=0}^{L-1} f\left(\frac{2\pi l}{L}\right) = 1$ we obtain

$$\sigma_{s,W}^2 = \sigma_s^2 \left(P_{ov} + (1 - P_{ov})\cos\left(\frac{4\pi\Delta}{L}\right)\right) + (1 - P_{ov})\sin^2\left(\frac{2\pi\Delta}{L}\right)$$

B.2 VECTOR ROTATIONS

As exposed, the center of mass vector of the watermarked attribute A_t results from a linear combination of three vectors, each of which corresponds to rotated versions of the original center of mass vector $V^{A,i}$ (resp. $V^{B,i}$).

In $\mathbb{R}^2$, the rotation of a vector given its Cartesian coordinates (x, y) and a rotation angle α is defined by a rotation matrix $\Omega(\alpha)$ such as:

$$\Omega(\alpha) = \begin{pmatrix} \cos(\alpha) & -\sin(\alpha) \\ \sin(\alpha) & \cos(\alpha) \end{pmatrix} \tag{B.1}$$

The Cartesian components of the rotated vector (x', y') are calculated such as:

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \cos(\alpha) & -\sin(\alpha) \\ \sin(\alpha) & \cos(\alpha) \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \tag{B.2}$$

Notice also that:

$$\sin(-a) = -\sin(a)$$
$$\cos(-a) = \cos(a)$$

Thus, the vector V' resulting from the addition of two vectors V_α and $V_{-\alpha}$ corresponding to rotated versions of the same vector V in α and $-\alpha$ respectively results in:

$$V' = V_\alpha + V_{-\alpha} = (\Omega(\alpha) + \Omega(-\alpha))V = 2\cos(\alpha)V \tag{B.3}$$

This result is applied in equation (3.7) in order to obtain the vector $V_W^{A,i}$ (resp. $V_W^{B,i}$).

Chapter C
Neyman-Pearson Lemma

Let us consider the following scenario: We have a dataset $X = (x_1, ..., x_n) \in \mathbb{R}^n$, where $\{x_i\}_{i=1,...,n}$ are independent and identically distributed (i.i.d.) random variables, and we want to determine the parameters θ of the p.d.f. these variables follow. We have two hypothesis:

- The hypothesis H_0 that the variables in X are i.i.d. according to a p.d.f. $f(x|\theta_0)$
- The hypothesis H_1 that the variables in X are i.i.d. according to a p.d.f. $f(x|\theta_1)$

The Neyman-Pearson lemma states that the optimal decision rule, i.e., the rule that maximizes the correct selection probability $P(\text{Select } H_1|H_1)$, when performing such an hypothesis test is the likelihood-ratio test $\Lambda(X)$ [Neyman and Pearson, 1933], where $\Lambda(X)$ is defined such as:

$$\Lambda(X) = \frac{L(X|\theta_1)}{L(X|\theta_0)} \lessgtr \eta \tag{C.1}$$

where $L(X|\theta_1) = \prod_{i=1}^{n} f(x_i|\theta_1)$ and $L(X|\theta_0) = \prod_{i=1}^{n} f(x_i|\theta_0)$ are the likelihood functions and η is the decision threshold defined so as to obtain:

$$P(\Lambda(X) \geq \eta \mid H_0) = \alpha \tag{C.2}$$

where α is the probability of false alarm, i.e., the probability of rejecting H_0 when it is the correct hypothesis.

Notice that in some cases, the use of the natural logarithm of the likelihood function, i.e., the log-likelihood, is of interest in order to reduce the complexity of the calculation, notably by transforming the products in sums. This is possible due to the monotonically increasing nature of the natural logarithm. We have then:

$$\Lambda(X) = \sum_{i=1}^{n} \log \frac{f(x_i|\theta_1)}{f(x_i|\theta_0)} \lessgtr \eta \tag{C.3}$$

Chapter D

Calculation of the detection scores and thresholds

In this appendix we present additional calculation on the detection scores and thresholds exposed in section 3.3.2 of Chapter 3.

D.1 SOFT-DECISION BASED DETECTION

The proposed detection score is based on the log-likelihood ratio between the p.d.f of the extracted angle value β_i conditionally to the embedded symbol s_i^j:

$$Sc^j = \sum_{i=1}^{N_g} s_i^j \cdot Sc_i^j = \sum_{i=1}^{N_g} s_i^j \cdot \log \frac{p(\hat{S}_i | K_s^{emb} = K_s^j, s_i^j = 1)}{p(\hat{S}_i | K_s^{emb} = K_s^j, s_i^j = -1)} \tag{D.1}$$

The corresponding p.d.f are such as:

$$p(\hat{S}_i | K_s^{emb} = K_s^j, s_i^j) \sim \begin{cases} \mathcal{N}(P_{DB_j}(-\frac{4\pi\Delta}{L}), \sigma^2_{\beta_i^{merge}}), & \text{if } s_i^j = -1 \\ \mathcal{N}(P_{DB_j}(\frac{4\pi\Delta}{L}), \sigma^2_{\beta_i^{merge}}), & \text{if } s_i^j = 1 \end{cases} \tag{D.2}$$

Then, we can operate in order to obtain the detection scores. For one extracted angle β_i we have:

$$\begin{aligned} Sc_i^j &= \log \left[\frac{e^{-(\beta_i - P_{DB_j}\frac{4\pi\Delta}{L})^2 / 2\sigma^2_{\beta_i^{merge}}}}{e^{-(\beta_i + P_{DB_j}\frac{4\pi\Delta}{L})^2 / 2\sigma^2_{\beta_i^{merge}}}} \right] = \frac{(\beta_i + P_{DB_j}\frac{4\pi\Delta}{L})^2 - (\beta_i - P_{DB_j}\frac{4\pi\Delta}{L})^2}{2\sigma^2_{\beta_i^{merge}}} \\ &= 2P_{DB_j}\beta_i \frac{4\pi\Delta}{L\sigma^2_{\beta_i^{merge}}} \end{aligned} \tag{D.3}$$

Thus, for the whole sequence $\hat{S}$, the detection score is calculated as:

$$Sc^j = \sum_{i=1}^{N_g} s_i^j \cdot 2P_{DB_j}\beta_i \frac{4\pi\Delta}{L\sigma^2_{\beta_i^{merge}}} \tag{D.4}$$

As exposed, based on this score, the decision about the presence of a database in a data warehouse can be made automatic by defining a decision threshold Z accordingly to a false

detection probability P_{FA}, i.e., identifying the database while it is not present and vice et versa. Due to the central limit theorem, the detection threshold Z can be derived, for a given P_{FA} and n possible identification sequences (with $n \geq |\mathsf{U}'|$), from:

$$P(Sc^j > Z | j \notin \mathsf{U}') = \frac{P_{FA}}{n} = \frac{1}{2}\mathrm{erfc}\left(\frac{Z}{\sqrt{2\sigma_{SC}^2}}\right) \tag{D.5}$$

We need to calculate the variance of the detection score σ_{SC}^2 in the case that j is not present in the data warehouse. In such a context, we can assume that β_i and s_i^j are independent. Given that $\mathsf{var}(s_i^j) = 1$, we have:

$$\sigma_{SC}^2 = \mathsf{var}\left(\sum_{i=1}^{N_g} s_i^j \cdot 2P_{DB_j}\beta_i \frac{4\pi\Delta}{L\sigma_{\beta_i^{merge}}^2}\right) = \left(2P_{DB_j}\frac{4\pi\Delta}{L\sigma_{\beta_i^{merge}}^2}\right)^2 N_g \sigma_{\beta_i^{mix}}^2 \tag{D.6}$$

Then, Z can be calculated from (D.5) as:

$$Z = \sqrt{2N_g\sigma_{\beta_i^{mix}}^2 (2P_{DB_j}\frac{4\pi\Delta}{L\sigma_{\beta_i^{merge}}^2})^2} \cdot \mathrm{erfc}^{-1}(2P_{FA}/n) \tag{D.7}$$

D.2 INFORMED DETECTION

The detection score is defined according to our knowledge on the p.d.f of the merged databases.

$$Sc^j = \sum_{i=1}^{N_g} Sc_i^j = \sum_{i=1}^{N_g} \log \frac{p(\hat{S}_i | K_s^{emb} = K_s^j, s_i^j)}{p(\hat{S}_i | K_s^{emb} \neq K_s^j)} \tag{D.8}$$

The corresponding p.d.f are such as:

$$p(\hat{S}_i | K_s^{emb} = K_s^j, s_i^j) \sim \begin{cases} \mathcal{N}(P_{DB_j}(-\frac{4\pi\Delta}{L}), \sigma_{\beta_i^{merge}}^2), & \text{if } s_i^j = -1 \\ \mathcal{N}(P_{DB_j}(\frac{4\pi\Delta}{L}), \sigma_{\beta_i^{merge}}^2), & \text{if } s_i^j = 1 \end{cases}$$

$$p(\hat{S}_i | K_s^{emb} \neq K_s^j) \sim \mathcal{N}(0, \sigma_{\beta_i^{mix}}^2) \tag{D.9}$$

Then, we can operate in order to obtain the detection scores. For one extracted angle β_i

we have:

$$\begin{aligned} Sc_i^j &= \log\left[\frac{\frac{1}{\sigma_{\beta_i^{merge}}} e^{-(\beta_i - s_i^j P_{DB_j}\frac{4\pi\Delta}{L})^2/2\sigma^2_{\beta_i^{merge}}}}{\frac{1}{\sigma_{\beta_i^{mix}}} e^{-(\beta_i)^2/2\sigma^2_{\beta_i^{mix}}}}\right] \\ &= \log\left(\frac{\sigma_{\beta_i^{mix}}}{\sigma_{\beta_i^{merge}}}\right) + \frac{\sigma^2_{\beta_i^{merge}}\beta_i^2 - \sigma^2_{\beta_i^{mix}}(\beta_i - s_i^j P_{DB_j}\frac{4\pi\Delta}{L})^2}{2\sigma^2_{\beta_i^{mix}}\sigma^2_{\beta_i^{merge}}} \\ &= \log\left(\frac{\sigma_{\beta_i^{mix}}}{\sigma_{\beta_i^{merge}}}\right) + \frac{\beta_i^2(\sigma^2_{\beta_i^{merge}} - \sigma^2_{\beta_i^{mix}}) + 2s_i^j\beta_i\sigma^2_{\beta_i^{mix}}P_{DB_j}\frac{4\pi\Delta}{L} - \sigma^2_{\beta_i^{merge}}(P_{DB_j}\frac{4\pi\Delta}{L})^2}{2\sigma^2_{\beta_i^{mix}}\sigma^2_{\beta_i^{merge}}} \end{aligned} \tag{D.10}$$

Once again, due to the central limit theorem, the detection score of a database that has not been uploaded to the data warehouse is normally distributed. Thus, one can compute a detection threshold Z for a given false alarm probability such as:

$$Z = \sqrt{2\sigma^2_{SC}} \cdot \mathrm{erfc}^{-1}(2P_{FA}/n) + m_{SC} \tag{D.11}$$

In this case, additionally to the calculation of σ^2_{SC}, a term m_{SC} has to be calculated in order to obtain Z. This is due to the fact that the mean of the detection scores is not zero (see D.10). In the case that the database j is not present in the data warehouse, we can assume that β_i and s_i^j are independent. Then m_{SC} is calculated as:

$$m_{SC} = \mathrm{E}\left[\sum_{i=1}^{N_g} Sc_i^j\right] = N_g \log\left(\frac{\sigma_{\beta_i^{mix}}}{\sigma_{\beta_i^{merge}}}\right) + \mathrm{E}\left[\sum_{i=1}^{N_g} \frac{\sigma^2_{\beta_i^{merge}}\beta_i^2 - \sigma^2_{\beta_i^{mix}}(\beta_i - s_i^j P_{DB_j}\frac{4\pi\Delta}{L})^2}{2\sigma^2_{\beta_i^{mix}}\sigma^2_{\beta_i^{merge}}}\right] \tag{D.12}$$

Given that $\mathrm{E}[s_i^j] = 0$ and each β_i is i.i.d. with $\mathrm{E}[\beta_i] = 0$ for $j \notin \mathsf{U}'$, we have:

$$m_{SC} = N_g\left[\log\left(\frac{\sigma_{\beta_i^{mix}}}{\sigma_{\beta_i^{merge}}}\right) - \frac{(P_{DB_j}\frac{4\pi\Delta}{L})^2}{2\sigma^2_{\beta_i^{mix}}} + \left(\frac{\sigma^2_{\beta_i^{merge}} - \sigma^2_{\beta_i^{mix}}}{2\sigma^2_{\beta_i^{merge}}}\right)\right] \tag{D.13}$$

For the calculation of σ^2_{SC}, we will make use of the following statistics result. Given a Gaussian random variable ψ, the variance of ψ^2, i.e., $\sigma^2_{\psi^2}$, is such as:

$$\sigma^2_{\psi^2} = 4\sigma^2_{\psi}\,\mathrm{E}[\psi]^2 + 2\sigma^4_{\psi} \tag{D.14}$$

Given that $\mathrm{E}[\beta_i] = 0$ for $j \notin \mathsf{U}'$, we calculate σ^2_{SC} as:

$$\sigma^2_{SC} = N_g\sigma^2_{\beta_i^{mix}}\left[\left(\frac{P_{DB_j}}{\sigma^2_{\beta_i^{merge}}}\frac{4\pi\Delta}{L}\right)^2 + \left(\left(\frac{\sigma^2_{\beta_i^{merge}} - \sigma^2_{\beta_i^{mix}}}{2\sigma^2_{\beta_i^{mix}}\sigma^2_{\beta_i^{merge}}}\right)^2 2\sigma^2_{\beta_i^{mix}}\right)\right] \tag{D.15}$$

Résumé

Avec l'évolution des technologies de l'information et des communications, l'accès à distance, la collecte et la gestion de données au sein d'entrepôts voire tout simplement de bases de données sont aujourd'hui l'enjeu d'intérêts économiques et stratégiques tant pour les entreprises que pour les institutions publiques. Dans un tel contexte, les fuites, le vol ou encore la dégradation volontaire ou non de l'information représentent un danger réel qui nécessite des méthodes de protection nouvelles et plus appropriées que l'offre actuelle, par ailleurs limitée. Les travaux réalisés dans cette thèse de doctorat ont pour objectif d'intégrer le tatouage ou «watermarking» à la famille d'outils existantes pour la protection de bases de données, avec pour cas d'étude le domaine de la santé. Dans son principe, le tatouage consiste à insérer un message ou une marque de façon imperceptible dans une base de manière à pouvoir, par exemple, déterminer son origine, connaître le dernier utilisateur à y avoir accédé ou vérifier son intégrité. Un avantage majeur du tatouage par rapport aux autres outils est qu'il laisse l'accès à l'information tout en la gardant protégée. Cependant, il faut s'assurer que la distorsion introduite par le tatouage ne perturbe pas l'interprétation de l'information dans la base. Une première partie de ces travaux a porté sur le tatouage réversible et robuste de bases de données. La propriété de réversibilité garantit la récupération des données originales après avoir retiré la marque insérée. Cette propriété permet de relâcher les contraintes d'imperceptibilité et aussi d'autoriser la mise à jour de la marque. Une deuxième partie de ces activités de recherche a focalisé sur la minimisation de la distorsion liée à l'insertion de la marque, y compris dans le cas réversible. L'objectif est de pouvoir garder la marque dans la base et assurer ainsi une protection continue de celle-ci. La solution proposée s'appuie sur une modélisation ontologique de la sémantique de la base qui traduit les relations clé à préserver entre les valeurs des attributs. Ainsi guidée, notre approche évite l'apparition de valeurs incohérentes ou improbables qui pourraient perturber l'interprétation de la base comme faciliter l'attaque de la marque.

Mots-clés : Information médicale, Bases de données, Sécurité, Tatouage réversible, Contrôle Sémantique de la distorsion, Ontologie.

Abstract

With the evolution of information and communication technologies, data gathering and management into data warehouses or simple databases represent today economical and strategic concerns for both enterprises and public institutions. Remote access and storage pools are now a reality allowing different entities to cooperate and reduce expenses. In that context, data leaks, robbery as well as innocent or even hostile data degradation represent a real danger needing new protection solutions, more effective than the existing ones. The work conducted during this Ph.D. thesis aims at the integration of watermarking in the family of existing database protection tools, considering the healthcare framework as case of study. Watermarking is based in the imperceptible embedding of a message or watermark into a database, watermark which allows us, for instance, to determine its origin as well as identifying its last user or verifying its integrity. A major advantage of watermarking in relation to other mechanisms is the fact that it enables to access the data while keeping them protected. Nevertheless, it is necessary to ensure that introduced distortion do not perturb the interpretation of the information contained in the database. A first part of this work has focused on reversible and robust database watermarking. The reversibility property ensures the recovery of the original data once the embedded sequence has been extracted. As defined, it allows us to relieve the constraints of watermark imperceptibility as well as to perform an update of the mark. A second part of these research activities has dealt with the minimization of the semantic distortion produced by the watermark embedding process, even in the lossless case. Indeed, even in this latter case, there is an interest in minimizing the introduced distortion in order to keep the watermark into the database, ensuring thus a continuous protection. The solution we propose is based on an ontological modeling of the database semantics which represent the relationships between attributes that should be preserved. By doing so, our approach avoids the appearance of incoherent and unlikely values which could perturb the interpretation of the database as well as indicate the presence of a watermark to a potential attacker.

Keywords : Health information, Databases, Security, Lossless watermarking, Semantic Distortion Control, Ontology.

TELECOM Bretagne

n° d'ordre : 2014telb0347

Télécom Bretagne

Technopôle Brest-Iroise - CS 83818 - 29238 Brest Cedex 3

Tél : + 33(0) 29 00 11 11 - Fax : + 33(0) 29 00 10 00