Protection of encrypted and/or compressed medical images by means of watermarking

Sahar Haddad

UNIVERSITE BRETAGNE LOIRE / MATHSTIC

IMT Atlantique
Bretagne-Pays de la Loire
École Mines-Télécom

THESE DE DOCTORAT DE

L'ÉCOLE NATIONALE SUPERIEURE MINES-TELECOM ATLANTIQUE
BRETAGNE PAYS DE LA LOIRE - IMT ATLANTIQUE

ECOLE DOCTORALE N° 601
Mathématiques et Sciences et Technologies de l'Information et de la Communication
Spécialité : *Signal,Image,Vision*

Par

Sahar HADDAD

Protection of Encrypted and/or Compressed Medical Images by Means of Watermarking

Thèse présentée et soutenue à Brest le 09/07/2020
Unité de recherche : LaTIM (Laboratoire de Traitement de l'Information Médicale)
Thèse N° : 2020IMTA0184

Rapporteurs avant soutenance :

Philippe Carré	Professeur, Xlim
Nacim Betrouni	Chargé de Recherche, Lille Neuroscience & Cognition – INSERM U1172

Composition du Jury :

Président :	Henri Maître	Professeur Émérite
Examinateurs :	Jean Claude Nunes	Maître de Conférences
	Philippe Carré	Professeur, Xlim
	Nacim Betrouni	Chargé de Recherche, Lille Neuroscience&Cognition–INSERM U1172
	Dalel Bouslimi	Product owner
Dir. de thèse :	Gouenou Coatrieux	Professeur, IMT Atlantique.
Co-dir. de thèse :	Alexandre Moreau-Gaudry	Professeur, TIMC Grenoble.

Invité

Michel Cozic	Dir. R&D, MEDECOM.

À mes parents
À ma sœur
À la mémoire de ma grand-mère
À tous ceux qui me sont chers, ...

Remerciements

Le travail présenté dans cette thèse a été réalisé au sein du département ITI de l'IMT Atlantique Bretagne-Pays de la Loire.

Par ces quelques lignes, je tiens à remercier toutes les personnes qui ont participé de près ou de loin au bon déroulement de cette thèse, en espérant n'avoir oublié personne.

Je tiens à adresser en premier lieu mes plus chaleureux remerciements à mon directeur de thèse Professeur Gouenou Coatrieux. Les conseils qu'il m'a prodigué, la patience, la confiance qu'il m'a témoignée ont été déterminants dans la réalisation de mon travail de recherche.

Je remercie également le professeur Alexandre Moreau-Gaudry, mon co-directeur de thèse pour son encadrement tout au long de cette thèse, ses conseils, et ses encouragements.

Ma très vive reconnaissance va à Monsieur Michel Cozic, directeur R&D, MEDECOM. Je vous remercie d'avoir cru en moi et en mes capacités, pour vos précieux conseils, et pour le temps que vous m'avez accordé ces années.

J'exprime tous mes remerciements à Professeur Henri Maitre d'avoir accepté de présider le jury d'examen. Je lui adresse mes sentiments les plus respectueux.

Je tiens à remercier Messieurs Philippe Carré et Nacim Betrouni pour avoir bien voulu rapporter sur mon travail de thèse et pour toutes leurs précieuses remarques qui m'ont beaucoup aidé à améliorer la qualité de ce travail.

Je remercie également Madame Dalel Bouslimi et Monsieur Jean-Claude Nunes d'avoir accepter de participer à mon jury de thèse.

Je tiens à remercier vivement tous les membres de ma famille, qui m'ont énormément aidée et tenue tout au long de ces années par leur soutien moral continu.

Pour conclure, je souhaite exprimer mes remerciements à tous mes amis et mes collègues pour leur aide et leur soutien.

Abstract

The rapid growth of information and communication technologies have offered new possibilities to store, access and transfer medical images over networks in between practitioners, for a second opinion for example, or between the physician and the patient him-/herself. In that context, data leaks, robbery as well as innocent or malicious data manipulations represent a real danger needing new protection solutions, more effective than the existing ones, especially in terms of data confidentiality, reliability control and data traceability.

The work conducted during this Ph.D. thesis aims at the combination, or even the fusion of different medical image security mechanisms. Among these mechanisms, one can find image encryption that transforms on the basis of an encryption key a plain-text image into an incomprehensible encrypted one. If this technique guarantees the confidentiality of the data, it offers a protection of the type *"a priori"* because, once deciphered, a data is no longer protected. Watermarking has been proposed as a complementary mechanism to the *"a priori"* protection solutions - called as an "*a posteriori* protection" - leaving access to information while keeping it protected. Basically, watermarking is defined as the invisible embedding of a message (e.g. digital signatures, users' ID, access rights) in the image by imperceptibly modifying its pixels' values. Such a message can be then used so as to verify whether the image has been modified, and/or whether it has been illegally redistributed.

The deployment of these security mechanisms in the healthcare domain must take into account its specificities. Notably, because medical images constitute large amounts of data, they are most often encoded in lossy or lossless compressed form to minimize or reduce transmission and storage costs. Thus, it becomes desirable to develop protection mechanisms by taking into account that medical images are compressed.

The first part of this work focused on joint watermarking-compression of medical images so as to be able to verify image integrity and authenticity without decompressing it, with the help of watermark that can be extracted without decompressing the image compressed bitstream, even partially. We proposed a solution whose principles remain the same whether applied to JPEG-LS or JPEG images.

In the continuity of this work, we were interested in verifying the reliability of compressed and encrypted images while maintaining their confidentiality. Its main principle is based on the insertion of two messages, containing security attributes, in the image during the image compression and encryption. Each of them is only accessible in one domain: the compressed domain or the encrypted domain, without having to decompress or decrypt the image, even partially. Note that all of these approaches introduce low image distortion and provide sufficient insertion capacity to go beyond controlling the image reliability. They also save computation time, because there is no need to decrypt nor to decompress the image to verify its authenticity and integrity. Moreover, they have been developed to be compliant to the DICOM standard, by combining JPEG-LS/JPEG compression, AES encryption in CBC mode, and watermarking based on bit substitution.

A second part of these research activities focused on the reversible watermarking of encrypted medical images. The reversibility property guarantees the recovery of the original image after removing the inserted watermark. We have developed an original watermarking scheme allowing a reversible insertion of a message in an encrypted image, based on a new robust histogram shifting reversible watermarking modulation. This message is accessible in both encrypted and clear domains (i.e. from the decrypted image), and can be used to control the reliability of the image, if or not it is encrypted.

Finally, in order to validate the distortion introduced by watermark embedding in the proposed solutions, we have implemented a "psychovisual" assessment protocol for medical CT image watermarking in collaboration with MEDECOM, CIC-IT, the Imaging Pole and the University Clinic of Radiology and Medical Imaging (CURIM) of the Grenoble-Alpes University Hospital. The objectives of this study are twofold: i) to determine the levels of distortion considered acceptable by the experts and ii) to tackle the problem of the acceptability of the watermark which by its presence should not change the medical image diagnosis.

Keywords **-** Medical imaging, Security, Watermarking, Encryption, JPEG-LS Compression, JPEG Compression.

Résumé

L'évolution rapide des technologies du multimédia et des communications s'exprime dans le domaine de la santé par la mise à disposition de nouveaux moyens de partage et d'accès distant aux données de l'imagerie des patients. Dans un tel contexte, la question de la sécurité des données est particulièrement sensible, notamment en termes de confidentialité, authenticité, intégrité et traçabilité.

Ces travaux de thèse ont trait à combinaison voir la fusion de différents mécanismes pour la protection des images médicales. Parmi ceux-ci, on trouve le chiffrement qui transforme sur la base d'une clé de chiffrement un texte en clair en un texte chiffré incompréhensible. Si cette technique garantie la confidentialité des données, elle offre avant une protection de type *"a priori "* car, une fois déchiffrée, une donnée n'est plus protégée. Le tatouage a été proposé comme un mécanisme complémentaire aux solutions de protection *"a priori "*, en laissant l'accès à l'information tout en la gardant protégée. Dans son principe, il s'appuie sur une distorsion contrôlée des données à protéger pour y dissimuler un message ou une marque (i.e. des attributs de sécurité – ex. signature numérique, codes d'authenticité – preuves de l'origine d'une image et de son attachement à un patient donné) qui peut ensuite aider à identifier leur origine (traçabilité) ou savoir si elles ont été modifiées (intégrité).

Le déploiement de ces mécanismes de sécurité dans le domaine de la santé doit prendre en compte les spécificités de ce domaine. Notamment, du fait que les images médicales constituent de grands volumes de données, elles sont le plus souvent encodées sous forme compressée avec ou sans pertes afin de minimiser ou de réduire les coûts de transmission et de stockage. Ainsi, il devient souhaitable de développer des mécanismes de protection en tenant compte du fait que les images médicales sont compressées.

La première partie de ces travaux a porté sur le tatouage-compression conjoint d'images médicales de manière à pouvoir vérifier l'intégrité et l'authenticité d'une image sans la décompresser, à l'aide d'un message tatoué que l'on peut extraire sans décoder, même partiellement, le flux binaire. Nous avons proposé une solution dont les principes restent les mêmes qu'elle soit appliquée à des images JPEG-LS ou JPEG. Dans la continuité de ces travaux, nous nous sommes intéressés à vérifier la fiabilité des images compressées et chiffrées tout en maintenant leur confidentialité. Son principe général est fondé sur l'insertion de deux messages contenant des attributs de sécurité dans l'image durant la compression et le chiffrement de cette dernière. Chacun d'eux n'est accessible que dans un domaine : le domaine compressé ou le domaine chiffré, sans avoir à décompresser ou déchiffrer l'image, même partiellement. Toutes ces approches introduisent une faible dégradation de l'image et offrent une capacité d'insertion suffisante pour aller au-delà du contrôle de la fiabilité de l'image. Elles permettent de gagner en temps de calcul, car il n'est pas nécessaire de déchiffrer ou décompresser l'image pour vérifier son authenticité et son intégrité. Elles ont par ailleurs été développées pour être compatible avec le standard DICOM, en combinant la compression JPEG-LS/JPEG, le chiffrement AES en mode CBC, et le tatouage fondé sur la substitution de bits.

Une deuxième partie de ces activités de recherche a focalisé sur le tatouage réversible d'images médicales chiffrées. La propriété de réversibilité garantit la récupération de l'image originale après avoir retiré la marque insérée. Nous avons développé un schéma de tatouage original permettant d'insérer une marque de manière réversible dans une image chiffrée, sur la base d'une nouvelle modulation de tatouage par décalage d'histogramme robuste. Ce message est accessible dans les deux domaines chiffré et spatial (i.e. quand l'image est déchiffrée), et peut être utilisé pour vérifier la fiabilité de l'image, qu'elle soit ou non chiffrée.

Enfin, dans le but de valider les solutions proposées vis-à-vis de la distorsion introduite par le tatouage, nous avons mis en place un protocole d'évaluation "psychovisuel" du tatouage en collaboration avec le CIC-IT et le Pôle d'Imagerie et la Clinique Universitaire de Radiologie et Imagerie Médicale (CURIM) du CHU Grenoble-Alpes. Les objectifs de cette étude sont doubles: i) déterminer les niveaux de distorsion considérés comme acceptables par les experts et ii) aborder le problème de l'acceptabilité de la marque qui par sa présence ne doit pas changer les pratiques des radiologues.

Mots clés - Imagerie médicale, Sécurité, Tatouage, Chiffrement, Compression JPEG-LS, Compression JPEG.

Contents

List of Figures

List of Tables

General Introduction

Medical imaging refers to techniques and processes used to create images of various parts of the human body for diagnostic and treatment purposes. Being well-known through various radiological imaging modalities such as: X-ray radiography, Magnetic resonance imaging (MRI), Computed Tomography (CT); medical imaging also includes ultrasound modalities, as well as in the visible spectrum like in ophthalmology and dermatology. Nowadays, medical imaging is an essential component of the care pathway, adding value in each stage it is involved in. Medical imaging contributes to better and more accurate diagnoses from the start and, through continuous monitoring allowing better care decisions and more effective treatments and results, in general. In addition, along with the introduction and the exponential development of telecommunication technologies in the healthcare domain, medical images are cross-exchanged in right time allowing new medical practices participating in the connection between physicians and other caregivers as well as the patient through, for example telediagnosis, teleconsultation and tele-expertise services. To sum up, medical images within the modern health information infrastructure present significant benefits.

However, while the recent advances in information and communication technologies provide new means to access, handle and exchange medical images, they also compromise their security. Medical images can be intentionally or unintentionally manipulated both within the secure medical system environment and outside. Hardly a day goes by without an article in the press featuring hacking of medical data or security breaches [1]. Due to the fact that medical imaging plays more and more an important role in the health system, questions of medical images' security are of utmost importance. Indeed, as any medical records, a medical image is a sensitive piece of data in terms of confidentiality, integrity (the alteration of an image can bring misdiagnosis) and so on.

Security requirements of medical information are mostly derived from strict and strong legislative and ethic regulations, that professionals are obliged to respect (e.g the French medical ethic code, the General Data Protection regulation (GDPR)[2]). This requires four mandatory features to satisfy: confidentiality, reliability, availability, and traceability. Confidentiality indicates that only authorized users can access medical data. Reliability corresponds to the confidence one can have into a piece of data. Basically, data reliability relies on the outcomes of: i) integrity - a proof that a piece of information has not been modified by a non-authorized user; and, ii) authenticity - a proof that certifies a piece of data belongs to the correct patient and is issued from the right source. Being sure that a piece of information is reliable, health professionals will use it in complete trust for diagnosis purposes. Availability defines the ability of authorised persons to use the health information system in the normally scheduled situations of access and practice. Data traceability stands on the proofs of the reliability of data, i.e. proofs of data authenticity and integrity, all along their life cycle. Hence, it is easy to understand the reasons of these needs: any violation of these security principles endangers the patient care and health, while also having consequences for the healthcare professional or institution. Furthermore, security is the base of a trustful relationship between patients and the health care system (practitioners, administration, *etc.*). As example, being sure that his or her say are confidentially stored, a patient will talk about his or her health problem without any shadows.

In order to counter-fight these threats, several protection mechanisms have been proposed. In an information system, such as the one at the hospital, the mechanism stands on the definition and deployment of a security policy. To do so, one can for instance follows ISO 27005 [3]. Such a standard asks for a security risk analysis of the information system, against various threats (accidents - e.g., material failures, natural phenomena, negligence; errors - e.g., mistyping, transmission errors; attacks and misappropriations - e.g., frauds, piracy, blackmailing), and helps to identify the appropriate security mechanisms to counteract possible threats. A non-exhaustive list includes access control, user rights management and encryption which are helpful for confidentiality while

digital signatures will ensure data integrity. However, these security solutions offer an *a priori* protection or in other words, once they are bypassed or more simply when the access to data content is granted, data are no longer protected. Here comes the interest of an "*a posteriori*" protection, a kind of protection watermarking ensures [4]. Basically, when it is applied to images, watermarking is defined as the invisible embedding of a message into a host image by imperceptibly modifying its gray values; such a message can be used so as to verify whether the image has been modified as well as whether it has been illegally redistributed. Watermarking leaves access to the data while maintaining them protected by the message.

On the other side, the ever-growing quantities of information needing to be managed have to be taken into account while protecting medical images. In fact, on a daily basis, large amount of medical images are acquired. Note that these images, in turn, constitute large volumes of data: for instance, a typical CT heart exam can yield a data size up to 1 Gigabyte. Compression of medical images is thus largely exploited to optimize costs of storage and of transmission. Images can be lossy or lossless compressed. If in the latter case no information loss occurs, that is to say that the original data are exactly recovered from the compressed data bitstream (e.g., using JPEG-LS, Lempel–Ziv compression standards), lossy compression tends to be more and more accepted (e.g., using JPEG or lossy JPEG2K). From this standpoint, it becomes desirable to develop solutions that can give access to watermarking-based security services (e.g. integrity and authenticity control) in the compressed domain.

In the medical domain, the communication, data format, storage, retrieval, visualization and printing of medical imaging information are specified by the international medical imaging standard DICOM (Digital Imaging and Communications in Medicine) [5]. DICOM takes into account different security aspects in its part 15. To ensure image confidentiality, while being DICOM compliant, the triple DES or AES encryption can be used. Image integrity can be provided with the help of the DSA digital signature. Such a signature will be located in the DICOM file header of the image. As a consequence, developing new watermarking-based security services for compressed or protected medical images should also be compliant to DICOM.

This is the context of our research works conducted during the three years of PhD. This thesis is structured as follows. Chapter 1 provides some general definitions about the main domains we addressed in order to position the problems we focused on. We will thus come back on: medical imaging modalities - the way medical images are produced; medical image information systems - the way medical images are shared and stored; the security needs associated to medical images in such open environment. We will also focus on the actual security tools, both general and specific to medical images, underlying their weakness and highlighting the interest of watermarking as a complementary security mechanism to these solutions. Then, we will introduce the main principles of image watermarking, its properties as well as its techniques. A state of the art of existing medical image watermarking methods concludes this chapter.

Chapter 2 is devoted to joint watermarking-compression of medical images. Existing methods, beyond the medical domain, are reviewed before focusing on the method proposed for medical image protection. The weaknesses of such methods led us to focus on a solution that allows accessing to watermarking-based security services from compressed data, an original topic. More clearly, as medical images constitute large volumes of data stored and distributed in a compressed form, we were interested in the access to the security services directly from the compressed image without having to decompress it, even partially. From this system, we have derived two schemes: the first embeds a watermark during the JPEG-LS compression of the image [6, 7] while the second watermarks an image during its JPEG compression. Note that both schemes are DICOM-compliant as JPEG-LS and JPEG are part of the DICOm standard. Furthermore, with our schemes watermark extraction is independent to the decompression process; that is to say, image decompression can be performed with the common standard decompression algorithm. Performance of these schemes are experimentally validated on large image data set, showing their suitability for different security objectives like integrity control and data authenticity.

In Chapter 3, we propose an amelioration of the method proposed in the previous chapter adding the possibility to access to various watermarking-based security services from the encrypted-compressed image and from the decrypted-compressed image bitstream [8]. This scheme comes from the interest of offering an *a priori* an *a posteriori* image protection. The originality of this proposal is twofold. First, as already said, it allows accessing to watermarking-based security services from both encrypted and compressed image bitstream without having to decrypt or to decompress them, even partially. Second, it combines the bit-substitution watermarking modula-

tion with JPEG-LS and the AES block cipher algorithm in its cipher block chaining (CBC) mode, in a single operation. It is important to notice that with our scheme, it is possible to decipher and decompress the image with the common AES and JPEG-LS algorithms. More clearly, decryption, decompression as well as message extraction processes are conducted independently without having to be modified or adapted. With such a capability, our scheme is DICOM compliant. The performance of this scheme in terms of embedding capacity and image quality distortion is theoretically analyzed and experimentally verified, demonstrating the possibility of access to various watermarking-based security services ranging from authenticity control to data traceability as well as the good quality of the watermarked-decompressed-decrypted images. Note also that the proposed scheme saves the computational complexity as it does not require decryption and decompression to verify the image reliability in the encrypted domain and the compressed one, respectively.

Chapter 4 is devoted to lossless or reversible image watermarking of encrypted images. We propose a novel data hiding approach which allows embedding, in a reversible manner, a message into an encrypted image; a message that can be accessed whether the image is encrypted or decrypted. To do so, the proposed solution relies on the insertion of a pre-watermark into the image before its encryption. This "pre-watermarking" step is conducted with the help of a robust reversible histogram shifting watermarking modulation we propose; modulation that has error-correction capabilities. Message insertion (resp. extraction) is then commonly conducted into (resp. from) the encrypted image. It is the impact of this message insertion process onto the "pre-watermark" that gives us access to the message into the clear domain, i.e. after the decryption process. Due to the robustness of the embedded pre-watermark, the errors introduced into it can be corrected in the clear domain, thereby allowing then to restore the original image. With our approach, the watermark process is independent of the knowledge of the encryption key. One just has to know the watermarking key for message embedding and extraction. Reciprocally, message embedding/extraction processes are completely independent from encryption/decryption. The feasibility of our approach is demonstrated considering the Trivium stream cipher and a robust histogram shifting modulation that we developed. Experiments conducted on some natural images and medical images confirm the efficiency of our approach to make available a message in both clear and encrypted domains and to correctly recover the original image in the clear domain.

Finally, in Chapter 5 we present a psychovisual study on the watermarking impact on CT (Computer Tomography) images. The first objective of this study is to confirm that for CT images, a loss of information related to watermarking is acceptable. The second objective is to determine, for a watermarking algorithm, the range of its parameters' values that can be used without any significant reduction of the image visual/diagnostic quality. To do so, we select in our study the least significant bit substitution modulation as a watermarking algorithm. Then, on a lung thoracic CT image database, we detail the implementation of two protocols for the subjective evaluation of image quality. The first protocol is designed to perform an in-house test with MEDECOM and IMT-Atlantique staffs, with as purpose to identify an embedding strength threshold above which the watermarked image appears visibly different from its original version. The second protocol was set up for a psychovisual study with radiologists to assess the diagnostic quality of watermarked images. All the results obtained from the first protocol are analyzed and compared with results obtained from objective quality metrics.

Chapter 1

Security of Medical Images

In recent years, advance in information and communication technologies have radically changed our daily lifestyle and have boosted all economic sectors including healthcare. They offer new possibilities to store, access and transfer data over networks in between practitioners, for a second opinion for example, between the physician and the patient. Medical data are rapidly growing, in both volume and complexity as the sources of data keep on proliferating. According to a report by EMC Digital Universe [9], healthcare data will be about 2,314 Exabytes (one Exabyte equates to about 1 billion gigabytes) per year by 2020. Medical imaging, very essential in many steps of patient cares (e.g. diagnosis, patient follow-up), largely contributes to this evolution with more and more accurate image modalities. Such an evolution also leads to a situation in which patients' health data are confronted to new security threats all along their life cycle in terms of confidentiality, integrity and privacy, but not only. For instance, in 2016, Los Angelos-based Hollywood Presbyterian Medical Center (LAHPC) was attacked by the Locky ransomware. In order to retrieve the access to its system, LAHPC has to pay nearly 160 K€. To respond such threats and due to the sensitive nature of medical data for patients' life, many regulations have been progressively established in order to impose on healthcare professionals to ensure the security of the data of their patients. At the same time, new security research topics have emerged due to the lack of solutions to reach these security objectives in such open environments (i.e. connected to internet). Let us take as example, cryptographic mechanisms and access control policies. Once these mechanisms bypassed or in other words, once the access to data is granted, data are no longer protected. Here comes the interest for a complementary security technique but still rather new in healthcare: "data watermarking".

This chapter hinges on three main parts. The first one presents the basic notions of what medical data are, and especially medical images, how they are distributed and the security risks they are submitted to. In the second part, we go through the security needs and actual security mechanisms used to protect medical images. We introduce watermarking in the third part and how it can advantageously complete the previous techniques to better protect data against leaks and malevolent manipulations.

1.1 Medical Imaging

1.1.1 Image modalities

In terms of diagnosis, common imaging types include radiology, ultrasound imaging (US), computerized tomography (CT), magnetic resonance imaging (MRI), and positron emission tomography (PET). Radiology is the oldest but one of the most frequently used imaging modalities. It uses X-ray attenuation properties and maps out the cumulated absorption process of an X-ray during its path along the tissues. To create a radiograph, a patient is positioned so that the part of the body being explored is located between an X-ray source and an X-ray detector. To improve X-ray image quality, a contrast agent can be used. CT imaging uses X-rays to produce a 3D image of the anatomy of the internal organs, bones, soft tissue and blood vessels within the body by acquiring several projections at different angles and reconstructing the 3D volume using tomography. X-ray properties make it very useful for medical diagnosis and treatment. It is widely used in clinical routines for the study of bones, to examine breasts (mammography) or lungs as well as in cardiol-

Modality	Organ	Image size	Depth (bits)	File size
Radiography	Thorax	2,060× 2,060	16	8 MB
CT	Abdomen	512× 512	16	250 MB
	Brain	512× 512	16	150 MB
	Heart	512× 512	16	1 GB
MRI	Abdomen	512× 512	16	15 MB
	Brain	512× 512	16	10 to 60 MB
	Heart	512× 512	16	50 MB
US	Standard	512× 512	8	12.5 MB/sec
PET	Heart	128× 5128	16	24 MB

Table 1.1: Typical properties of images acquired in clinical routine through different modalities and on different organs.

ogy for coronary angiography. Regarding MRI, it is used to observe the anatomy of body organs; their lesions as well as their performance that cannot be well seen using X-rays or CT scans. MRI is used to diagnose strokes, tumors, spinal cord injuries, aneurysms and brain function. US uses sound waves rather than ionizing radiation. High frequency sound waves are transmitted from the probe to the body via the conductive gel. These waves then bounce when they reach the different structures of the body and create a diagnostic image. Because of the minimal risk associated with it, US imaging is the first choice in pregnancy, but its applications are so vast: cardiac, spine, internal organs, and emergency diagnosis. PET is a type of nuclear medicine technique in which tracers are used for diagnosis disease to efficiently distinguish between benign and malignant tumors in single imaging.

Let us now emphasize some of their characteristics: for instance, all the medical imaging modalities offer image sequences either spatial (e.g. slices in CT, MRI, PET) or temporal (e.g. US, X-ray). Furthermore, the contrast between the organs or the characteristics of the organs is very variable: if the bones are clearly visible in CT images, this modality is not very discriminating for the soft tissues hence the need to inject a product of contrast (the best example concerns the angiography, the possibility of seeing the vascular network). MRI provides a better understanding of soft tissue differences. US images are considered difficult images because of the so-called speckle noise. Noise is intrinsically present in all images: physical noise of course due to the sensors but also to the diffusion and the diffusion effects (US, PET, etc.). Artifacts are common: they can come from the reconstruction itself, as in a scanner, when several heartbeat cycles are used to obtain a sequence of images in a time. The movement of the organs (heartbeat, breathing or displacement of the patient during the image acquisition) can also provide a filtering and therefore a smoothing of the borders of the organs.

In fact, unlike natural images, which are generally encoded on 8 bits, medical images have a larger range of gray values. The magnitude of each pixel or voxel is encoded on 16 bits while in radiology, images are often encoded on 12 bits. The difference between images issued from different modalities can also be noticed in images' size, as given in Table 1.1. The image depth influences the visualization and thus interpretation of radiologists, even if the eye is unable to differentiate as many tones. Indeed, a study done in [10] has compared the performance in terms of interpretation time and radiologists' diagnosis reliability when images are projected on 11-bit screens and 8-bit screens using various types of images. Authors concluded that 11-bit screens could improve observer performance.

To resume, these various imaging modalities are complementing each other by highlighting

different properties of normal and abnormal structures and tissues of the human body. This is the basis of multimodal imaging. There exist several ways to deal with it. A joint system can be designed by coupling in the same device two or more modalities: PET-CT is one example where CT brings the anatomical frame totally absent in PET. Another way consists of performing distinct image acquisitions, separated over time, and then registering the data through geometric transformations, rigid or elastic.

After being acquired, medical images are consequently reviewed for clinical analysis, diagnosis, and treatment as part of the patient's care plan. The collected information can be used to identify any anatomical and physiological abnormalities, to map the progress of treatment as well as for research purposes. Having digital access to the latest version of a patient's medical images, clinical reports and medical history can speed up and improve care and avoiding redundant testing. Digital access can also improve patient safety and save both the facility, its time and money. Herein, some technologies were developed to provide an efficient mechanism in dealing with the outsourced images, either for remote consultations inside the hospital, or outside, or combining both environments.

1.1.2 Medical image information system

1.1.2.1 Hospital Information System & Radiology Information System

In healthcare organization, different categories of information systems exist: systems for the general practitioners and the specialists (cardiologists, ophthalmologists, dentists, etc.); systems for hospital units such as radiology, functional exploration, laboratory; hospital information systems. Hospital Information System (HIS) is an integral component of any hospital. It provides the context within which collecting, processing, analysis and reporting of medical information take place. More clearly, it is designed to access patient information, reports from various services, and billing information. This electronic patient records give physicians an overview of the care provided to the patient. HIS can be composed of various software components with specific extensions as well as of a large variety of medical speciality sub-systems. However, HIS rarely give to medical images, nor do they have a user interface that can display images and navigate a large, complex, multi-image study. This is the role of the Radiology Information System (RIS) that is designed to place radiology orders, to receive interpretations, and to prepare bills for patients. This system is especially useful for tracking radiology imaging orders and billing information, and is often used in conjunction with PACS (Picture Archiving and Communication System) [11] that constitutes the core of radiography technical units.

1.1.2.2 Picture Archiving and Communication System (PACS)

PACS is a combination hardware and software systems that are used to acquire, store, distribute, and retrieve medical images. It facilitates the handling of digital radiology images so that they can be accessed and viewed by a variety of health professionals in different locations and settings, enhancing patient care and improving operational efficiencies for healthcare professionals.

As summarized in Figure 1.1, a PACS is constituted of four major components: image acquisition devices such as MRI and CT, communication networks, PACS archive and server for the storage and retrieval of patients' images and reports, and integrated display workstations for interpreting and reviewing images. In fact, this technology replaces the need for hard-copy films and management of physical archives, enabling radiologists in different physical locations to review the same data at the same time and allows health professionals to manage and follow the workflow of a patient exam. Furthermore, PACS offers an electronic platform for images interfacing with other medical information systems such HIS and RIS.

In addition to the mentioned medical imaging systems, there exist complementary technologies that attempt to improve the efficiency, accuracy, speed and access of medical imaging systems. One of the most remarkable technologies is Telemedicine.

1.1.2.3 Telemedicine

This technology [12, 13] corresponds to the remote delivery of health services, such as health assessments or consultations, over networks. It allows physicians to assess, diagnose and treat patients using common technologies, such as videoconferencing and smartphones, without the need for an in-person visit in order to provide care at less cost to the patient, especially in emergency

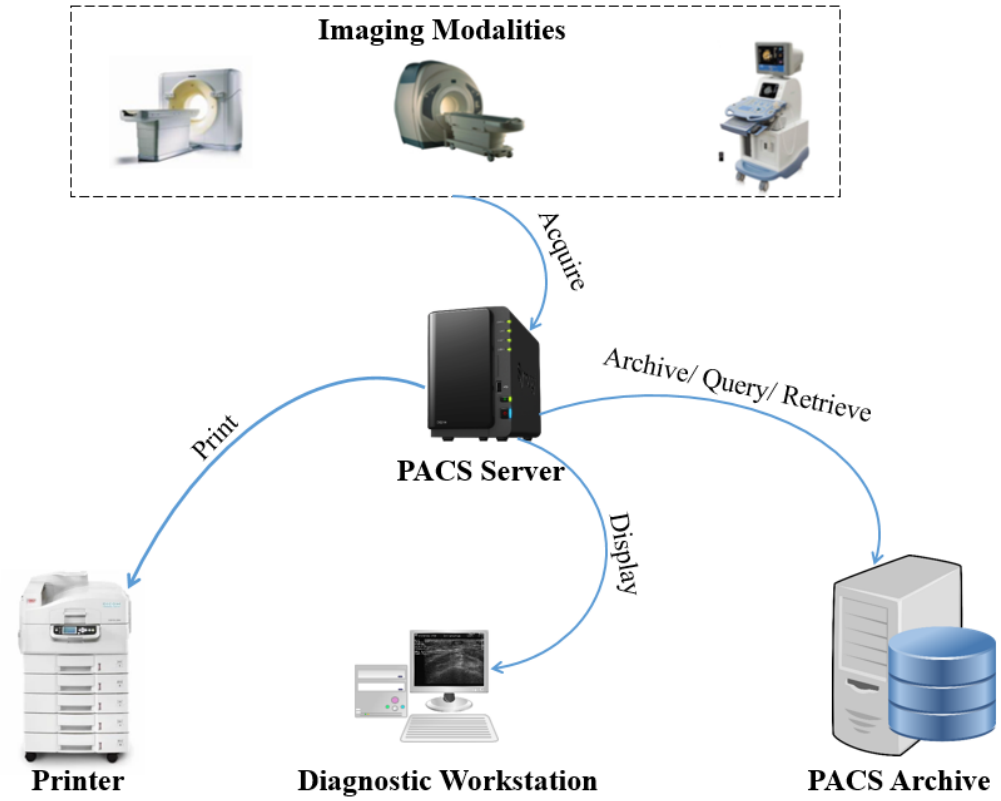


Figure 1.1: Picture Archiving and Communication System

cases. Telemedicine can be categorized into three classes; which are: i) remote patient monitoring, ii) interactive telemedicine and iii) store-and-forward. The first class allows for instance patients with chronic diseases to be tracked and monitored at home with mobile medical devices that collect data related to health (e.g. blood glucose level, blood pressure), patient behaviour (verification he/she follows treatment procedure).

Practitioners can remotely review acquired data instantly. This can also empower patients to better manage their health and participate in their health-cares. Interactive telemedicine, called also as "synchronous telemedicine", is a real-time video visit where patients and health providers are able to exchange information and communicate between each other. The patient may be at home or in a healthcare center. Regarding the last class also known as "asynchronous" telemedicine, it regroups techniques by which healthcare providers share patient medical data like reports, images with another physician, radiologist, or specialist at another location even across long distance to collaborate for remote diagnosis.

Healthcare innovation and virtualization continue to drive the expansion and strengthening of hospital service lines. However, telemedicine technologies are designed for specific healthcare provider or organization, with static requirements and specifications; which make them difficult to scale to changing needs or extended applications. Moreover, another obstacle to the use of telemedicine is the prohibitive up-front costs for healthcare providers and patients, including investment in expensive equipment and costly network infrastructure and specialized training. As revealed by [14], the cost of implementing a data-center system to store and manage medical data costed, in 2017, from $25 millions up to $10 billions for 26 hospitals. Therefore, alternative and innovative ways to provide more flexible and affordable telemedicine services are of high interest.

Cloud-based platforms for telemedicine [15] have addressed such a problem. It is essentially the use of remote servers hosted on Internet. More clearly, it allows health professionals to flexibly store and process massive data remotely, without a need to purchase and maintain their own infrastructure, unlike physical servers. Basically, Cloud-computing allowing multiple users geographically separated from one another to manage, store and share massive amount of data with high availability.

Diagnosis is also concerned with the development in the healthcare domain. As known, misdiag-

nosis costs unnecessary additional testing, results in delayed treatment plans and reduced survival or remission rates compared to what it might have happened if the disease had been correctly captured and identified earlier. That is why and due to the availability of large-scale labelled (i.e. annotated) image datasets, some technology innovators are trying to address these issues by experimenting with artificial intelligence (AI) and machine learning. Computer-aided diagnosis system can work on colossal amounts of data - much faster and more precisely than health professionals - to discover patterns and predictions to improve the diagnosis of diseases, inform treatment plans and improve public health.

1.1.2.4 Transmission standards

Transmission of medical data between different users and systems is complicated for two reasons: i) medical information systems use different machine platforms and ii) medical images are created from various imaging modalities from distinct manufacturers. Here comes the need for universal normal data formats and correspondence protocols that enable interfacing between two or more medical systems.

Health Level seven (HL7) is the standard of reference for managing non-imaging data such as test results, patient demographics, and billing information. Basically, HL7 [16] is a set of international standards for exchange, management and integration of clinical and administrative electronic data in healthcare centers. It allows thus the interoperability between various medical information systems (e.g. HIS-RIS interfaces).

The Digital Imaging and Communications in Medicine (DICOM) is the standard [5] for handling medical images. It was originally developed by the American College of Radiology and the National Electrical Manufacturers Association in 1985 to improve the interoperability between medical imaging systems. DICOM provides detailed specifications on how to format and exchange medical images and their associated information, both inside and outside the healthcare organization (e.g. telemedicine). DICOM interfaces are available for facilitating the connection between any combinations of the following categories of digital imaging devices:

- Image acquisition equipment (e.g. Ultrasound imaging, CT imaging, MRI, X-ray imaging);
- Image processing devices and workstations for displaying images;
- Image archive;
- Print devices.

As previously mentioned, PACS can resolve the problems of images storage and transmission but the ever-growing quantities of information needing to be managed also have to be taken into consideration. In fact, on a daily basis, large amount of medical images are acquired using various imaging modality devices. For instance, a typical CT heart exam can yield a data size up to one Gigabyte (see Table 1.1 in Section 1.1.1). One hospital can thus produce at least 27,000 Terabytes of medical imaging data per year [17]. Furthermore, according to a report by EMC Digital Universe [9], healthcare data will be about 2,314 Exabytes (one Exabyte equates to about 1 billion Gigabytes) per year by 2020. Medical image compression can therefore be extremely useful to optimize costs of storage, transmission and management of these data. Many of compression standards have been thus included in DICOM as compressed data formats for the exchange of medical images.

1.1.2.5 Image compression

Image compression reduces storage space, transmission time and bandwidth requirements. Basically, it takes advantage of the image information redundancy to minimize the number of bits needed to represent the image. We give in Figure 1.2 an overview of the main steps of an image compression chain. The first step consists in transforming the original image from its representation in the spatial domain (i.e. pixels) into a separate type of representation (e.g. prediction-errors, transformed coefficients) to reduce image signal redundancies. In essence, three types of redundancy in images can be identified [18]. Beyond the fact that the statistical distribution of pixels' values varies; there are pixels of more occurrences than others. There also exists interpixel redundancy due to the high correlation between nearby pixels within the image; i.e. they have very similar intensities as those of their neighboring pixels. This type of redundancy can be reduced when the

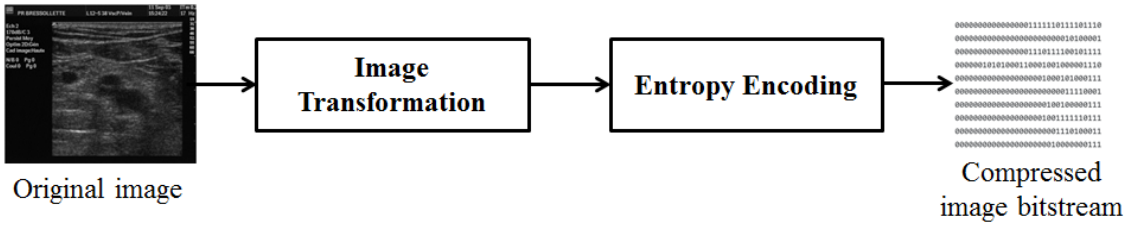


Figure 1.2: General scheme of image compression.

2-D representation of the image is transformed into a representation of differences between pixels. This is the case of compression standards based on prediction such as JPEG-LS [8]. But in general, the main idea of transforming the image is to find another representation where the signal energy is compacted onto some transformed and statistically independent coefficients. That is the case of JPEG and JPEG2000 that are based on the Discrete Cosine Transform or the Discrete Wavelet Transform, respectively. Notice that these compression standards also combine prediction to image transform so as to improve the compression rate (i.e. the ratio between the sizes of the compressed image and its uncompressed form). Furthermore, this image transformation process can introduce some information loss by taking into account the properties of the human visual system (HVS) properties [19]. More clearly, the human eye is more sensitive to lower frequencies (e.g. flat areas in the image) that to the higher one (e.g. textural regions). It can be hence considered as a low-pass filter. Therefore, such high frequency information or image details can be discarded without affecting or changing the visual quality of the image. JPEG and JPEG2000 achieve this goal by more or less quantizing DCT or wavelet coefficients, respectively. The harder the quantization is , the more information loss occurs and the more efficient the compression is.

The image transformation step is then followed by an entropy encoding step, the purpose of which is to find a more compact encoding representation of the transformed data with no information loss. One basic idea is to assign to a coefficient value a *"codeword"*, the size of which depends on the probability of the coefficient value in the image. By attributing shorter codes to the more probable coefficients' values, it is possible to reduce the global size of the image representation. Among entropy encoding compression techniques, one can find arithmetic coding [20], Huffman encoding [21], and Golomb-Rice encoding [22]. We will come back on these aspects in the case of JPEG-LS and JPEG with more details in Chapter 2.

Based on these various redundancy reduction types, compression techniques can be discriminated into two classes:

- *Lossless image compression* – it is also called reversible or noiseless compression [23]. It allows the representation of the image with the smallest possible number of bits without information loss. More clearly, the original image is exactly recovered after its decompression. This type of compression is based on reducing the coding and/or interpixel redundancies in the image. However, due to its modest compression rate, lossless compression techniques are of limited use (e.g. it can be used in few applications with stringent requirements such as healthcare domain).

- *Lossy image compression* – it allows some loss of information as long as the recovered image is perceived to be identical to the original one [24]. This loss of information comes from the removal of psychovisual redundancies by quantization process known as an irreversible process. This type of compression can achieve much higher compression rate compared to lossless methods but at the cost of a decrease in quality.

For medical images, the use of lossy compression techniques is of major concern. They offer a significant space gain but could change the diagnostic interpretation of the image and result in injury to a patient. It is complicated to define the amount of distortion accepted that could preserve the reliability of the diagnosis of the decompressed image. Therefore, the adoption of lossy compression was not permitted [25]. Nevertheless, several professional organizations have issued guidelines and standards supporting the safe use of lossy compression in clinical practice. For example, the American College of Radiology released a standard [26] that allows the compression of medical images using only the algorithms defined in the DICOM standard such as JPEG,

Anatomical region	Modalities									
	Radiology		CT		US		MRI		NI	
	JPG	J2K	JPG	J2K	JPG	J2K	JPG	J2K	JPG	J2K
Body	30:1	30:1	15:1	10:1	12:1	12:1	24:1	24:1	11:1	11:1
Chest	30:1	30:1	15:1	15:1	-	-	24:1	24:1	11:1	11:1
Musculoskeletal	30:1	20:1	15:1	15:1	12:1	12:1	24:1	24:1	11:1	11:1
Neuro-radiography	-	-	12:1	8:1	-	-	24:1	24:1	11:1	11:1
Breast	25:1	25:1	-	-	12:1	12:1	24:1	24:1	11:1	11:1

Table 1.2: Maximum lossy compression ratios recommended by the Canadian Associations of Radiologist Standard. NI corresponds to the Nuclear Medicine, JPG to JPEG and J2K to JPEG2000.

JPEG2K. Still, it does not highlight any statements on "the type or amount of compression that is appropriate to any particular modality, disease, or clinical application to achieve the diagnostically acceptable goal". Canadian Association of Radiologists has recently published the recommended compression ratios [27] for lossy techniques. Note that these recommendations were based on objective and subjective quality assessments and have involved many experts. If we refer to this study, the allowable compression ratio does not only vary depending on the used compression technique, but it also largely depends on the characteristics of the image; characteristics that are related to the image modality as well as the nature of the organ being explored. Table 1.2 provides the maximum compression ratios recommended for JPEG and JPEG2000 for the specific modalities and anatomical areas.

To sum up, medical images play an important role in diagnosis, patient follow-up and screening. This is why they are more and more shared in-between health professionals (e.g. telemedicine applications, cloud-based medical imaging applications) and also re-used and combined with machine learning techniques so as to develop tools for diagnosis aid support. But, the ever-growing quantities of medical imaging needing to be managed have to be taken into consideration. Therefore, medical images should be stored and distributed in compressed form in order to optimize their costs of storage, transmission and management. However, with the widespread applications of medical images, possibilities for distant access, processing and distribution of medical images over open environments have increased the chances of leaks, losses and alterations of the medical information. As a consequence and due to their sensitive nature, medical images security is of a major concern .

1.2 Protection of Medical Images

1.2.1 Medical image risks and threats

As stated previously, medical images are more and more manipulated in open environments, where they are accessed by different users from different places. This has plagued medical information with several breaches. According to Moody's Investors Service report, hospitals are a prime target for hackers and pirates. As reported by Protenus [1], in USA about 15 M records were breached in 2018, where hacking was the cause of 44% of the total number of breaches throughout the year. This is to say that 56% of the total breaches are caused by other types of breaches. Risks can be thus classified into two categories [28]; which are innocent and malicious disruptions.

1.2.1.1 Innocent disruptions

Based on the report done by CLUSIF (Club de la Sécurité de l'Information Français) in 2018 [29], this category can be discriminated into two classes

- *Accidents*, as for instance:
  - Total or partial destruction of hardware, software (earthquake, fire, explosion, ...);
  - Dysfunction of hardware, software or technology environment;
  - Negligence or failure/lack of technical staff responsible for handling or maintenance of the system.

  These risks are always present and hospitals or medical centers can only attempt to limit their consequences. The definition of safety standard and guidelines can contribute to minimize such issues. CLUSIF reported that in 2018, among 151 hospitals, 17% of hospitals lost essential services (e.g. water and power cuts) and that 19% of them experienced errors in conception of software.

- *Errors*, the most of them being:
  - Input errors;
  - Errors in the transmission of data by the information system;
  - Error handling functions of the information system;
  - Error resulting from the incorrect use of the information system

  Manipulation errors can take multiple forms. The most classical example is the assignment of patient records to another patient, leading thus to errors in diagnosis, therapy and/or surgery. Loss or alteration of information during the transmission can be identified in this type of disruption. For example, a simple degradation of images during their transmission can hide pathology, which leads to error in diagnosis. Again, these risks are always present, and can be mitigated with the help of good practice guidelines as well as user training.

1.2.1.2 Malicious disruptions

Nowadays, digital security risks are of high level in the healthcare industry, which comes from or is exacerbated by human factors. Malicious manipulations can arise from hacking, malware, physical sabotage and materiel theft to insider threats. In fact, hacking can be defined as the unauthorized access to an information system to gain information or cause disruption. For example, in Singapore in 2018 hackers who were targeting the prime Minister have stolen medical files of 1.5 million patients (about 25% of the population). In its side, malware (i.e., malicious software) refers to programs designed to infiltrate information systems without users' consent. These programs include threats such as viruses and ransomwares. For example, on May 2019, Indiana-based Talley Medical Surgical Eyecare Associates notified the US department of Health and Human Services of a potential malicious attack of 106K patient records, caused by a ransomware attack on its system. In France, CLUSIF reported that, among 151 French public hospitals (constituted of over 100 beds), the information system of 44% has been infected by viruses. Material sabotage and theft of are also present in hospitals. As reported by the Irving, TX-based healthcare supply chain network VHA Inc. [30], found that about $52 million worth of items are stolen each year from hospitals. Based on the study done by CLUSIF, 7% among 151 french hospitals experienced physical sabotage. Insiders' threats are also important issues created by inadvertent (naïve and careless insiders) or malicious employee actions. A study done by MediaPro [31] found that 37 percent of healthcare employees pose an outright risk to their organizations, meaning their actions could cause a breach of privacy or a security incident. More clearly, the inadvertent employees do not intend to do harm, but they still unintentionally respond to phishing emails, lose their laptops or send unencrypted emails, for example. In addition, we find employees that engage in malicious use of confidential data. The study [32] reported that 18% of healthcare employees are willing to sell confidential data to unauthorized parties. To sup up, the consequences of these risks are not negligible, as they concern an individual and his health. According to Dr. Sung Choi, a researcher at Vanderbilt university, over 2,100 patient deaths per year are caused by data breaches [33].

As a conclusion, these risks create an urgent need for the protection of medical information. That is why, many countries attribute legal and ethical weight to this question; acknowledging patient rights and thus obliging healthcare professionals and medical centers to ensure the security of medical data, including medical images.

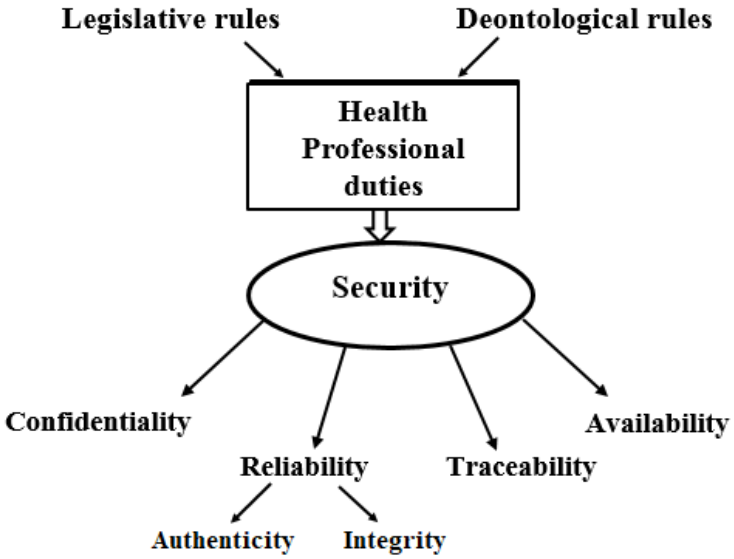


Figure 1.3: Information security.

1.2.2 Security requirements in healthcare: ethics and legislation

Medical data are subject to deontological and legislative rules that require their security in order to build and maintain trust in the patient-health professional relation-ship. Indeed, trust is a critical factor that influences the quality of care, patient health, and the medical system as a whole. In other words, without trust the patient will not give to the practitioner a clear view of his/her situation. That is why ethics are considered a crucial and important branch in medicine guiding good medical practice. Among the different deontological ethic codes, one can find the "Hippocratic Oath" [34], and the French medical ethic code (article R. 4127-73) [35], which insist on the fact that the duty of keeping a medical secret is for the whole life; it does not stop with the end of the professional activity. However, in France, the concept of medical secret does not exist. It derives from the professional secret and it refers to the principle of not disclosing confidential information about the patient. Divulging a medical secret or breaking the professional secret lead to criminal convictions as determined by the different international and national legal norms and laws. In Europe, the General Data Protection Regulation(GDPR) [2] has been designed to harmonize data privacy laws across European countries to maintain reasonable and appropriate administrative, technical and physical safeguards to ensure data protection, including medical data. Healthcare organizations that fail to ensure security of patients' data, risk fines of up to 20M €. Recently, Barreiro Montijo hospital has been fined € 400K for violating the GDPR [36]. It allowed indiscriminate access to an excessive number of users while being incapable to ensure continued confidentiality, integrity, and availability of medical treatment systems and services. Notice also, as part of GDPR, the Information Commissioner's Office (ICO) must be notified of any security breaches that are *"likely to result in a risk to people's rights and freedoms"* within 72 hours of breach occurring. In US, health professionals must take care of the Health Insurance Portability and Accountability Act (HIPPA) [37]. HIPPA is a federal law invoked to protect patients' data and well-being by imposing regulatory standards on any organization with a hand in healthcare industry. Beyond these legislative and deontological rules, one must also consider regulations of general scope like for database and software protection, illegal intrusion in information systems [38]. Furthermore, there exist national and international recommendations which provide implementation guidance. That is the case of IHE (Integrating the Healthcare Enterprise) that is an international standardized medical information integration organization established to address integration problems among medical information systems. Furthermore, IHE adopts existing standards such as HL7 and DICOM format to fulfill specific clinical needs regarding medical information sharing as well it provides security recommendations. Note that HL7 and DICOM standards on their side include security constraints (see part 15 of DICOM standard for instance).

In Figure 1.3, we sum up the security needs for patient data any health professional has to consider when considering deontological and legislative frameworks. These ones correspond to

data:

We define now these security engagements before explaining how they can be ensured in Section 1.2.2:

- *Confidentiality* – it refers to the prevention of information disclosure to non authorized individuals or systems. Only authorized users, in the normally scheduled situations, have access to data.

- *Reliability* – Reliable data can be used by a healthcare professional in a total trust. It is based on the outcomes of data:

  - *Authenticity* – It is the proof that a piece of information belongs to the right patient and is delivered from the correct source.

  - *Integrity* – The purpose of integrity control is to find out whether any modification has been done upon the medical image. In other words, it assures that the image is accurate and has not been altered.

- *Traceability* – The concept of reliability can be extended to traceability when it becomes possible to trace data along their life cycle. Tracing a piece of information along its distribution is an important issue. It is defined as the capability of identifying all the users or systems that have accessed, transferred, modified or deleted a piece of information during its transmission from its source to its final user or in a given period of time.

- *Availability* – it corresponds to the ability of a medical information system (including equipments, software and communication channels) to be used under the normal conditions of access. Its absence may lead to a partial or total destruction of the information or a denial of access. This may be caused by different attacks such as accidents, errors or malicious disruptions.

Once security needs identified, the next question is how to achieve these goals. This is a part of security policy that is more and more deployed by hospitals due to legislative constraints, of course, but also for a better functioning of their information system and for the image of the healthcare organization as a whole.

1.2.3 Security mechanims

Beyond these legislative and deontological rules, there also exist , such as rules stated by ISO/27799 standard [3] which specifies a set of detailed control for managing health information security and provides best practice guidelines on how to protect images in terms of confidentiality, integrity, availability, authenticity and traceability.

As previously mentioned, as medical images are stored and accessed in different systems, they are exposed to different security breaches; i.e. hacking and other unlawful forms of processing (see Section 1.2.1). Healthcare organizations have to implement appropriate technical measures to ensure a level of security. Nowadays, there exist security standards that specify a set of detailed rules and requirements to be satisfied by a healthcare information system. These rules precise how security services must be deployed and how medical data can and cannot be accessed, and by whom, etc. Among these standards, one can find the ISO 2700X, the latest being the ISO 27005, which are based on the principle of continual improvement of data protection, known as the model PDCA (Plan, Do, Check, Act). In France, recommendations of ANSSI (Agence Nationale de la Sécurité des Systèmes d'Information) and ASIP Santé (Agence des Systèmes d'Information Partagés de Santé) for the healthcare domain have to be considered too.

As risks in healthcare are multiplying, both in number and costs, risk analysis has to be at the forefront. It allows the identification of the requirements and objectives of security that an information system must achieve. This analysis can be done by means for example of the standard EBIOS (Expression of Needs and Identification of Security Objectives in english) [39], which provide a list of risks discriminated into different classes based on the required security applications (i.e. confidentiality, authenticity, integrity and traceability); FMEA (Failure Mode and Effect Analysis) [40] or MEHARI (MEthode Harmonisée d'Analyse de RIsques) [41] that is a complete model of risk analysis. A risk analysis process is based on the following steps illustrated in Figure 1.4:

Figure 1.4: Overview of risk analysis process.

- *Context establishment* – that identifies activities and actors, data workflow and the components of the system within the scope of the risk analysis.

- *Hazard identification/Prioritization* – consists in the identification of something with the potential to cause harm. Hazards can be identified from different sources such as EBIOS, by discussing with users (department Chiefs, and employees), analyzing patient complaints and satisfaction survey results, incident reporting system, and so on.

- *Risk evaluation* – it is about understanding and evaluating the gravity of a risk by underlying its causes, analyzing its probability and impact associated-with and identifying its level or in other worlds computing the "risk score" [42]. The risk risk may be of different impacts on data and system resources in terms of confidentiality, reliability, traceability and availability. The purpose of computing the risk score is to prioritize the risks from "most to least" critical so as to determine the appropriate level of training and control measures necessary for effective mitigation.

- *Risk management* (or risk mitigation) – In order to minimize/counter the identified risks, existing protection tools and devices are deployed.

The used security tools deal with four aspects; which are protecting patients' confidentiality, controlling data reliability, traceability as well as ensuring data availability. Ignoring any of these aspects may cause a number of problems. However, a "unique" security mechanism cannot ensure all these security objectives at the same time.

Protection means can be classified into physical and logical mechanisms. The former mechanisms concern the materials and essentially aim at counteracting unauthorized physical access; theft/robbery and natural risks such as fire, flooding. In such cases, an early warning fire and smoke detection systems can be used in different areas of the hospitals. To improve the physical/environmental safeguards, the use of video surveillance, security staff, intrusion detection systems, innovative architectural (e.g. use of access badge) are also suggested to avoid external agents and unauthorized access to data centers.

In practice, the efficiency of physical protection is arguable, as health organizations are open structures in which patients and their accompanists are mixed with health professionals with an easy access to equipments. This is why, logical protection mechanisms are important to be deployed. Hence, to have more comprehensive view of the existing logical security tools for images, we propose to present them depending on three main security issues.

1.2.3.1 Confidentiality

Four different categories of risks related to disclosure and misappropriation should be considered: 1) illicit access to medical images, 2) mistakes − intentional or not − in data manipulation (file copy, diffusion of information to a non-authorized third, etc.), 3) interception in transmission, and 4) viruses. To mitigate such threats, one can find five main mechanisms:

a) *User authentication* - It is the fundamental method of security for protecting medical data being stored in a medical information system. Usually, two steps are required: i) identification; where a user provides a proof of his/her identity (e.g. login name, smart card) to respond to the question "who is he?", ii) authentication, where the user proves his identity using a password. This solution can be associated with a token, which ensures a unique connection per user. Once the user connected, the token is assigned to the computer in which he/she is logged on.

b) *Access control* – Once the user has provided a correct login name and password and accessed the system, it is mandatory to control the actions he/she can perform by defining user access rights. However, the design of access control systems is very complex. There exist many implementations of an access control model; according to the appropriate access control policies that define the rules that only authorize certain people to access certain information to organize access privileges. Among the different access control models, one can find:

- The Role-Based Access Control (RBAC) [43] or the Organization-Based Access Control (OrBAC) [44] that associates rights to groups of users according to their roles within the organization (i.e. it is based on a hierarchy system where the person with the greater responsibility has access to more resources throughout the healthcare system);
- The Mandatory Access Control (MAC) model [45] that gives the delegation to a third-party (or a central) authority to determine what information can be accessed and by whom;
- The Discretionary Access Control (DAC) [46] in which the owner of a file is responsible for access control on this file and is granted the ability to restrict access to it based on the users' identity or a membership in certain group.

Nevertheless, a model can also be hybrid and include more than one model in order to tackle the more heterogeneous needs of an organization.

Furthermore, to avoid indiscreet or malicious access, the implementation of automatic log-out procedure for users in case of prolonged use is thus desirable. However, the use of such a model cannot block all the possible attacks by itself, as a user can by-pass its imposed restrictions. Hence, it is necessary to implement other security tools to ensure confidentiality, for instance, firewalls, anti-viruses and data encryption.

c) *Firewalls* - they are designed to strengthen the security of an organization's networks and the information that resides on the network. Basically, it is defined as a system that allows filtering the incoming and outgoing network traffic based on a set of access control rules. There exist different forms of firewalls that can be implemented both internally (e.g. an equipment, local computer, intranet) and externally (e.g. the internet) to protect the organization from any type of threats to the information the network could face [47]. Furthermore, it can accomplish other various functions such as address translation (i.e. it allows to keep an internal address space unknown and inaccessible from the outside) or act as an application proxy which interprets each request to internet of an application (commands, requests and responses) so as to verify whether these information exchanges follow the security policy rules.

d) *Anti-viruses* - Viruses represent a major threat to the security of a system and of its data [48]. They can be introduced into the information system in many ways: through a connection to an open network as well as by simply uploading data from a CD-ROM or an USB key. Herein, an anti-virus policy should be defined in terms of prevention, detection and isolation of viruses and system restoration (backup management). Prevention of viruses is based not only on testing all storage units but also on Internet connections, databases and programs that can be imported. Detection consists in simply controlling the virus throughout the

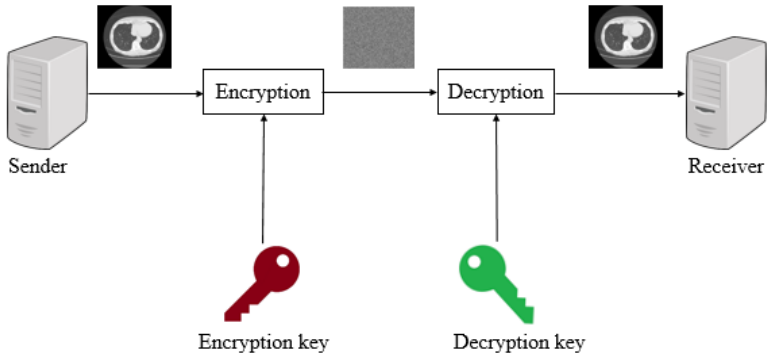


Figure 1.5: General scheme of image encryption.

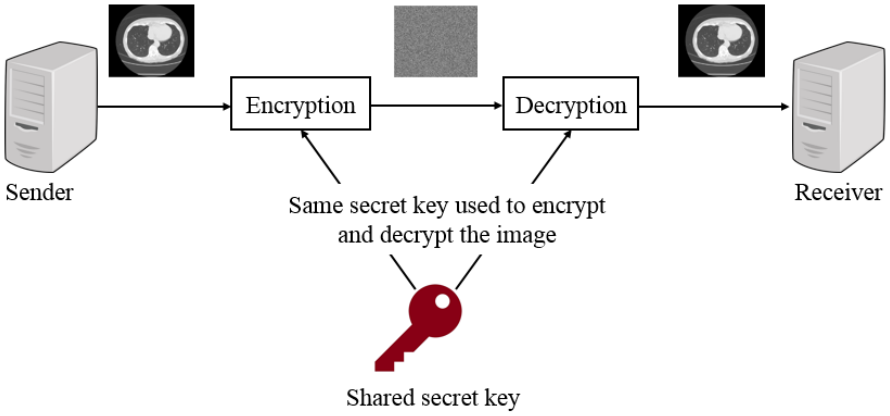


Figure 1.6: Overview image of symmetric encryption.

information system. According to the isolation principle, suspicious memory units should be disconnected. System restoration goes from virus removal performed by an anti-virus to the reformatting of the storage units and reinstallation of the information system.

e) *Image encryption* - it aims at maintaining data confidentiality. As depicted in Figure 1.5, it consists in transforming a clear plaintext image into a non-interpretable cipher-text image by means of an encryption algorithm parameterized by an encryption key. Encryption algorithms can be grouped according to various characteristics: i) symmetric algorithms (with a secret key) and ii) asymmetric algorithms (with private and public keys). Secret key systems allow encryption and decryption with the same key as shown in Figure 1.6. As a consequence, the sender and the receiver should hold the secret key to be able to encrypt and decrypt image data. Symmetric encryption algorithms can be also divided into two encryption types: block-cipher systems and stream-cipher systems. The former one encrypts a fixed size of n-bits of data – known as a block – at one time. The sizes of each block could be 64 bits, 128 bits, and 256 bits. 3-DES (triple Data Encryption Standard) [49], AES (Advanced Encryption Standard) [50], and Blowfish are some of the commonly used encryption algorithms that fall under this group. For their parts, stream-cipher algorithms encrypt 1 bit or byte of plain-text image, at one time. Here, one can find Trivium [51], RC4 [52], RC6, etc.

Regarding the asymmetric encryption systems, as shown in Figure 1.7, they are based on a pair of keys. One of them is known by the world at large (the public key) and is used to encrypt data while the other is confidential (the private key) and exploited to decrypt the image. By doing so, only the owner of the private key can decrypt the image, ensuring thus confidentiality. The most widespread asymmetric algorithm is RSA [49] and is recommended by the DICOM standard.

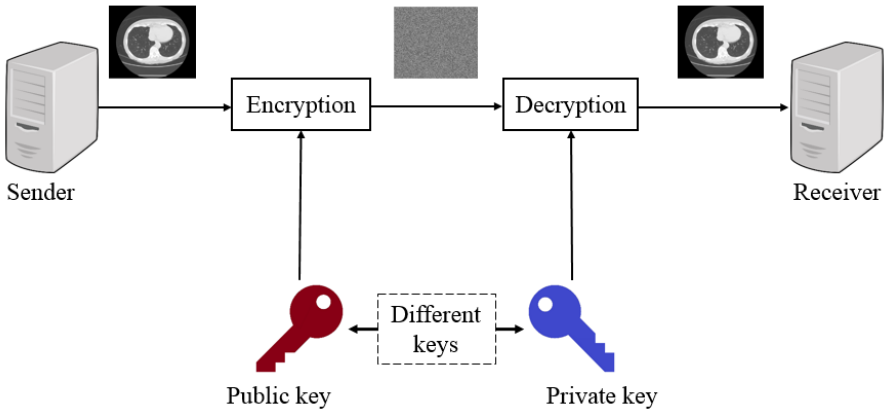


Figure 1.7: General principles of asymmetric encryption.

1.2.3.2 Reliability Control

Reliability control corresponds to the second most important property after confidentiality because, to attack it, information must be accessed beforehand. It is based on the outcomes of the data integrity and authenticity.

In DICOM (Digital Imaging and COmmunications in Medicine), there exists a digital signature (DS) profile. Thiss one combines RSA with the RIPEMD-160, MD5, or SHA-1 hashing functions. One of these functions is first used to generate a Message Authentication Code (MAC), which is computed over the whole image or over some specific local regions of the image. This MAC is a digest of the data, encoded on a few number of bits. For instance, SHA-256 requires that the maximum size of the input data be of $2^{64} - 1$ bits and produces a hash of 256 bits. the SHA-256 "strength" is of 128 bits against collision attacks. This digest is then encrypted using RSA parameterized with the private key of the emitter. By doing so, the resulting signature can only be decrypted using the public key of the emitter, ensuring thus the non-repudiation. At the verification stage, any slight difference between the recomputed DS/MAC and the decrypted one will be proof of the image integrity violation. Note that, in DICOM, such a digital signature is stored in the DICOM image file. Once the signature deleted, the image is no more protected in terms of integrity.

Regarding authenticity control, the DICOM standard makes use of Unique IDentifiers (UIDs) to uniquely identify DICOM images. These UIDs are present in the DICOM image file header and they refer to the acquisition machine, health institution, patient, date and time of examination, etc. Notice that patient identifier is sometimes defined in a legislative way. For instance, in France, a national health identifier is provided to each patient since 2007 (Article L1111-8 of the public health code [53]). Such identifiers must be used with care as they may induce privacy issues.

1.2.3.3 Availability

A violation of data availability is simply equivalent to making its access impossible. Hardware or material problems can be minimized by means of maintenance contracts, prevention of natural damage, staff training, etc. Logical risks can result from a violation of data integrity or from a logical dysfunction at the access control level, etc. Herein, to ensure availability, medical information systems often have redundant components, known as fault-tolerance systems. So if one component fails or is experiencing problems the system (e.g. in case of malware detection), will switch to a backup component. The main objective is thus to limit the impact of the risks on patients' health.

1.2.3.4 Traceability

Classic traceability techniques are based on log-files that record the history of all interactions in the medical information system. Later solutions are based on audit trails the purpose of which is to track all system activities, by generating their date and time stamps; detailed listings of what was viewed, for how long, and by whom; and logs of all modifications to medical records. System administrators can also specify which reports were printed, the number of screenshots taken, the

exact location, and even the computer used to submit a request. Alerts are often configured to report suspicious or unusual activity, such as reviewing information about a patient who is not treating or attempting to access information that one is not authorized to view. However, audit trails do not prevent unintentional access or disclosure, but it can be used to deter the would-be violators.

In last year, Blockchain technology [54] has gained considerable attention in the healthcare domain. It focuses on ensuring patient health management at a superior level at all times. Basically, Blockchain provides a safe and secure platform enabling not only tracking each step of a user's interaction with this platform, but also improving traceability of falsified data. However, Blockchain runs complex algorithms which in turn require large amount of computing power. Hence, due to its complexity, its transactions can take a while to process.

As exposed, the above security mechanisms are designed to deal with one or more security objectives or specific risks. However, they do not offer a "never-failing" protection as accessed and consulted medical data can sometimes escape these measures of control. In fact, once the access to data is granted or once data are cut off from their ancillary data (i.e. file header) that contain proofs of integrity and authenticity (e.g. digital signature, proof of image origin), they are no more protected and can be altered or illegally redistributed straightaway. Several questions arise then about the security of the image once decrypted. Here comes the interest for an "*a posteriori*" protection that allows the user to access the image content while keeping it protected – a kind of protection the watermarking techniques are best suited.

1.3 Watermarking as a complementary security mechanism

By definition, watermarking technology offers "*a posteriori*" protection: it allows accessing the image information while keeping it protected by a watermark inserted or dissimulated into it. In the following, we come back on watermarking fundamentals, and on how it is applied on medical images.

1.3.1 Fundamentals of image watermarking

1.3.1.1 Definition

Digital watermarking appeared during the 90's. It consists in the embedding of a message or a watermark into the "host" image by imperceptibly modifying its gray values, without disturbing its normal use or interpretation. It is a concept closely related to steganography, in that they both hide a message inside a host document but differ in their purpose. With steganography, the host image has no importance. It is merely used as a cover to hide the message for secret communications while watermarking aims at protecting the "host" image by means of the embedded message. Indeed, watermarking was originally designed for copyright protection of multimedia contents [55] through the embedding of the owner identifier. Since then, watermarking has been extended for different security objectives in various application domains. As this report is concerned with the healthcare domain, several cases making use of watermarking can be identified [56]:

- Authenticity of images, by the embedding of data confirming the source of image and to which patient it belongs. For instance, in [6], it is proposed to embed a unique authenticity code that refers to the image origin (e.g. patient, date of image acquisition, user).
- Integrity control of images. The basic idea is to insert a digital signature of the image [57] or a summary of it [58] that can be used to detect whether an image has been altered or falsified, identify which parts of the image can still be used for diagnosis purposes and, at least, identify the nature of the alteration (is it a transmission error or a malicious attack?).
- Addition of meta-data such as elements related to an authorization policy and to the patient consent [59].

As defined, watermarking can be defined as a complementary security mechanism to those described above (in Section 1.2.2), as the watermarked image can be accessed and interpreted while maintaining its protection by means of a message.

It should be known that, beyond security services, watermarking has also been proposed in order to enrich the image content or to provide new image functionalities, by inserting some elements about semantic description of the image [60].

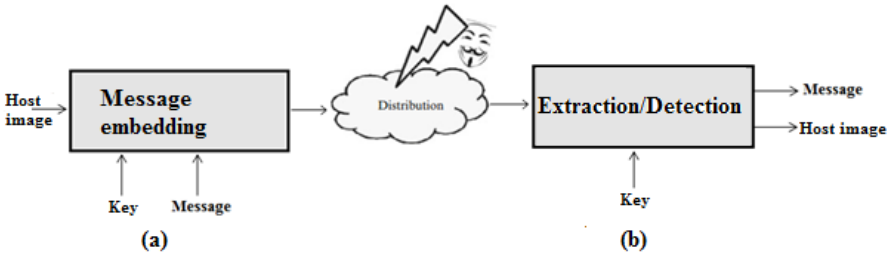


Figure 1.8: Overview of a watermarking technique.

1.3.1.2 How does watermarking actually works?

As shown in Figure 1.8, the general scheme of watermarking is composed of two main stages: a) embedding (or protection), and b) extraction/detection (or verification). The embedding process, as given in Figure 1.8 (a) is performed by the modification or the alteration of the image under the principle of controlled distortion; i.e. by imperceptibly modifying the image pixels' gray value or the transformed image coefficients (e.g. DCT (Discrete Cosine Transform, DWT (Discrete Wavelet Transform)). This modification is performed based on a secret watermarking key K_s. This key is usually used to secretly select regions in the image where the message could be embedded and/or to construct the message itself.

During the verification stage (see 1.8 (b)), with the help of the secret watermarking key K_s, the message can be extracted or simply detected depending on the used watermarking method. Here, two models of detection/extraction can be distinguished according to the need to have access or not to the original image; which are known as non-blind and blind watermarking schemes, respectively.

1.3.1.3 Watermarking properties

Each watermarking technique is described based on various properties:

- *Robustness* - a watermarking method is said "robust" [61] when the watermark can resist some image modifications, being innocent or malicious. Malicious disruptions want to suppress or modify the embedded watermark for "illegal" use of the image (e.g. distribution to unauthorized users, illegal copy). "Innocent" image modifications are those considered as authorized in the application framework. They usually correspond to image processing like lossy compression (e.g. JPEG, JPEG2000), filtering, geometric transformation, etc. Conversely, one can find "fragile" watermarking methods where once the image is altered or falsified, the watermark will be subsequently modified. Note that fragile watermarking techniques allow integrity control of images; i.e. based on the watermark, it will be possible to verify whether the image has been altered. There also exist semi-fragile schemes in which the watermark has been designed to survive certain image processing but not others that are considered as illegal in the application framework.

- *Capacity* – it expresses the embedding rate that is to say the number of bits of message embedded per pixel (*bpp*) of an image. It gives an indication of the message length that can be embedded in an image. The larger is the message length, the more watermarking-based security services can be provided.

- *Imperceptibility* – watermarking is a process that alters the image to add a message into it, therefore it can degrade the perceptual quality of the image. Various objective assessment measures of image quality can be used to assess it such as: PNSR [62] (Peak-Signal-to-Noise Ratio) and MSE (Mean Squared Error) which are appropriate in the case of a uniform image distortion; SSIM (Structural SIMilarity) [62] and UQI (Universal Quality Index) [63] that take into account some visual human properties. Note that watermark imperceptibility is very important for medical images. No obvious difference between the watermarked and the original image must be noticed so watermarking methods must keep the degradation of the image quality to a minimum. Different psychovisual masking strategies have been proposed so as to better preserve image quality. Most of them have been designed for natural public images. They take into account the Human Visual System (HVS) in the conception of a watermarking algorithm. Basically, they first select the suitable sites for the imperceptible

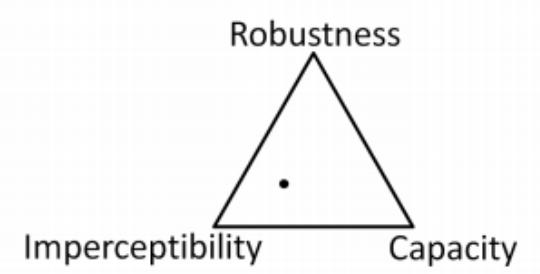


Figure 1.9: Graphical representation of the antagonism between the three canonical watermarking properties. A high performance in terms of two of them typically implies a very low performance in terms of the third one, as represented by the black dot in the figure.

embedding of the watermark, then they attribute a specific watermarking strength to apply to the selected sites before finally inserting the watermark. Because natural images differ from medical images, specific watermarking methods have been proposed for medical images (see Section 1.3.2).

- *Reversibility* – some watermarking methods allow the exact reconstruction of the original image from the watermarked one after having extracted and removed the message. These methods are called reversible or "lossless" watermarking techniques [64]. These methods are most of the time fragile and devoted to integrity control. In general, watermarking schemes are lossy or irreversible [6, 8].

- *Complexity* – it is an indication of the calculation time needed for embedding and extraction. The algorithm complexity plays an important role in case of images of large volume.

- *The need of the original image for message detection/extraction* - if during the verification stage, the original image is not required to extract the watermark, the watermarking method is called "blind". Otherwise, it is called as "non-blind". Note that integrity control applications are not possible with non-blind watermarking methods; the question of knowing whether the original image has been modified has no sense in such a case.

- *Security* - the basic idea behind this property is that the access to the embedded watermark is impossible without the knowledge of the secret key as for cryptography.

A watermarking method cannot provide all the above properties at the same time. Each one establishes a balance between these properties. Let us consider the trade-off between the basic watermarking characteristics, which are capacity, robustness and imperceptibility. As shown in Figure 1.9, an increase in capacity comes with imperceptibility and robustness costs. A stronger watermark can better resist against alterations (innocent or malevolent modifications), but its presence will be more obvious in the image and the capacity will be reduced. Robustness is essential in authenticity and traceability control, where the considered watermarking modulation should ensure that embedded identifiers remain detectable after innocent or malicious manipulations of image data. In this case, as the embedded watermark is visible in the image due to its robustness, some HVS-based models are applied on the image so as to select positions to be watermarked and to identify the watermark amplitude in each position in order to increase the imperceptibility of the message. On the contrary, if we consider for instance integrity control, the watermark is fragile and the embedding capacity (i.e. the length of the watermark that can be hidden into the image) should be high in order to be able to detect and locate the tamper location in the image. Thus, from a general point of view, to an application framework, corresponds a set of watermarking properties.

1.3.1.4 Watermarking techniques

Watermarking techniques can be classified in many ways. Nevertheless, two types of algorithm are generally identified. The first group involves additive methods and the second one covers substitutive methods.

Additive methods

Their first step consists in generating , from a given binary message $M = \{m_i\}_{i=1,...,N}$, a watermark that is a pseudo-random uniformly distributed signal $W = \{w_i\}_{i=1,...,N}$. The generated watermark W is then added to the image or to a transformation of it $I = \{I_i\}_{i=1,...,N}$ (e.g. DCT, DWT), leading to the watermarked image $I_w = \{I_{wi}\}_{i=1,...,N}$ as follows:

$$I_{wi} = I_i + \beta\left(I_i\right) W \tag{1.1}$$

where β is a strength watermark parameter that allows controlling the robustness/imperceptibility compromise. This parameter depends on image characteristics. More clearly, its value is usually computed based on psychovisual mask that selects the adequate parts of the image where to embed the watermark and the amplitude of this later to optimize the trade-off between imperceptibility and robustness.

With this kind of methods, the presence of the watermark is checked by correlation techniques, which implies the orthogonal nature of the pseudo-random sequences W. More clearly, let us consider the original host image I, its watermarked version I_w with its watermark W. Let us also denote I'_w the possibly attacked watermarked image, and W' the extracted watermark. Assuming that the watermark is orthogonal to the image, the detection of W in I'_w stands on the correlation coefficient γ given by:

$$\gamma =< I'_w, W >=< I + W', W >=< I, W > + < W', W > \approx < W', W > \tag{1.2}$$

Hence, the computation of γ comes in the computation of the correlation between the embedded and the extracted watermark. the decision of whether W is present or not in I'_w depends on a threshold value [65, 66] such that:

$$\begin{array}{l} \text{if } \gamma \geq T, \text{ then the watermark is present} \\ \text{if } \gamma < T, \text{ then the watermark is absent} \end{array} \tag{1.3}$$

By working on a block or a set of pixels or of coefficients, it is possible to embed a binary sequence. In this case, each set is watermarked with a message that encodes a specific symbol (e.g. '0' or '1').

Substitutive methods

Such a modulation replaces or substitutes some features of the image (e.g. pixels, coefficients) by some others issued from dictionaries that encode the desired symbols to embed. Message extraction takes place with a simple re-reading or interpretation of the image features that have been substituted. Among various substitutive watermarking modulations, one can find the Least-Significant-Bit (LSB) substitution and the Quantization Index Modulation (QIM).

- LSB substitution modulation [67] – it is the simplest substitutive method. Basically, it replaces the least significant bit of each pixel or transformed coefficient x_i of the image to be watermarked by the message bit $b_i \in \{0, 1\}$. Given for instance a pixel x_i encoded on 8 bits as $'10001010'$ and a message bit b_i, then the resulting watermarked pixel $x_{iw} =' 1000101b'_i$. The extraction of the message relies on the interpretation of the parity of the watermarked image features. Note that this method provides high embedding capacity with low image quality distortion, but it is "fragile" to any image modification.

- QIM [68] - it is an extension of the previous modulation and relies on quantifying some image components independently (e.g., pixels, coefficients) or at once (e.g., group of pixels or transformed coefficients - vector quantization) according to a set of quantizers based on codebooks in order to embed the watermark. More clearly, QIM associates, to each message m_i issued from a finite set of q possible messages $M = \{m_i\}_{i=0,q}$, a codebook C_{m_i} such that:

  $$C_{m_i} \cap C_{m_j} = \emptyset, i \neq j \tag{1.4}$$

  To embed the message m_i, an image component x has to be substituted by its nearest element x_w in the codebook C_{m_i}. The insertion function is given by:

  $$x_w = Q_{m_i}\left(x, C_{m_i}\right) \tag{1.5}$$

  where $Q_{m_i}(.)$ determines the nearest elements x^w of x within C_{m_i}. The distance between x_w and x corresponds to the watermarking distortion.

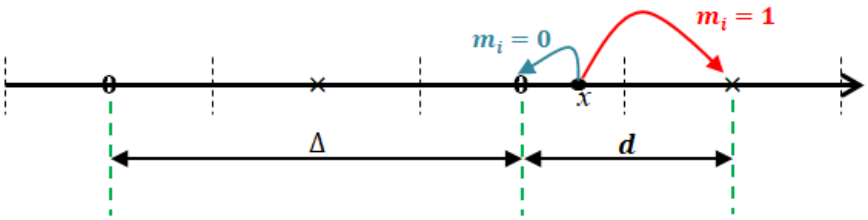


Figure 1.10: Example of two codebooks' cells in the mono-dimensional space (i.e. x is a scalar value) considering an uniform quantization of quantization step Δ. Symbols o and $\times$ denote cells' centers that encode 0 and 1 respectively. $d = \Delta/2$ establishes the measure of robustness to signal perturbations.

Let us consider the image components such as a vector of pixels $x \in \mathbb{N}^N$, while dividing the $\mathbb{N}^N$-dimensional space into non-overlapping cells of equal size. To satisfy (1.4), each cell is associated to a codebook C_{m_i}, $i = 0, .., q$. The insertion process is conducted as follows. If x belongs to the cell which encodes the message to be inserted, x_w (the watermarked version of x) corresponds then to the center of this cell; otherwise, x is moved to the center of the nearest cell that encodes the desired message.

For instance, given a binary message; i.e. $m_i \in \{0, 1\}$, two codebooks C_0 and C_1 are defined. These codebooks can be built up by uniformly quantizing a pixel x with a quantization step Δ, as illustrated in Figure 1.10. In this example, cells centered on crosses represent C_1 (i.e. $m_i = 1$) whereas cells centered on circles represent C_0 (i.e., $m_i = 0$). Thus x will be quantized to the nearest cross or circle in order to encode the message m_i.

To extract the message, the watermark reader has to determine the cell to which the received version x'_w of x_w belongs.

Notice that, unlike the LSB modulation, QIM is a robust substitutive watermarking method [61].

1.3.2 Watermarking of medical images

In their principles, watermarking methods proposed for medical images differ from "classic" or "conventional" ones (i.e. watermarking methods proposed for natural images) mainly in the management of imperceptibility. The visual fidelity of the watermarked image to its original counterpart constitutes the big issue in the healthcare domain. Watermarking a medical image without controlling the introduced distortion could lead to misdiagnosis with possible life threatening consequences. Hence, specific watermarking strategies have been proposed for medical images. One can distinguish three types of watermarking methods: common lossy watermarking, region of non-interest watermarking, and reversible or lossless watermarking.

1.3.2.1 Lossy watermarking techniques

These solutions directly apply conventional watermarking techniques, while paying careful attention in order to ensure that the watermark does not interfere with the diagnostic information. These solutions consider that if the use of lossy compression in medical imaging is accepted, as previously detailed in Section 1.3.1, lossy watermarking methods can also be applied on medical images, but under control. The first solutions of watermarking proposed for medical images are based on this reasoning. They are based on the LSB substitution modulation [56, 69] due to its slight damage of the original image and its large embedding capacity. Let us recall that these solutions are obviously fragile. That is why, robust watermarking [70, 71, 72, 73] schemes were developed. In order to preserve the image quality, these methods make use of a Human Visual System (HVS) model to adapt the watermark amplitude according the image local characteristics. However, these masking models [73, 74, 75] have been proposed for natural images and ahve not been validated for medical images, especially for X-ray, CT, ultrasound and MR images/ Developping such masking models in medical imaging remains a challenging task as it requires mobilizing a lot of expertise resources so as to conduct large subjective studies. To the best of our knowledge,

one solution to overcome this issue has been proposed in [76]; that proposed, in case of X-Ray imaging, to embed the watermark into the quantum noise. Such a noise is inherent to any medical image modality but with different intensities (the noise in ultrasound images is greater than in CT images). In [76], the noise is extracted by means of denoising method in transformed domain proposed by [77]. Once extracted, the noise is modulated so as to embed a message for various security services. Due to the fact that such a noise is of small amplitude, the embedded watermark is fragile.

1.3.2.2 Region Of Interest and Region Of Non-Interest watermarking

These watermarking techniques have been proposed to optimize the performance of the lossy methods in terms of robustness and embedding capacity without introducing a visual distortion in the image. It stands on the idea that medical image can be divided into two separate regions: Region Of Interest (ROI) [78] that includes the informative image features for diagnostic purposes and that should not be modified, and the Region Of Non-Interest (RONI) of less interest or more generally the areas that are not important for the diagnosis and where the watermark amplitude /robustness can be optimized. Such RONI usually corresponds to the black background of the image. These methods can achieve high robustness by embedding strong watermark in the RONI [79, 80]. However, these methods have some drawbacks:

- Robust watermark can hinder and annoy the radiologists in their interpretation of the image.
- The embedded message length depends on the RONI area size, which can be absent in certain medical images where ROI covers the whole image.
- If the ROI that constitutes the important informative area of the image is not watermarked/protected against malicious attacks, it facilitates ROI copy/past attacks (e.g. copying/pasting ROI or RONI from distinct images).

Most of RONI- watermarking methods [78, 81, 82, 83] work as follows: the regions of interest are preselected by experts [78] or using an automatic segmentation procedure. Then, the ROI LSBs' information [78] or a cryptographic watermark representing the hash of ROI of the image [82, 83] is embedded into the RONI in order to provide integrity. Along with this watermark, patient personal data and the hospital logo can be also embedded into RONI to implement authenticity.

1.3.2.3 Reversible watermarking

As stated above, the embedding of a watermark, no matter how minor or trivial the modifications are, there is still a risk to bias the radiologists' interpretation and diagnostic. From here comes the idea of reversible watermarking. As stated in Section 1.3.1, the reversibility property ensures the possibility to remove the embedded watermark, and thereby the exact recovering of the original image. Unlike other methods, reversible watermarking techniques also allow the update of the content of the watermark in order to be able to trace the images without introducing additional distortion; As we will see in more details in Chapter 4, these methods are based on either Difference Expansion (DE) [84], histogram shifting [85, 86], or combination of them [87]. In their vast majority, reversible methods are fragile. Once the watermarked image damaged, the watermark is lost and cannot be removed from altered image parts. This is why these methods are appropriate for integrity control. Moreover, the desire to maximize the capacity often leads to highly visible watermarks, hence the watermark has to be removed before the image can be used. We can here notice that once the watermark is removed, the image is unfortunately no longer protected as for cryptography. Nevertheless and as we will see in Chapter 4, these methods are still of interest in the medical field.

1.4 Conclusion

As we saw in this chapter, medical images are issued from various sources and manipulated in highly open environments. They are transferred and shared between different users inside and outside the medical centers, in case of telemedicine or when the PACS is externalized into the cloud. In such a framework, medical images can be modified accidentally, as for example during communication, or malevolently with the introduction or removal of signs of pathology. These risks together with the

sensitive nature of medical images have led to the definition of strict deontological and legislative rules that oblige health professionals to ensure the security of patients' data. Focusing on medical data in daily medical practice, a special interest should be given to their reliability. As seen, this one is at the basis of trust a physician can have in the data he/she handles and on which he/she is going to take a medical decision. By definition, data reliability is based on the outcomes of data: i) integrity – proof that the image has not been modified by non-authorized users, and ii) authenticity – proof that the image belongs to the correct patient and issued from the right source. Being able to provide the proof of data reliability all along their life cycle leads to traceability; a major concern when data are reused by machine learning algorithms so as to develop new diagnosis aid systems.

We have also pointed out that the deployment of security solutions for medical images requires the implementation of various security mechanisms. Indeed, there does not exist a "unique" mechanism that can control the access to data while ensuring its reliability during its life cycle. These techniques should be complementary and coherent to provide a high level of security while not perturbing the daily clinical practice. But, in essence, most of security mechanisms are oriented to data confidentiality and to the protection of the medical information system. In addition, very few are devoted to integrity; these ones being based on the digital signature principles. All these mechanisms provide an *a priori* protection. Once data decrypted or dissociated from their digital signature, they are no longer protected. In this context, watermarking, an *a posteriori* protection, is of main interest. It allows to the data while maintaining it protected. Even though watermarking provides a non-trivial contribution, there is an interest to go further and to see how it can be combined with cryptography. The idea is to take advantage of both technologies and to ensure an *a priori* and *a posteriori* protection at once.

Nevertheless, the deployment of security mechanisms in the healthcare domain should take into account its specificities. In particular and as we have seen, medical images constitute large volumes of data: for instance, one CT image file size can be up to 1GB. For question of efficiency, medical images should be compressed in order to reduce costs of storage and transmission. Verifying the security of medical while keeping them compressed, that is to say without decompressing them even partially constitutes one of the main objective of this Ph. D thesis. In order to gain in processing time, we propose to verify the reliability of medical images directly from their compressed bitstreams. This will be detailed in the next chapter. In Chapter 3, we will study how jointly conduct encryption, compression and watermarking in order to ensure the confidentiality of images while being able to control their reliability, encryption in the compressed and encrypted domains.

As also discussed, the reversibility property of watermarking is of special interest in the healthcare domain as the original image can be restored once the watermark is extracted, leading thus to various applications such as integrity and authenticity control. In chapter 4, we will illustrate how to merge reversible watermarking with encryption so as to verify the security of an image in its both encrypted and clear (i.e. decrypted) forms. In particular, we introduce a novel robust histogram shifting modulation.

The evaluation of the distortion introduced by watermarking is another important issue in medical imaging. This will be the subject of Chapter 5, where we introduce a preliminary study which aims at measuring the impact of the watermarking process on the visual and diagnostic quality of medical images. The first objective of this study is to demonstrate the fact that the information loss sue to the embedding of a watermark is not necessarily linked to a loss of quality. Beyond, the main objective is to identify the optimal watermark distortion so that diagnostic value is not compromised.

Chapter 2

Reliability Control of Compressed Medical Images

As mentioned in Chapter 1, in recent years, medical information system are more and more integrated and connected across local, regional and institutional boundaries in order to offer infinite ways of making progress toward better health. For example, it becomes possible to share medical data between physicians for remote diagnosis or a second opinion, as well as for patients to access to their medical records stored in the Cloud or to discuss with the health professionals. However, such advances in information and communication technologies also led to a situation where data are confronting many security and privacy threats. Medical images are obviously concerned. Recent image processing based on deep learning demonstrates that they can be easily manipulated. In [88], authors makes data increase for mammogram classification by introducing finding in images of healthy women. Any misuse of such data could lead to irreversible consequences for the patient. More generally, in 2018, [33] estimates that more than 2,100 patients annually die because of data breaches at hospitals.

Different security mechanisms have been proposed in order to maintain data confidentiality and/or control their reliability all along their life cycle. Among them, one can find digital watermarking. By definition and when it is applied on images (see Section 1.3.1), watermarking corresponds to the embedding of a binary message/watermark (e.g. image origin, user identifier) by imperceptibly modifying the image pixel or coefficient values. Such a message can be used so as to verify whether the image has been modified or whether it has been illegally redistributed. Furthermore, one of its major interests is that watermarking provides a protection independent of the file format storage of the information. It leaves access to the information while maintaining it protected by the embedded watermark. It is an *a posteriori* protection mechanism that completes other existing data protection tools (e.g. encryption, access control). Its development in the medical domain is of strong interest, but several constraints have to be considered. On of them is that medical images are usually stored under a compressed form.

Indeed, the growth and the advent of medical imaging have resulted in growing demand to transmit and store images digitally. In fact, medical images constitute very large volume of data. One hospital can produce at least 27,000 terabyte of medical imaging data per year [17]. Herein, compression obviously plays an important role. It is not only used to decrease the requirements of storage, but also to decrease the entire time of transmission, image processing and so on. As seen in Chapter 1, many lossy and lossless image compression techniques have been included in the DICOM standard as compressed data formats for medical image storage and exchange. If with lossless compression the original image can be exactly recovered after its decompression, lossy image compression induces information loss, which under some constraints remains negligible while allowing increasing compression rate. Notice that such a technique is widely used in medical applications where a less amount of compromise on image quality is tolerable [27].

From this standpoint, it becomes thus desirable to develop solutions that can give access to watermarking-based security services (e.g. integrity and authenticity control) in the compressed domain i.e. available from the image compressed bitstream. To achieve this goal, three main issues should be considered when working with medical images. The first one stands on the capacity to embed a message without causing quality degradation of the image. The second is to identify where and how the compressed bitstream can be modified without changing its semantic; that is

to say guaranteeing that the bitstream can be decompressed. The third one relies on the capability to extract the message from the compressed image bitstream without having to decompress it nor to parse it, even partially.

2.1 Combination of compression and watermarking

Existing solutions that combine watermarking and compression can be discriminated depending on the compression-watermarking processual sequence: watermarking before compression (WBC) [89, 90, 91, 92], watermarking of the compressed bitstream (CBW) [93, 94, 95] or joint watermarking-compression when watermarking is conducted during the compression process (JWC) [96, 97, 98]:

- *WBC schemes* − the message is embedded before image compression and is only available in the decompressed domain. In essence, these methods watermark the quantized coefficients of the image Discrete Cosine Transform (DCT) or of the image Discrete Wavelet Transform (DWT), depending on the subsequent image compression algorithm, that is to say JPEG or JPEG2000, respectively. In [89], two human visual system models (visual and edge entropy models) are used to select DCT blocks suitable for the embedding, i.e. blocks that cause the least image distortion. The message is then embedded into certain DCT coefficients of the middle frequency based on a psychovisual threshold to enhance robustness and imperceptibility. In [90], the message is inserted into the wavelet coefficients of the image by applying spread spectrum modulation onto the most significant part of the host image without introducing noticeable degradation. The authors of [91] proposed to embed the watermark into the DCT coefficients of the image based on texture and luminance masks so as to adapt the watermark amplitude and maintain it imperceptible. In these methods, the message is designed to be robust to a lossy compression algorithm by embedding it into the appropriate transformed domain (i.e. Discrete Cosine Transform (DCT) - JPEG; Discrete Wavelet Transform (DWT) - JPEG2K). However, the watermark is only available in the uncompressed domain. To make these operations possible the compressed image bitstream has to be pre-processed before being encrypted. This pre-process reorganizes and modifies the compressed wavelet coefficients in order to: i) facilitate the identification of watermarkable coefficients in the encrypted image bitstream, and ii) ensure the watermarking process does not modify elements of the compressed bitstream that will make impossible image decompression.

- *CWB methods* - This strategy consists in embedding the message directly into the image compressed bitstream or after its partial decompression. Watermark extraction is usually conducted in the same domain as the embedding process. In [93], it is the Variable Length Coding (VLC) table of the JPEG file header that is watermarked. In order to extract the message, the reader has to compare the original VLC table to the watermarked one. This scheme is thus non-blind and once the image is decompressed, it is no longer protected. In [94], after Huffman decoding of the JPEG image file; that is to say after a partial decompression of the image, the watermark is inserted into the quantized AC coefficients. In order to preserve the file size, the authors proposed to maintain the same coefficient category (i.e. coefficient magnitude) for the watermarked coefficients. The message is extracted from the watermarked quantized AC coefficients by following the same steps of the embedding. The authors of [95] went further. After JPEG2000 compressing the image, the bitstream is encrypted and a watermark is inserted into the least significant bit planes of the encrypted coefficients of the middle wavelet frequency in order to guarantee a good quality of the watermarked image. The watermark detection is conducted from the decrypted JPEG2000 compressed image bitstream using a correlation detector along with the *a priori* known pseudo-random sequence. To sum up, in this class, the watermark detection is always conducted in the same domain as the embedding process.

- *JWC schemes* - where the message is embedded during the compression of the image. The main benefit of such a strategy is that compression is no longer considered as an attack from the watermark robustness point of view. In [96], the insertion of the message is performed during the image the wavelet coefficients quantization step of JPEG2000 image compression. Message extraction can be achieved either during the inverse quantization step of JPEG2000 decompression of from the decompressed watermarked image. The main benefit of such a

strategy is that compression is no longer considered as an attack from the watermark robustness point of view. The scheme proposed in [97] involves watermarking during JPEG compression. It allows watermark embedding into the middle or high frequency (quantized or not) DCT coefficients based on an additive watermarking. However, this scheme is not blind; that is to say, the original image is required to be able to extract the watermark. [98] addresses joint watermarking-JPEG-LS. It is based on the near-lossless mode of JPEG-LS, which induces some information loss in addition to the distortion introduced by watermarking. At the same time, with this scheme, the message can only be read from the decompressed watermarked image or during the decompression process. More clearly, accessing the message without decompressing the bitstream is not possible. It should be noticed that, in this class, it is not possible to extract the watermark from the compressed image bitstream. More clearly, to verify the image security, the watermark reader has to decompress totally or partially the watermarked image.

With methods of the first class, message extraction is always conducted in the decompressed domain. That is not the case of the two last classes, where the watermark reader has to decompress the image bitstream, totally or partially or to compare the watermarked image to the original one for message extraction. As we will see, that is not the case of the solution we propose.

During this PhD, we propose a new joint watermarking-compression scheme we first propose for JPEG-LS and by next we extend its principles to JPEG. Its originality stands on the fact that it allows accessing to watermarking-based security services from compressed image bitstream without having to decompress it, even partially. The system we propose combines, in a single operation, bit-substitution watermarking modulation with JPEG-LS, in a first scheme, and with JPEG in a second one. It is important to notice that with our schemes, it is possible to normally decompress the image with the common JPEG-LS or JPEG algorithms. More clearly, decompression as well as message extraction processes are conducted independently. We did not modify nor adapt them for a proprietary protection mechanism. With such a capability, our scheme is fully DICOM-compliant. Furthermore, this solution makes possible to trace images and control their reliability (i.e. authenticity and integrity) directly from compressed domain. Note also that the proposed scheme saves the computational complexity as it does not require decompression to verify the image reliability in the compressed domain.

In this chapter, the two compression standards, JPEG-LS and JPEG, are first presented. In the second part, we will give an overview of the method proposed to access to watermarking-based security services from the image com-pressed bitstream without decompressing it, even partially. Then, we will detail how watermarking is jointly conducted with JPEG-LS and with JPEG compression for verifying the image reliability. Finally, we will discuss both schemes' performance in terms of image quality and embedding capacity.

These works were published in [6] and in [7].

2.2 JPEG-LS and JPEG standards

2.2.1 JPEG-LS

JPEG-LS is a lossless image compression standard proposed by the International Standards Organization ISO/IEC JTC [99]. DICOM allows its use in order to compress medical images. JPEG-LS is based on the Low Complexity Lossless Compression for Images (LOCO-I) algorithm [100]. This one relies on a pixel prediction based on a local contextual statistical model.

The main steps of JPEG-LS are described in Figure 2.1. As it can be seen, an image is sequentially processed pixel by pixel. JPEG-LS encodes one pixel x accordingly to two modes: the regular-mode or the run-mode.

2.2.1.1 JPEG-LS compression

Local activity analysis

To decide encoding a pixel x in the regular-mode or in the run-mode, JPEG-LS first analyses the local activity around it. To do so, three local gradients $\{g_i\}_{i=1..3}$ are computed from its immediate causal neighborhood constituted of previously encoded pixels (i.e. pixels a, b, c and d in Figure 2.1), such that:

$$\{g_1 = d - b, g_2 = b - c, g_3 = c - a\} \tag{2.1}$$

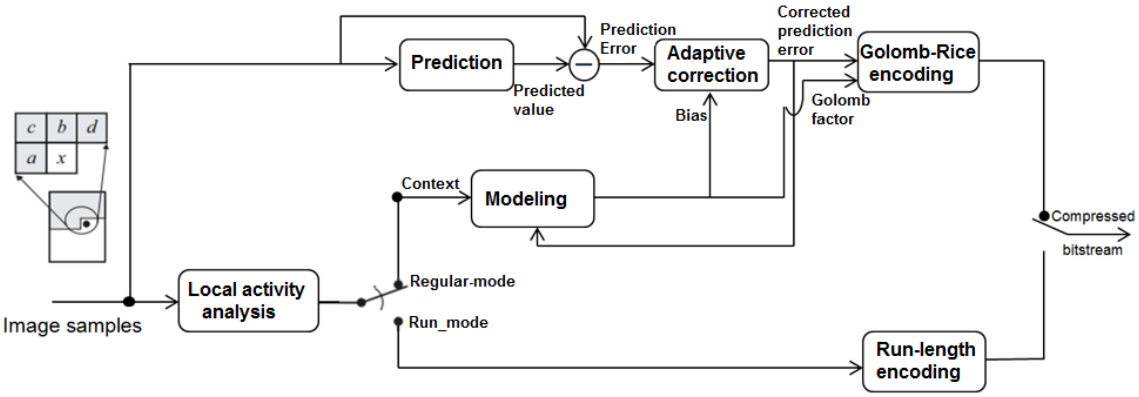


Figure 2.1: JPEG-LS general scheme

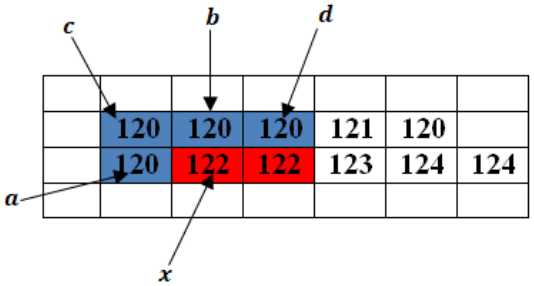


Figure 2.2: An example of a pixel sequence that will be encoded in run-mode, where the local gradients computed from $\{a, b, c, d\}$ are null and the sequence length is of 2 pixels.

If all local gradients are null then the run-mode is activated, otherwise the regular-mode is chosen.

Run-mode encoding

JPEG-LS run-mode is based on run-length encoding (RLE). It encodes the number of times the previous pixel has been repeated. This mode stops when a pixel of a different value is encountered or when the end of an image line is met. An example of a pixels' sequence that will be encoded in run-mode is given in Figure 2.2. This run length encoding mode is based on an inverse unary encoder in order to code the repetition number. For example, if the repeated sequence length is of 2 pixels, the inverse unary code of 2 corresponds to '110'.

Regular-mode encoding

As depicted in Figure 2.1, the regular-mode encoding of one pixel x relies on three steps:

- Prediction
- Adaptive correction
- Golomb-Rice encoding

Prediction

The causal neighborhood of x is used to estimate its gray value. To do so, JPEG-LS exploits the edge-detecting predictor:

$$\hat{x} = \begin{cases} \min(a,b) & \text{if } c \geq \max(a,b) \\ \max(ab) & \text{if } c \leq \min(a,b) \\ a+b-c & \text{otherwise} \end{cases} \tag{2.2}$$

where $\hat{x}$ is the estimated value of x. The prediction-error, that will be processed, is then calculated as $e = x - \hat{x}$.

Adaptive correction

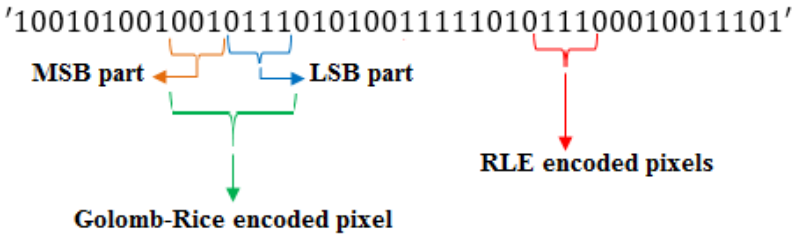


Figure 2.3: JPEG-LS compressed bitstream example.

Because the edge-detecting predictor only provides integer values, it introduces a prediction bias. This bis corresponds to a translation of the prediction-error e. In order to compensate such a systematic offset, a context-dependant term is computed so as to shift back the prediction-error for each image pixel. To do so, JPEG-LS assigns to x a context value Q evaluated from its associated local gradients according to the following procedure:

a) Local gradients $\{g_i\}_{i=1...3}$ are quantized into $\{q_i\}_{i=1...3}$ accordingly to a non-uniform scalar quantization defined in the JPEG-LS standard;

b) JPEG-LS computes the context Q of x, defining it as an index value, such that:

$$Q = 81q_1 + 9q_2 + q_3 \tag{2.3}$$

This context value is subsequently used to correct the prediction-error bias as well as to parameterize the Golomb-Rice encoder. For a given pixel x and its context Q, this term corresponds to the prediction-error mean $\bar{e}$ of pixels belonging to the same context Q as x.

Golomb-Rice encoding

The next step in the regular-mode consists in encoding the corrected prediction-error with the help of the Golomb-Rice encoder (GRE). Due to the fact, GRE only encodes non-negative values, the corrected prediction-error is modified into a "mapped prediction-error" $\tilde{e}$ such that the positive and negative values are mapped to even and odd positive integers of $\tilde{e}$, respectively. Hence, the parity of $\tilde{e}$, or equivalently its least significant bit indicates the corrected prediction-error sign. Then, the resulting mapped prediction-error is GRE encoded. Basically, GRE encodes a positive integer value p into two parts that are concatenated. These two parts of p will be referred in the sequel as: the *Most Significant Bit part* (p_{MSB}) and the *Least Significant Bit part* (p_{LSB}). Notice that these two parts are of great importance in our proposal. The MSB-part corresponds to the quotient of the Euclidean division of p by 2^k, where k is the non-negative Golomb-Rice factor. This quotient is unary encoded, i.e. by a sequence of '0' ended by '1', the number of '0' being the quotient value. This part is then concatenated to the LSB-part that corresponds to the binary encoding of the division reminder, i.e. an integer value encoded on k bits. To sum up, in the case of $p = 53$, $k = 4$, then the Golomb-Rice encoding p is such as $'00010101'$, where p_{MSB} code is '0001' and p_{LSB} code is '0101'. Note that the value of Golomb-Rice factor, i.e. k, is also pixel's context dependent. Notice also that the Golomb-Rice MSB-part codes follow a geometric distribution making it highly suitable for situations in which the occurrence of small prediction-errors' values is significantly more likely than large values.

2.2.1.2 JPEG-LS decompression

JPEG-LS decompression is accomplished by performing the same steps as for compression. To decode one pixel, its local context is re-computed from its causal neighborhood already decompressed. Based on this context, the decoding mode is chosen. If the regular-mode is selected, the predicted value $\hat{x}$ of x is computed and its bias corrected as in the encoding stage. Then, the pixel's Golomb-Rice factor k is calculated. The mapped prediction-error is decoded from the compressed image bitstream and the pixel value x is worked out. Otherwise, the run-mode is activated and JPEG-LS decodes the number of repetitions of the previous decoded pixel.

The example given in Figure 2.3 illustrates the JPEG-LS encoding of a sequence of pixels. It can be seen that without any additional information it is rather impossible to clearly identify a pixel in the compressed bitstream. Nevertheless, there is a high probability that a sequence '0X1'; where X is a sequence of '0', indicates the position of a regular-mode encoded pixel. As we will

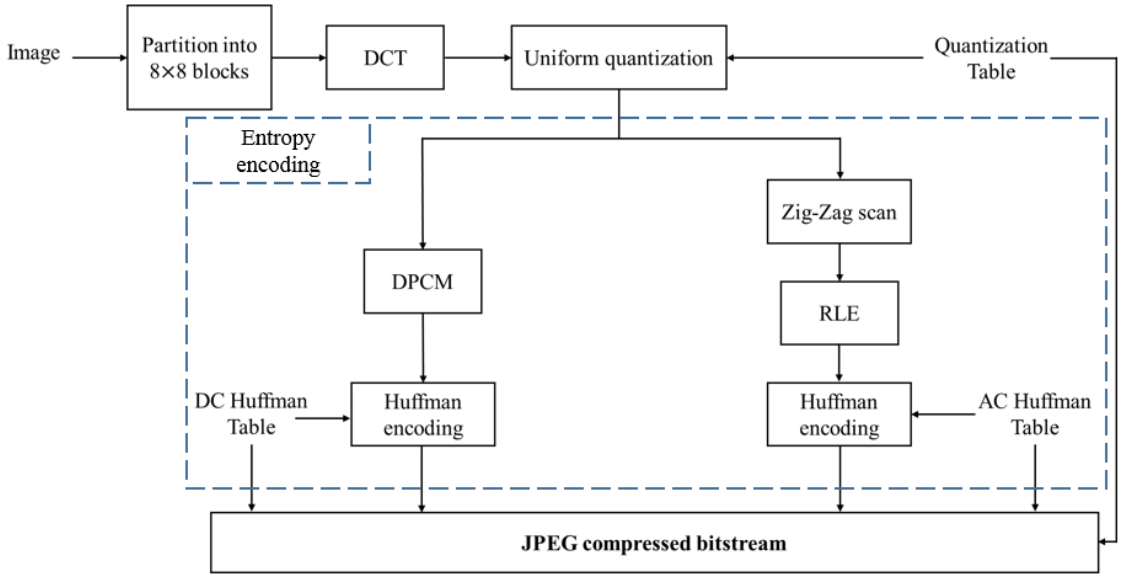


Figure 2.4: JPEG compression general scheme.

see in Section 2.3, we will take advantage of such property so as to make a watermarked message available in the compressed domain.

2.2.2 JPEG

JPEG (Joint Photographic Experts Group) [101] is a widely used image compression standard. It supports two different coding processes: lossy and lossless.

In this work, we are interested in the "baseline JPEG" (i.e. lossy JPEG compression in its sequential mode). Its main steps are given in Figure 2.4. Considering gray-scale images, first of all, JPEG divides the image into non-overlapping 8×8 blocks. Each block is then sequentially processed: transformed using the Discrete Cosine Transform (DCT), quantized, and prepared for the entropy encoding. We detail these steps in the following differentiating the compression and decompression stages of JPEG.

2.2.2.1 JPEG Compression

DCT transformation

Due to the fact that, in the spatial domain, the information of an individual pixel is relatively small but highly correlated to its neighbors (i.e. one pixel value can be predicted from its neighboring ones), image data are transformed into the frequency domain so as to eliminate this correlation. To do so, each pixel $B(x, y)$ of a 8×8 pixel block B is transformed from the spatial to frequency domain based on the Discrete Cosine Transform (DCT):

$$B(u,v) = \frac{2}{N} c(u)c(v) \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} B(x,y) \cdot \cos\left[\frac{\pi}{N} u \left(x + \frac{1}{2}\right)\right] \cdot \cos\left[\frac{\pi}{N} v \left(y + \frac{1}{z}\right)\right] \tag{2.4}$$

where $\begin{cases} c(w) = \frac{1}{\sqrt{2}}, & \text{if } w = 0 \\ c(w) = 1, & \text{otherwise} \end{cases}$

$B(0,0)$ is the DC coefficient (DC as in "*Direct Current*"), which refers to the average brightness in the block while $B(u, v)$, where $u, v > 0$, are the AC coefficients (AC from "*Alternating Current*").

The interest of DCT is not only to decorrelate the image signal, but also to concentrate its energy in a small number off coefficients. More clearly, coefficients of low spatial frequencies carry most of the signal energy (i.e. top left region as given in Figure 2.5). Even if high frequency coefficients are set to zero, the resulting reconstructed block (i.e. after applying the inverse DCT) will be quite close to the original block [102].

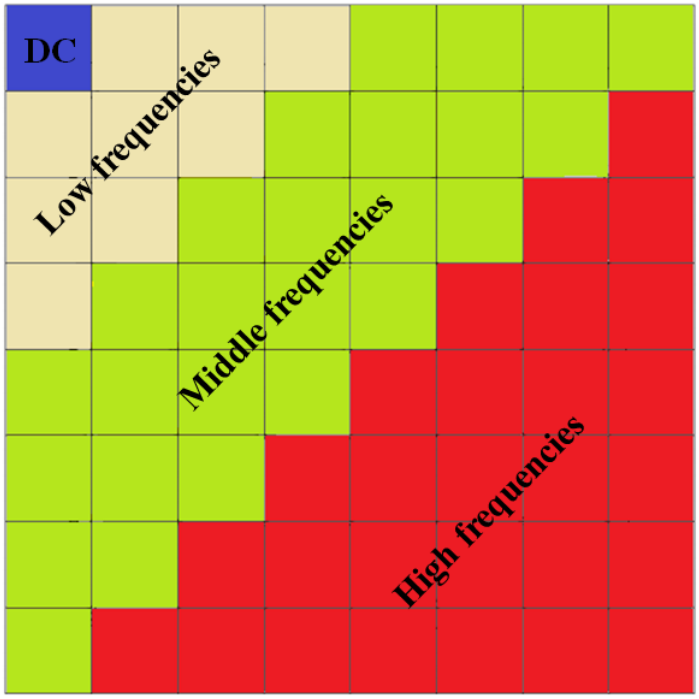


Figure 2.5: Low, Middle, and High frequencies distribution in a DCT block.

16	11	10	16	24	40	51	61
12	12	14	19	26	58	60	55
14	13	16	24	40	57	69	56
14	17	22	29	51	87	80	61
18	22	37	56	68	109	103	77
24	35	55	64	81	104	113	92
49	64	78	87	103	121	120	101
72	92	95	98	112	100	103	99

Figure 2.6: An example of a quantization table.

JPEG quantization

The objective of this step is to take advantage of the human visual system and to optimize the compression rate by suppressing the information where human eyes are less sensitive to. Indeed, human vision can be considered as a low pass filter, being less sensitive to high frequency variations. To conduct this task, DCT coefficients are quantized according to psychovisual-dependent quantization tables. These default tables can be also scaled to vary the compression ratios. More clearly, JPEG allow the user to specify a range of values for the scaling factor, by specifying a compression metric called quality factor Q_f (0 (lowest quality) $< Q_f \leq 100$ (best quality)):

$$B_q(u,v) = \left\lfloor \frac{B(u,v)}{q(u,v)} \right\rfloor Q_f \tag{2.5}$$

where $B(u,v)$ is the original DCT coefficient, $q(u,v)$ is the quantization step associated to the spatial frequency (u,v) and issued from table Q, and $B_q(u,v)$ is the quantized DCT coefficient. Note that the quantization table has to be stored in the header of the JPEG image file.

As it can be seen in Equation 2.5, this process introduces an irreversible information loss, due to integer rounding. It is not possible to recover $B(u,v)$ from $B_q(u,v)$. Figure 2.6 gives an example

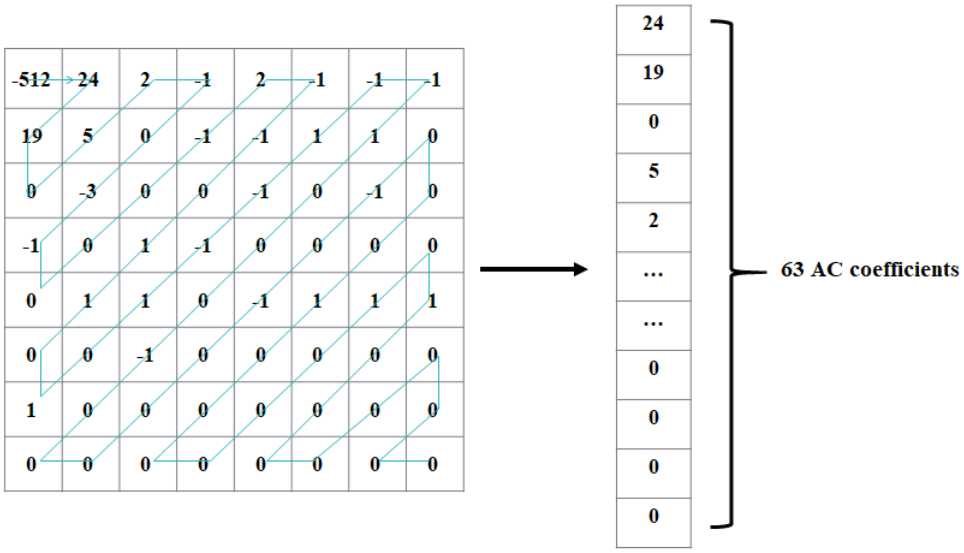


Figure 2.7: Ordering of a 2D matrix into 1D vector according to a zig-zag scan.

of a quantization table. As shown, the quantization steps for high frequencies are higher than for low frequencies.

JPEG entropy encoding

After the quantization step, DC and AC coefficients are managed differently. The redundancy between DC coefficients of adjacent blocks is exploited using prediction so as to reduce the information entropy to encode. To do so, the Differential Pulse Code Modulation (DPCM), which computes the difference *diff* between DC coefficients of the current and the previous blocks.

Regarding AC coefficients, they are first ordered in a 1×63 vector according to a zig-zag scan. The objective of this process is not only to pass from a 2D structure to a 1D vector easier to handle, but also to ensure that consecutive coefficients are close in terms of spatial frequencies, and consequently in values. Another consequence of such a zig-zag scan along with quantization, is that this 1D vector will end with a more or less long sequence of null coefficients (see example in Figure 2.7). In order to take advantage of this ordering and to optimize the compression rate, JPEG next applied RLE (Run Length Encoding). RLE encodes the 1D vector coefficient vector in (*run*, *value*) pairs; where *run* corresponds to the number of zeros preceding a non-zero AC coefficient, and value corresponds to the value of the non-zero AC coefficient.

Finally, Huffman entropy encoding is applied onto RLE encoded AC and on DC prediction-errors. In essence, Huffman encoding is based on variable-length table it derives a table from the estimated probability for each possible value of the source symbol. This table associates to one symbol a codeword, a sequence of bits, the length of which depends on the symbol probability. Shorter codewords are associated to symbol of highest probability and longer ones to more rare symbols. Notice that no codeword is the prefix of another one, otherwise decoding will not be possible.

As defined, Huffman encoding requires computing this table, but to reduce the algorithm complexity and skip this latter step, JPEG standard fixed four tables (two tables used for DC coefficient: one of Luminance (Y) and the other of Chrominance, and two tables for AC coefficients: one of Luminance (Y) and the other of Chrominance) that are reasonable approximations of the statistics of most images and that define the mapping between the variable-length codes and the code values. Note that these tables are stored in the header of any JPEG image file.

In order to encode one DC prediction-error, the difference value *diff* is represented as ($size_{diff}$, val_{diff}); where $size_{DC}$ is the number of bits needed to encode the value of *diff* val_{diff} in binary. Then, $size_{diff}$ is encoded the corresponding JPEG DC-Huffman table. Table 2.1 provides the JPEG DC-Huffman luminance table of gray-scale images. It can be seen that *diff* magnitudes $size_{diff}$ are grouped into 12 categories and a unique Huffman codeword is associated to each category. A codeword c_i in the Huffman table is not the prefix of any other codeword; that is

Table 2.1: Huffman - Luminance (Y) - DC

$size_{diff}$	Codeword	val_{diff}	Code
0	00	0	–
1	010	-1, 1	0,1
2	011	-3,-2, 2,3	00,01,10,11
3	100	-7,. . . , -4, 4,. . . , 7	000,. . . ,011,100,. . . ,111
4	101	-15,. . . , -8, 8,. . . , 15	0000,. . . ,0111,1000,. . . ,1111
.	.	.	.
.	.	.	.
11	111111110	-2047,. . . ,-1024, 1024,. . . , 2047	. . .

to say, no other codeword in Huffman table starts with c_i in order to correctly decode *diff*.The difference value val_{diff} is then binary encoded and concatenated to the Huffman code of $size_{diff}$. For instance, given a difference value $diff = -512$. Thus, $size_{diff}$ is equal to 10 bits and $val_{diff} = -512$. According to DC-Huffman table, the codeword that correspond to $size_{diff} = 10$ is of $'1111110'$ and the binary code of $val_{diff} = -512$ is of $'0111111111'$. The final code of *diff*, that corresponds to the concatenation of $size_{diff}$ and val_{diff} is of $'1111110011111111'$.

Regarding AC coefficients, the resulting (*run*, *value*) pairs of RLE are Huffman encoded according to the following procedures:

- $[(run, size_{AC}), val_{AC}]$; where $size_{AC}$ represents the number of bits needed to encode the non-zero AC of value val_{AC}.
- $(run, size_{AC})$ is next encoded based on the corresponding AC-Huffman table . We give an example of such table in case of gray-scale images in Table 2.2.
- Finally, the Huffman encoded version of $(run, size_{AC})$ is concatenated to the binary code of val_{AC}.

It is important to notice that in the case the block ends with zeros (i.e. null AC coefficients), this sequence of zeros is encoded with the codeword $EOB =' 1010'$, where EOB refers to End Of Block.

Note that the Huffman encoding is a variable length algorithm, which assigns the shortest codes to the most frequently values. The JPEG standard provides up to four Huffman tables that define the mapping between the variable-length codes and the code values.

To more clarify the entropy encoding, we give an example, in Figure 2.8, of Huffman entropy encoding of a zig-zag scanned quantized block.

2.2.2.2 JPEG Decompression

JPEG decompression is accomplished by applying the preceding steps in reverse order. First of all, based on the Huffman tables from the JPEG image-file header, entropy and RLE decoding are performed. The resulting data are then rearranged accordingly to the zig-zag scan procedure to reconstruct 8×8 quantized DCT blocks. After that, de-quantization is applied on blocks. More clearly, each quantized coefficient $B_q(u, v)$ is multiplied by the corresponding quantization step value. Let us recall that this value is also stored in the JPEG file header. Finally, IDCT (Inverse Discrete Cosine Transform) is applied so as to get access to decompressed pixels $B(x, y)$ of the

Table 2.2: Huffman - Luminance (Y) - AC

run/$size_{AC}$	**Codeword**	**run/$size_{AC}$**	**Codeword**
0/1	00	3/1	111010
0/2	01	4/1	111011
0/3	100	0/6	1111000
0/0 (EOB)	1010	1/3	1111001
0/4	1011	5/1	1111010
1/1	1100	6/1	1111011
0/5	11010	0/7	11111000
1/2	11011	...	...
2/1	11100	F/A	1111111111111110

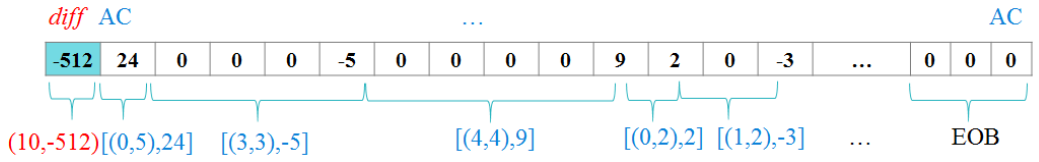


Figure 2.8: Example of a Huffman encoding of the zig-zag scanned vector; where EOB corresponds to the "End of Block" (i.e. a sequence of zeros until the end of the block). The resulting bitstream corresponds to the Huffman code of the resulting data after the DPCM and RLE processes on the DCT block.

8×8 block. IDCT is given by:

$$\mathrm{B}(x,y) = \frac{1}{4}\sum_{u=0}^{7}\sum_{i=0}^{7} c(u)\cdot c(v)\cdot B(u,v)\cdot \cos\left[\frac{\pi}{8}u\left(x+\frac{1}{2}\right)\right]\cos\left[\frac{\pi}{8}v\left(y+\frac{1}{2}\right)\right] \tag{2.6}$$

where $\begin{cases} c(w) = \frac{1}{\sqrt{2}}, & \text{if } w = 0 \\ c(w) = 1, & \text{otherwise} \end{cases}$

It is important to notice that the decoding process needs to parse the JPEG file header to extract the Huffman and quantization tables so as to be able to identify, in the compressed bitstream, bits that correspond to the DCT coefficients. Without these pieces of data, it is rather impossible to identify a DCT coefficient directly in the compressed bitstream. Nevertheless, as already said in Section 2.2.2.1, the standard Huffman tables have specific codewords, where no codeword corresponds to a prefix of another one in the same table and shorter codewords correspond to more frequent DCT coefficients. Hence, we will take advantage of the standard JPEG Huffman codewords in order to make a watermark accessible from the JPEG compressed image bitstream without decompressing it, even partially.

2.3 Proposed Joint Watermarking-Compression Scheme

The purpose of our scheme is to give access to watermarking-based security services directly from the compressed image bitstream without decompressing it, even partially. To do so, we propose to

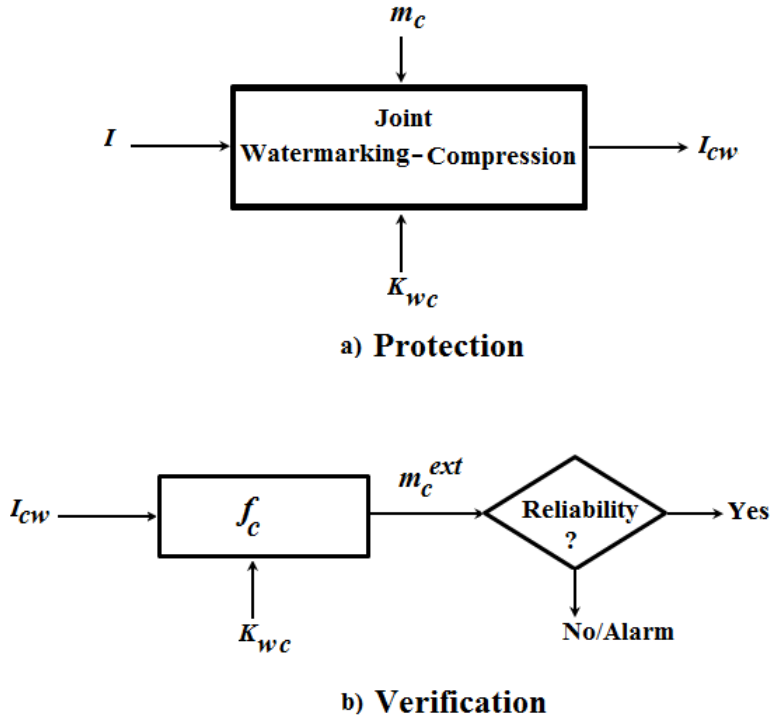


Figure 2.9: General scheme of the proposed joint watermarking-compression for image protection.

merge image compression and watermarking in a single operation, under the constraint that decompression can be normally conducted; that is to say without the need of a proprietary decompression algorithm in order to be compliant to the DICOM standard.

2.3.1 System Architecture and Basic Principles

The basic architecture of our scheme is given in Figure 2.9. At the protection stage (see Figure 2.9(a)), the watermarking process or equivalently message embedding is performed during the encoding of the image I. Such a joint watermarking-compression allows the insertion of a message m_c that will be available from the compressed image bitstream I_{wc} without having to decompress it, even partially. The embedded message m_c contain security attributes that assess the image reliability (see Sub-Section 2.3.5). The embedding and the extraction of m_c depend on a secret watermarking key K_{wc} in order to provide an additional security to our scheme.

At the verification stage (see Figure 2.9(b))), the image reliability can be verified by accessing to the extracted m_c^{ext} with the help of the extraction function f_e. Note that the watermarked-compressed image I_{wc} can be decompressed in a regular way. That is to say JPEG image decompression is performed with no modification of its algorithm; watermarking is completely transparent to JPEG. This is the main interest of DICOM system that are not watermarking compliant.

In the following, we expose how to deploy our DICOM-compliant scheme in the case of JPEG-LS and JPEG, while considering the bit-substitution watermarking modulation. Let us recall that such a watermarking modulation replaces one bit of an integer p by one bit of the message to embed.

2.3.2 Joint Watermarking-Compression (JWC) Scheme

2.3.2.1 Message embedding

As mentioned in Chapter 1 (Section 1.1.2), compression methods stand on an image transformation followed by an entropy encoding. Both Huffman and Golomb-Rice entropy encoders are variable-length encoders, which replace a fixed-length input symbol by the corresponding variable-length prefix codeword. Hence, we take advantage of this property to embed the message m_c and to make it accessible from the compressed image bitstream without decompressing it. Our basic idea is to

fix a reference sequence R (i.e. label) among the variable-length prefix codewords related to the entropy encoder and embed the message into image pixels or coefficients their prefix codewords correspond to the reference sequence R. This later has to be known from the watermark reader in order to extract the message m_c. More clearly, the reference sequence R will labialize the position of the message bits in the image watermarked-compressed bitstream.

Nevertheless, different constraints has to be considered so as to decide which bit that will be watermarked and to guarantee an error-free extraction of m_c:

(i) The embedded message should be extracted from the compressed image bitstream without computing any additional parameters. The watermark reader has just to identify the reference sequence R so as to extract the watermark.

(ii) It is possible to embed a message bit after each reference sequence R.

(iii) A sequence of bits equals to R should correspond to a watermarked pixel/coefficient in the watermarked-compressed bitstream. We will come back to these constraints with more details for each compression standard (i.e. JPEG-LS and JPEG) in Section 2.3.3 and 2.3.4.

2.3.2.2 Message extraction

Based on the secret watermarking key K_{wc}, the extraction function identifies the reference bit-sequence R in the compressed-watermarked image bitstream I_{wc} and extract the message m_c^{ext}, without neither decompressing the compressed image bitstream nor computing additional parameters.

2.3.3 Joint Watermarking-JPEG-LS Compression (JWJLS) Scheme

As exposed in section 2.2.1, JPEG-LS sequentially encodes pixels depending on their local context and accordingly to two encoding-modes: the regular and the run modes. To roughly sum up, the run-mode is selected to compress sequences of pixels in flat areas while the regular-mode is used to encode the others. We recall that this latter relies on a prediction and context-based coding of pixel prediction-errors using the Golomb-Rice encoder. As shown in Section 2.1, it is rather impossible to clearly identify the bits of one pixel in the bitstream, but there is a high probability that a sequence of bits $'0X1'$, where X is sequence of '0', corresponds to a regular-mode encoded pixel. We take advantage of this property so as to embed the message m_c and to make it readable from the image compressed bitstream without decompressing it. Our basic idea is to embed m_c in pixels the mapped prediction-error of which has a Golomb-Rice MSB-part encoded such that $\tilde{e}_{MSB} =' 0X1'$ (see Section 2.1) by modifying its Golomb-Rice LSB-part using watermarking bit-substitution modulation. In the sequel, $R =' 0X1'$ corresponds to the "reference sequence".

In order to decide which bit of the pixel's Golomb-Rice LSB-part can be substituted by one bit of m_c , let us recall that the LSB-part is encoded onto k bits; k being the pixel Golomb-Rice factor (see Section 2.2.1.1). k depends on the pixel's context and is thus unknown from the watermark reader unless it decompresses the bitstream. As a consequence, under the constraint not parsing the bitstream, it is not possible to identify the bits of the pixel's Golomb-Rice LSB-part. Nevertheless, it is possible to embed one bit of m_c by substituting the higher order bit of the Golomb-Rice LSB-part by one bit of the message m_c. More clearly, let us consider a watermarkable (i.e. to be watermarked) pixel p, i.e. a pixel the $\tilde{e}_{MSB}$ of which equals the reference sequence arbitrary fixed to $'0001'$, with a Golomb-Rice factor $k = 4$ and a mapped prediction-error as $\tilde{e} =' 00011010'$; where $\tilde{e}_{MSB} =' 0001'$ and $\tilde{e}_{LSB} =' 1010'$. $\tilde{e}$ will be watermarked into $\tilde{e}_w =' 0001b010' = \tilde{e}_{w-MSB}\tilde{e}_{w-LSB}$ where b is one bit of the message m_c, i.e. $b \in \{0, 1\}$.

However, as mentioned is Section 2.3.2, two main issues have to be considered so as to guarantee the error-free extraction of m_c: i) the possibility to embed one bit after each reference bit-sequence $R =' 0X1'$; ii) a sequence of bits equals to R should correspond to a watermarked pixel in the watermarked compressed image bitstream. Regarding the former issue, a watermarkable pixel should hold an LSB-part $\tilde{e}_{LSB}$, i.e. its Golomb-Rice factor should be non-null, otherwise message insertion is not possible. In addition, due to the fact that the last bit of $\tilde{e}_{LSB}$ informs the JPEG-LS decoder of the prediction-error sign, k should be strictly greater than 1. To overcome this issue, our JWEC algorithm modifies the pixel $\tilde{e}_{MSB}$ by suppressing one '0' (i.e. $'0X1'$ is changed into $'0X1'$). The pixel is thus turned into a non-watermarked pixel.

The second issue imposes the constraint that all sequences $'0X1'$ in the protected compressed bitstream correspond to a watermarked pixel. Otherwise, the watermark will be desynchronized extracting bit values that are not part of m_c. More clearly, the extracted message will be longer than the original one. This kind of problem occurs when a non-watermarkable pixel (i.e. $\tilde{e}_{MSB} \neq' 0X1'$) is preceded by a sequence of bits $'0S'$, where S is a sequence of '0', such that $'0S'\|\tilde{e}_{MSB} =' 0X1'$; '$\|$' being the concatenation operator. To solve it, our scheme considers this pixel as watermarkable and one bit of m_c is embedded into its $\tilde{e}_{LSB}$. Such a problem may also occur in the mapped prediction-error LSB-part, i.e. $\tilde{e}_{LSB}$, of pixels the Golomb-Rice factor k of which is greater than the reference sequence length (i.e. $k > |'0X1'|$ bits). Here, our JWEC algorithm commutes one bit of the pixel $\tilde{e}_{LSB}$ so as to avoid the problem.

Regarding the watermarking extraction function f_c (see Figure 2.9), the watermark reader has just to identify the reference sequence $\tilde{e}_{MSB} =' 0X1'$ in the compressed bitstream I_{cw} and to read the immediate following bit value to extract m_c. In order to ensure the security of our scheme, that is to say securing the access to m_c, we suggest secretly selecting watermarkable pixels with the help of the watermarking key K_{wc}.

2.3.4 Joint Watermarking-JPEG Compression (JWJPG) Scheme

As exposed in Section 2.2.2, JPEG sequentially encodes 8×8 DCT blocks based on the predefined standard JPEG tables (i.e. quantization and Huffman tables). Hence, it will be rather impossible to identify a DCT coefficient in the compressed domain without decompressing the JPEG image bitstream. In other word, without extracting the Huffman tables from the JPEG file header and parsing the JPEG bitstream depending on these tables, none can identify a DCT coefficient, and a pixel, by next. Nevertheless, there is a high probability that the Huffman codewords can be identified from the JPEG compressed bitstream. As previously said in Section 2.2.2, these codewords are present in Huffman tables in the JPEG image header. Hence, by taking the advantage of these properties, we proposed to embed during the JPEG image encoding, a message m_c that will be accessible from the image compressed bitstream without neither extracting the Huffman tables from the JPEG image-file header nor decompressing it, even partially.

To do so, the fundamental of the proposed method is to watermark DCT coefficient values during their encoding preceded by a pre-fixed Huffman codeword, as defined above as reference sequence R that will be used to identify the watermarked bits during the verification stage.

In order to decide which coefficients their Huffman codewords will be used to labialize or to identify m_c, let us recall that the DC coefficients correspond to the average of all coefficients in a 8×8 DCT block and its modification will thus introduce high distortion in the image. That is why we propose to watermark non-null AC coefficients encoded as $[(run_R, size_{AC_R}), val_{AC_R}]$. As described in Section 2.2, the first part $(run_R, size_{AC_R})$ correspond to a Huffman codeword that should not be modified. The second part val_{AC_R} that defines the non-null AC coefficient value will be thus watermarked based on the bit-substitution watermarking modulation.

Therefore, the reference sequence R can be one of the codewords given in Table 2.2 or a concatenation of Huffman encoded DCT coefficients that certainly finish with an AC coefficient Huffman codeword so as to be able to watermark its val_{AC_R}.

After identifying the reference sequence R that will labialize the watermark m_c, let us now define which bit of val_{AC_R} will be watermarked. Unlike Golomb-Rice encoder used by JPEG-LS, the Huffman codewords give an idea about the size of the encoded DCT coefficients. Hence, we suggest substituting the lower order bit of the non-null AC coefficient value preceded by the reference sequence R, by one bit of the message. More clearly, let us consider the following example of a reference sequence $R = EOB\| \left(\text{size}_{diff} = 1, val_{diff} = 1\right) \| \left(run = 0, \text{size}_{AC} = 1\right) =' 1010010100'$; where: '$\|$' is a concatenation operator, $(\text{size}_{diff}, val_{diff})$ corresponds to the Huffman encoded DC coefficient; and $(run, size_{AC})$ is the Huffman codeword of the AC coefficient. Herein, one bit $b \in \{0, 1\}$ of m_c will be embedded into the low order bit of val_{AC}. The watermarked non-null AC_w coefficient value is thus as $'1010010100b'$.

However, one main constraint has to be checked so as to guarantee an error-free extraction of the message m_c. It imposes that all binary sequences equal to the reference sequence R in the watermarked compressed JPEG image bitstream correspond to watermarked coefficients, otherwise the watermark reader will be desynchronized due to the extraction of bits that are not part of m_c. As in the case of our joint watermarking-JPEG-LS compression (JWJLS) scheme, this kind of problem occurs when a non-watermarkable AC' coefficient (i.e. $(run, size_{AC'}) \neq R$)

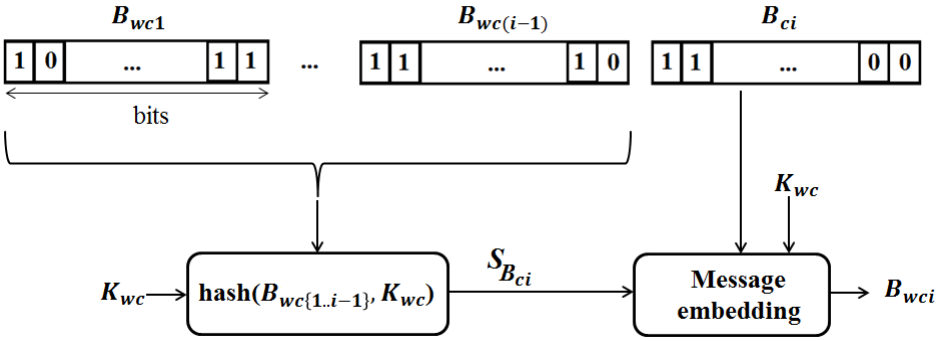


Figure 2.10: Computation and embedding of $S_{B_{ci}}$ into the compressed block B_{ci}, where $B_{wc\{1..i-1\}}$ are the previous compressed-watermarked blocks of N bits, K_{wc} is the secret watermarking key and B_{wci} is the compressed-watermarked block.

is preceded by a binary sequence S, such that $S \| (run, size_{AC'}) = R$. Herein, another problem has to be considered; which is the possibility, or not, of embedding a bit of the message after the formed reference sequence. It is possible to embed one message bit when $size_{AC'}$ is greater or equal to $size_{AC_R}$ (i.e. $size_{AC_R}$ corresponds to the size of AC coefficient adopted by the reference sequence R). Hence, a bit of message m_c is embedded into the AC' coefficient value $val_{AC'}$ at bit position b_i of which corresponds to $b_i = (size_{AC'} - size_{AC_R})$ so as to not desynchronize the watermark reader during the verification stage. Let us take an example of a reference sequence $R = EOB \,\|(\text{size}_{diff_R} = 1, val_{diff_R} = 1)\| \,(run_R = 0, \text{size}_{AC_R} = 1) =' 1010\ 0101\ 00'$ and the Huffman codeword of the non-watermarkable AC' such as $(run = 0, size_{AC'} = 3) =' 100'$ concatenated to $S =' 1010010'$. Herein, the bit-index of $val_{AC'}$ that will be watermarked is $b_i = size_{AC'} - size_{AC_R} = 3 - 1 = 2$, and as $val_{AC'} =' 110'$, the watermarked $val_{AC'_w}$ is $'b10'$, where b is a bit of the message m_c.

Regarding the situation where it is impossible to embed a message bit, this problem occurs when the $S \| (run, size_{AC'}) = R$ but $size_{AC'}$ is lower than $size_{AC_R}$. Herein, the position of the message bit to be embedded falls into a Huffman codeword that should not be modified. Therefore, to circumvent this problem, we proposed to break this sequence, by commuting one bit of S, a bit that should belong to a val_{AC} or a val_{diff}. Let us consider an example where the reference sequence $R = (run_R = 0, \text{size}_{AC_R} = 5) =' 11010'$ and the Huffman codeword of the non-watermarkable AC' such as $(run = 0, size_{AC'} = 2) =' 01'$ concatenated to $S = val_{AC} =' 1101'$. As given, $size_{AC'}$ is lower than $\text{size}_{AC_R} = 5$. Hence, the Least Significant Bit of S should be permuted; S will be of $S' =' 1100'$

Regarding the message extraction stage, like in the JWJLS scheme, the extraction function f_c first identifies the reference sequence R in the watermarked-JPEG compressed image bitstream I_{cw} and reads the message m_c bit value. In order to secure the access to m_c, we suggest encrypting the Reference sequence R and the message m_c with the help of the watermarking key K_{wc}.

2.3.5 JWC Image Reliability Control From The Compressed Domain

As seen in Chapter1, in the healthcare domain, the confidence of practitioners into pieces of data they receive relies on their reliability, that is to say proofs that [56]: i) data have not been modified by an unauthorized user (integrity), and, ii) they belong to the correct patient and are issued from the correct source (authenticity). Non-reliable data should be automatically rejected by medical information system. We further propose an extension of the previous scheme so as to ensure image reliability in the compressed domain. Indeed, it inserts a fragile watermark in both encrypted and compressed domains. Any modification of the image, as well as of its compressed bitstream will alter the embedded messages.

Ensuring the authenticity of the image in both domains is rather simple. One just has to integrate in m_c an authenticity code AC; that is the combination of unique identifiers of the image, the patient and of the image transmitter, that we estimate of about 600 bits by combining the French National Identifier with the DICOM Unique Identifier [60].

Providing the proof of the image integrity is more complex. In the compressed domain, image

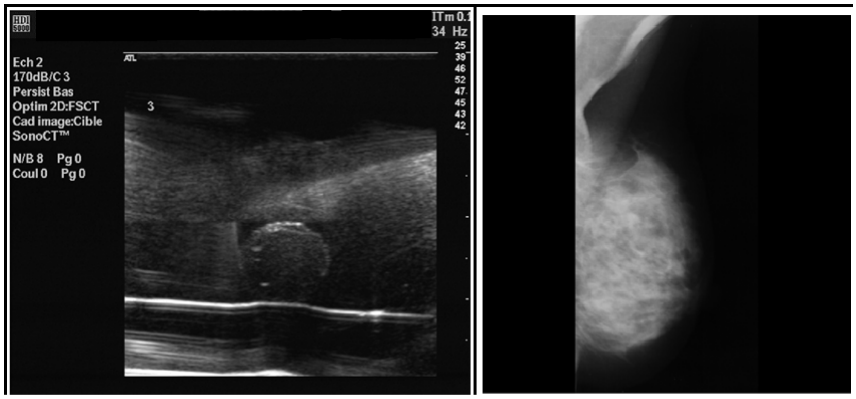


Figure 2.11: Samples of our image test sets: (a) Ultrasound and (b) Mammography images.

integrity can be controlled by means of a cryptographic hash computed from the image compressed bitstream elements that are not modified during the JWC watermarking process. At the verification stage, any differences between the extracted hash and the recomputed one will indicate a loss of data integrity [103]. In this work, we opted for the well-known 256 cryptographic Secure Hash Algorithm (SHA-256). Because such a hash function is very sensitive to data modification and due to the fact the watermarked-compressed bitstream evolves along with the insertion of m_c (see Section 2.3.2 and 2.3.4), we propose an iterative procedure, where the compressed image bitstream is divided into non-overlapping blocks of size N. As depicted in Figure 2.10, the hash $S_{B_{ci}}$ of all previous compressed-watermarked blocks $B_{wc\{1..i-1\}}$ is embedded into the i^{th} compressed block B_{ci}. $S_{B_{ci}}$ is given by:

$$S_{B_{ci}} = \text{hash}\left(B_{wc\{1..i-1\}}, K_{wc}\right) \tag{2.7}$$

where K_{wc} is the watermarking key based on which we secretly select bits of the hash that will be inserted into B_{ci}. Regarding the last compressed block B_{cf}, its integrity is controlled by the hash of the non-watermarked pixels in the same block along with the hash of the previous blocks. More clearly, the final hash $S_{B_{cf}}$ to be inserted into the block B_{cf} corresponds to:

$$S_{B_{Cf}} = hash\left(B_{wc\{1..f-1\}}, K_{wc}\right) \| hash\left(nw\left(B_{cf}\right), K_{wc}\right) \tag{2.8}$$

where: the function $nw\left(B_{cf}\right)$ provides the image bits that are not modified by the watermarking process and, $\|$ is the concatenation operator.

Note that because the SHA-256 "strength" is of 128 bits, if 1 bit of B_{wci} changes then, there is one-in-two chance that $S_{B_{ci}}$ commutes and to detect the bitstream tampering.

At the detection stage, the watermark reader will just have to compute the SHA-256 of the compressed blocks and use the secret watermarking key to determine which bits from the compressed block hash correspond to the embedded message bits. If image data are tampered, compressed blocks will not allow the corrected recovery of m_c, indicating thus the data cannot be used.

2.4 Experimental Results

Experiments were conducted on two sets of medical images of 8-bit depth: 200 Ultra-sound images of 576×690 pixels and over 300 mammography images from a public image database [104] of 1024×1024 pixels. Some samples of our image test set are given in Figure 2.11.

2.4.1 Capacity Rates

The capacity rate corresponds to the size of message that can be embedded into an image, being expressed in bit of message per pixel or per coefficient of image (bpp or bpc, respectively).

The following series of experiments have been performed considering one reference sequence for message embedding in both schemes (i.e. joint watermarking-JPEG-LS compression (JWJLS) and joint watermarking-JPEG compression (JWJPG)). In JWJLS, we opted for $R_1 =' 0X_1 1' =' 001'$,

Table 2.3: Embedding capacities in the compressed domain for two different reference sequences in case of JWJLS scheme.

	Ultrasound		Mammography	
	Average	Variance	Average	Variance
% Regular-mode encoded pixels	80%	13%	36.72%	4%
Embedding capacity for $'0X_11' =' 001'$ (bpp)	0.095	0.034	0.037	0.0094
Embedding capacity for $'0X_21' =' 0001'$ (bpp)	0.05	0.011	0.018	0.0027

and in JWJPG, $R'_1 = (run_{R_1}, size_{AC_{R_1}}) = (0,5) =' 11010'$. These reference sequences are chosen based on their high occurrence probability in order to achieve high embedding capacity. Note also that in all tests, the embedded binary message m_c is uniformly distributed and all watermarkable pixels were used in order to have an idea of the maximum capacity that can be achieved.

2.4.1.1 Capacity of joint watermarking-JPEG-LS compression (JWJLS) scheme

Let us first start with the JWJLS scheme. It is important to notice that the capacity of our scheme depends on the percentage of pixels that have been encoded during JPEG-LS regular-mode. The greater this percentage, the greater is the water-marking capacity. This can be seen in Table 2.3, where capacity values are given based on the use of two different reference sequences $R_1 =' 0X_11' =' 001'$ and $R_2 =' 0X_21' =' 0001'$. It can be noticed that the capacity is impacted by the choice of the reference sequence. In fact, as mentioned in Section 2, the Golomb-Rice MSB-part (the reference sequence in other words) follows a geometric distribution (i.e. the sequence $'001'$ is of greater occurrence than the sequence $'0001'$), hence it is normal that our JWEC reaches better embedding capacity in the compressed domain with R_1 than with R_2. The watermark capacity is also impacted by the percentage of regular-mode encoded pixels: the more percentage is, the more the embedding capacity can be achieved.

2.4.1.2 Capacity of joint watermarking-JPEG compression (JWJPG) scheme

According to the JWJPG scheme, we give in Figure 2.12 the trade-off between the embedding capacity average using two reference sequences of different sizes that correspond to $R'_1 = (run_{R_1}, size_{AC_{R_1}}) = (0,5) =' 11010'$ and $R = EOB \,\|\, (\text{size}_{diff_{R_2}}, val_{diff_{R_2}}) \,\|\, (run_{R_2}, \text{size}_{AC_{R_2}}) = EOB\|(1,1(-1))\|(0,1) =' 1010\ 010\ 1(0)\ 00'$, and the JPEG quality factor (i.e. Q_f), on our ultrasound image dataset. It can be seen that the capacity using R'_2 decreases and the one using R'_1 increases, along with the increase of the Q-factor. This can be explained by the increase of the number of non-null AC coefficients and by next the decrease of the number of zeros in the end of blocks, in other words decrease of the occurrence probability of "EOB" in the image when Q_f is as close to 100.

Moreover, it is important to notice that the capacity of our scheme depends on the used reference bit-sequence itself. As shown in Figure 2.12 and in Table 2.4, the embedding capacity when the reference sequence $= R'_1$ is larger than when the reference sequence $= R'_2$. This is due to the fact that shorter reference sequences are more probable than longer ones.

In both our JWC schemes, the offered embedding capacity is large enough compared to the requirements for verifying the reliability of the image that we estimate about 800 bits (a digital signature provided by the SHA-2: 256 bits and one authenticity code: 600 bits by combining the French National Identifier). As a consequence, image quality can be better preserved than in previous experiments by reducing the number of watermarked pixels. This can also contribute to the security of our scheme by secretly selecting subsets of watermarkable pixels and coefficients (i.e. identified by the reference sequence) for the embedding of m_c. On another hand, one can take advantage of this free space to embed some other pieces of information like some elements related to a security policy and to the patient consent [59] or any other metadata in relation with the patient health [60].

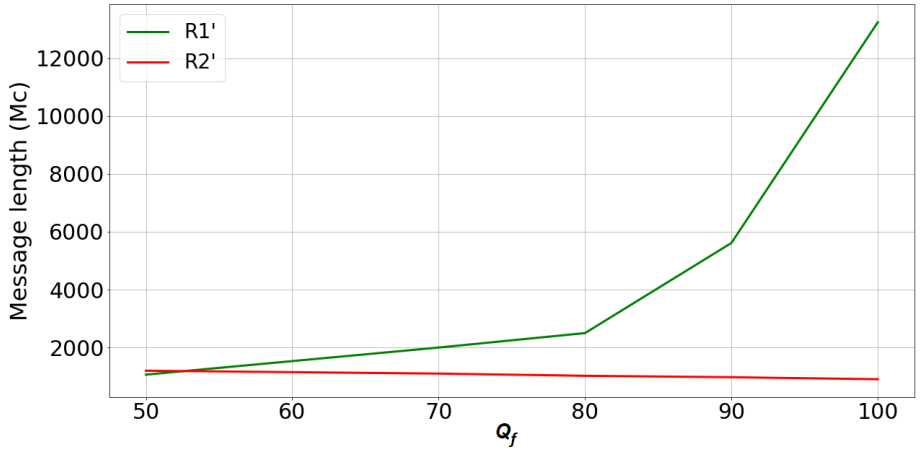


Figure 2.12: Trade-off between the message length and JPEG quality factor (Q_f).

Table 2.4: Embedding capacities in the compressed domain for two different reference-sequences in case of JWJPG scheme (Q-factor = 90).

	Ultrasound		Mammography	
	Average	Variance	Average	Variance
Embedding capacity for R'_1 (bpp)	0.012	6.10^{-4}	0.0087	$4.3.10^{-5}$
Embedding capacity for R'_1 (bpp)	$2.46.10^{-3}$	$7.8.10^{-7}$	$9.61.10^{-4}$	$8.5.10^{-8}$

2.4.2 Image Distortion

Due to the fact that our JWEC algorithm introduces on average the same distortion in each block, one can refer to the Peak-Signal-to-Noise-Ratio (PSNR) so as to evaluate the image distortion. Moreover, the structural similarity (SSIM) between the original image I and its decompressed-watermarked version I_{wd} is considered as better takes into account some human visual system properties.

PSNR is given by:

$$PSNR\left(I, I_{wd}\right) = 10\log_{10}\left(\frac{\left(2^d-1\right)^2}{MSE}\right) \tag{2.9}$$

with

$$\text{MSE}\left(I, I_{wd}\right) = \frac{1}{nm}\sum_{i=1}^{m}\sum_{j=1}^{n}\left[I(i,j) - I_{wd}(i,j)\right]^2 \tag{2.10}$$

where d corresponds to the image depth, and $n \times m$ is the image size.

On its side, SSSIM [105] is defined as:

$$SSIM = l\left(I, I_{wd}\right) \cdot c\left(I, I_{wd}\right) \cdot s\left(I, I_{wd}\right) \tag{2.11}$$

where $l(I, I_{wd})$ is the luminance comparison function, $c(I, I_{wd})$ is contrast comparison and $s(I, I_{wd})$ is structural comparison.

The compromise between embedding capacity and PSNR of JWJLS and of JWJPG schemes are given for both ultrasound and mammography image datasets in Figure 2.13, respectively. Obviously, quality distortion varies with the embedding capacity in both schemes. This compromise evolves depending on the image type (or modality) and on its content (e.g. texture of the image, flat areas). As given in Figures 2.13 and , in average, the obtained PSNR values for ultrasound and mammography images are of 52.44 dB and 55.53 db with JWJLS, and of 41 dB and of 44.23 dB with JWJPG, respectively. As it can be also noticed, the distortion introduced by the JWJPG is

Table 2.5: PSNR and compression ratio versus JPEG Q-factor for Ultrasound images.

Q-factor	Compression Ratio	$PSNR_c$	$PSNR_{wc}$
50	12:1	35.3054	34.16
70	10:1	37.7	36.6
80	8:1	40.03	37.5
90	5:1	44.18	41

larger than the one introduced by JWJLS. This is due to the fact that JPEG is a lossy compression standard and by applying it, already some information loss occurs. As the JPEG image quality is managed via the quality factors, we give in table 2.5 the trade-off between the Q_f, the compression ratio, $PSNR_c$ (i.e. the PSNR between the original and the decompressed JPEG images) and $PSNR_{wc}$ (i.e. the PSNR between the original and the watermarked-decompressed JPEG images). These results show that PSNR is not directly related to the watermarking process or in other words to the length of the message embedded into the image but strongly affected by the quality factor Q_f. In fact, as said before, the quality factor manages the trade-off between image quality and compression rate. Hence, the lower the quality factor is, the more loss of information occurs and by next the higher image quality distortion is.

To sum up, in both JWJLS and JWJPG schemes, the degree of the introduced distortion is small enough and does not endanger the diagnostic value of medical images. One can refer to the study on the impact of lossy image compression on medical images in [106], where it is reported that image distortion should be maintained in the range of 40 and 50 dB without endangering the diagnostic quality.

Resulting SSIM values, being greater that 0.98 also confirm the good quality of watermarked images [105]. As also illustrated in Figure 2.15, where we give two examples of JWC protected images along with their original versions, it can be noticed that we are so far from introducing a visible distortion in the image using both JWJLS and JWJPG schemes..

2.4.3 Algorithm Complexity, Compression Rate & Performance Comparison

As stated in Section 2.3.2, our scheme has sometimes to modify compressed image bitstream so as to ensure the error-free extraction of the message. This can be seen in Figure 2.16, where we provide JWJLS and JWJPG files' size evolution depending on the total embedding capacity in the case of ultrasound images. File size increased linearly with the capacity. As also shown, this increase is very small.

In terms of complexity, we verified that our scheme is about one and half time slower than only JPEG-LS compressing the image. The reason of this difference is obviously caused by the embedding of m_c in the compressed domain. However, our JWC gives access to watermarking-based security services in compressed domain without decompressing the image bitstream, even partially.

To the best of our knowledge, our joint watermarking-compression scheme is the first of its kind. Let us recall that it gives access to different security services (e.g. authenticity, integrity) in the compressed domain without having to decompress the compressed image bitstream, even partially nor counting any additional parameters for partial decompression. Moreover, watermarking is fully transparent to compression in both JWJLS and JWJPG schemes and can be hence decompressed by means of the common JPEG-LS or JPEG decoder, respectively. Nevertheless, as the proposed scheme induces an irreversible image information loss, we propose to compare it to the scheme presented in [98]. The authors of [98] embed thee message during the JPEG-LS encoding of the image and authentication is only carried out in the non-compressed domain unlike our proposal. Furthermore, our JWJLS scheme is the first that combines watermarking with JPEG-LS in its fully lossless mode contrarily to [98] that works with JPEG-LS in its near-lossless (i.e. lossy) mode.

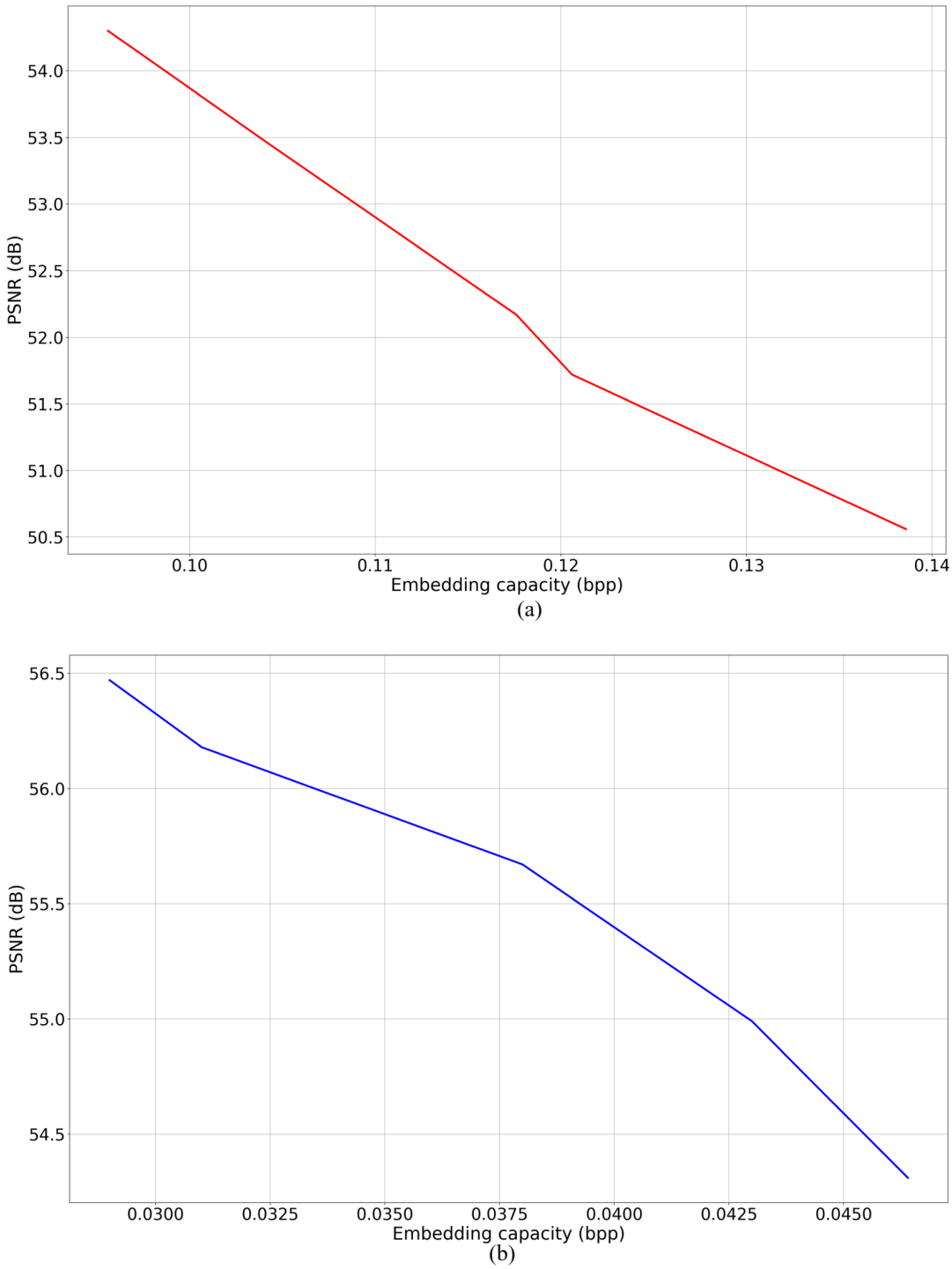
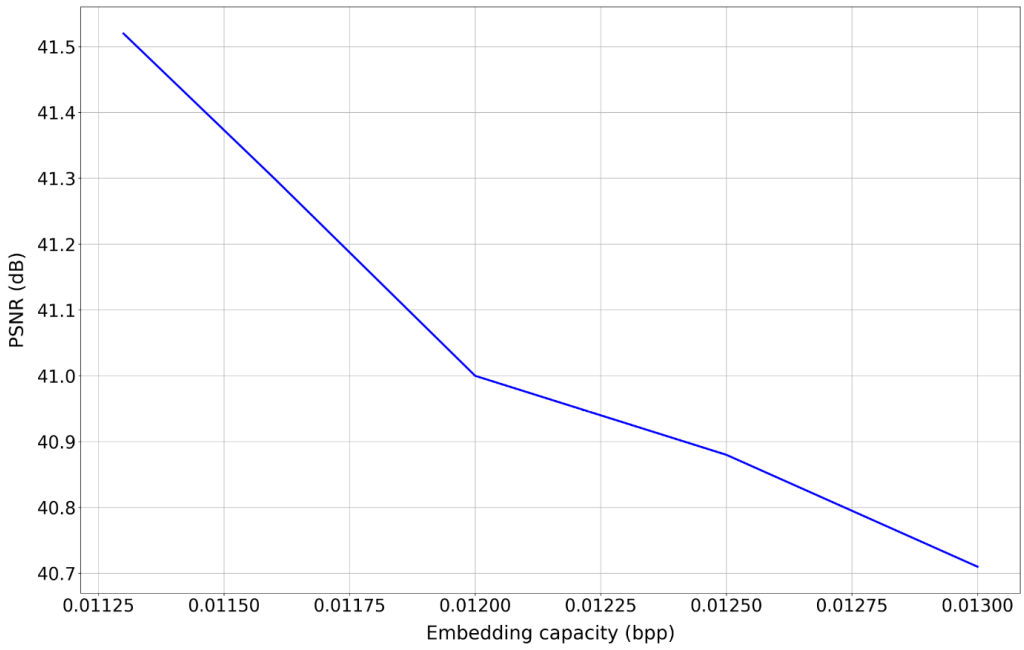


Figure 2.13: PSNR vs. total embedding capacity in the compressed domain in case of JWJLS scheme for: a) ultrasound images, and b) mammography images.

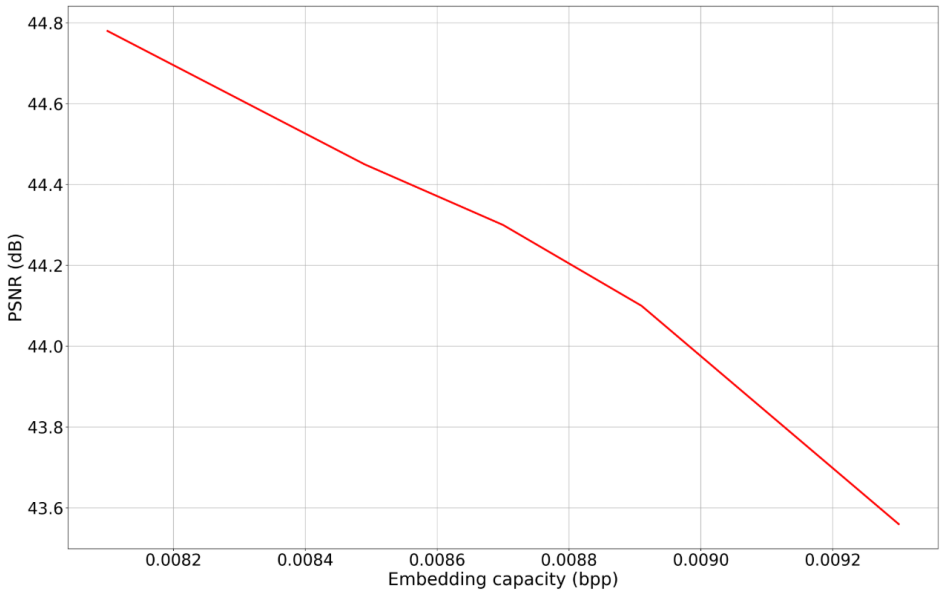
2.5 Conclusion

In this chapter, we have focused on the protection of compressed medical images by means of watermarking. As exposed, improving medical data security is a top healthcare priority. On the other hand, medical images constitute huge volumes of data. That is why, they are most of the time stored in a compressed form so as to reduce costs of storage and transmission.

Considering these requirements, we proposed two DICOM-compliant joint watermarking-compression systems. The first is based on the lossless JPEG-LS compression standard and the



(a)



(b)

Figure 2.14: PSNR vs. total embedding capacity in the compressed domain in case of JWJPG scheme for: a) ultrasound images, and b) mammography images.

second on the lossy JPEG compression. As discussed above, the proposed scheme originality stands of the fact that it allows verifying image reliability directly from the compressed image bitstream without decompressing it, even partially. Indeed, JPEG-LS or JPEG is combined to the bit-substitution watermarking modulation in a single operation so that message extraction can be performed in the image compressed domain without decompressing the compressed bitstream nor computing any additional parameter that can be used for partial image decompression. To do this, we take advantage of the variable-length encoders; i.e. Huffman and Golomb-Rice, used in JPEG and JPEG-LS, respectively. These encoders replace a fixed-length input symbol by the corresponding variable-length prefix codeword. Our basic idea thus is to fix a reference sequence among these variable-length codewords and to embed a message bit into the image pixels or coefficients their prefix codewords correspond to the reference sequence. This later has to be known

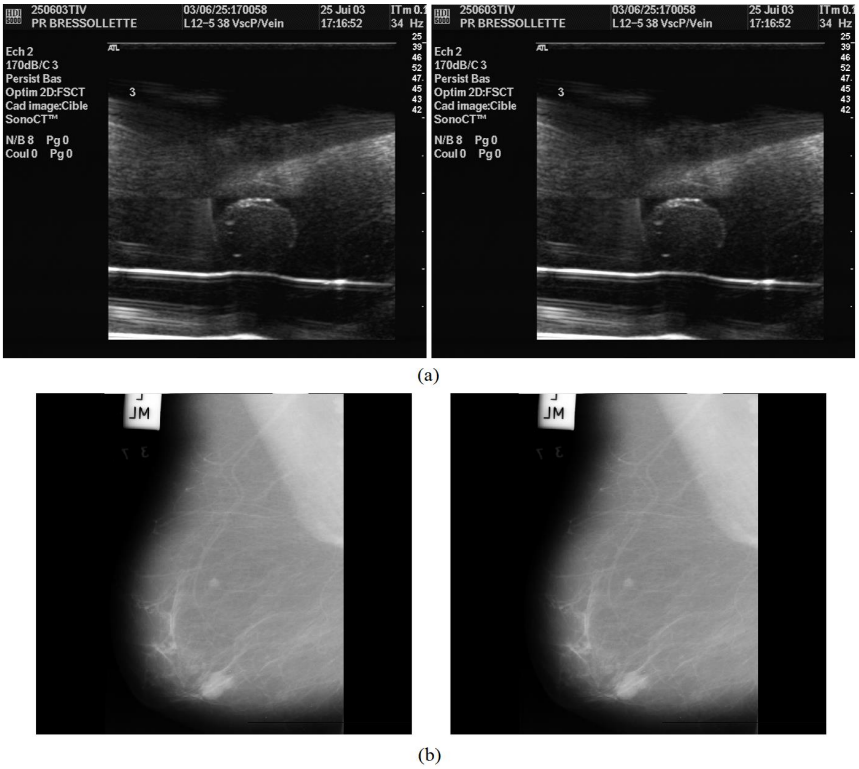


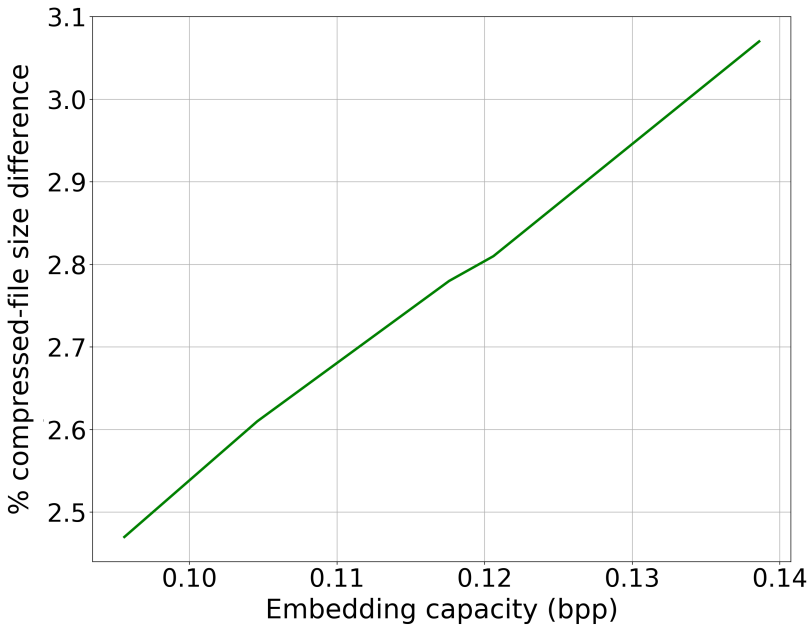
Figure 2.15: Original images (left) alongside with their decompressed-watermarked images (right) for both JWJLS (a) and JWJPG (b) schemes.

from the watermark reader so as to extract the message, as it labialize the positions of the message bits in the image watermarked-compressed bitstream.

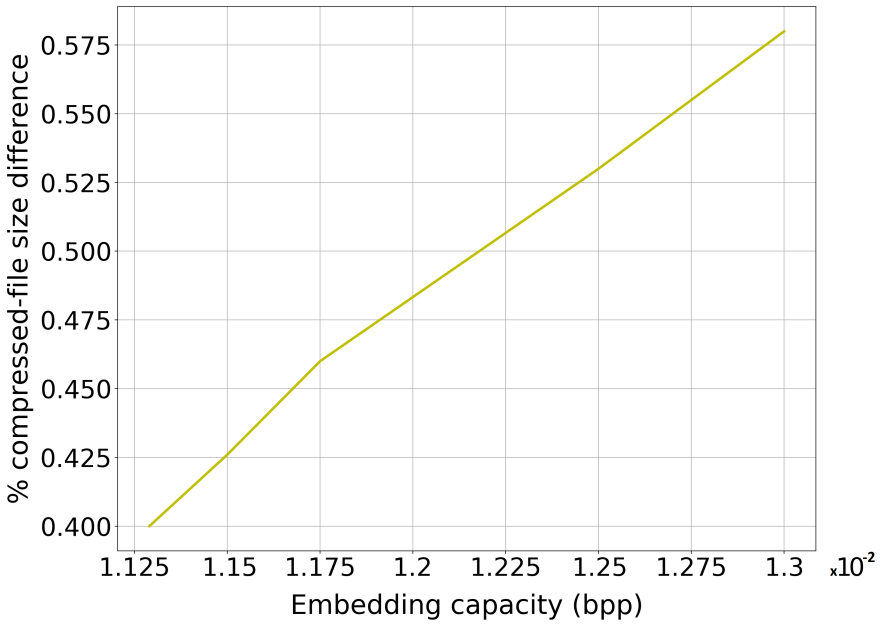
In order to guarantee normal image decompression and an error-free extraction of the watermark, some issues should be addressed during its embedding. The first issue is to identify where and how the compressed bitstream can be modified without changing its semantic so as to be able to reconstruct the image from its compressed bitstream. The second stands on the capability to embed a message without causing high quality degradation as we work on medical images.

By addressing these constraints, as discussed in Section 2.4, the offered watermarking capacity in both schemes is large enough to give access to different security services ranging from authenticity control to traceability. Regarding the watermarked image quality, it is small and does not endanger the diagnostic value of medical images.

As seen, being able to access to watermarking-based security services from the compressed image without decompressing it nor computing additional parameters is of major concern. Nevertheless, such a solution can be improved, essentially when we consider most of medical images are encrypted before being stored or distributed. Thus, it becomes desirable to access to watermarking-based security services from both encrypted and compressed domains. This issue will be dealt with in the following chapter. Furthermore, we will theoretically evaluate the performance of our scheme in terms of embedding capacity in both compressed and encrypted domains and image quality distortion, and we will analyze the security of the proposed scheme.



(a)



(b)

Figure 2.16: % file size evolution vs. embedding capacity on ultrasound image dataset in both JWJLS (a) and JWJPG (b) schemes.

Chapter 3

Joint Watermarking-Encryption-JPEG-LS for Medical Image Reliability Control in Encrypted and Compressed Domains

As previously concluded, there is a need to combine watermarking with confidentiality control tools in order to ensure a better protection of medical images. Regarding medical image confidentiality, one has to consider the Digital Imaging and Communications in Medicine (DICOM) standard that has been developed in order to facilitate the transmission handling of medical images [5]. DICOM also takes into account different security aspects in its part 15. To ensure image confidentiality, while being DICOM compliant, the triple DES or AES encryption can be used. Image integrity can be provided with the help of the DSA digital signature. Such a signature will be located in the DICOM file header of the image. It is important to notice that the kind of protection these security mechanisms offer is *"a priori"* [49], in the sense that once an image is decrypted or its digital signature is lost or deleted, the image is no longer protected. Here comes the interest of the *"a posteriori"* protection watermarking ensures [4]. Let us recall that watermarking leaves access to the data while maintaining them protected by the message. It has been recently shown that there is a great interest to combine watermarking with encryption so as to simultaneously achieve an *"a priori"*/ *"a posteriori"* protection [107]. By doing so, we refer to crypto-watermarking techniques the main purpose of which is to provide watermarking-based security services from encrypted data [103].

As seen in Chapter 1 and 2, the deployment of such crypto-watermarking protection in the healthcare domain needs to take into account its specificities, and in particular the fact that medical images are most of time stored in a compressed form so as to decrease storage costs and improve transmission efficiency. A hospital generates more than 27,000 Terabytes per year [17]. As a consequence, there is a need for crypto-watermarking-compression solutions. In this chapter, we extend our joint compression watermarking scheme into a joint encryption-compression-watermarking tool. This solution is the first of its kind. Its originality is twofold. First, it allows accessing to watermarking-based security services from both encrypted and compressed image bitstreams without having to decrypt or decompress them, even partially. In second, it is DICOM-compliant. If the protection we offer is proprietary, image decryption, decompression and watermark extraction processes can be conducted separately.

This chapter is organized as follows. First, we come back on the state of the art of crypto-watermarking and watermarking of compressed and non-compressed images in order to refine the originality of our proposal. Then, we recall the main principles of the AES cryptosystem and the JPEG-LS compression standard, before detailing how they can be combined with substitutive watermarking so as to provide a general joint watermarking-encryption-compression technique.Finally, experimental results and comparison with all the above methods are given and discussed.

Notice that the work we present in this chapter has been published in [8] and has been accepted for submission in IEEE Transactions on Information Forensics and Security (IEEE-TIFS).

3.1 Combining Watermarking, Encryption and Compression

Watermarking, encryption and compression have been considered in different ways but, as we will see, rarely all three together. We propose to consider four main categories of techniques: encryption & watermarking, watermarking & compression, encryption & compression, and watermarking & compression & encryption. We then sub-classify the corresponding methods from a technical point of view after having recalled the main applications or security services these methods have been proposed for:

a) *Encryption-watermarking methods* − They combine encryption and watermarking mechanisms and have been proposed with as main concern copyright protection in Video on Demand frameworks. As suggested in [108], they can however provide integrity and traceability security services. They can be differentiated depending on the domain from where the watermark or the message can be accessed:

- *In the clear domain* [103, 109, 110] − such methods require to decrypt the image before accessing to the watermark.
- *In the encrypted domain* [111, 112, 113] − these schemes jointly conduct encryption and watermarking processes or not. In the latter case, the watermark has to be removed so as to make possible image decryption.
- *In both encrypted and clear domains* [107, 114, 115] − these schemes usually take advantage of partial , invariant or homomorphic encryption.

It is important to notice that methods of the two last categories are not DICOM compliant. They need to reorganize the image bitstream or to parse the encrypted bitstream in a specific fashion so as to give access to the watermark. Beyond, none of them have been combined with image compression. As we will see, the solution we propose makes possible to extract a message from both encrypted and compressed bitstreams without having to reorganize or parse (i.e. decrypt or decompress) them, even partially.

b) *Watermarking-compression schemes* − Again, those ones mainly focus on the copyright protection of data. They can be discriminated depending on the compression-watermarking processual sequence:

- *Watermarking before compression* [90, 92] − the message is embedded before image compression and is only available in the decompressed domain. The watermark is usually designed to be more robust to a lossy compression algorithm than another by inserting the message into the appropriate transform domain (i.e. Discrete Cosine Transform (DCT) – JPEG; Discrete Wavelet Transform (DWT) – JPEG2K).
- *Compression followed by watermarking* [93, 116, 117, 118]− This strategy consists in embedding the message directly into the compressed bitstream or after a partial decompression of it. Watermark extraction is usually conducted in the same domain as the embedding process.
- *Joint watermarking and compression* [6, 98] − where the message is embedded during the compression process. The main benefit of such a strategy is that compression is no longer considered as an attack from the watermark robustness point of view.

With the first sub-class methods, message extraction is always conducted in the decompressed domain. That is not the case of the two last sub-classes, where the watermark reader has to parse the image bitstream or to compare the watermarked image to the original one for message extraction. As we will see, that is not the case of the solution we propose.

c) *Encryption-compression solutions* − one can distinguish three different classes:

- *Compression before encryption* [119, 120] − these schemes aim at improving the security of compressed images during their transmission over bandwidth-constrained channels.

- *Encryption before compression* [121, 122, 123] − such methods are worked-with when the storage service provider wants to gain memory space without knowing the encryption key.
- *Joint encryption and compression* [124, 125] − where both mechanisms are conducted simultaneously to gain in complexity, security and storage space.

As it can be seen, the only purpose of these schemes is to ensure data confidentiality.

d) *Encryption-compression-watermarking techniques* − to the best of our knowledge, two methods focusing on watermarking, encryption and compression, have been proposed. The technique in [95] embeds a watermark into an homomorphically encrypted JPEG2000 compressed image. Homomorphic encryption allows conducting some operations onto encrypted data with the guaranty that the decrypted result equals the one of the equivalent calculation conducted onto unencrypted data. In [95], authors use a stream-cipher with homomorphic properties so as to not expand the size of compressed data. They embed a pseudo-random sequence into the less significant bit planes of the encrypted middle wavelet coefficients of the image in order to ensure the watermark invisibility. The watermark detection is conducted from the decrypted JPEG2000 compressed image bitstream using a correlation detector along with the *a priori* known pseudo-random sequence. To make these operations possible the compressed image bitstream has to be pre-processed before being encrypted. This pre-process reorganizes and modifies the compressed wavelet coefficients in order to: i) facilitate the identification of watermarkable coefficients in the encrypted image bitstream, and ii) ensure the watermarking process does not modify elements of the compressed bitstream that will make impossible image decompression. Authors of [126] proposed a reversible watermarking scheme in the encrypted domain for labeling ciphered images for their storage in the Cloud. In this scheme, the JPEG-compressed image is rearranged, resized and re-encoded before being encrypted. The JPEG image header is encrypted based on a stream cipher algorithm (e.g. RC4) while Huffman encoded coefficients are encrypted using a permutation technique. After the image encryption, secret data are embedded into the Huffman table present in JPEG image header and into the quantized AC coefficients after their partial decompression using histogram shifting, a reversible watermarking modulation [85]. To extract the message, the encrypted-JPEG compressed bitstream should be partially decompressed. In order to allow image decryption and decompression, the modified Huffman table has to be replaced by the table predefined in the JPEG standard and permuted-watermarked AC coefficients have to be un-watermarked and de-permuted. As it can be seen, beyond these two works, there is no joint watermarking-compression-encryption algorithm, contrarily to the other above classes. If [95] and [126]algorithms are separable, they also require compressed data reorganization before and after data watermarking in the encrypted domain. Without such data reorganization, decryption and decompression processes are impossible. As a consequence, these schemes need proprietary decoding (decryption and decompression cannot be conducted regularly) and are not compliant with the DICOM standard. It can also be noticed that watermarking based services are only available in the encrypted domain for [126] and in the decrypted domain for [95].

In this chapter, we propose the first joint watermarking-encryption-compression (JWEC) scheme. Its originality is twofold. First, it allows accessing to watermarking-based security services from both encrypted and compressed image bitstreams without having to decrypt or decompress them, even partially. Second, it combines bit-substitution watermarking with JPEG-LS and the AES block cipher algorithm in its cipher block chaining (CBC) mode [127], in a single operation performed on the entire image (i.e. the image is entirely compressed and encrypted). It is important to notice that with our scheme, it is possible to decipher and decompress the image with the common AES and JPEG-LS algorithms. More clearly, decryption, decompression as well as message extraction processes are conducted independently without having to be modified or adapted, unlike [95] and [126]. With such a capability, our scheme is DICOM compliant. Furthermore and as we will demonstrate, our solution makes possible to trace images and control their reliability (i.e. authenticity and integrity) directly from both encrypted and compressed domains. Note also that the proposed scheme saves the computational complexity as it does not require decryption and decompression to verify the image reliability in the encrypted domain and the compressed one, respectively.

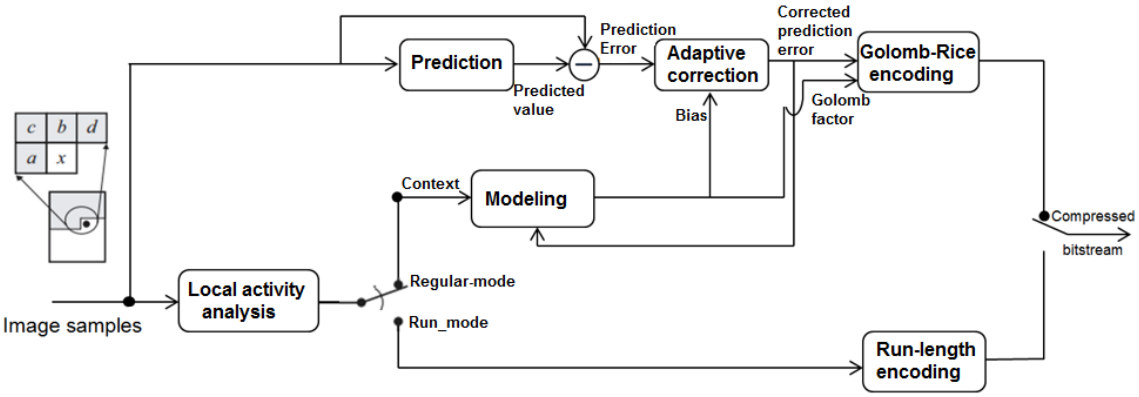


Figure 3.1: JPEG-LS general scheme.

3.2 AES and JPEG-LS in brief

As our scheme is combines watermarking with JPEG-LS and the AES cryptosystem, let us recall some of their basic principles our solution relies on.

3.2.1 Advanced Encryption Standard

AES is a symmetric key cryptosystem that encrypts blocks of data, typically 128-bit blocks (see Chapter 1, Section 1.2.3.1 about block cipher algorithms). Furthermore, AES works different modes of operation so as to adapt the algorithm for a specific application or to enhance its efficiency. To meet these requirements, the NIST (National Institute of Standards and Technology), in 2001, standardized five AES-modes of operations: ECB (Electronic Code Book), CBC (Cipher Block Chaining), CFB (Cipher FeedBack), OFB (Output FeedBack) and CTR (Counter).

In this work, we opted for the Advanced Encryption Standard (AES) in its cipher block chaining (CBC) mode in order to make our system compliant with the DICOM standard. The CBC mode refers to the manner in which plain-text blocks are processed at the encryption and decryption stages. Technically, each block of plain-text is XORed with the previous cipher-text block before being encrypted. More clearly, let us denote the encrypted version of the i^{th} block B_i as B_i^e. With the CBC mode, B_i^e is given by:

$$B_i^e = AES\left(B_i \oplus B_{i-1}^e, K_e\right) \tag{3.1}$$

where B_{i-1}^e is the previous encrypted block and K_e is the encryption key.

3.2.2 JPEG-LS compression

Our system hides a message that will be available in the image JPEG-LS bitstream without having to parse it, even partially. In this section, we just recall the core elements of JPEG-LS compression technique we already presented in Chapter 2 (Section 2.2).

JPEG-LS is a lossless image compression standard [99]. DICOM allows its use in order to compress medical images. JPEG-LS is based on the Low Complexity Lossless Compression for Images (LOCO-I) algorithm [100]. This one relies on a pixel prediction based on a local contextual statistical model.

The main steps of JPEG-LS are described in Figure 3.1. As it can be seen, an image is sequentially processed pixel by pixel. JPEG-LS encodes one pixel x accordingly to two modes: the regular-mode or the run-mode. To decide in-between these modes, JPEG-LS first analyses the local activity around x, based on its immediate causal neighborhood constituted of previously encoded pixels (i.e. pixels a, b, c and d in Figure 3.1), by calculating its corresponding local gradients (see equation ()).

If all local gradients are null then, the run-mode is activated, otherwise the regular-mode is chosen. JPEG-LS run-mode is based on run-length encoding (RLE). It encodes the pixel value followed by the number of times this one is repeated.

During the regular-mode, the causal neighborhood of x is used to estimate its gray value based on the edge-detecting predictor (see equation in ()). By next, JPEG-LS assigns to x a context

Q evaluated from its associated local gradients. This context value will subsequently be used to correct the prediction-error bias as well as to parameterize the Golomb-Rice encoder. More clearly, because the edge-detecting predictor only provides integer values, a prediction bias exists. This bias induces a translation of the prediction-error $e = x - \hat{x}$. In order to compensate such a systematic offset, a context-dependent term is computed so as to shift back the prediction-error for each pixel. For a given pixel x and its context Q, this term corresponds to the prediction-error mean $\overline{e}$ of the pixels of the same context Q as x, i.e. of same Q value.

The next step in the regular-mode consists in encoding the corrected prediction-errors with the help of the Golomb-Rice encoder (GRE). Due to the fact, GRE only encodes non-negative integer values, the corrected prediction-errors are modified into "mapped prediction-errors" $\tilde{e}$ such that the positive and negative values are mapped to even and odd positives integers of $\tilde{e}$, respectively. So the parity of $\tilde{e}$, or equivalently its least significant bit indicates the corrected prediction-error sign. Then, the resulting mapped-prediction errors are GRE encoded. Basically, GRE encodes a positive integer p into two parts that are concatenated. These two parts of p will be referred in the sequel as: the *Most Significant Bit part* (p_{MSB}) and the *Least Significant Bit part* (p_{LSB}). Notice that these two parts are of great importance in the proposed scheme. The MSB-part corresponds to the quotient of the Euclidean division of p by 2^k, where k is the non-negative Golomb-Rice factor. This quotient is unary encoded, i.e. by a sequence of '0' ended by '1', the number of '0' being the quotient value. This part is then concatenated to the binary encoding of the division reminder, i.e. an integer value encoded on k bits. Note that the value of Golomb-Rice factor, i.e. k, is also pixel's context-dependent.

JPEG-LS decompression is accomplished by performing the same steps as for compression (see Chapter 2 Section 2.2 for more details).

3.3 Proposed Joint Watermarking-Encryption-Compression (JWEC) Scheme

3.3.1 System architecture and basic principles

The purpose of our system is to ensure the confidentiality of an image I through encryption while giving access to watermarking-based security services in both encrypted and compressed domains. As shown in Figure 3.2, it relies on two main procedures: protection and verification. At the protection stage (see Figure 3.2 (a)), substitutive watermarking, JPEG-LS and AES in its CBC mode are jointly conducted so as to protect an image I. This procedure, we name as JWEC, for joint watermarking-encryption-compression, allows the insertion of two messages m_e and m_c that will be available or readable from the encrypted image bitstream I_{cwe} and from the compressed image bitstream I_{cw}, respectively, without having to parse them , even partially. Both messages contain security attributes that assess the image reliability (see Sub-Section 4.2.3). The embedding and the extraction of messages m_e and m_c depend on two distinct watermarking keys: K_{wc} in the compressed domain and K_{we} in the encrypted domain. On its side, AES is parameterized with the encryption key K_e.

At the verification stage (Figure 3.2 (b)), the image reliability can be verified by accessing to m_e or m_c with the help of two watermarking functions f_e and f_c, respectively. Notice that a JWEC protected image I_{cwe} can be decrypted and decompressed in a regular way. AES decryption and JPEG-LS decompression can be conducted independently with no modification of their algorithms. More clearly, our JWEC scheme does not need some proprietary decryption or decompression procedures. Watermarking is completely transparent to JPEG-LS and to AES. This is of goal interest for DICOM systems that are not watermarking-compliant.

In the sequel and for sake of simplicity, we first recall how to combine JPEG-LS with the least significant bit substitution watermarking modulation (see Chapter 2, Section 2.3.3). We then detail how to take into account AES encryption so as to build our JWEC scheme and the way it can be used to protect the reliability of an image in the encrypted domain.

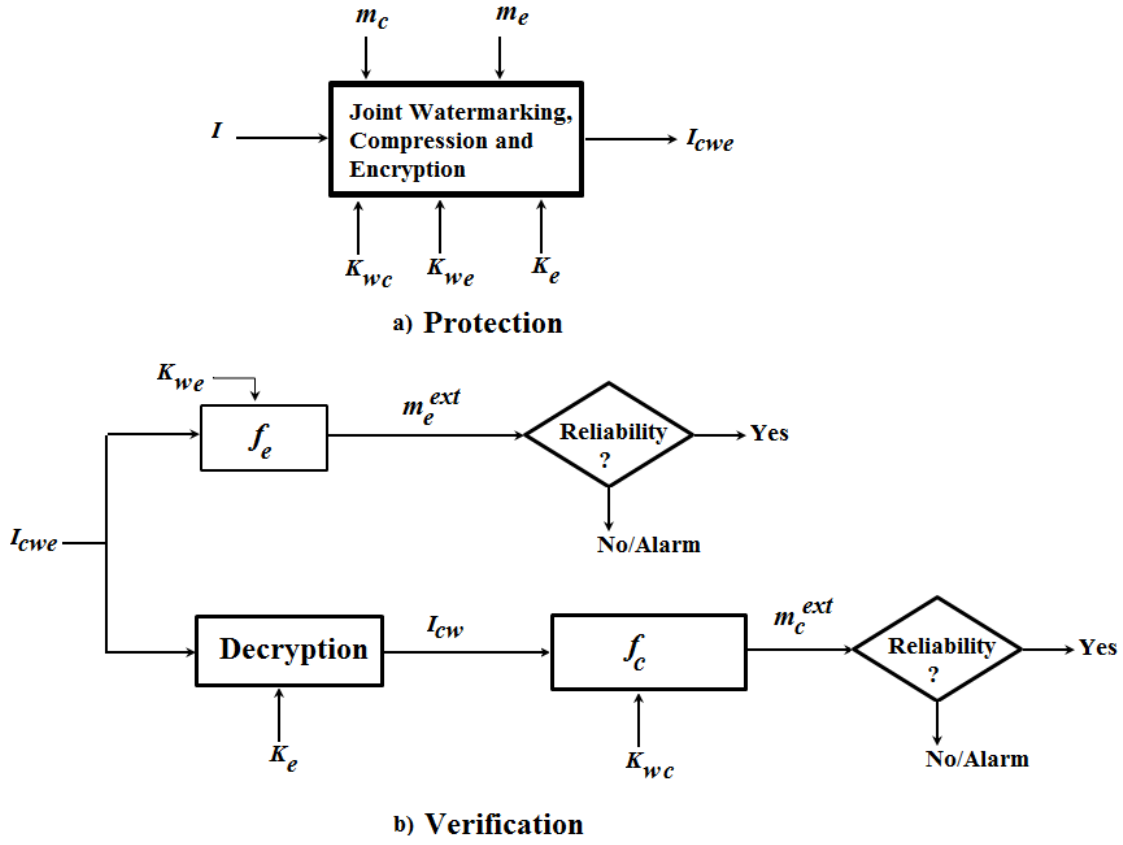


Figure 3.2: General architecture of the proposed system where: I, I_{cwe}, I_{cw}, I_{dw}, K_{wc}, K_{we}, K_e, m_c, m_e, m_c^{ext}, m_e^{ext} correspond to the original, the watermarked-encrypted-compressed, the watermarked-compressed and the decompressed images, the watermarking keys in compressed and encrypted domains, the encryption key, the embedded and extracted messages from compressed and encrypted domains, respectively.

3.3.2 Combination of watermarking, encryption and JPEG-LS

3.3.2.1 Embedding of m_c message available in the compressed domain - Joint JPEG-LS-Watermarking

As exposed in Section 3.2.2, JPEG-LS sequentially encodes pixels depending on their local context and accordingly to two encoding modes: the regular and the run modes. To roughly sum up, the run-mode is selected to compress sequences of pixels of same gray value while regular-mode is used to encode the others. We recall that this latter relies on a prediction and context-based coding of pixel prediction-errors using the Golomb-Rice encoder. As shown in Section 2.2.1, it is rather impossible to clearly identify the bits of one pixel in the bitstream, but there is a high probability that a sequence of bits $'0X1'$, where X is a sequence of '0', corresponds to a regular-mode encoded pixel. As already said in Chapter 2, Section 2.3.3, we take advantage of this property so as to make the message m_c readable from the image compressed bitstream without decompressing it. Our basic idea is to embed m_c in pixels the mapped prediction-errors of which have a Golomb-Rice MSB-part encoded such that $\tilde{e}_{MSB} =' 0X1'$, $'0X1'$ is called as "the reference sequence", by modifying its higher order bit of Golomb-Rice LSB-part using watermarking bit-substitution modulation (see Section 2.2.1). The used reference sequence has to be known from the watermark reader in order to extract the message m_c.

In order to guarantee the error-free extraction of m_c, two main constraints have to be considered : i) the possibility to embed one bit after each reference sequence $'0X1'$; ii) a sequence of bits equals to $'0X1'$ should correspond to a watermarked pixel in the watermarked-compressed image bitstream. As we have detailed and showed how to address these constraints in Chapter 2 Section 2.3.3, let us briefly remember what are these issues. Regarding the former issue, a watermarkable pixel should hold an LSB-part $\tilde{e}_{LSB}$, i.e. its Golomb-Rice factor should be non-null, otherwise

message insertion is not possible. In addition, due to the fact that the last bit of $\tilde{e}_{LSB}$ informs JPEG-LS decoder of the prediction-error sign, k should be strictly greater than 1. To overcome this issue; i.e. when $\tilde{e}_{MSB} =' 0X1'$ but $k \leq 1$, we propose to modify the pixel $\tilde{e}_{MSB}$ part: it suppresses one '0' from $\tilde{e}_{MSB}$ (i.e. $'0X1'$ is changed into $'X1'$). The pixel is thus turned into a non-watermarkable pixel. The second issue imposes that all reference sequences $'0X1'$ in the protected compressed bitstream correspond to the watermarked pixels. Otherwise, the watermark reader will be desynchronized by extracting bit values that are not part of m_c. This kind of problem occurs when a non-watermarkable pixel (i.e. $\tilde{e}_{MSB} \neq' 0X1'$) is preceded by a sequence of bits $'0S$, where S is a sequence of '0', such that $'0S'\|\tilde{e}_{MSB} =' 0X1'$, '||' being the concatenation operator. To solve it, our scheme considers this pixel as watermarkable and one bit of m_c is embedded into its $\tilde{e}_{LSB}$. As we will see in the experimental section, these solutions can attribute to the increase in the watermark capacity and the image distortion.

Regarding the watermarking extraction function f_c (see Figure 3.2), it has just to identify the reference sequence $\tilde{e}_{MSB} =' 0X1'$ in the watermarked-compressed image bitstream I_{cw} and to read the immediate following bit value to extract m_c. In order to ensure the security of our scheme, that is to say securing the access to m_c, we suggest secretly selecting watermarkable pixels with the help of watermarking key K_{wc}. It can be noticed that, after decompression, as long as the protected image has not been modified, one just has to JPEG-LS re-compressed it to retrieve the message m_c in the compressed domain.

Theoretical capacity and image quality distortion

The embedding capacity of our JWEC in the compressed domain, it means the size in bits of m_c, depends on the probability of occurrence of the reference sequence $'0X1$. It has been shown in [128] that unary coding is an optimally efficient encoding for the following discrete probability distribution:

$$\mathrm{P}\left('0X1'\right) = 2^{-(|X|+2)} \tag{3.2}$$

where $|X|$ is the number of '0' the sequence X contains.

In our case, as all pixels are not necessarily encoded in the regular-mode, the optimal embedding capacity in bits of message per pixel of image (bpp) of our JWEC in the compressed domain is given by:

$$C_{m_c} = 2^{-(|X|+2)} R \tag{3.3}$$

where R is the ratio of regular-mode encoded pixels in the image.

One can also estimate the image distortion this joint watermarking-JPEG-LS solution introduces. According to [129], for a given reference sequence $'0X1'$, the Golomb-Rice factor k follows a geometric distribution of parameter p. It is thus possible to compute the mean squared error (MSE) between an image I of $n \times m$ pixels, and its watermarked-decompressed counterpart I_{wd}:

$$\begin{array}{c} \mathrm{MSE}_c\left(I, I_{wd}\right) = \frac{1}{nm} \sum_{i=1}^{m} \sum_{j=1}^{n} \left[I(i,j) - I_{wd}(i,j)\right]^2 = \\ \frac{1}{nm} \sum_{k=0}^{k\,\max -1} 2^{2k} N_k \end{array} \tag{3.4}$$

where $k_{\max} = 8$ for 8-bit depth image; N_k is the number of pixels the Golomb-Rice factor of which equals k. When C_{m_c} (given by (6)) of pixels are watermarkable, $N_k = C_{m_c}\,\mathrm{P}(k)$.

In that case, the average distortion is such that:

$$\begin{aligned} \overline{MSE_c\left(I, I_{wd}\right)} &= E\left[\frac{1}{nm} \sum_{k=0}^{k_{\max}-1} 2^{2k} N_k\right] \\ &= \frac{1}{p} \frac{C_{m_c}}{nm} \frac{\left(4^{k_{\max}-1}\right)}{3} \end{aligned} \tag{3.5}$$

where $E[x]$ is the expected value of x. Note that the expected value of a geometrically distributed value is $E[k] = 1/p$.

3.3.2.2 Embedding of m_e message available in the encrypted domain - proposed JWEC scheme

The embedding of the message m_e works similarly as the previous one. The watermarked-compressed bitstream is modified or, equivalently, again watermarked so as to make m_e available

from the watermarked-encrypted-compressed image bitstream. Our idea is to introduce a distortion in the compressed image bitstream in order one can extract m_e from the AES encrypted data using watermarking extraction function f_e.

Let us consider B_{wci} is the i^{th} block of consecutive bits of a watermarked-JPEG-LS compressed bitstream. In the case of AES, such a block is of 128-bit long (see Section 3.2.1). m_e will be made available in the encrypted domain by modifying B_{wci} into B^{we}_{wci} such that

$$f_e\left(B^{we}_{wci}, K_{we}\right) = f_e\left(AES\left(B^{w}_{wci}, K_e\right), K_{we}\right) = m_e \tag{3.6}$$

where K_e is the AES encryption key, B^{we}_{wci} is the AES encrypted version of B^{w}_{wci} and K_{we} is the secret watermarking key in the encrypted domain.

One simple solution to give access to m_e without having to parse the watermarked-encrypted-compressed bitstream of a JWEC protected image consists in secretly selecting one bit of each watermarked-encrypted-compressed block B^{we}_{wci} that will correspond to one bit of m_e. More clearly, B_{wci} will be modified into B^{w}_{wci} such that the secretly selected bit of B^{we}_{wci} equals one bit of m_e. The watermark extraction function f_e of our JWEC system is thus defined as:

$$f_e\left(B^{we}_{wci}, K_{we}\right) = AES\left(B^{w}_{wci}, K_e\right)_j \tag{3.7}$$

where $AES(.)_j$ corresponds to the j^{th} bit of the AES block. The choice of the value of j for each AES encrypted block depends on the secret watermarking key K_{we}.

Nevertheless, to make this scheme works, two main constraints have to be considered. First, the embedding of m_e should not interfere with the one of m_c. In our JWEC scheme, the way one watermarked-JPEG-LS compressed block B_{wci} is modified into its version B^{w}_{wci} is an iterative procedure. It modifies some mapped prediction-errors belonging to B_{wci} until the bit of m_e is embedded, i.e. by verifying equation 3.7. To do so, the bit-substitution watermarking modulation is applied to the second lower order bit of the LSB-part of some prediction-errors that are not used for the embedding of m_c (i.e. pixels the Golomb-Rice MSB-part of which differ to the reference sequence $'0X1'$). The choice of the second lower order bit stands on the fact that the least significant bit encodes the prediction-error sign and should not be modified (see Section 3.1). This procedure guarantees that the embedding of m_e does not interfere with the one of m_c. Beyond, as in JPEG-LS the modification of one bit or equivalently of one pixel impacts the encoding of the following pixels, the embedding of m_e cannot be conducted separately from the one of m_c. The second constraint to consider is that a compressed-watermarked block B_{wci} should at least possess one pixel encoded in the regular-mode in order to be able to embed one bit of m_e. When a block only contains run-mode encoded pixels, these repeated sequences should be broken; by introducing one or more several regular-mode encoded pixels. To sum-up, the procedure we propose to watermark one block B_{wci} is iterative and relies on a buffer memory to manipulate a block as it is necessary to come back to the JWC encoding of pixels so as to ensure the embedding of m_e without interfering with the one of m_c. To give an example, let us assume B_{wci} is constituted of N pixels or equivalently, N Golomb-Rice encoded prediction-errors: $B_{wci} = \{(\tilde{e}_{MSB}, \tilde{e}_{LSB})_u\}_{u=1\ldots N}$. The procedure we propose to insert m_e is as given as follows:

Algorithm 1: Embedding of m_e

```
u = N ;
% Start with the last encoded pixel in the AES block ;
B^we_wci = AES(B^w_wci, K_e);
% encrypting the watermarked − compressed block
while AES(B^w_wci, K_e)_j ≠ m_ei do
    %m_ei corresponds to i^th bit of m_e
    if ẽ_MSB,u ≠' 0X1' then
        bit_substitution(ẽ_LSB,u);
        reencode(B^w_wci);
        B^we_wci = AES(B^w_wci, K_e);
    end
    u = u − 1;
end
```

where: the function *bit_substitution*() substitutes one bit of the prediction-error LSB-part ($\tilde{e}_{LSB,u}$); and *reencode*() corresponds to the JPEG-LS re-encoding of the modified 128-bit block.

Theoretical capacity and image quality distortion

The capacity of our JWEC in the encrypted domain depends on the AES block size ($N = 128$ bits), the image size ($n \times m$) and depth (d) as well as the JPEG-LS compression rate (τ). For a given image, the number of bits one can embed (or the number of blocks in the image, in other words) is such that

$$C_{meB} = \frac{nmd}{N\tau} \quad \text{bits} \tag{3.8}$$

or, equivalently, in bit of message per pixel of image

$$C_{m_e} = \frac{d}{N\tau} \quad bpp \tag{3.9}$$

It should be known that, as mentioned in [130] and [131], the optimal compression ratio of the JPEG-LS for medical images is of $2:1$ (i.e. $\tau = 2$).

Considering the AES cryptosystem as perfect, the diffusion property states that the change of one bit of the clear-text will lead to an AES cipher-text with at least 50% different bits. There is thus one-in-two chance that the change of one bit in a block leads to the correct value of the bit of m_e (see 3.7). The probability of being able to embed one bit after t tests is defined as $P(t) = 1 - (1/2)^t$. This probability converges rapidly to 1 with the increase of t. On average, one bit of m_e is inserted in one block within 2 tests. Our JWEC process is thus 2 times slower than simply compressing then encrypting the image.

Similarly to the embedding of m_c in the compressed image bitstream, we can estimate the distortion this second watermarking process injects into the image so as to embed m_e. Let us consider a JWEC block of N bits. Because of the diffusion property of AES [132], as already said, there is one chance over two that a pixel change in the block commutes one bit of the AES block, and because this embedding modifies the second lower order bit of the Golomb-Rice LSB part of pixels, the average MSE in the whole image is such that

$$\overline{MSE_e}\,(I, I_{wd}) = \frac{2^2 l_{\text{mean}} C_{meB}}{nm} \tag{3.10}$$

where, l_{mean} is the number of changes to make so as to encode one bit of m_e. Because of the AES property, $l_{mean} = 2$ and C_{meB} is given by 3.8.

More generally, because watermarking processes in the encrypted and in the compressed domains do not rely on the same pixels, it is possible to estimate the total average distortion that our JWEC can introduce in an image by summing 3.5 and 3.10, that is to say:

$$\overline{MSE_T} = \left(\frac{1}{p}\frac{C_{m_C}}{nm}\frac{\left(4^{k_{\max}-1}\right)}{3}\right) + \left(\frac{2^2 l_{\text{mean}} C_{meB}}{nm}\right) \tag{3.11}$$

Due to the fact that our JWEC algorithm introduces on average the same distortion in each block, one can refer to the Peak-Signal-to-Noise-Ratio (PSNR) so as to evaluate the image distortion rather than MSE. Let us consider an image of d bit-depth and let us denote I and I_{wd} the original image and its watermarked-decompressed-decrypted version, respectively, the theoretical average PSNR is given by

$$PSNR_{\text{mean}}\,(I, I_{wd}) = 10\log_{10}\left(\frac{\left(2^d - 1\right)^2}{\overline{MSE_T}}\right) \tag{3.12}$$

3.3.3 JWEC Image Reliability Control in both encrypted and compressed domains

In the healthcare domain and as seen in Chapter 1, the confidence of practitioners into pieces of data they receive relies on their reliability, that is to say proofs that [56]: i) data have not been modified by an unauthorized user (integrity), and, ii) they belong to the correct patient and they are issued from the correct source (authenticity). Non-reliable data should be automatically

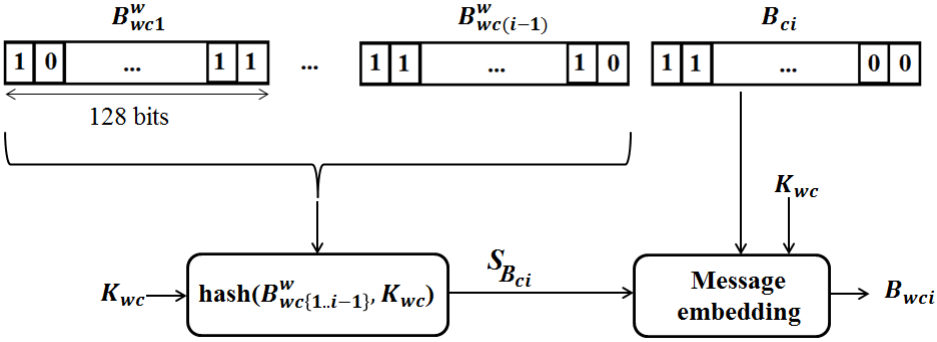


Figure 3.3: Computation and embedding of $S_{B_{ci}}$ into the compressed block B_{ci}, where $B^w_{wc\{1...i-1\}}$ are the previous compressed-watermarked blocks of 128 bits, K_{wc} is the secret watermarking key and B_{wci} is the compressed-watermarked by m_c block.

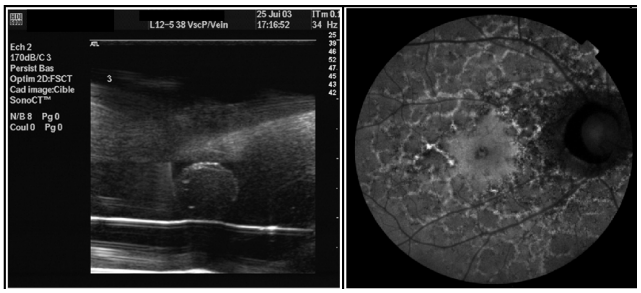


Figure 3.4: Samples of our image test sets (ultrasound and Retina images, respectively).

rejected by medical information system. We thus propose an extension of the previous JWEC scheme so as to ensure image reliability in the compressed and in the encrypted domains. Indeed, it inserts fragile watermarks in both encrypted and compressed domains. Any modification of the image, of its compressed bitstream as well as of its encrypted bitstream will alter the embedded messages.

Ensuring the authenticity of the image in both domains is rather simple. One just has to integrate in m_c and in m_e an authenticity code AC; that is the combination of unique identifiers of the image, the patient and of the image transmitter (see [133] for more details).

Providing the proof of the image integrity is more complicated. In the compressed domain and by taking into account the scheme proposed for integrity control in Chapter 2, image integrity can be controlled by means of a cryptographic hash computed from the image compressed bitstream elements that are not modified during the JWEC watermarking process. At the verification stage, any differences between the extracted hash and the recomputed one will indicate a loss of integrity. In this work, we opted for the well-known 256 cryptographic Secure Hash Algorithm (SHA-256). Because such a hash function is very sensitive to data modification and due to the fact the watermarked-compressed bitstream evolves along with the insertion of m_c and m_e (see Section 3.3), we propose an iterative procedure. Let us consider the previous JWEC system based on 128-AES. As depicted in Figure 3.3, the hash $S_{B_{ci}}$ of all previous compressed-watermarked blocks $B^w_{wc\{1..i-1\}}$ is embedded into the i^{th} compressed block B_{ci}, unlike in the joint watermarking-compression scheme proposed in Chapter 2 where the hash was computed onto compressed blocks watermarked by m_c, only. $S_{B_{ci}}$ of the JWEC-based reliability control is thus given by

$$S_{B_{ci}} = \text{hash}\left(B^w_{wc\{1,i-1\}}, K_{wc}\right) \tag{3.13}$$

where K_{wc} is the secret watermarking key based on which we secretly select bits of the hash, that will be inserted into B_{ci}.

Regarding the last compressed-block, its integrity is controlled by the hash of the non-watermarked pixels in the block along with the hash of the previous blocks. More clearly, the final hash $S_{B_{cf}}$ to be inserted in the block B_{cf} corresponds to :

Table 3.1: Comparison between theoretical and experimental embedding capacities expressed in bit of message in pixel (bpp) – μ and σ correspond to average and standard deviation of capacity values

Image type	Retina		Ultrasound	
Image size	627×643		576×690	
Image depth (bits)	8		8	
%Regular-mode encoded pixels	80%		75%	
Theoretical embedding capacity of m_c (bpp) (see (7))	0.093		0.1	
Experimental embedding capacity of m_c (bpp)	μ	σ	μ	σ
	0.14	0.012	0.095	0.034
Theoretical embedding capacity of me (bpp) (see (13))	0.031		0.031	
Experimental embedding capacity of me (bpp)	μ	σ	μ	σ
	0.03	0.002	0.0286	0.0017

$$S_{B_{cf}} = \text{hash}\left(B^{w}_{wc(1,f-1)}, K_{wc}\right) || hash\left(nw\left(B_{cf}\right), K_{wc}\right) \tag{3.14}$$

where: the function $nw\left(B_{cf}\right)$ provides the image bits that are not modified by the watermarking process and, $||$ is the concatenation operator.

To control the image integrity from the image encrypted-compressed bitstream, we propose to modify the extraction function f_e that gives access to m_e. Rather than extracting the message from bits of the AES bitstream, we propose to extract them from the SHA-256 hash of AES blocks. More clearly, the extraction function f_e used to extract m_e from an AES encrypted B^{we}_{wc} is defined as:

$$f_e\left(B^{we}_{wci}, K_{we}\right) = \text{hash}\left(AES\left(B^{w}_{wci}, K_e\right), K_{we}\right)_j = h_j \tag{3.15}$$

where h_j corresponds to the j^{th} bit of the SHA-256 hash of B^{we}_{wci}. j is chosen based on the secret watermarking key K_{we}. Because the SHA-256 "strength" is of 128 its, if 1 bit of B_{wci} changes then, there is one-in-two chance that h_j commutes and to detect the bitstream tampering. Notice that, this JWEC-based reliability control induces the same image distortion as the previous JWEC scheme (see Section 3.2.2). It is nearly 2 times slower than simply compressing then encrypting an image. It is of same complexity as the our general JWEC scheme.

At the detection stage, the watermark reader will just have to compute the SHA-256 of the encrypted blocks and use the secret watermarking key to determine which bit from the encrypted block corresponds to the embedded message bit. In the case encrypted data are tampered, ciphered-block will not allowed the corrected recovery of m_e , indicating thus the data cannot be used.

3.4 Experimental Results of The Proposed JWEC-based Reliability Control

Experiments were conducted on two sets of medical images of 8-bit depth : 1200 Retina images of 627×643 pixels, and 100ultrasound images of 579×690 pixels. Some samples of our image data set are given in Figure 3.4.

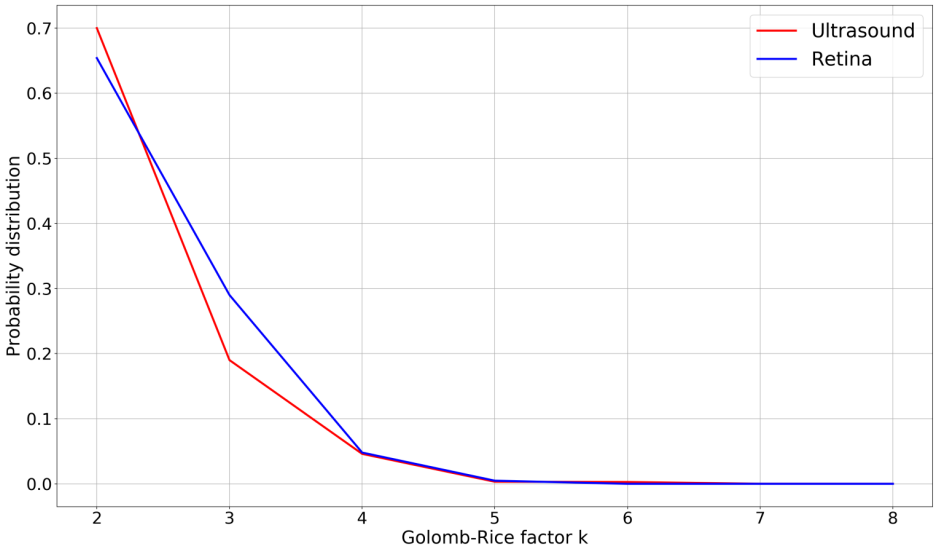


Figure 3.5: Probability distribution of the Golomb-Rice factor corresponding to watermarked pixels (computed on ultrasound and Retina images).

3.4.1 Capacity Rates

As already mentioned in Section 4.2.2, message capacity in the compressed domain depends on the distribution probability of the chosen reference sequence $'0X1'$ and on the percentage of regular-mode encoded pixels. In the following tests, the reference sequence $'001'$ has been considered for the embedding of m_c.

Table 3.1 provides the capacity rates we achieved in both encrypted and compressed domains. It can be seen that the theoretical capacity given by 3.3 is close to the practical one for ultrasound images. For Retina images, the difference stands on the fact that the watermark can be embedded into the bitstream of pair of consecutive prediction-errors (see Section 4.2.2). Ultrasound images provide the lowest embedding capacity values. This can be explained by the fact they contain large black background encoded in run-mode. The embedding capacity in the encrypted domain depends on the encrypted block size (N), the image depth (d) and the JWEC compression rate (τ). As we considered the AES cryptosystem in its CBC mode, the block size is limited to $N = 128$ bits. In practice, as indicated in Table 3.1, resulting capacities in the encrypted domain are very close to the theoretical one given by 3.9. This latter gives an upper insertion capacity limit. For Retina images, one can at least embed a message of 56 *Kbits*.

We also give in Table 3.2, the embedding capacity values considering two different reference sequences;$'0X1' =' 001'$ and $'0X1' =' 0001'$ for Retina images. It confirms that the watermark capacity is impacted by the choice of the reference sequence. It is normal that our JWEC reaches better embedding capacity in the compressed domain with $'0X1' =' 001'$ than with $'0X1' =' 0001'$.

The offered embedding capacities in both domains are large enough compared to the requirements for verifying the reliability of the image that we estimate about 1 *Kbits* (a digital signature provided by the SHA-2: 256 bits and one authenticity code: 600 bits by combining the French National Identifier with the DICOM Unique Identifier [133]). As a consequence, image quality can be better preserved than in previous experiments by reducing the number of watermarked pixels. This can also contribute to the security of our scheme by secretly selecting subsets of watermarkable pixels (i.e. identified by the reference sequence $'0X1'$) for the embedding of m_c and encrypted blocks from which m_e will be extracted. On another hand, one can take advantage of this free space to embed other pieces of information like some elements related to a security policy and to the patient consent [59] or any other metadata in relation to the patient health [60].

3.4.2 Distortion

The Peak-Signal-to-Noise-Ratio (PSNR) and the Structural SIMilarity (SSIM) between the original image I and its decrypted-decompressed-watermarked version I_{wd}, were considered so as to evaluate the image distortion. As already said in Chapter 2, if PSNR is appropriate in the case of a uniform

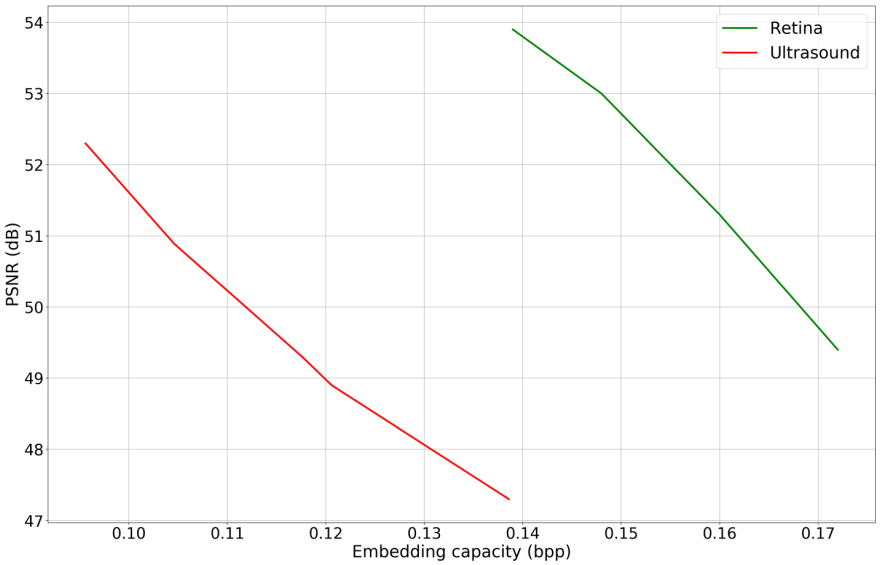


Figure 3.6: PSNR vs. total capacity considering both encrypted and compressed domains.

Table 3.2: Embedding capacities in the compressed domain for two different reference sequences. Tests done on Retina images

	Average	Variance
Embedding capacity for $'0X1' =' 001'$ (bpp)	0.14	0.012
Embedding capacity for $'0X1' =' 0001'$ (bpp)	0.04	0.0038

image distortion, SSIM [105] takes into account some visual human properties. It is defined as

$$SSIM = l\left(I, I_{wd}\right) \cdot c\left(I, I_{wd}\right) .s\left(I, I_{wd}\right) \tag{3.16}$$

where $l\left(I, I_{wd}\right)$ is the luminance comparison function, $c\left(I, I_{wd}\right)$ is contrast comparison and $s\left(I, I_{wd}\right)$ is the structural comparison function.

As discussed in Section. 3.2.2, the distortion introduced by our JWEC scheme depends not only on the considered reference sequence $'0X1$ but also on the Golomb-Rice factor k and the number of modulated pixels l for the insertion of m_e. In the case of $'0X1' =' 001'$, it is possible to compute the theoretical value of $PSNR_{mean}$ (see Equation (3.12)). Figure 3.5 confirms that the Golomb-Rice factor k follows a geometric distribution of parameter p the value of which can be estimated for both ultrasound ($p = 0.7$) and Retina images ($p = 0.6$). We have also verified that, in order to embed m_e, two pixels are in average modulated per JWEC block (i.e. a block of $N = 128$ bits compressed with a ratio $\tau = 2 : 1$. In consequence, the theoretical mean PSNR value ($PSNR_{mean}$) between the original and the watermarked-decompressed-decrypted images is of 54.01 dB, and of 54.11 dB for Retina and ultrasound images, respectively.

The compromise between the embedding capacity and PSNR of our JWEC scheme is given in Figure 3.6 for both used image data set. Obviously, capacity varies with the image distortion. This compromise evolves depending on the image type (or modality) and of its content (e.g. the percentage of regular-mode encoded pixels, the texture of the image). In average, the obtained PSNR values for ultrasound and Retina images are of 49.8 dB and 51.65 dB, respectively. The gap between the previous theoretical values has the same explanation as the difference between theoretical and experimental capacity values. Several bits of the watermark in the compressed domain are embedded in pairs of consecutive pixels' prediction-errors the bitstream of which gives access to the reference sequence $'0X1'$ (see end of the first sub-section of Section 3.2.2). More generally, our system has sometimes to introduce other distortions so as to guarantee the error-free extraction of both messages m_e and m_c. Beyond, the degree of such distortion is small and does not endanger the diagnosis value of the medical images. One can refer to the study on the

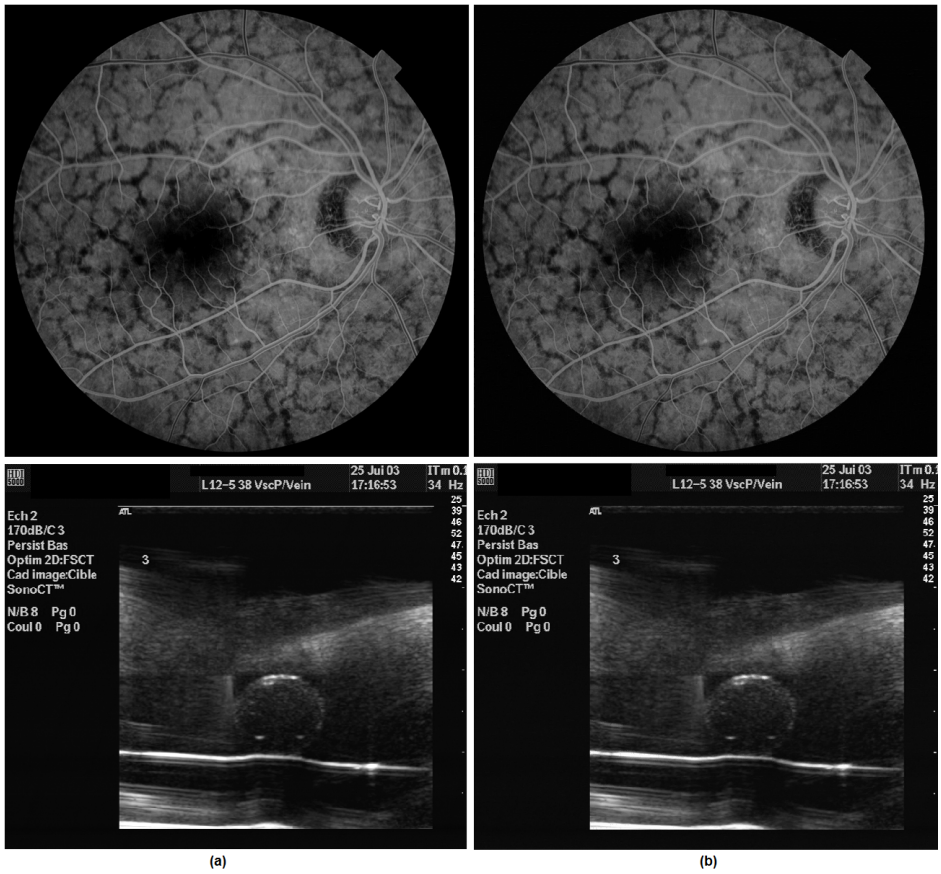


Figure 3.7: Image samples from our data set (a) Original images (b) decompressed-decrypted-watermarked images.

impact of lossy image compression on medical images in [106], where it is reported that image distortion should be maintained in the range of 40 and 50 dB to establish good diagnosis. To sum up, even though our JWEC scheme induces an image information loss, it well preserves the diagnosis value of images. Resulting SSIM values, being greater than 0.98 also confirms the good quality of watermarked images [105]. As also illustrated in Figure 3.7, where we give two examples of JWEC protected images along with their original versions, it can be noticed that we are so far from introducing a visible distortion in the image and the diagnosis value of images is preserved.

3.4.3 Distortion-Capacity Performance and Comparison

We compare in Table 3.3 our JWEC solution to methods of the state of the art dealing with encryption, watermarking and compression, independently of the medical imaging domain. As it can be seen, our scheme is the first joint watermarking-encryption-compression algorithm. All other joint methods only consider encryption & watermarking, or watermarking & compression or encryption & compression. By merging encryption, compression and watermarking in a single process, our scheme is neither commutative nor separable. Nevertheless, as the proposed JWEC does not rely on data reorganization, it makes watermarking completely transparent to decryption and decompression. Hence, decompression and decryption can be conducted separately and regularly (i.e. using the common AES and JPEG-LS) making our scheme compliant to DICOM. Only a few watermarking & encryption schemes can do the same ([107], [103], [115], [6] and [98]). Note that watermarking & compression schemes that work with JPEG, JPEG 2000, or JPEG-LS will obviously be compliant to DICOM. Our scheme also offers watermarking-based services from the encrypted and compressed domains without having to decrypt or to decompress, even partially, the image data. None of its competitors from the literature provides this functionality. Moreover, our method is fully compliant to JPEG-LS, while [98] works with a near-lossless version of JPEG-LS [98] and only allows message extraction from the spatial (i.e. pixel values) domain unlike our JWEC that gives access to a message in the compressed domain. As stated in Section

Table 3.3: Comparison of methods from the state of the art and the proposed JWEC scheme. Non-Independent decryption or non-independent decompression indicate that the solution depends on some data pre/post-processing so as to decrypt or to decompress the data, respectively.

		Data Encryption		Data Compression		Message embedding			Message extraction				Decryption		Decompression		Medical imaging
Class	Method	Entirely	Partially	Entirely	Partially	Joint	Separable.	Commutative	Spatial	Partially decompressed	compressed	Encryption	Independent	Non-independent	Independent	Non-independent	
Encryption-watermarking methods	[107]	X				X			X			X	X				X
	[103]	X				X			X				X				
	[109]	X					X		X				X				
	[110]	X					X		X				X				
	[111]	X					X					X		X			
	[112]	X					X					X	X				
	[113]		X				X					X		X			
	[114]		X					X	X			X		X			
	[115]	X					X		X			X	X				X
Watermarking-compression schemes	[90]				X		X		X						X		X
	[92]				X		X		X						X		
	[93]			X			X			X						X	
	[116]			X			X			X						X	
	[117]			X			X		X						X		
	[118]			X			X			X						X	
	[6]			X		X				X	X				X		X
	[98]			X		X			X						X		X
Encryption-compression solutions	[119]	X		X									X			X	
	[120]	X		X									X			X	
	[121]	X		X										X	X		
	[122]	X		X									X		X		
	[123]	X		X										X		X	
	[124]	X		X									X		X		
	[125]																
Encryption-compression-watermarking techniques	[95]	X					X									X	
	[126]						X									X	
	JWEC scheme	**X**		**X**		**X**					**X**	**X**	**X**		**X**		**X**

3.2.2, if a decompressed JWEC protected image has not been modified, the watermark reader just has to JPEG-LS re-compress it to retrieve the message from the compressed image bitstream. Our scheme is however fragile in the sense that, if data are modified, message recovery is not possible. To go further, the recovery of the message in the encrypted domain requires the knowledge of the encryption key so as to re-encrypt the compressed bitstream. This can be a constraint.

Because our JWEC algorithm induces an irreversible image information loss, we propose to compare it to schemes presented in [95] and [126] from the same class (see Table 3.3). In these schemes, encryption, watermarking and compression are separable in the sense that these operations are conducted in a cascade manner but not in one operation. They however rely on a proprietary decryption or decompression algorithms, due to the fact that data have to be reorganized at one point. Furthermore, they give access to the watermark in only one image domain.

Regarding the capacity/distortion compromise, in case of the well-known test image "*Lena*",

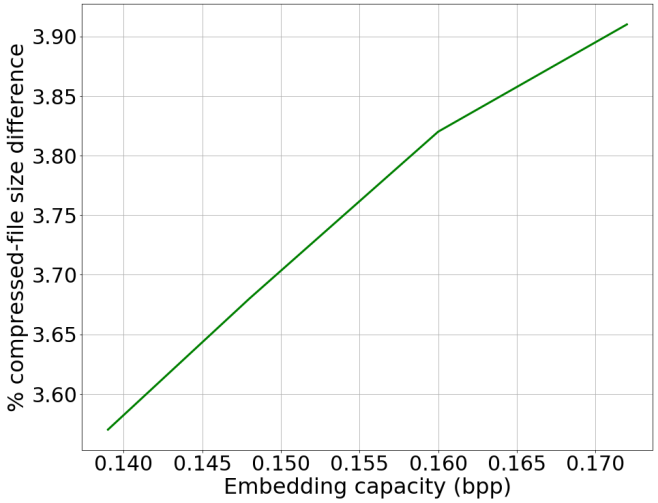


Figure 3.8: % file size evolution vs. embedding capacity on Retina images.

our proposal provides higher PSNR and embedding capacity values. For a capacity of 0.2 *bpp* (or equivalently 52757 bits), it achieves a PSNR of 42.7 *dB* while [95] reaches a maximum embedding capacity of 0.1 *bpp* for a PSNR of 35.22 *dB* and [126] reaches a capacity of 0.007 *bpp* for a PSNR of 38 *dB*.

3.4.4 Algorithm Complexity and Compression Rate

As stated in Section 3.2.2, our scheme has sometimes to modify compressed image bitstream so as to ensure the error-free extraction of messages. This can be seen in Figure 3.8, where we provide JWEC file size evolution depending on the total embedding capacity (i.e. considering the embedded messages in both encrypted and compressed domains) in the case of Retina images. File size increased linearly with the capacity. Notice also that this increase is very small.

In terms of complexity, we verified that our scheme is about 2 times slower than compressing then encrypting the image in a cascade manner. The reason of this difference is obviously caused by the embedding of m_e in the encrypted domain. However, our JWEC gives access to watermarking-based security services in both encrypted and compressed domains while ensuring the confidentiality of the image, without parsing or decoding data bitstream, even partially.

3.4.5 Security Analysis

The proposed JWEC scheme allows ensuring confidentiality of images while providing watermarking-based security services, through two messages m_c and m_e. It can however face different cryptographic and watermarking attacks.

3.4.5.1 Cryptographic Attacks

Our scheme is implemented with the AES block-cipher algorithm in its CBC mode. Because we do not intrinsically modify it, its security performances against common cryptographic attacks are preserved [134]. Indeed, even if K_{we} or m_e are known by the attacker, he/she has no additional means than a regular cryptographic attack to get k_e or to have an idea about the watermarked-compressed image bitstream.

3.4.5.2 Watermarking Attacks

Three classes of watermarking attacks should be considered [135]:

- *Unauthorized embedding* − where an attacker wants to modify the embedded messages or insert his own ones. In this case, if he/she does not know neither the encryption key nor the watermarking keys (K_{we} and K_{wc}), such a tamper will be detected easily. Without the watermarking key K_{we} , the attacker cannot insert fake message m_e that will be considered

as valid. Even if he/she succeeds, the watermarked-compressed bitstream as well as the image will not be properly decoded. Thus, if the attacker knows the watermarking key in the encrypted domain, an alarm will be raised in the compressed domain thanks to the message m_c that will be altered too. In the case that the attacker knows K_e and K_{we} , he/she can decrypt the image but his/her capability to modify or insert a new message m_e is impossible without the knowledge of the watermarking key in the compressed domain (K_{wc}). Indeed, modifying even one bit of the image watermarked-compressed bitstream will make impossible the JPEG-LS decompression. If the attacker now decompresses the protected image and tries to JWEC it again, he/she still has to ensure the embedding of a valid message m_c; the validity of which depends on the watermarking key in the compressed domain.

- *Unauthorized detection/extraction of messages* − where an attacker wants to detect or extract the embedded messages. With our JWEC scheme, the location of the messages' bits in both domains depends on secret watermarking keys. Without these keys, the attacker cannot distinguish the bits that correspond to the watermarks. To reinforce security, we recommend encrypting the messages before being embedded into the image.

- *Unauthorized removal attack* − in such case, the attacker tries to remove the embedded messages from the image. As already said in Section 3.3.3, when the received image is stated unreliable, i.e. it is not possible to verify its integrity and its origins; it has to be directly rejected by the medical information system. As a consequence, valid watermarks should be present in the encrypted and compressed domains.

3.5 Conclusion

In this chapter, we have descried the first joint watermarking-encryption-compression system, which guarantees an *a priori* and *a posteriori* protection of medical images. Our JWEC gives access to watermarking-based security services in the compressed domain as well as in the encrypted one, without having to decrypt or to decompress the corresponding the corespondent image bitstream, even partially. Additionally, we have shown how such a JWEC algorithm can be used for verifying the reliability of an image. Beyond, offered embedding capacity is large enough so as to allow various watermarking-based security services ranging from authenticity to traceability and patient consent management. Even though our scheme introduces an image information loss, this one is small and the diagnosis value of images is preserved. Our JWEC system is about two times slower than simply JPEG-LS compressing the image and encrypting it, but it provides watermarking-based security functionalities. Furthermore, image decryption and decompression processes are not modified making our scheme compliant or transparent to the DICOM standard. A JWEC protected image can be deciphered and decompressed by a non-compliant watermarking system, using AES and JPEG-LS normally.

Due to the need to combine watermarking with confidentiality control tools in order to ensure a better protection of medical image, our future works focus on extending the joint watermarking-encryption-compression scheme to JPEG compressed domain. As described in Chapter 2, it is possible to combine watermarking and JPEG compression in a single operation so as to extract the message directly from the compressed image bitstream, without decompressing even partially. Hence, it will be possible to combine this scheme with an encryption algorithm so that message extraction will be performed in both JPEG compressed and encrypted domains.

As discussed in Chapter 1, the reversibility property is of special interest in the medical context as the original image can be completely recovered once the watermark extracted, leading thus to applications like integrity and authenticity control and traceability. In the next chapter, we will detail a lossless watermarking scheme, which allows the embedding, into the encrypted image, of a message that can be extracted from both encrypted and clear (i.e. spatial) domain, as well as the error-free recovery of the original image.

Chapter 4

Reversible Image Crypto-Watermarking based on Robust Histogram Shifting

In the previous chapters, we addressed the protection of compressed images, encrypted or not, by means of lossy or irreversible watermarking. As stated in Chapters 1, 2 and 3, such a distortion can impact image interpretation. Herein, it is interesting to use reversible watermarking, which guarantees the exact recovery of the original image from its watermarked version by removing the embedded watermark. Furthermore, it makes possible to update the watermark content at any time without adding new image distortions.

Reversible watermarking methods are generally based on the Difference Expansion (DE) [136] and/or the Histogram Shifting (HS) [85] modulations. To roughly summarize DE, this one divides the image into pairs of pixels and embeds bits of the secret by expanding their differences so as to create a virtual LSB bit; one bit of the message. At the reading stage, the reader recomputes the pixels' differences, read the virtual LSB bit values so as to extract the message and expands back pixels differences so as to recover the original image. Regarding, we will come back on its details in the sequel, as we propose an original extension of it. It is also important to notice that these reversible watermarking modulations are in essence fragile, in the sense that the watermark will not survive to any image modifications. We propose a solution in order to make HS robust to some image modifications.

In this chapter, we are interested in being able to reversibly watermark encrypted images. As medical images are sensitive data, they are usually stored in a encrypted fashion in order to ensure their confidentiality. When managing such data, it is desirable to embed additional data (e.g. time stamps, labels) without accessing to their contents so as to trace them or protect them in terms of integrity, without having to decrypt the data. In this context, it is interesting to allow message extraction from the encrypted and decrypted image. For instance, one server can embed in the encrypted image the identity of the user who requested it with the capability to retrieve the user identity from the decrypted image so as trace herhim. This can be helpful in order to verify the reliability of an image this one being encrypted or not.

Herein, we propose a new crypto-watermarking approach based on "pre-watermarking" which allows embedding in a reversible manner a message in encrypted data; a message that can be extracted before and after image decryption. Its main originality stands on the insertion of a reversible pre-watermark that is robust to the insertion of a message in the encrypted domain. To do so, we propose a robust version of the well-known histogram shifting modulation. Due to its robustness, the original pre-watermark can be correctly recovered, allowing thus restoring the original image. The proposed scheme encrypts the entire image while watermarking and encryption/decryption processes are conducted separately. More clearly, message insertion and extraction (resp. encryption and decryption) do not require neither the knowledge of the encryption key (resp. watermarking key or other extra parameters). Notice also that, unlike actual similar solutions, our scheme does not require the reorganization of image pixels or bitstream. More clearly, the image is commonly decrypted and can be used by the user. The decryption algorithm is not proprietary.

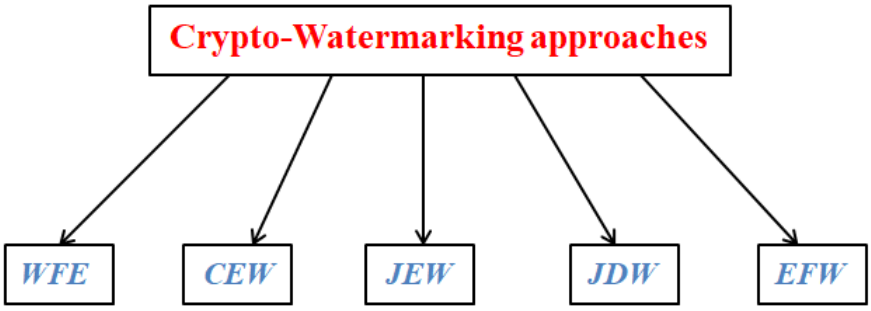


Figure 4.1: The five categories of crypto-watermarking approache combinations; where WFE corresponds to Watermarking Followed by Encryption, CEW to Commutative Encryption-Watermarking, JEW to Joint Encryption-Watermarking, JDW to Joint Decryption-Watermarking and EFW to Encryption Followed by Watermarking.

This chapter is organized as follows. First, we come back on the state of the art of crypto-watermarking schemes so as refine the originality of the objectives of our proposal. Then, we present some preliminaries about stream cipher encryption, we use in our scheme, as well as on the principles of the Histogram Shifting (HS) modulation before detailing our robust version of HS. The third par is devoted to the general principles of our approach and illustrates one possible implementation of it using the Trivium stream cipher along with the LSB substitution watermarking modulation and our robust HS modulation. Some experimental results considering natural and medical image datasets are then provided so as to support possible applications of our scheme as, for instance, how it can be used for verifying the image integrity and authenticity in both clear and encrypted domains.

4.1 Crypto-Watermarking Schemes

As seen in Chapter 3, for the last few years, various crypto-watermarking approaches have been proposed in order to offer both "*a priori*" and "*a posteriori*" protection. The basic objective of such crypto-watermarking schemes is to guarantee at the same time confidentiality of data through encryption while providing watermarking-based security services. These security services can be made available in the clear domain and/or in the encrypted domain. Five categories of crypto-watermarking approaches can be distinguished according to the way encryption and watermarking mechanisms are combined (see Figure 4.1).

- *Watermarking Followed by Encryption* (WFE) [137, 138] – where the host signal is watermarked and by next encrypted. With such a method, the embedded message is not available from the encrypted signal.

- *Commutative Encryption and Watermarking* (CEW) [139, 140, 141] – with these schemes, the ordering of watermark insertion and image encryption processes has no influence: the same encrypted-watermarked image will be obtained. These methods are usually based on partial encryption [140] or homomorphic encryption [139]. In the former case, image data are divided into two sets: one set is used for watermark embedding and the other set for encryption. By doing so, there is no risk that watermarking interferes with encryption/decryption processes. [141] proposes a commutative scheme where a bit-plane subset of the image representation in a suitable transformed domain is watermarked while the remaining bit-planes are encrypted. Due to the fact that correlation between watermarked and encrypted subsets can introduce some security breaches, authors propose to pseudo-randomly select the representation domain among a wide set of parametric transformation to enhance scheme security. To make this possible, the parametric discrete Tree-Structured Haar transform has been used, which depends on a secret set named Discontinuity Point Vector (DPV) that defines the splitting scheme. Homomorphic schemes are referred as "invariant schemes" as they allow conducting some operations onto encrypted data with the guaranty that the decrypted result

equals the one of the equivalent calculation conducted onto unencrypted data. However, these methods can be only realized with specific encryption algorithms that matched with a watermarking algorithm limiting thus their applicability. The authors of [139] propose an algorithm based on the Paillier homomorphic cryptosystem and the Patchwork watermarking to realize commutative watermarking/encryption. The Patchwork watermarking uses the statistical characteristics of the carrier data to embed the watermark. To sum up, even if these commutative encryption and watermarking schemes allow accessing to the watermark from both encrypted and clear domains, they present many inconveniences. One may consider that CEW schemes based on partial encryption do not fully ensure confidentiality. It could be possible to use non-encrypted data to attack the system. Regarding homomorphic encryption-based CEW schemes, it is important to notice that beyond its high computation complexity, homomorphic encryption induces a huge expansion of data. A clear text message may have an encrypted version of more than 2048 bits (e.g. Paillier cryptosystem [142]).

- *Joint Encryption/Watermarking* (JEW) [107, 143, 144] - In these schemes, watermarking and encryption operations are conducted jointly. For instance, in [143], two messages conveying some security attributes are embedded during the encryption process. Each message is only available in one domain in order to ensure watermarking services in both domains. It is important to notice that in general, even though encryption and message embedding are jointly conducted, decryption and message extraction can be performed independently. One weakness of these approaches is that most of the time, they insert read-only messages. More clearly, it is not possible to modify the messages in the encrypted bitstream. Modifying the message will not allow the correct decryption of the image.

- *Joint Decryption/Watermarking* (JDW) [145, 146] - where a fingerprint is embedded during the decryption process. These schemes are inspired by the well-known cipher algorithm Chameleon [147]. The basic idea of JDW scheme is to insert a watermark during the decryption process. Such a strategy allows essentially reducing time computing and complexity on the server side.

- *Encryption Followed by Watermarking* (EWF) [111, 112, 113, 115, 148, 149, 150, 151, 152, 153, 154, 155] - The solution we propose belongs to this category of methods. Five main strategies can be distinguished. The first one consists in embedding the message in the encrypted domain using a classical reversible watermarking modulation. In [112, 148, 149], some bits of the encrypted image are losslessly compressed and replaced by their compressed versions along with the message. As a consequence, the message is only available in the encrypted domain and the image can only be decrypted after extracting the message. The image decryption is thus proprietary too. These schemes remain limited depending on the efficiency of the crypto-system. In theory, a perfect cryptosystem maximizes entropy leading to an un-compressible bitstream. The second strategy takes advantage of the differences of the statistical properties of the encrypted and decrypted images for watermark embedding and removal [111, 113]. For instance, [111] embeds one message bit per encrypted image block by flipping the three LSB planes of half of the encrypted pixels. To extract the message, the recipient first decrypts the image and computes some correlation measurement in each pixel block. If the correlation is high, the block has not been modified and a bit '1' is extracted. On the contrary, the block has been modified and a bit '0' is extracted. To recover the original block, the recipient flip-back the LSBs of the encrypted block, and decrypt the block again. Notice that with this second EWF strategy, the embedded message can be only extracted from either the clear or the encrypted domain. Notice also that the original image is sometimes recovered with some errors due to the fact that correlation fails to detect the correct message bit. The third strategy exploits homomorphic and/or probabilistic encryption [150, 151]. For instance, [151] first independently Paillier encrypts the image pixels. Then, by taking advantage of the additive homomorphism property of the used cryptosystem, the authors apply pixel expansion so as to embed the message. More clearly, considering a pixel value p, this one is changed into the value $2p+b$; where b is a bit of the message. These operations are easy to perform in the Paillier encrypted domain. To recover the original pixels' values and to extract the message, the encrypted image needs to be decrypted first. The forth strategy, which can be referred under the concept of "vacating room before encryption", consists in: 1) generating an embedding room by embedding LSBs of certain pixels into other pixels using traditional reversible watermarking,; 2) encrypting the image and; 3) embedding the message

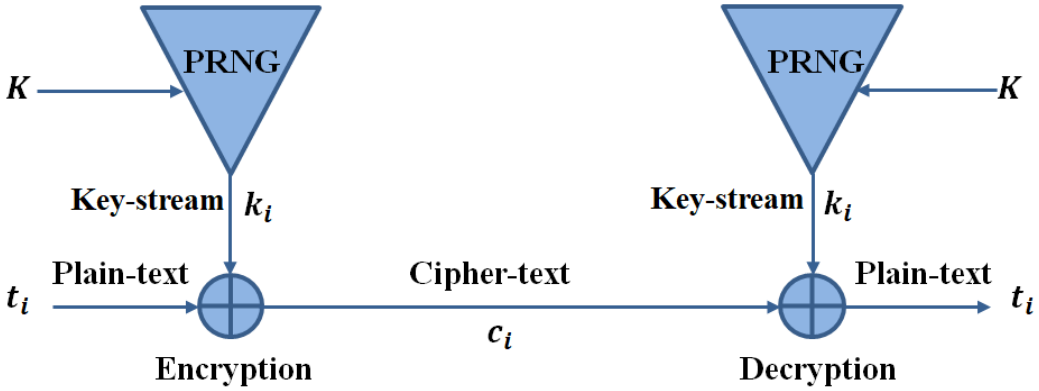


Figure 4.2: Encryption/decryption processes of a stream cipher algorithm with a secret key K. t_i, c_i and k_i correspond to the plain-text bits/bytes, the cipher-text bits/bytes and the secret bits/bytes keystream respectively. k_i is issued by a Pseudo-Random Number Generator (PRNG).

by replacing the vacated LSBs in the encrypted bitstream. The main inconvenient of such a scheme is that image data have to be reorganized before being encrypted, in order to make possible the identification of the vacated space in the encrypted domain. As example, this one can be placed at the beginning of the encrypted image bitstream [156]. Furthermore, the image is in essence partially encrypted. The last strategy [115, 152, 153] are based on the embedding of a pre-watermark in the plain-text image before encrypting it. The basic idea is to next embed the message in the encrypted image and to ensure that the decryption errors, related to this insertion process, impact the pre-watermark so as to be able to extract it from the clear/decrypted domain. Message extraction in the clear domain is done by comparing the extracted pre-watermark to the original one. The extraction in the encrypted domain depends on the modulation used to insert the message. However, these methods are not reversible, in the sense that original image cannot be exactly recovered after having extracted the message. To sum up, none of the above reversible watermarking-based EFW schemes allow the extraction of a message, embedded into the encrypted domain, from both encrypted and clear domains. In addition, they do not entirely encrypt the image and/or, image decryption is only possible after the encrypted image bitstream reorganization. The solution we propose does not suffer of these issues. We sum up in Table 4.1 the above state of art and compare it to our EFW solution independently of the medical imaging domain.

4.2 Encryption and Watermarking Primitives

4.2.1 Stream Cipher

As described in Chapter 1, two kinds of encryption algorithms can be distinguished: block cipher and stream cipher algorithms. Contrarily to block cipher algorithms which operate on large blocks of plaintext, stream cipher algorithms; like RC4 [157] and Trivium [51], manipulate stream of bits/bytes of plaintext.

As given in Figure 4.2, stream cipher algorithms combine the bits/bytes of plaintext $T = [t_1, \dots t_i, \dots, t_n]$ with a secret keystream of bits/bytes $K = [k_1, \dots k_i, \dots, k_n]$ issued by a pseudo-random number generator (PRNG), through a typically XOR operation. The keystream generation depends on the secret encryption key K_e. Thus, bits/bytes of cipher-text $C = [c_1, \dots c_i, \dots, c_n]$ are usually defined as:

$$c_i = t_i \oplus k_i \tag{4.1}$$

where $\oplus$ represents the XOR operation.

The way the PRNG produces the bit/byte keystream is specific to the stream cipher algorithm. Some of the main advantages of this type of algorithms are they are easy to implement and operate at a higher speed than block cipher algorithms making them appropriate to the encryption of huge volume of data. Furthermore, it has maximum entropy because 0's and 1's are generated with equal probability [158], that is to say there is any form of correlation within the encrypted image.

Table 4.1: Comparison of methods from the state of the art and the proposed EFW scheme. Non-Independent decryption indicates that the solution depends on some data pre/post-processing so as to decrypt the data, respectively.

		Data Encryption		Watermarking		Message extraction domain		Decryption		Medical imaging
Class	Method	Entirely	Partially	Reversible	Irreversible	Spatial	Encrypted	Independent	Non-independent	
Watermarking Followed by Encryption (WFE)	[137]	X			X	X		X		
	[138]	X		X		X			X	X
Commutative Encryption and Watermarking (CEW)	[139]	X			X	X	X	X		
	[140]		X	X		X	X		X	
	[141]		X		X	X	X		X	
Joint Encryption/ Watermarking (JEW)	[107]	X			X	X	X	X		X
	[143]	X			X	X	X	X		X
	[144]		X	X		X	X		X	X
Joint Decryption/ Watermarking (JDW)	[145]	X			X	X			X	
	[146]	X			X		X		X	
Encryption Followed by Watermarking (EFW)	[111]	X		X			X	X		
	[112]		X	X			X	X		
	[113]		X	X			X		X	
	[115]	X			X	X	X	X		X
	[148]	X		X			X	X		
	[149]	X		X			X	X		
	[150]	X		X		X			X	
	[151]	X		X			X		X	
	[152]	X			X	X	X	X		X
	[153]	X			X	X	X	X		X
	Proposed Method	**X**		**X**		**X**	**X**	**X**		**X**

In this work, we opted for the Trivium stream cipher. Basically, Trivium is a synchronous stream cipher algorithm designed to generate up to 2^{64} bits of random binary sequence from an 80-bit secret key and an 80-bit initial value (IV). This algorithm consists of 2 phases: i) the internal state of the cipher is initialized using the key and the IV, then ii) the state is repeatedly updated and used to generate the random binary sequence. Trivium was designed to be compact in constrained environments and fast in applications that require a high throughput as mentioned in [51].

The choice of this algorithm stands on the fact that it belongs to the eSTREAM portfolio of recommended stream cipher algorithms suitable for different applications and due to its provable security and high software/hardware implementation efficiency [159].

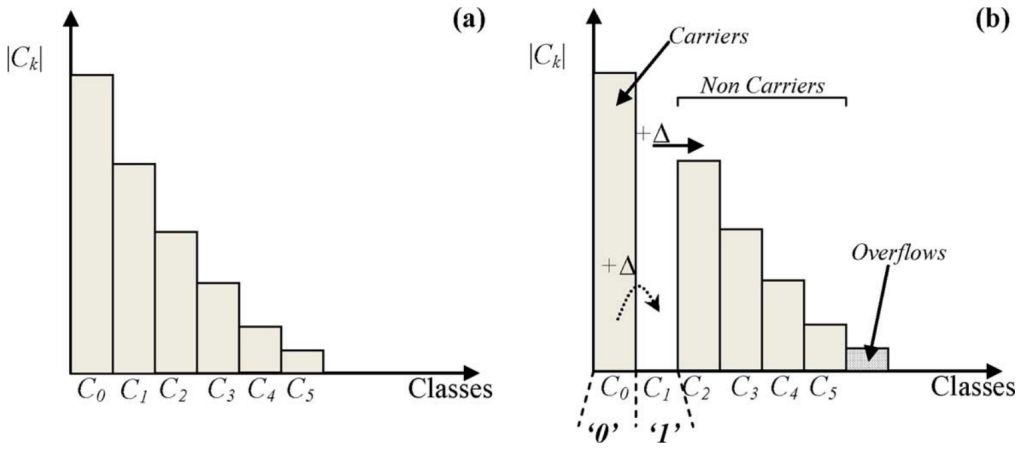


Figure 4.3: Basic principle of the Histogram Shifting modulation. (a) original histogram (b) histogram of the watermarked data.

4.2.2 Histogram Shifting Modulation

4.2.2.1 General Principle

Proposed by Ni *et al.* [160] and as illustrated in Figure 4.3, the histogram shifting (HS) modulation, consists in shifting a range of the signal histogram according to a fixed magnitude Δ to create a 'gap' near the histogram maxima; which corresponds to C_1 in Figure 4.3. Samples of the signal (e.g. pixels) with values associated to the class of histogram maxima, are then shifted to the gap or kept unchanged to encode '1' or '0', respectively. These samples are referred as "carriers" and samples of other classes are called "non-carriers". At the extraction step, the watermark reader just has to interpret the message from classes' samples and to shift back samples' values. Nevertheless, in order to restore exactly the original signal, the watermark reader needs to be informed of the samples values of which have not be shifted at the insertion stage due to the fact their values are at the extremities of the sample dynamic range. These samples are called overflows or underflows (see overflows' example in Figure 4.3). This requires the embedding of some extra data (a message overhead) along with the message to inform the watermark reader. Such a message overhead reduces the watermark capacity. This overhead generally corresponds to a location map the components of which inform the reader whether samples are original or have been shifted. Another strategy is to build a vector, one component of which indicates to the reader if a sample of value in the value in the overflow class has been or not shifted.

4.2.2.2 Prediction-Error Histogram Shifting (PEHS)

In the traditional form of HS, the embedding capacity is limited by the cardinality of peak point class. As a consequence, the prediction-error-based HS has been introduced, in order to go beyond the low embedding capacity of HS in the spatial domain, by exploiting the similarity of neighboring pixels. The idea is to construct a prediction error histogram. This also facilitate the identification of the carrier classes as they are centered on zero.

PEHS thus relies on two main steps: 1) Computation of the histogram of the image prediction-errors, and 2) Watermarking of prediction-errors based on HS modulation. The former step consists in calculating the prediction-error e_{ij} of each image pixel p_{ij}, such that

$$e_{ij} = p_{ij} - \hat{p}_{ij} \tag{4.2}$$

where $\hat{p}_{ij}$ is the predicted value of p_{ij}. Different predictors have been proposed in the literature. In this work we opted for the very classic one, where $\hat{p}_{i,j}$ is derived from the four nearest neighbor pixels of $p_{i,j}$ [85]:

$$\hat{p}_{i,j} = \frac{(p_{i-1,j} + p_{i,j+1} + p_{i+1,j} + p_{i,j-1})}{4} \tag{4.3}$$

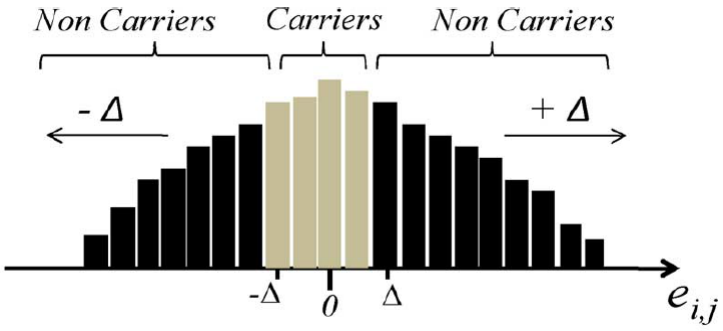


Figure 4.4: Histogram Shifting applied on prediction-errors.

As illustrated in Figure 4.4, prediction-errors are then HS modulated to insert a message as follows. The "non-carrier" classes which regroup prediction-errors that do not belong to the carrier-class $C_c = [-\Delta, \Delta[$ are first shifted of $+/-\delta$ as:

$$e_{ij}^w = \begin{cases} e_{ij} - \Delta & \text{if } e_{ij} < -\Delta \\ e_{ij} + \Delta & \text{if } e_{ij} \geq \Delta \end{cases} \tag{4.4}$$

The prediction-errors within the carrier-class C_c are used for the embedding of the binary message $M = \{b_i\}_{i=1,...N}$, $b_i \in \{0, 1\}$. A bit b_i is embedded into e_{ij}, $e_{ij} \in C_c$ such that

$$e_{ij}^w = \begin{cases} e_{ij} & \text{if } b_i =' 0' \\ e'_{ij} - \Delta & \text{if } b_i =' 1' \text{ and } e_{ij} \in [-\Delta, 0] \\ e_{ij} + \Delta & \text{if } b_i =' 1' \text{ and } e_{ij} \in [0, \Delta[\end{cases} \tag{4.5}$$

Notice that embedding a message in the prediction-error domain implies: i) modulating the image pixels; and ii) the possibility of underflows or overflows in the clear domain. To address this problem, the embedding of an image overhead is required.

4.3 Robust Histogram Shifting

PEHS offers very interesting performance in terms of embedding capacity and image quality preservation. However, as most of reversible watermarking methods, it is fragile. Indeed, any modification of watermarked image prevents the correct extraction of the message and consequently the restoration of the original image. We detail in what follows the proposed Robust Histogram Shifting (RHS) considering the general case where HS is applied to any image characteristics (e.g. pixels, DCT coefficients) and the case where it is applied to image prediction-errors.

4.3.1 General Principle of Robust HS

The objective of our robust HS scheme is to embed a robust message. Message robustness is defined as the ability to correctly extract the embedded message after an image modification or alteration. In our work, if we look at an image sample watermarked using HS modulation (see Section 4.1.2), this one is said robust to an attack if it remains in the same class (carrier or non-carrier) that encodes the same bit value. In the case of a sample image attack, leading to the introduction of a distortion δ, three possible situations can happen:

(i) A message-bit error: where a carrier sample that encodes $'0'$ becomes carrier that encodes $'1'$, and vice versa.

(ii) A carrier deletion: where a carrier sample encoding $'1'$ is turned into a non-carrier sample.

(iii) A carrier injection: when a non-carrier sample becomes a carrier sample that encodes $'1'$.

Notice that carrier injections and deletions lead to a desynchronization between the message embedder and reader. In fact, in the case of a carrier deletion, the message reader will extract a message shorter than the embedded one.

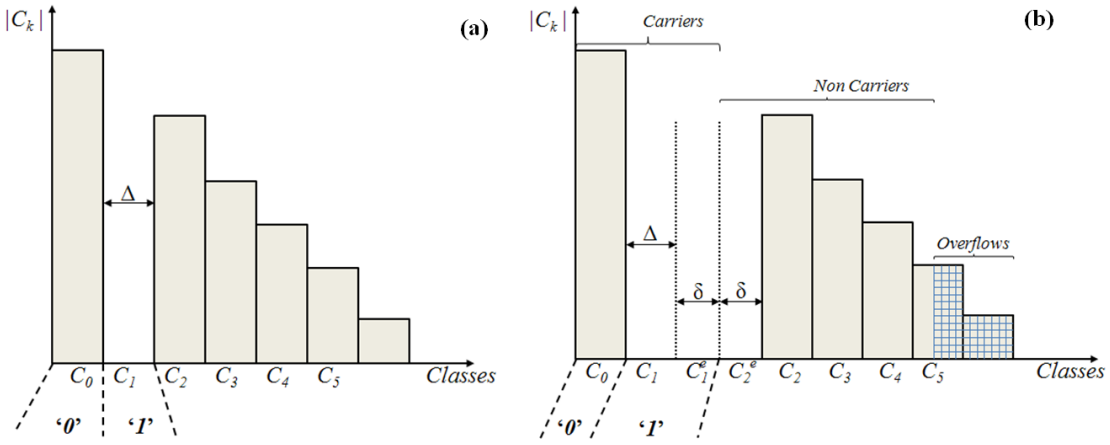


Figure 4.5: Histogram of the watermarked data in the case of: (a) the classic Histogram Shifting modulation, (b) our robust HS modulation.

To circumvent the desynchronization problems, supposing that the maximum distortion that can be introduced into an image sample is known and is of δ; where $\delta < \Delta$ and Δ is the shifting magnitude, the solution we propose consists in creating two gaps, both of size δ (C_1^e and C_2^e in Figure 4.5.b), in-between the carrier class that encodes $'1'$ (C_1 in Figure 4.5) and the non-carrier class (C_2 in Figure 4.5). By doing so, after an attack of maximum amplitude δ, a non-carrier sample remains non-carrier or falls in the gap near the non-carrier classes, i.e. C_2^e in Figure 4.5.b and a carrier sample encoding $'1'$ may fall in the gap near the carrier class encoding $'1'$, i.e. gap in C_1^e in Figure 4.5.b, or remain carrier but encoding $'0'$, leading thus to a message-bit error. Therefore, at the extraction stage, if a watermarked image sample belongs to the gap C_1^e (*resp.* C_2^e), it is considered as a carrier encoding $'1'$ (*resp.* non-carrier). Thus, this solution (i.e. the creation of the gaps) allows avoiding the desynchronization problems.

Regarding message-bit errors that can occur after an image attack, we propose to use an error-correcting code before message embedding to overcome this problem. Notice that our solution does not guarantee the error-free message extraction. Indeed, the detection/correction of message-bit errors depends not only on the correction ability of the adopted error-correcting code, but also on the statistical distributions of injected errors and carrier samples (i.e., the number of carrier samples should be greater than the one of injected errors). Notice that even if our strategy makes the message more robust against attack, it increases the size of the message overhead, which avoids the reader confusing underflow and overflow samples with non-carriers samples.

4.3.2 Robust Prediction-Error Histogram Shifting (RPEHS)

Similarly to HS, the main problems to solve in building robust PEHS are to counteract synchronization and message bit error issues. The solution we adopted to circumvent these issues is similar to robust HS in the spatial domain. Indeed, to overcome the synchronization problem, two gaps of size 2δ ($[2\Delta, 2(\Delta+\delta)[$ and $[-2(\Delta+\delta), -2\Delta[$ in Figure 4.6(a)) are created between the carrier and non-carrier classes. In order to avoid the message-bit error problem, we also propose to encode the message using an error-correcting code before its embedding. To make our proposal clearer, let us detail how message embedding is conducted. As illustrated in Figure 4.6, each non-carrier prediction-error e_{ij} is shifted of $+/-2(\Delta+\delta)$ as follows:

$$e_{ij}^w = \begin{cases} e_{ij} - 2(\Delta+\delta) & \text{if } e_{ij} < 0 \\ e_{ij} + 2(\Delta+\delta) & \text{if } e_{ij} > 0 \end{cases} \tag{4.6}$$

To embed a bit b_i of the message M, a carrier prediction-error e_{ij} is modulated such that:

$$e_{ij}^w = \begin{cases} e_{ij} & \text{if } b_i =' 0' \\ e_{ij} - \Delta & \text{if } b_i =' 1' \text{ and } e_{ij} \in [-\Delta, [0 \\ e_{ij} + \Delta & \text{if } b_i =' 1' \text{ and } e_{ij} \in [0, \Delta[\end{cases} \tag{4.7}$$

Note that M corresponds to the concatenation of the message overhead M_{over} and a message

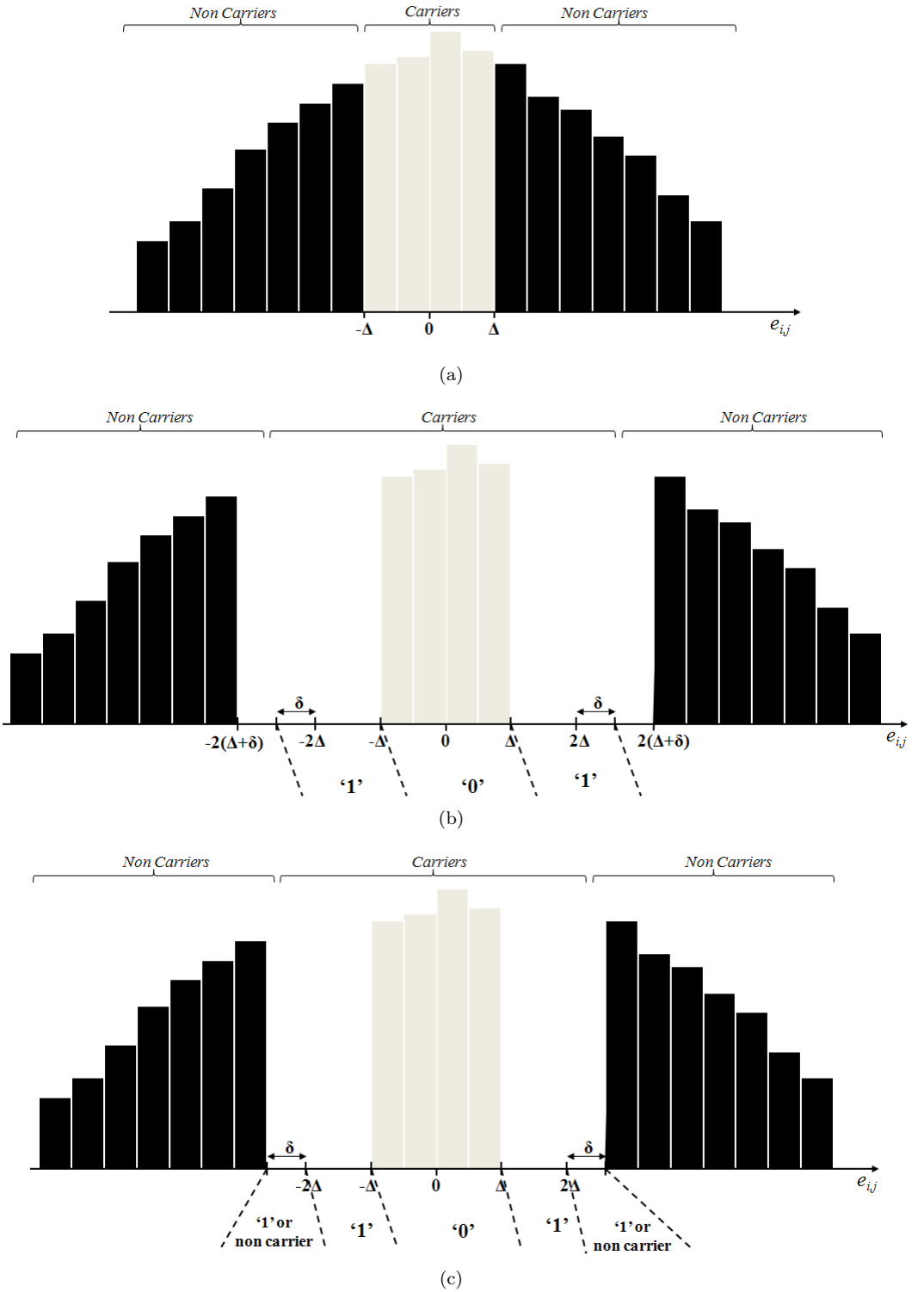


Figure 4.6: Basic principle of Robust Prediction-Error Histogram. (a) original prediction-error histogram (b) histogram of the watermarked prediction-errors (c) histogram of the watermarked prediction-errors when the attack is known and corresponds to the permutation of the LSBs of image pixels.

M_s encoded with an error-correcting code err_code:

$$M = err\_code\left(M_{sg}\right), \quad \text{where} \quad M_{sg} = M_{over} \parallel M_s \tag{4.8}$$

Let us now consider $\hat{e}_{ij}^w$ a possible attacked version of e_{ij}^w, $\hat{e}_{ij}^w = e_{ij}^w + / - \theta$, where $\theta \leq \delta$. Hence,

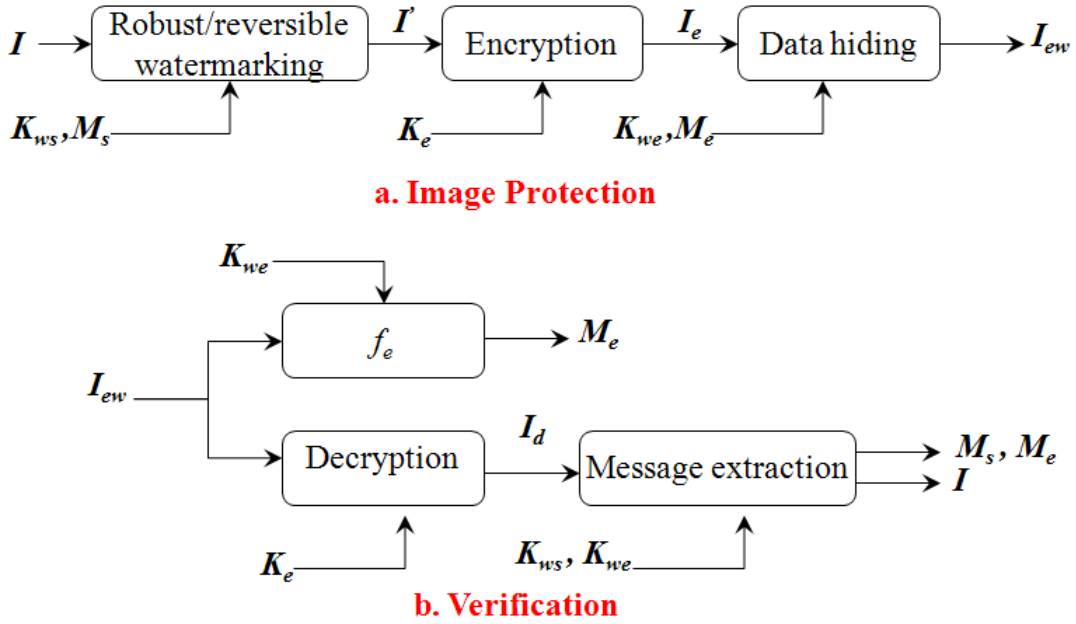


Figure 4.7: General Architecture of the system associated to our approach. I, I_w, I_e, I_{ew}, I_d, K_e, K_{ws} and K_{we} denote the original image, the watermarked image, the encrypted image, the encrypted-watermarked image, the watermarked-decrypted image and the encryption key and watermarking keys in both clear and encrypted domains, respectively. M_s and M_e are the embedded messages in the clear and encrypted domains, respectively. f_e is the watermarking extraction function in the encrypted domain.

the detected bit $\hat{b}_i$ is given by

$$\hat{b}_i = \begin{cases} 1 & \text{if } \hat{e}^w_{ij} \in [-2\Delta - \delta, -\Delta[\cup[\Delta, 2\Delta + \delta[\\ 0 & \text{if } \hat{e}^w_{ij} \in [-\Delta, \Delta[\end{cases} \tag{4.9}$$

The extracted message $\hat{M}$ is then decoded and corrected so as to obtain M_{sg}

Note that creating gaps introduces a distortion of 4δ into the non-carrier image prediction-errors. This distortion can be minimized if the attack or the type of the distortion is known. In our case, as the introduced distortion corresponds to the permutation of image pixels' LSBs by RPEHS message embedding, it can be minimized of 2δ. In fact, instead of creating two gaps of 2δ between carrier and non-carrier classes, we will only create two gaps of size δ in $[-2\Delta - \delta, -2\Delta[$ and in $[2\Delta, 2\Delta + \delta[$, as given in Figure 4.6(c). By doing so, after an attack (i.e. a permutation of image pixels' LSBs), a non-carrier prediction-error and a carrier encoding $'1'$ can fall in the same gap ($[-2\Delta - \delta, -2\Delta[$ and $[2\Delta, 2\Delta + \delta[$ in Figure 4.6(c)). Hence, the watermark reader only has to permute the LSB of the corresponding pixel in order to obtain the corrected prediction-error.

4.4 Reversible Watermarking of Encrypted Images

4.4.1 Basic Principles and General Architecture

Our approach allows embedding within an encrypted image a message M_e that is available in both encrypted and clear domains (i.e. in the encrypted and decrypted images). To do so, as illustrated in Figure 4.7, its principle is based on the insertion into the image I of a predefined watermark, a pre-watermark W before image encryption process. This insertion process is conducted by embedding a binary message M_s using our robust and reversible watermarking scheme and a secret watermarking key K_{ws}. The least significant bits of the pixels of the obtained pre-watermarked image I' correspond to the pre-watermark W. I' is then encrypted into I_e. A message M_e is then embedded into I_e using a common watermarking scheme and a watermarking key K_{we}, leading to the watermarked encrypted image I_{ew}. By definition, M_e can be extracted from I_{ew} using an extraction function f_e. Furthermore, the message M_e is available from the decrypted image I_w

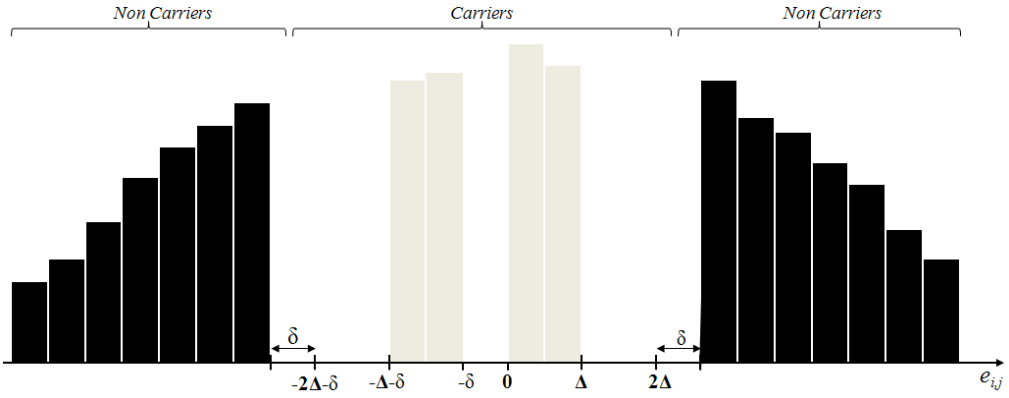


Figure 4.8: Basic principle of our RPEHS modified for ensuring the exact recovery of the original image.

based on the fact that the insertion of M_e modifies the embedded pre-watermark W. In fact, due to the robustness of M_s, the extracted pre-watermark W_{ext} can be corrected to recover the original W. The differences between the recovered pre-watermark W_r and the extracted one allow encoding M_e into the decrypted image and the extraction of M_e during the verification process from the decrypted image I_d. Moreover, as M_s is embedded using a reversible watermarking scheme, it can then be removed to restore the original image. As it can be seen, in our approach, message extraction and image decryption are considered as two independent processes. It is important to notice that all of pre-watermark embedding, encryption and data hiding processes have not to be conducted at the same time, i.e. jointly. As example, images can be pre-watermarked and encrypted and stored before being watermarked at a later time.

In the following, we present one implementation of our approach based on a stream cipher algorithm and our RPEHS (see Section 4.3.2).

4.4.2 Message Embedding using RPEHS

As said above, this step allows embedding into an image I a pre-watermark W by inserting a binary message M_s robust against the insertion of a second message M_e into the image after encryption using the LSB substitution modulation. With this technique, the image distortion is limited to the permutation of the image LSBs and the value of $\delta = 1$. Under this constraint, our objective is to embed W using the robust PEHS so that, after the insertion of M_e , the message M_s and the original image can be correctly recovered. As seen in Section 4.3.2, with our RPEHS, the original image cannot be recovered after permuting the image LSBs. So, we propose to modify the RPEHS to ensure the original image recovery and the embedded message M_s, and consequently the embedded pre-watermark W. To do so, we shall ensure correcting the LSBs of the pixels that correspond to the non-carriers and carrier prediction-errors. To be able to correct the pixel LSBs that correspond to the non-carriers errors, we propose to encode those LSBs using an error correcting-code and insert the redundancy bits *redu* into the carrier prediction-errors. On the other hand, pixel LSBs that correspond to the carrier prediction-errors carry the message encoded with an error-correcting code. Nevertheless, the use of this code allows correcting the embedded message, but not the pixel LSBs that carry it. In fact, the carrier prediction-error of positive value encoding '0' can become negative, but remains encoding the same bit '0', and vice versa. This allows introducing an error into the image LSBs, but not into the embedded message. To circumvent this problem, we propose to create a gap of size $\delta = 1$ between the positive and negative carriers that encode '0'. To summarize the way the modified version of RPEHS works, let us consider the process the message embedder follows. After computing image prediction errors, it builds the message M_s to be embedded such that:

$$M = \text{err\_code}\left(M_{osr}\right), \text{ where } M_{osr} = M_{over} \| \text{ redu } \| M_s \tag{4.10}$$

where M_{over} corresponds to the message overhead, *redu* are the redundancy bits and M_s is the secret message encoded with an error-correcting code.

The message M is then embedded as follows. As illustrated in Figure 4.8, a non-carrier prediction-error $e_{ij}$is firstly shifted of $2(\Delta+\delta)$ if $e_{ij} < -\Delta$ and of $2\Delta+\delta$ if $e_{ij} \geq \Delta$. A carrier prediction-error e_{ij} of a negative value is also shifted of δ for creating a gap between the positive and negative carrier prediction errors. The shifted version e^s_{ij} of a prediction-error e_{ij} is thus given as follows

$$e^s_{ij} = \begin{cases} e_{ij} + 2\Delta + \delta & \text{if } e_{ij} \geq \Delta \\ e_{ij} - 2(\Delta+\delta) & \text{if } e_{ij} < -\Delta \\ e_{ij} - \delta & \text{if } -\Delta \leq e_{ij} < 0 \\ e_{ij} & \text{if } 0 \leq e_{ij} < \Delta \end{cases} \tag{4.11}$$

A shifted carrier prediction-error e^s_{ij}, $e^s_{ij} \in [0, \Delta[\cup[-\Delta-\delta, -\delta[$, is modulated as follows for embedding one bit b_i of M

$$e^w_{ij} = \begin{cases} e^s_{ij} & \text{if } b_i =' 0' \\ e^s_{ij} - \Delta & \text{if } b_i =' 1' \text{ and } e^s_{ij} \in [-\Delta-\delta, -\delta[] \\ e^s_{ij} + \Delta & \text{if } b_i =' 1' \text{ and } e^s_{ij} \in [0, \Delta[\end{cases} \tag{4.12}$$

Notice that modulating the prediction error e_{ij} implies adding $\left(e^w_{ij} - e_{ij}\right)$ to its corresponding pixel p_{ij}. The watermarked version of p_{ij}, p^w_{ij}, is thus computed as:

$$p^w_{ij} = p_{ij} + \left(e^w_{ij} - e_{ij}\right) \tag{4.13}$$

Furthermore, in order to improve the embedding capacity, we propose a preprocess that allows removing the over/underflows, and consequently the overhead to embed. This process is done before the robust PEHS message insertion. Considering in the image N_{over} and N_{und} pixels that if watermarked lead to an overflow or an underflow, two thresholds T_{min} and T_{max} can be identified such that

$$T_{\min} = \max_{n=1,\ldots,N_{und}} \left(\hat{p}_{ij}\right); T_{\max} = \min_{n=1,\ldots,N_{\text{over}}} \left(\hat{p}_{ij}\right) \tag{4.14}$$

where $\hat{p}_{ij}$ is the predicted value of p_{ij} (see Section 4.2.2.2). Therefore, the N_{over} pixels (resp. N_{und}) are preprocessed by subtracting $2\Delta+\delta$ (*resp.* adding $2(\Delta+\delta)$). By doing so, image pixels do not lead to an underflow or overflow if watermarked using our robust PEHS. In such a case, the message M (see Equation 4.10) is defined as

$$M = \text{err\_code}\left(M_{sr}\right), \quad \text{where} \quad M_{sr} = \quad \text{redu}\|M_s \tag{4.15}$$

At the extraction stage, the extractor needs to know the histogram shifting Δ, δ, T_{min}, T_{max}, and the parameters of the adopted error correcting code. Those parameters form the watermarking key K_{ws} that should be transmitted to the extractor enabling him to extract the message M_s and recover the original image.

Herein, our objective is to theoretically determine the embedding capacities in clear domain, i.e. the maximum size of the message M_s one can embed into the image. As stated above, this capacity depends on the number N_c of carrier prediction errors, those for which $|e_{ij}| < \Delta$. Indeed, the maximum size of M (see Equation 4.15), which corresponds to the total amount of bits that can be embedded into the image prediction errors is equal to N_c, i.e. $|M| = N_c.N_c$ can be calculated based on the probability density function of e_{ij}. Let us assume that e_{ij} follows a centered normal distribution $\mathcal{N}\left(0, \sigma_e^2\right)$ (see Figure 4.9). Therefore, the probability P_c that e_{ij} is a carrier is defined as:

$$P_c = \Phi\left(\frac{\Delta}{\sigma_e}\right) - \Phi\left(-\frac{\Delta}{\sigma_e}\right) \tag{4.16}$$

where $\Phi(.)$ is the cumulative distribution function of a normal distribution, defined as

$$\Phi\left(\frac{\Delta}{\sigma_e}\right) = \frac{1}{\sigma_e\sqrt{2\pi}} \int_{-\infty}^{\Delta} e^{\frac{-t^2}{2\sigma_e^2}} dt = \frac{1}{2}\left(1 + \text{erf}\left(\frac{\Delta}{\sigma_e\sqrt{2}}\right)\right) \tag{4.17}$$

Considering N_e the number of carrier and non-carrier prediction errors in an image of $L \times N$ pixels. As an image prediction error corresponds to the difference between a pixel and its predicted

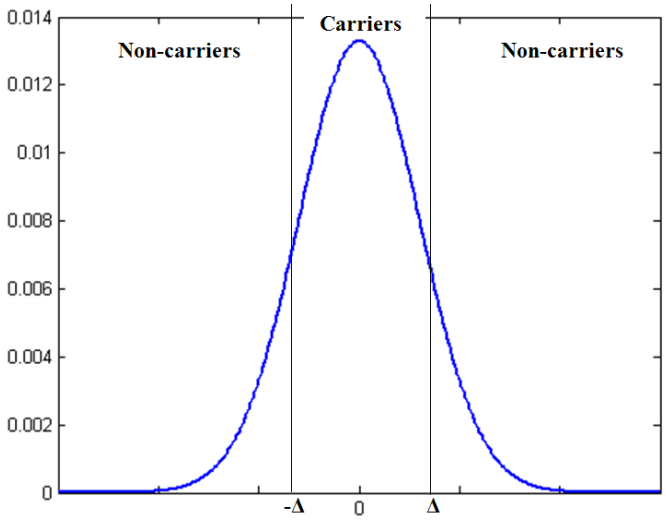


Figure 4.9: Image prediction error distribution.

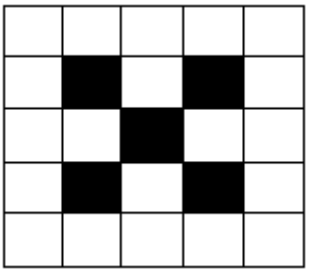


Figure 4.10: Positions (in black) of pixels our RPEHS on which is applied.

value derived from its four neighbor pixels, N_e is equal to the number of pixels the sum of the horizontal and vertical positions of which is even (see Figure). Therefore, we have:

$$|M| = N_c = N_e \times P_c \tag{4.18}$$

As given in Equation 4.15, M is formed by the concatenation of the secret message M_s and the redundancy bits of the LSBs of the non-carriers *redu*, $M_{sr} = redu \| M_s$), encoded with an error correcting code $C(n, k, d)$. The maximum size of M_{sr} is given by:

$$|M_{sr}| = \frac{N_c k}{n} \tag{4.19}$$

In order to get the maximum capacity of our system in the clear domain, i.e. the maximum size of M_s, the size of the redundancy bits *redu* resulting from encoding the LSBs of non-carriers prediction errors with C should be subtracted from $|M_{sr}|$. The size of *redu* is such that

$$|redu| = \left| \frac{N_{nc}}{k} \right| (n - k) \tag{4.20}$$

where N_{nc} is the number of non-carrier prediction errors calculated as:

$$N_{nc} = N_e - N_c = N_e \left(1 - P_c\right) \tag{4.21}$$

Therefore, the maximum size of M_s is given as:

$$|M_s| = |M_{sr}| - |redu| \tag{4.22}$$

4.4.3 Encryption Step

During this step, each watermarked pixel p_{ij}^w is encrypted into p_{ij}^e using a stream cipher algorithm (noted E) initialized with the encryption key K_e:

$$p_{ij}^e = E\left(p_{ij}^w, K_e\right) = p_{ij}^w \oplus k_i \tag{4.23}$$

where $[k_1, \ldots, k_i, \ldots, k_{l\times n}]$ is the secret keystream of bytes generated by the adopted stream cipher to encrypt the pre-watermarked image.

4.4.4 Data Hiding in The Encrypted Domain

This step can be made in a differed time. It enables a user to embed a message M_e into the encrypted image pixels $\left\{p^e_{ij}\right\}_{i=1,\ldots,l,j=1\ldots h}$ without knowing its original content. In the sequel, M_e corresponds to a sequence of bits $\{b_j\}_{j=1,\ldots,q}$ uniformly distributed. Let us consider $C(n,k,d)$ the error-correcting code adopted by our RPEHS.$(n,k,d)C$ acts on a block of k bits of input data to generate a codeword of n bits of output data. d is its minimum distance stating that the code C can detect $(d-1)$ errors and correct $d_c = \left\lfloor \frac{d-1}{2} \right\rfloor$ errors per codeword. In order to be able to correct the errors that will be introduced by this process in the clear domain, only d_c pixels per each set of n image pixels that are modified/protected by our robust PEHS are modulated according to the following process for embedding M_e. Firstly, the encrypted version of the N_e pixels that are modulated by our robust PEHS for embedding M_s are identified. Those pixels are then split into N_s non-overlapping sets of $N_{\text{modc}} = n^2$ pixels $\{I^e_i\}_{i=1,\ldots N_s}$.

One bit b_j of M_e is then embedded by set I_i as follows. I_i is firstly divided into two subsets of $N_{sub} = N_{\bmod c}/2$ pixels, I_{1i} and I_{2i}, where $I_{ki} = \left\{I^j_{ki}\right\}_{j=1,\ldots,N_{sub},k=1,2}$, I^j_{ki} represents the j^{th} pixel of the subset I_{ki}. b_j is then inserted into each subset so as I_{1i} (*resp.* I_{2i}) carries b_j in the encrypted domain (*resp.* the clear domain).

Bit insertion in the encrypted domain

b_j is embedded into I_{1i} by substituting the LSB of the d^{th} pixel I^{d1}_{1i} by b_j. The watermarked version I_1wi of I_{1i} is given as follows

$$I_{1wi} = \left\{I^j_{1wi}\right\}_{j=1,\ldots N_{sub}} \tag{4.24}$$

where

$$I^j_{1wi} = \begin{cases} I^j_{1wi} = I^j_{1i} \quad \text{if } j \neq d1 \\ I^{d1}_{1wi} = 2\left[\frac{l^{d1}_{1i}}{2}\right] + b_j \text{ otherwise} \end{cases} \tag{4.25}$$

The choice of the rank $d1$ depends on the secret watermarking key K_{we}. As a consequence, the extraction function f_e we use in the encrypted domain in order to extract one bit from one encrypted watermarked pixel subset I_{1wi} is such that

$$b_j = f_e\left(I_{1wi}\right) = LSB\left(I^{d1}_{1wi}\right) \tag{4.26}$$

Bit insertion in the clear domain

b_j is embedded into I_{2i} by modifying or not the LSB of one of its pixels according to the value of b_j. More clearly, if b_j equals 1, we modify the LSB of one randomly-selected pixel I^{d2}_{2i}. In such a case, the watermarked version of I_{2i}, I_{2wi}, is given by

$$I_{2wi} = \left\{I^j_{2wi}\right\}_{j=1,\ldots N_{sub}} \tag{4.27}$$

where

$$I^j_{2wi} = \begin{cases} I^j_{2wi} = I^j_{2i} \quad \text{if } j \neq d2 \\ I^{d2}_{2wi} = 2\left|\frac{I^{d_2}_{2i}}{2}\right| + \overline{LSB\left(I^{d2}_{2l}\right)} \text{ otherwise} \end{cases} \tag{4.28}$$

The choice of the rank $d2$ also depends on the secret watermarking key K_{we}.

Rather, if $b_j =' 0'$, I_{2i} is kept unchanged: $I_{2wi} = I_{2i}$.

The errors introduced by this insertion process into the LSBs of the decrypted version of I^j_{2wi}, I^j_{2di}, will give access to the message in the clear domain. Indeed, in the case where $b_j =' 1'$, one LSB of I_{2i} is modified for obtaining I_{2wi}. This implicitly commutes the corresponding bit of the pre-watermarked version of I_{2i}, and consequently introduces an error into I^j_{2di}. Therefore, if an error is detected into I^j_{2di}, the embedded bit is $'1'$, $'0'$ otherwise.

Regarding the embedding capacity in the encrypted domain, as mentioned above, in order to ensure recovering the embedded pre-watermark and the original image, only one bit of M_e is

embedded by set I_i of N_{modc} encrypted image pixels. Therefore, the maximum size of M_e equals to the number of image sets:

$$|M_e| = N_s = \frac{1}{2}(\frac{N_e}{N_{modc}}) \tag{4.29}$$

Regarding the algorithm complexity, the proposed RPEHS does not require high computational complexity because it is based on simple algebraic calculations: Equations (4.3), (4.11), (4.12), (4.13), (4.14), (4.15), (4.25), and (4.28).

4.4.5 Message Extraction & Image Recovery

As mentioned previously, unlike the pre-watermark M_s that can be only extracted from the clear domain, the message M_e can be extracted in both encrypted and clear domains.

4.4.5.1 Message extraction in the encrypted domain

As during the insertion process (see Setion 4.4.4), after identifying the N_{mod} pixels, those pixels are firstly split into sets of N_{modc} pixels, $\{I_{wi}\}_{i=1,...,N_s}$ according to K_{we}. Each set I_{wi} is then divided into two subsets of N_{sub} pixels, I_{1wi} and I_{2wi} according to the secret watermarking key in the encrypted domain K_{we}. Then, the extraction function f_e (Equation 4.26) is used to extract from each I_{1wi} one bit of the message M_e.

4.4.5.2 Message extraction in the clear domain

Extraction of the message M_s & image recovery

In addition to the extraction of the message M_s and the recovery of the original image, this step allows constructing a binary vector map V that indicates if there is an error in the image pixels or their prediction-errors, and consequently in the embedded pre-watermark W. We will see in the next subsection how this map will be used to extract the message M_e.

Let us first detail how the message M_s is extracted and the original image is reconstructed. To do so, the watermark reader computes the histogram of prediction errors of the decrypted image. As a first step, he corrects the prediction errors that fall in the 3 gaps, i.e. $[-2(\Delta+\delta), -2\Delta-\delta)[\cup[-\delta, 0[\cup[2\Delta, 2\Delta+\delta[$. As the image modification is limited to the permutation of the image pixel LSBs, if a prediction error falls in a gap, this means that the LSB of its corresponding pixel is permuted. Therefore, to correct a prediction error $\hat{e}_{ij}$ of a pixel p_{ij} that belongs to a gap, the extractor permutes the LSB of p_{ij} and re-computes its prediction-error. This latter is then attributed to $\hat{e}_{ij}$. Moreover, a bit '1' is attributed to V_{ij} in order to indicate that the pixel p_{ij} has been modified during the message M_e insertion in the encrypted domain. After correcting the image prediction errors, the attacked version $\hat{M}$ of M is extracted. In fact, one bit $\hat{m}_i$ of $\hat{M}$ is extracted from each carrier prediction error $\hat{e}_{ij}$, $\hat{e}_{ij} \in [-2\Delta-\delta, 2\Delta[$ as follows:

$$\widehat{m}_i = \begin{cases} 0 & \text{if } \hat{e}_{ij} \in [-\Delta-\delta, \Delta[] \\ 1 & \text{otherwise} \end{cases} \tag{4.30}$$

The extracted message $\hat{M}$ is then decoded using the adopted error-correcting code (see Equation 4.15). During this decoding step, if a message bit error is detected, this means that there an error in the pixel p_{ij}. To correct it, the LSB of p_{ij} is permuted. At the same time, $'1'$ is attributed to V_{ij} to indicate that there was an error in the pixel p_{ij}. Meanwhile, knowing the non-carrier prediction-errors and the parameters of the error correcting code, the extractor can retrieve the size of the redundancy bits *redu* and extracts then it along with M_s from the decoded and corrected value of the message M.

The next step consists in correcting the LSBs of the pixels that correspond to the non-carrier prediction errors. This is done based on *redu*. Again, note that when an error is detected in the LSB of a non-carrier pixel p_{ij}, $'1'$ is attributed to the binary map V_{ij}. After correcting those LSBs, the non-carrier prediction errors are re-computed.

Finally, the original prediction error histogram is built by shifting back the prediction errors as follows:

$$e_{ij}^{o} = \begin{cases} \hat{e}_{ij} + 2(\Delta+\delta) \text{ if } \hat{e}_{ij} < -2(\Delta+\delta) \\ \hat{e}_{ij} + \delta \text{ if } -\Delta-\delta \leq \hat{e}_{ij} < -\delta \\ \hat{e}_{ij} - \Delta \text{ if } \Delta \leq \hat{e}_{ij} < 2\Delta \\ \hat{e}_{ij} - 2\Delta - \delta \text{ if } \hat{e}_{ij} \geq 2\Delta + \delta \end{cases} \tag{4.31}$$

Then, based on the knowledge of T_{min} and T_{max}, the extractor identifies the pixels that are pre-processed and corrects them by applying the reverse process.

Extraction of the message M_e

The errors introduced into the embedded pre-watermark M by the insertion process of M_e into the encrypted domain give access to M_e in the clear domain. In fact, as explained in section 4.4.4, one bit of M_e is embedded into each subset of pixels I_{2wi} by introducing an error within one of their LSBs. As the encryption algorithm used here is of stream cipher type, the modification of the LSB of one encrypted pixel implicitly commutes the corresponding bit of its decrypted version. Therefore, if one bit error is detected into the decrypted version I_{2di} of I_{2wi}, then $b_j =' 1'$, $'0'$ otherwise. In order to verify if there are bit errors in I_{id}, we exploit the binary map V produced by the previous step and the watermarking key K_we. Based on K_we, the N_{sub} bits of V that correspond to the pixels of I_{2di} are selected $\{\zeta_i\}_{i=1,\ldots,N_{sub}}$. The bit ζ_i indicates that there is error in the LSB of the i^{th} pixel of I_{id} if its value equals $'1'$.

Therefore, a bit b_j of M_e is extracted as follows:

$$b_j = \begin{cases} 0 \text{ if } \sum_{i=1,\ldots N_{sub}} \zeta_i = 0 \\ 1 \text{ else} \end{cases} \tag{4.32}$$

4.5 Experimental Results

These tests have been performed over several natural gray-scale images (an example of image data set in Figure 4.11) and medical image test sets issued from different modalities: 320 mammography images [104] and over 100 ultrasound images (see Figure 4.12).

4.5.1 Capacity Rates

As already mentioned in Section 4.4.2, capacity rate in clear domain depends on the distribution of carriers in the image ($|N_c|$), the shifting magnitude (Δ) and the error-correcting code parameters (n, k). Note that experimental embedding capacities are equal to the theoretical ones as messages are only embedded in predefined locations and based on the used system's parameters (Δ, n and k).

The selection of the value of Δ actually depends on the distribution of the carriers and the non-carriers. In other word, it depends on the image properties. We give in Figure 4.13, the prediction-errors' histograms in "Airplane" and "Peppers" images. It can be seen that the distribution of the prediction-errors in "Airplane" is different from the one in "Baboon". In this later, the distributions of prediction-errors are close to each other whereas in Airplane, the histogram is significantly concentrated around 0. This can be explained by the fact that in image with more smooth areas, the consistency of the adjacent pixels in a local area is strong. Thus, higher accuracy prediction-error around $'0'$ is obtained, and |carriers| >> |non-carriers|. As a consequence, Δ can be of a small value. However, this is not the case for textured images where the shifting magnitude Δ has to be of bigger value so as to be able to embed the redundancy bits of non-carriers and consequently extract M_s and M_e with no error and correctly recover the original image. This is can be experimentally confirmed. As given in Table 4.2, the watermarking capacity in textured images is lower than the one in images with more flat/smooth areas.

Furthermore, as given in Figure 4.14 with a larger value of the shifting magnitude Δ, a higher embedding rate in the clear domain can be ensured. For instance, the achieved capacities of M_s are equal to 7857 and 40015 bits when $\Delta = 2$ and $\Delta = 4$, respectively. Note that only M_s depends on the shifting magnitude Δ whereas M_e depends on the error-correcting code parameters and the image size.

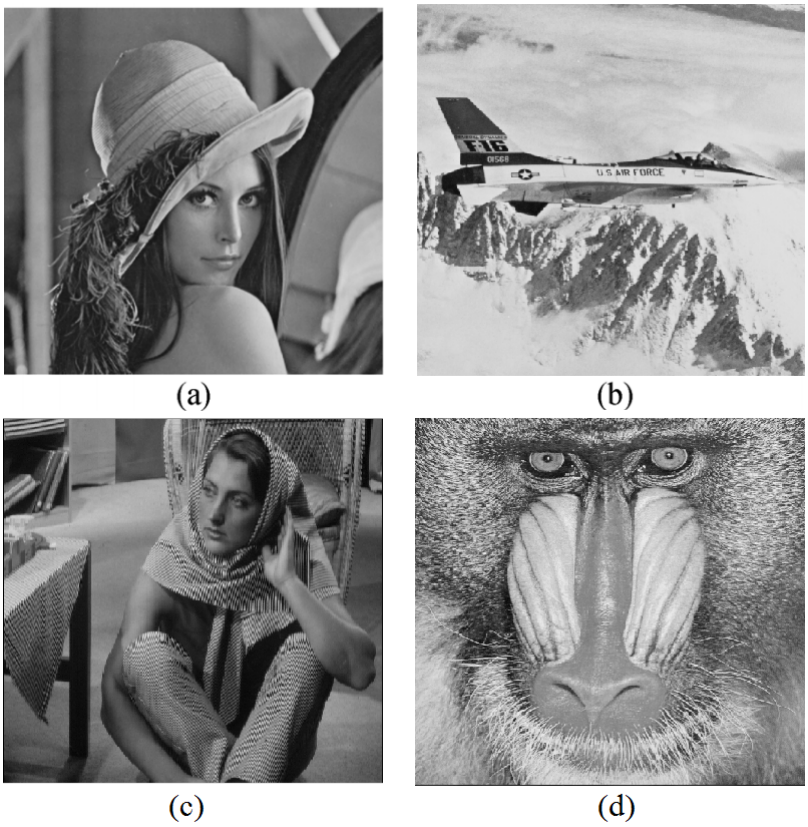


Figure 4.11: Natural test images of 512×512 pixels : (a) Lena, (b) Airplane, (c) Barbara, (d) Baboon.

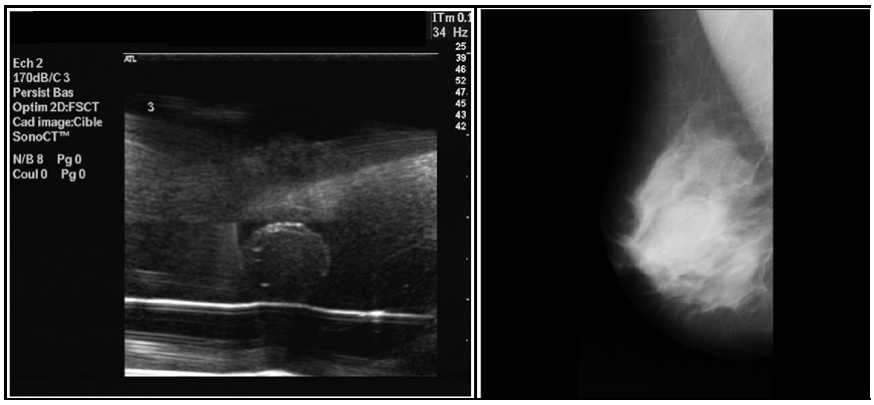
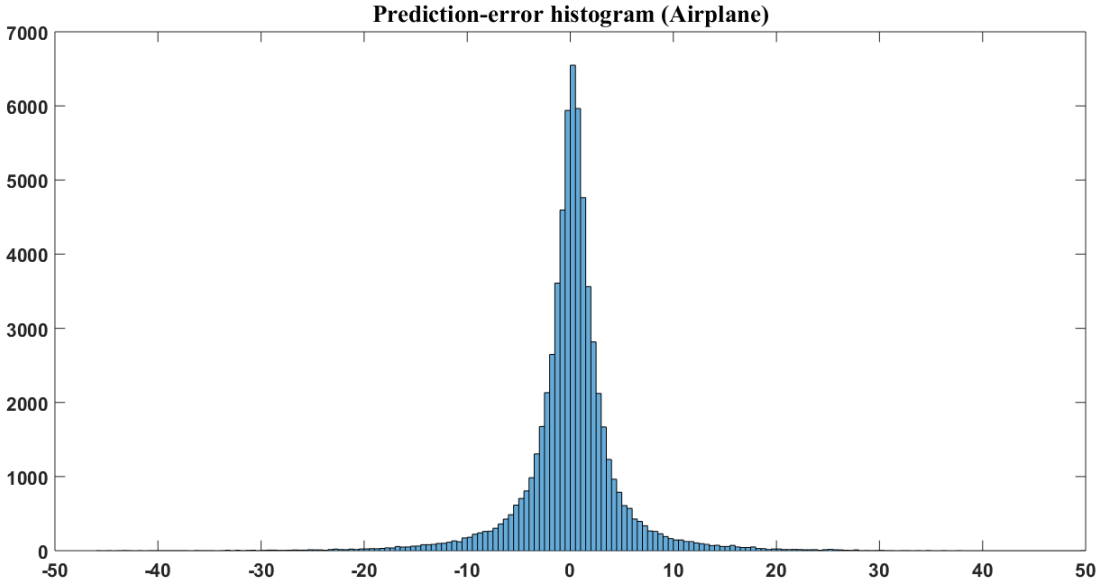
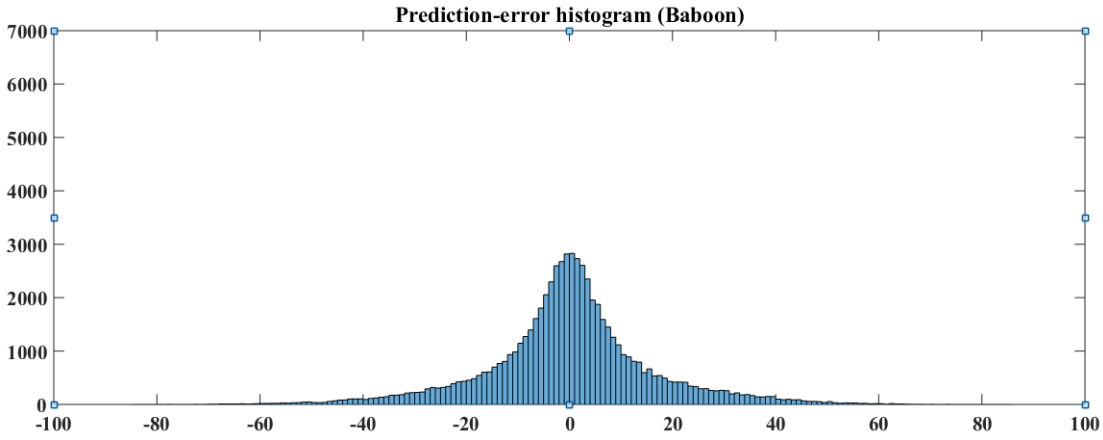


Figure 4.12: Samples of our image test sets (ultrasound, mammography, respectively).

Let us now consider the impact of the error-correcting code parameters on the proposed approach. Figure 4.15 gives an idea about the effects of these parameters (i.e. n and k), on the watermarking capacity in both encrypted and decrypted domains. Obviously and as shown in Fig. 10, $|M_e|_{cr(7,4)} > |M_e|_{cr(15,11)}$ but $|M_s|_{cr(7,4)} < |M_s|_{cr(15,11)}$, where $cr(.)$ corresponds to the error-correcting code function and $|x|$ gives the length in bits of x. Hence, when n and k are of small values, our system is able to reach a higher embedding rate in the encrypted domain as one bit of M_e is embedded into each block of n^2 pixels (see Section 4.4.4). But, in the clear (i.e. decrypted) domain, message embedding increases when $k > d$, or in other words, when the length of redundant bits decreases. Nevertheless these parameters should be adjusted in order our system remains robust against the embedding in the encrypted domain.



(a)



(b)

Figure 4.13: Distribution of prediction-errors in Airplane (a) and Baboon (b) images.

4.5.2 Image Distortion

As we propose in this chapter a reversible scheme the original image can be recovered without any distortion, after extracting the embedded watermark. Nevertheless, the visual quality of the watermarked-decrypted image should be evaluated. The Peak-Signal-to-Noise-Ration (PSNR), the structural similarity (SSIM) and the Universal Quality Measure (UQI) between the original image I and its decrypted-watermarked version I_{wd} were considered so as to evaluate the image distortion. On its side, SSIM and UQI are considered to be correlated with the human visual system. As exposed in Section 4.4, the distortion introduced by our RPEHS is obviously related to the shifting of pixels' values.

The performance of our RPEHS in terms of distortion under different values of shifting magnitude (Δ) on "Airplane" image is given in Figure 4.16. Obviously, the bigger Δ is, the lower the achieved PSNR value is or, in other words, the larger the introduced distortion is. For instance, for "Airplane", when $\Delta = 2$, the obtained PSNR value is of 41 dB while, the obtained PSNR between the original and the watermarked-decrypted image is equal to 38.5 dB by setting $\Delta = 4$.

Table 4.3 shows the PSNR values and the correlation coefficient between the encrypted and the decrypted images. PSNR values are very low (near to 0) so it is rather difficult to detect the content of the original image from its encrypted counterpart. The correlation coefficients between encrypted original images are almost zero, so no information can be recognized from the encrypted

Table 4.2: Embedding rate comparisons of the proposed method with some the state of the art methods

Embedding capacity (bpp)	[154]	[155]	[150]	[109]	*Proposed*	
					M_e	M_s
Lena	0.07	0.062	0.15	0.004	0.01	0.17
Barbara	0.05	0.001	0.15	0.001	0.01	0.088
Airplane	0.15	0.039	0.19	0.001	0.01	0.29
Peppers	0.06	0.009	0.15	0.004	0.01	0.075

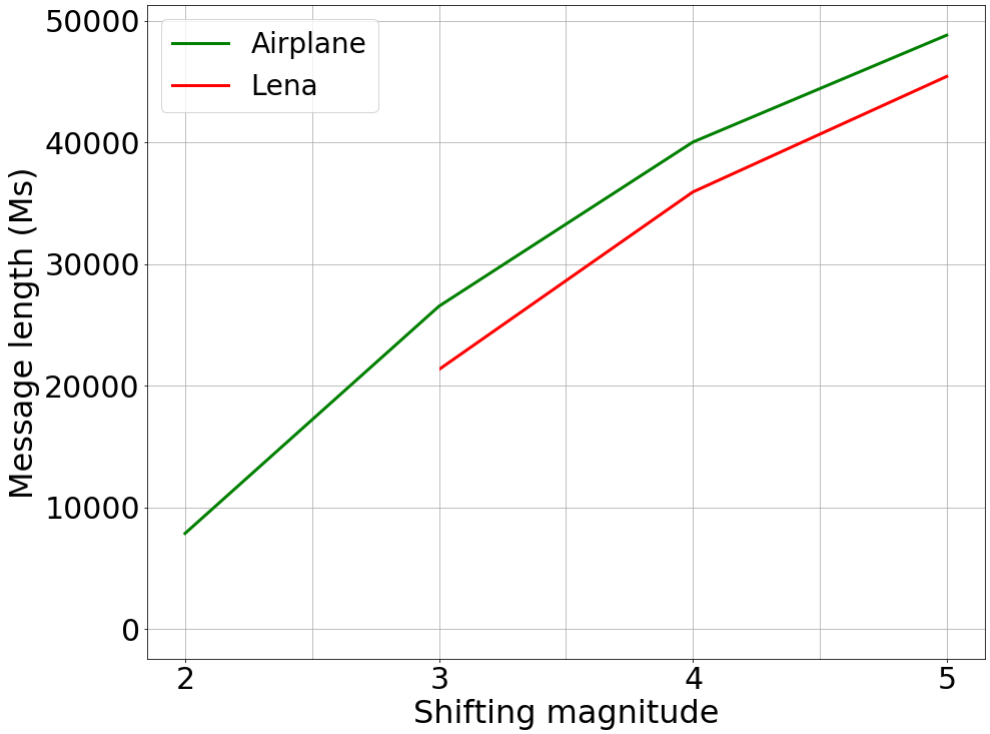


Figure 4.14: Tests on "Airplane" and "Lena" images: Embedding rate in clear domain depending on the shifting magnitude Δ values.

images without decrypting the image.

Figure 4.17 shows the encrypted and decrypted images of “Lena” image that were generated by the proposed RPEHS. Figure 4.17(c) is the watermarked-decrypted image, in which the PSNR value is of 41 dB and the SSIM is of 0.998. The reconstructed image is given by Figure 4.17(d), which is completely the same as the original image in Figure 4.17(a). As it can also be seen, there are no visible differences between the watermarked-decrypted image (in Figure 4.17(c)) and the original one (in Figure 4.17(a)) which means the embedded secret data are imperceptible.

For all test images, obtained values of SSIM and UQI are greater than 0.99 and 0.89, respectively, confirming the good quality of watermarked images [105].

For medical images, the results reported in Table 4.4 are somewhat equivalent to those obtained for standard test gray-scale images. Both embedding capacity and introduced distortion increase with the increase of the shifting magnitude (Δ). Beyond, the degree of the distortion in these images is small and does not endanger their diagnosis value. One can refer to the study on the impact of lossy image compression on medical images in [106], where it is reported that image

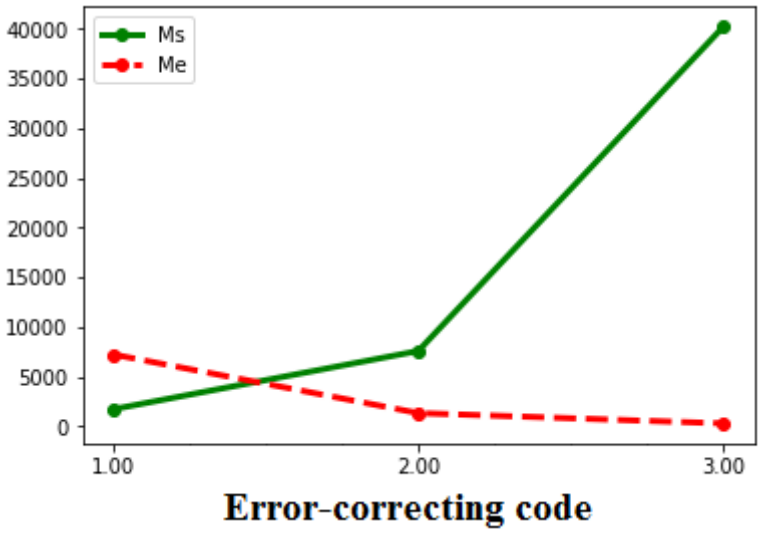


Figure 4.15: Tests on "Airplane" image: Embedding rate in both encrypted and clear domains depending on the error-correcting code parameters where "1", "2" and "3" corresponds to the repetition code (3, 1), Hamming (7,4) and Hamming (15,11), respectively.

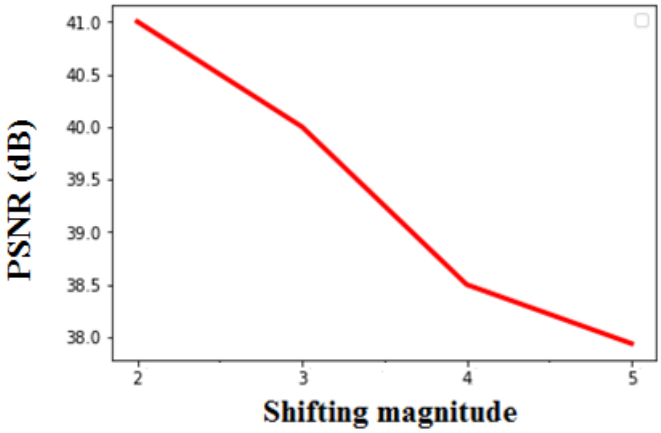


Figure 4.16: Tests on "Airplane" image: PSNR according to the shifting magnitude (Δ).

distortion should be maintained in the range of 40 dB and 50 dB to establish good diagnosis.

4.5.3 Capacity-Distortion Performance Comparison

In this Section, we provide some comparisons between the proposed method and different techniques from state-of-the-art that allow extracting the secret data and recovering the original image without any loss of information. We also selected 5 standard natural images which are "Lena", "Barbara", "Airplane", "Peppers" and "Boat" in order to evaluate their embedding rates and image distortion. As given in Table 4.2, compared to [109, 150, 154, 155], our method achieves better embedding rate, especially when the test image contains less textured regions (e.g. boat and airplane). For instance, in Lena, M_e – message available in both encrypted and decrypted domains – is of 1328 bits and M_s (i.e. message available in the decrypted image) is of 21326 bits. In the case of Lena test image, the introduced distortion quantified with PSNR is of 39.8 dB that can be considered as an acceptable value while in [109] for example, 256 bits are embedded into the encrypted image and with a larger quality distortion ($PSNR = 37.9dB$). Note that all these techniques carried the message in only one domain; i.e. the message can be extracted from either the plain-text image or the encrypted one. For example, in [155], data extraction must be carried out before image decryption so as to be able to decrypt the image and recover its original version.

Table 4.3: PSNR and correlation coefficients between encrypted and plain-text images

Image	Peppers	Airplane	Baboon	Lena
PSNR (dB)	8.876	8.01	9.54	9.46
Correlation coefficient	$8.67.10^{-4}$	$-2.21.10^{-4}$	0.0012	$4.63.10^{-4}$

Table 4.4: Embedding rate of medical images when the message is encoded with Hamming (7, 4)

Images	Shifting magnitude	Embedding capacity (bpp)		PSNR (dB)
		M_e	M_s	
Ultrasound	$\Delta = 1$	0.01	0.1	43.6
	$\Delta = 2$	0.01	0.27	40
Mammography	$\Delta = 1$	0.01	0.12	44.2
	$\Delta = 2$	0.01	0.29	41.05

4.6 Discussion & Possible Applications

If we consider the scenario given in Figure 4.18 where a content owner outsources certain images, already encrypted and pre-watermarked, to a cloud server, this later – who is not granted to access to the image content – can verify images' reliability and at the same time, embed into the encrypted images some additional data (e.g. image owner and her/his identifiers, time stamps). On the other side, images can be downloaded and decrypted by a legal user holding the decryption key. Hence, original image content can be recovered without any error after image decryption and message extraction processes using the secret watermarking key. Notice that it is very important to extract the additional data (M_e), embedded by the cloud server, from the decrypted images in order to trace the data. This can be helpful in case of illegal redistribution or of image content falsification. Similarly, the embedded pre-watermark in the clear domain by an authorized user (M_s) are used to control the reliability of data and trace them.

The proposed scheme can be used in different applications. For instance, in healthcare, medical data are required to be reliable so as to be used in complete trust by the practitioner for diagnosis or therapeutic purposes. Non-reliable data should be rejected by the medical information system [56]. In this case, if the purpose of the healthcare application is just the verification of data authenticity (i.e. embedding of the patient ID and/or image ID) then the capacity requirements are not so important. But, if the purpose is to allow various security services ranging from integrity to traceability, then the capacity needs substantially increase. Fortunately, the offered capacities, in both domains, are large enough compared to the requirements for verifying image reliability (i.e. authenticity and integrity control) that we estimate less than 1 Kbits (i.e. a digital signature provided by the SHA-1: 160 bits and one authenticity code AC: 600 bits by combining the French National Identifier with the DICOM Unique Identifier [133]).

Consequently, we opted for the embedding, in both domains, of a message composed of the identifier of the image origin along with a digital signature; an integrity proof so as to be able to determine whether the image has been modified during its transmission. So, the message available either in the clear or encrypted domain, T, is such that:

$$T =< AC \| RSA_{K_pr}(H(f(I)) > \tag{4.33}$$

where: '||' is the concatenation operator, H corresponds to the SHA cryptographic hash function, $f(.)$ is a function that identifies pixels that will not be modified by the watermarking process, RSA is the well-known asymmetric encryption algorithm parameterized with the private key K_pr of the image provider.

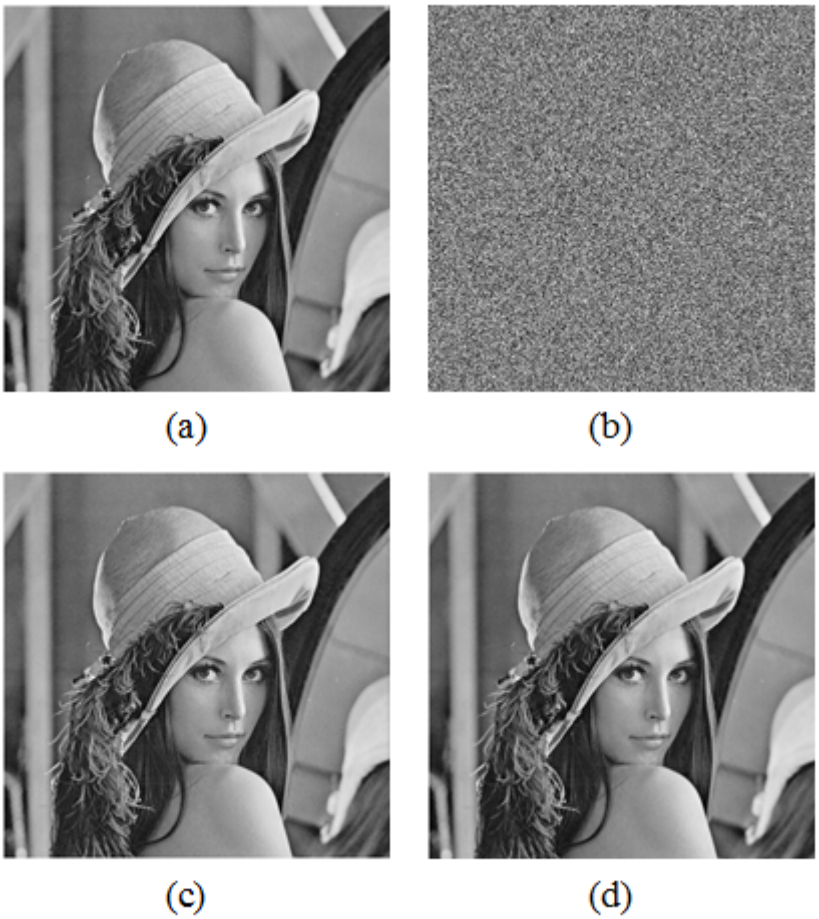


Figure 4.17: Simulation results of the proposed scheme on "Lena": (a) original image; (b) watermarked-encrypted image, with $\Delta = 3$ and message encoded with Hamming (7, 4) the error correction code; (c) decrypted-watermarked image where $PSNR = 41\ dB$ and $SSIM = 0.998$; (d) reconstructed image.

In the following, we will discuss the security of this scheme in terms of confidentiality and reliability. We first start by analyzing our method considering cryptographic attacks; which aims at breaking data confidentiality, and then focus on watermarking attacks.

In this work, encryption operations are performed using the Trivium algorithm. Trivium offers a security level of 80 bits. Because we only exploit its semantic security, there is no access to private parameters like private key and user's data. Indeed, even if K_{we} or m_e are known by the attacker, he/she has no additional means than a regular cryptographic attack to get K_e or to have an idea about the clear image content.

There exist different types of watermarking attacks: unauthorized detection/extraction of messages, unauthorized embedding, and unauthorized removal attack. In both encrypted and clear domain, the message embedding and the reliability verification of the image depend on the corresponding watermarking keys K_{we} or K_{ws}. Without the knowledge of these secret keys, it is extremely difficult for an attacker to know the image partitioning and to distinguish the bits of messages from the others. Even though the attacker has an idea about the message structure, he/she can only try an exhaustive search until he/she finds a valid message; which does not mean that it corresponds to the image. Furthermore, as our scheme is fragile, if the attacker tries to modify arbitrarily or falsify the watermark, an alarm will be raised in both encrypted and clear domains thanks to the message M_e and M_s that will be altered too. Hence, valid messages should then be present in all images in their clear and encrypted domains to ensure the image reliability.

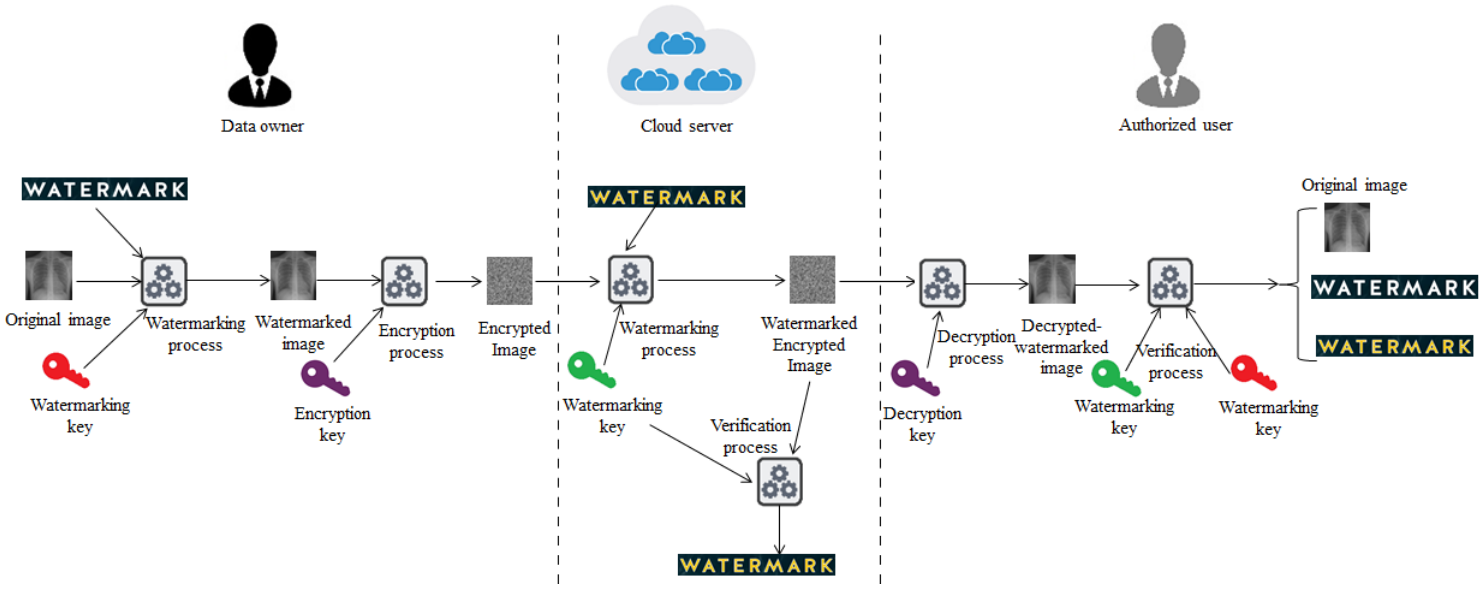


Figure 4.18: Example of a scenario using the proposed scheme.

4.7 Conclusion

In this chapter we have presented a new lossless watermarking scheme that gives access to different watermarking-based security services in both encrypted and clear domains and allows the error-free reconstruction of the original image. In fact, it allows the embedding, in the encrypted image, of a message that will be accessible from both clear and encrypted domains. The originality of our proposal stands on the insertion of a robust pre-watermark in the image before its encryption, that will make the message embedded in the encrypted image accessible from the decrypted image and makes the insertion/extraction processes independent of the encryption/decryption processes, and vice versa. Experimental results show that the achieved capacities in both domains are large enough to support various security services.

As noticed, in reversible watermarking schemes once the watermark is removed, the image is unfortunately no longer protected as for cryptography. That is why non-reversible watermarking is of high interest. However, embedding a watermark into a medical image, in an irreversible manner, can unfortunately sometimes introduce, or hide pathology. The consequences of this risk are not negligible, as they concern an individual and his health. That is why we propose in the next chapter to evaluate the introduced quality degradations caused by watermarking in order to identify a range of watermarking parameters that could preserve the image diagnostic and visual quality.

Chapter 5

Assessment Protocol of Watermarking Impact on CT Images

As mentioned earlier, the choice of watermarking scheme depends on tradeoff in-between different properties so as to respond the intended application, and the pursued security purposes by just focusing on the three main basic properties: watermarking capacity, robustness and imperceptibility. For instance, if data authenticity and traceability are of major interest in an application framework where an attacker could falsify the data so as to break its protection, the watermarking scheme should be robust against such a modification in order to ensure the correct detection/extraction of the embedded message. Considering another example, where watermarking is used to control data integrity with the ability to detect, localize and restore tampered parts of the image, such an application will need huge watermarking capacity. These both situations will obviously have an impact on the watermark imperceptibility. As said in Chapter 1, imperceptibility is very important for medical images as these later hold decisive information about patient's health. Thus, watermarking should not compromise the image interpretation. Defining the acceptable amount of distortion that can preserve the diagnosis accuracy of the watermarked image is a complex problem and one of the main issues in the deployment of watermarking in medical imaging.

Medical image quality can be assessed and analyzed depending on two criteria [161]: fidelity and intelligibility. The former one corresponds to the perceptual similarity between the watermarked and the original images. The latter relies on the preservation of the same diagnostic interpretation of the image. These two metrics demonstrate the complexity of such assessment that requires the intervention of radiologists. At the same time, the automatization of watermarking distortion-control is also an open debate. If solutions based on psychovisual masking exist for watermarking of general public images (i.e. natural images), no equivalent approach has really been proposed for medical imaging data or even for simulated medical images such as phantom database [162]. Nevertheless, one can refer to the work of Karasad *et al.* [76] that allows hiding a watermark in the sensor noise of X-Ray images. However, this method is fragile due to the fact that sensor noise is of low amplitude. There is thus an interest to develop a solution that better controls image distortion in order to increase watermark robustness and/or capacity.

Assessment of the quality of watermarked images is performed by means of both objective and subjective measures. Objective metrics are based on the use of quantitative quality measures (e.g. PSNR, SSIM, UQI) and have often been used in medical image watermarking literature. However, whether an objective measure on image quality is efficient or not, it depends strongly on its accordance with subjective measures. These later are run by a group of radiologists who rate the diagnostic quality as well as the visual quality of the watermarked image so as to assess the legal risk of medical image watermarking. Unfortunately, this type of studies is difficult to arrange [25]. It is complicated to give a standard amount of distortion. Indeed, this amount does not only depend on the watermarking technique, but also on the image characteristics being studied that is to say: the acquisition modality and its parameterization, the anatomical object, the different expression of the possible lesions that could be seen, and so on. Furthermore, this type of study is time-consuming and expensive. That is why few subjective quality assessment methods of medical images have been done in literature and most of these studies [27, 106, 163, 164] evaluate the distortion caused by lossy compression techniques and, to the best of our knowledge, only two studies [161, 165] are done to assess the quality of watermarked medical images.

In this chapter, we are interested in measuring the impact of the watermarking process on the diagnostic quality and thus on the interpretation of the medical watermarked images. To do so, we implemented a subjective study protocol that first aims to provide an optimal watermarking parameter that will ensure the preservation of diagnostic information in all watermarked image data-set. More clearly it attempts to quantify the level of loss that may be acceptable. Second, it relies on the determination of a link between subjective and objective measures so as to control the image distortion cause by one watermarking modulation. As stated above, for practical reasons and to reduce the complexity of such a study, this one is limited to lung thoracic CT images encoded on 16 bits as well as to the LSB substitution watermarking modulation. In order to more reduce the number of tests and limit the study to a small number of radiologists, this work was conducted in several stages. In a first, we carried out two pre-studies which aim to reduce the range of watermarking parameters' values. The first pre-study aims at identifying the "maximum" watermarking parameters not to exceed, on the basis of objective measurements in comparison with the results of the Canadian Association of Radiologists (CAR) (study on lossy compression of thoracic CT images [27]). The second pre-study consists in a subjective experiments where IMT Atlantique and MEDECOM staff – non-experts persons – are invited to assess the visual quality of the watermarked images and identify a subset of "threshold values" of watermarking parameters most likely to introduce visible distortion into the image in order to limit duration of tests with radiologists to this range of values. The last stage thus corresponds to the subjective study with experts so as to evaluate both diagnostic and visual qualities of watermarked lung thoracic CT images. Unfortunately, this stage has not yet been started due to the unavailability of radiologists and time constraints.

This chapter is divided into four parts. In the first part we detail what is a subjective image quality assessment. In the second part, we will identify the chosen image test set, and the selected watermarking method. The third part deals with both objective and subjective pre-studies so as to identify the range of watermarking parameters' values that will be adopted by the subjective study with radiologists. Furthermore, a link between subjective pre-study results and objective quality measures is determined so as to quantify the level of information loss that may be acceptable. The last part is however only devoted to the definition of the operational implementation of the subjective quality assessment study because the results of tests with radiologists are not available.

5.1 Subjective Quality Assessment of a watermarking scheme

To analyze the visual and diagnostic quality of a medical image, experts have to assess the diagnosis legibility on a particular set of images. Radiologists have to do this task in adapted conditions, similar to those applied during clinical routine, allowing them to observe all the necessary details. This type of evaluation is desirable but of large complexity of implementation and costs (duration of tests, availability of experts, equipment, etc.). The number of images used and of radiologists participated in the study must be sufficient for a statistical evaluation. For instance, images must be evaluated by at least three experts selected from amongst senior radiologists specialized in the analysis of the organ being studied [166]. That is why, we find few subjective quality assessment studies [27, 163, 164]. Most of them evaluated the impact, on medical images' visual and diagnostic quality, of the loss of information caused by lossy compression and, more rarely, by watermarking. regarding lossy image compression, one can refer, for example, to the study realized by the Canadian Association of Radiologists [27]. It published a standard to validate the use of lossy compression under certain conditions and for specified examination types. For instance, they recommend compression ratios of 15:1 using JPEG and JPEG2000 on chest CT and from 10:1 to 15:1 on body CT images. These recommendations were based on data collected during a large-scale study done by 100 radiologists across Canada, on 23 different sessions, of 60 to 80 images for each session. To the best of our knowledge, only two works [161, 165] focused on the evaluation of the impact of the distortion introduced by watermarking in medical images. [165] is a subjective study that aims at establishing the link between the visibility of the watermark embedded in the Region of Non-Interest (RONI) – the black background of MRI images in this case – the interference it can induces for image interpretation. Indeed, even if the watermark is located in the black background of the image, the change of the background gray-scale, making it clearer for instance, may be uncomfortable or "non-natural" for physicians. The objective of this study is thus to determine distortion threshold to satisfy while inserting a robust watermark in the DCT domain. The study done in [161] showed that ROI (Region Of Interest) can also be an area for watermarking without

altering the clinical diagnosis. This study was done by three radiologists on 21 X-ray, 27 CT and 27 ultrasound images watermarked based on the LSB modulation. Experts were then asked to comment the images' quality and to give a clinical diagnosis for each image in order to compare it with the original one. The results of this study have shown that LSB watermarking modulation applied to various medical images in both ROI and RONI is possible and preserves image quality for clinical purposes. To sum-up, as shown, these subjective studies are expensive and time-consuming as they require large image dataset and many experts to conduct these researches.

Although, because of the lack of bibliographical references on this subject, a current approach that consists in comparing the level of distortion introduced by a watermarking method to the one caused by lossy compression is of interest. The underlying assumption is to consider that both watermarking and compression induce a similar information loss. By doing so, it becomes possible to take advantage of the subjective studies like the one conducted by the Canadian Associations of Radiologists. Notice also that it has never shown that such hypothesis is valid.

In this chapter, we hence propose to evaluate the imperceptibility of the watermark embedded into the image according to the observations of experts so as to define the amount of distortion accepted that could preserve the diagnosis accuracy of watermarked images. However, some critical points have to be taken into account for the deployment of such an evaluation protocol:

- *Selection of image test set* - the nature of medical image is very diverse, even when the modality and organ are restricted. Furthermore, conducting large-scale studies with radiologists for researches is thus highly impractical, time-consuming and expensive. It is indeed very complicated to gather different images to define a standard amount of distortion caused by watermarking enabling us to judge the result on all medical images issued from the various modalities, for a given organ and pathology. Hence, the choice of the test image data set will be linked to a limited and specific modality and pathology at the same time.

- *Variability of the evaluators* (i.e. the experts) - an evaluation protocol depends not only on the skill level of the radiologists, but also on their practice and the conditions under which the evaluation is conducted. It is highly recommended that the evaluators are specialized in the analysis of the organ and the pathology being studied. Moreover, the more motivated the radiologists are on the watermarking subject, the more efficient and productive the evaluation will be.

- *Conditions of the evaluation* - in order to avoid drifts during the statistical analysis of the subjective evaluation, the conditions of observations during the tests must be standardized and remains close enough to the conditions of real medical clinical practice. Indeed, any change may bias the analysis of test results. For example, changes of contrast and/or luminosity as well as zooming are not allowed. The same observation parameterization as during first image interpretation will be considered. Moreover, environmental conditions (e.g. an ambient room light less than 5 lux [167], observer-screen distance of 1m) have to be the same for all evaluators. We will come back on these elements in Section

Therefore, all these conditions make the subjective study time-consuming, expensive and "unique" due to the fact that it focuses on analyzing one specific pathology using one image modality.

5.2 Materials And Methodology

5.2.1 The Choice of Image Modality and Organ to be Studied

In this study, we were interested in evaluating the visual and diagnostic quality of thoracic Computed Tomography (CT) scans for lung cancer diagnosis purposes (i.e. detection of nodules); a use-case proposed by the CHU Grenoble-Alpes' (CHUGA) radiologists who have strong expertise in the domain.

Lung cancer is the most commonly occurring cancer in men and the third most commonly occurring cancer in women worldwide [168]. France had the 13 highest rate of lung cancer in 2018. The French public health body and the French national cancer institute have reported that the rate of lung cancer among women has increased by 5 % every year since 1990, and now 45% of all cancers among women in France are lung cancer and in men, lung cancer is the main cause of death. For patients with lung disease, chest CT is especially effective to assess disease progression

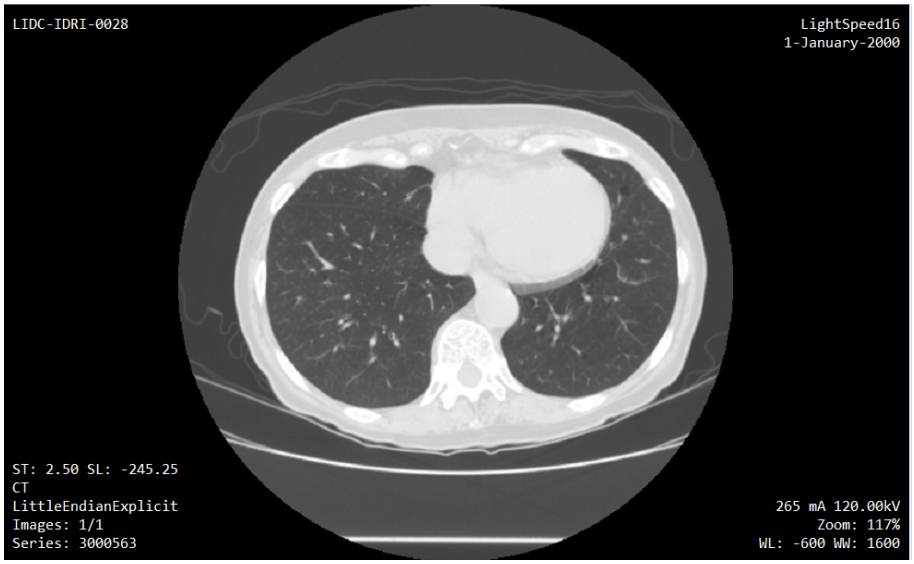


Figure 5.1: An example of a lung thoracic CT image.

or treatment response. Basically, the diagnosis of lung cancer in CT scans consists in the detection and interpretation of lesions in the form of pulmonary nodules within the CT data. These nodules will show up as a spot that are roughly spherical with round opacity and a diameter of up to 30 mm. After the identification of lung nodule, the radiologist will look at its size, shape, and general appearance. Lung nodules are usually about 3 millimeters to 30 millimeters in size. A larger lung nodule, of 30 millimeters or larger, is more likely to be cancerous than is a smaller lung nodule [169]. In such a context, it may be possible that the watermarking masks nodules, introduces, or even changes their sizes. Hence, "lung thoracic CT scan" can be a representative example. A lung thoracic CT image is given in Figure 5.1.

5.2.2 Watermarking Technique

As mentioned above, we opted for Least Significant Bit (LSB) substitution watermarking modulation applied to pixels, for different reasons:

- It is the simplest method to implement, to embed and extract a message into an image.
- We have adopted this watermarking modulation for all our proposed protection schemes.

LSB substitution modulation is a watermarking technique that has been presented in Chapter 1 Section (1.3.1). To avoid problems of notation we propose to recall it here in detail. It consists in substituting the β least significant bits of each pixel P_i of an image of depth d by a sequence of bits of the message m to be embedded. β corresponds to the watermarking strength (where $1 \leq \beta \leq d$). As a consequence, the watermarked version P_i^w of the i^{th} pixel P_i in the image is given by:

$$P_i^w = \left[\frac{P_i}{2^\beta}\right] \cdot 2^\beta + \sum_{k=0}^{\beta-1} m_{j+k} 2^k \tag{5.1}$$

where m_j is the j^{th} of the message m.

In order to extract the message, the watermark reader has to read the β LSB(s) of the watermarked image pixels P_i^w.

To increase the detection difficulty of the message, a secret key can be used to control the location into which the message is going to be embedded. But in this study, all bit-planes will be considered. Figure shows the impact of LSB substitution watermarking modulation in case of lung thoracic CT images for different values of β. As seen, the watermark is not necessarily visible. Also, for an insertion strength = 6, the texture in the image appears more contrasted than in the previous images. It is therefore necessary to determine for which values of β, the watermark remains invisible and does not alter the structures of the image or other information essential for the interpretation of the radiologist.

5.3 Objective and Subjective Pre-Studies

As mentioned above, in order to limit tests with radiologists, a pre-study is done before. The pre-study includes two stages: an objective and a subjective selection of the watermarking parameters.

5.3.1 Objective pre-protocol: preselection of the watermarking parameters

The goal of this first pre-study is to make a preselection of the watermarking parameters (i.e. the watermarking strength β) that will be respected by next in the subjective pre-study with MEDECOM and IMT Atlantique staff (security team), i.e. non-expert assessors. We conducted this task objectively on the basis of the Canadian Association of Radiologists (CAR); i.e. a subjective study for lossy JPEG2000 compression for CT images [27].

Our idea is to pre-select watermarking parameters; i.e. the values of the watermarking strength β of the LSB substitution watermarking modulation, based on an objective comparison between the distortion introduced by watermarking and lossy JPEG2000 compression by means of several image quality measures. More clearly, we will only keep watermarking parameters' values that induce an objective distortion closer to the one induced by lossy JPEG2000 and considered as allowable by CAR. This later has shown that a JPEG2000 compression ratio of 1:15 can be performed without compromising the diagnostic value of the lung thoracic CT image. We decided not to consider JPEG compression as it is less efficient than JPEG2000 in terms of image quality preservation/compression rate compromise.

In this study, we opted for the following quality objective measures:

- PSNR (Peak-Signal-to-Noise Ratio) [62] - it is one of the most used metrics for assessing the perceptual quality of the watermarked or decompressed image. PSNR is a good criterion when the distortion is uniformly over the whole image. But, if this hypothesis is not verified, it is better to use other quality measurements. PSNR is defined as:

$$PSNR\left(I, I_d\right) = 10\log_{10}\left(\frac{\left(2^d-1\right)^2}{MSE}\right) \tag{5.2}$$

with

$$\mathrm{MSE}\left(I, I_d\right) = \frac{1}{nm}\sum_{i=1}^{m}\sum_{j=1}^{n}\left[I(i,j) - I_d(i,j)\right]^2 \tag{5.3}$$

where I, I_d correspond to the original and the distorted d-depth images, respectively of size $[n, m]$.

- UQI (Universal Quality Index) [63] - it has been designed to model any image distortion as a combination of three factors: loss of correlation, luminance distortion and contrast distortion:

$$UQI\left(I, I_d\right) = \frac{4cov_{II_d}\mu_I\mu_{I_d}}{\left(\sigma_I^2+\sigma_{I_d}^2\right)\left(\mu_I^2+\mu_{I_d}^2\right)} \tag{5.4}$$

where μ_I, μ_{I_d}, σ_I, σ_{I_d}, cov_{II_d} are the mean of the original image I, the mean of the distorted image I_d, the variance of I, the variance of I_d and the covariance of I and I_d.

- SSIM (Structural Similarity) [62] - it is an objective metric for assessing perceptual image quality, working under the assumption that human visual perception is highly adapted for extracting structural information (e.g. texture, edges) from a scene. Quality assessment is thus based on the degradation of this structural information. To do so, SSIM separates the task of similarity measurement between the original image and its distorted version into three comparisons: luminance l, contrast c and structural information s.

$$\mathrm{SSIM}\left(I, I_d\right) = l\left(I, I_d\right)c\left(I, I_d\right)s\left(I, I_d\right) = \frac{\left(2\mu_I\mu_{I_d}+c_1\right)\left(2\sigma_I\sigma_{I_d}+c_2\right)\left(2cov_{I_d}+c_3\right)}{\left(\mu_I^2+\mu_{I_d}^2+c_1\right)\left(\sigma_I^2+\sigma_I^2+\sigma_I^2+c_2\right)\left(\sigma_I\sigma_{I_d}+c_3\right)} \tag{5.5}$$

where c_1, c_2 and c_3 are constants used to avoid instability when the denominator might approach zero.

- MSSIM (Mean Structural Similarity) - it is applied to calculate the average SSIM in small local area. It thus consists of luminance, contrast and structure correlation components.

$$\mathrm{MSSIM}\,(I, I_d) = \frac{1}{M}\sum_{j=1}^{M}\mathrm{SSIM}\,(I_j, I_{dj}) \tag{5.6}$$

  where M is the number of local window in the image, I_j and I_{dj} are the small images at the j^{th} window corresponding to the original image and the distorted one, respectively.

- NQM (Noise Quality Measure) - it is a human visual system-based metric. It was originally used to assess the quality of images degraded by noise injection. Nevertheless, it has shown acceptable results in thee presence of other types of degradation. Moreover, it takes into account contrast measure as well as the luminance, size, etc.

We considered a test data set of 200 lung thoracic CT images [170] for this objective pre-study, which consists in:

- Compressing images using lossy JPEG2000 with a ratio of 15, the maxi-mum ratio recommended by the CAR standard.
- Watermarking images using the LSB substitution modulation for different embedding strengths β; $\beta \in \{1, .., 12\}$.
- Determining the amount of distortion caused by compression and watermarking based on the quality objective metrics given above.

The objective measures' values obtained for LSB substitution modulation and lossy JPEG2000 compression are given in Table 5.1. As we can see in this table, the distortion introduced by the substitution modulation of LSBs is close to the distortion introduced by JPEG2000 compression until $\beta \geq 6$ and very different beyond. Based on this result, the insertion force values we retain for the future tests are: $\beta \in \{2, 3, 4, 5, 6\}$. However, the size of this range of preselected values still remains high, hence a second pre-study to reduce this range.

5.3.2 Subjective Pre-Protocol: Selection of watermarking parameters

The purpose of this pre-study is to identify the watermarking parameters or strengths most likely to introduce visible distortion into the image. These identified parameters will be then used in the study with the radiologists. This pre-protocol involves IMT Atlantique and MEDECOM staff.

Before detailing this pre-study, we first present the different ways to conduct such a subjective protocol in order to justify our implementation choice.

5.3.2.1 Subjective protocol models

Subjective evaluation is considered as the most reliable way to judge the quality of an image because it involves experts in the domain. The subjective assessment methods can be distinguished into three main models:

- *Single stimulus categorical rating* - that assesses the quality of an image without any information about the original one.
- *Double stimulus categorical rating* - where the observer is asked to judge the similarity between two images.
- *Comparative methods* - the purpose of these methods is to compare between two or more images according to their quality.

Table 5.1: Quality measure values obtained for LSB substitution watermarking modulation and for lossy JPEG 2000 with a compression ratio of 15.

Quality measure		JPEG2K Ratio=15	$\beta = 1$	$\beta = 2$	$\beta = 3$	$\beta = 4$	$\beta = 5$	$\beta = 6$	$\beta = 7$	$\beta = 8$
PSNR (dB)	**Mean (variance)**	79.34 (3.215)	99.34 ($4.4.10^{-5}$)	90.63 ($8.4.10^{-5}$)	83.54 ($8.57.10^{-5}$)	77.03 (3.0710^{-4})	70.76 ($4.6.10^{-3}$)	64.75 ($3.3.10^{-3}$)	58.56 ($6.3.10^{-3}$)	52.54 ($4.1.10^{-3}$)
UQI	**Mean (variance)**	0.79 (3.10^{-5})	0.998 ($3.11.10^{-6}$)	0.988 ($1.18.10^{-4}$)	0.95 ($1.2.10^{-3}$)	0.93 ($2.6.10^{-5}$)	0.86 ($5.7.10^{-4}$)	0.77 ($1.8.10^{-3}$)	0.7 ($2.5.10^{-3}$)	0.64 ($3.5.10^{-4}$)
SSIM	**Mean (variance)**	1 ($\cong$ 0)	1 ($\cong$ 0)	1 ($\cong$ 0)	1 ($\cong$ 0)	1 ($\cong$ 0)	0.9999 (3.10^{-6})	0.9997 (10^{-5})	0.998 ($\cong$ 0)	0.995 ($\cong$ 0)
MSSIM	**Mean (variance)**	0.98 ($2.9.10^{-3}$)	0.9998 ($\cong 0$)	0.998 ($\cong$ 0)	0.993 ($1.6.10^{-6}$)	0.98 ($8.2.10^{-5}$)	0.93 ($6.9.10^{-5}$)	0.854 ($2.5.10^{-4}$)	0.75 ($5.1.10^{-4}$)	0.64 ($5.8.10^{-4}$)
NQM (dB)	**Mean (variance)**	51.75 ($1.9.10^{-2}$)	67.41 (2.69)	60.26 (0.259)	56.91 (0.26)	50.53 (0.15)	43.8 (0.225)	36.68 ($1.1.10^{-2}$)	29.9 ($1.3.10^{-2}$)	22.3 (3.10^{-2})

Quality measure		$\beta = 9$	$\beta = 10$	$\beta = 11$	$\beta = 12$	$\beta = 13$	$\beta = 14$	$\beta = 15$	$\beta = 16$
PSNR (dB)	**Mean (variance)**	46.23 ($8.2.10^{-3}$)	39.9 ($1.2.10^{-2}$)	34.43 (10^{-2})	28.46 (0.13)	21.84 ($1.9.10^{-2}$)	15.43 ($4.6.10^{-3}$)	9.24 ($1.2.10^{-3}$)	5.36 ($3.5.10^{-3}$)
UQI	**Mean (variance)**	0.61 (8.10^{-3})	0.56 ($1.9.10^{-3}$)	0.5 ($1.3.10^{-3}$)	0.41 ($2.4.10^{-3}$)	0.2 ($1.5.10^{-3}$)	0.067 ($3.1.10^{-4}$)	0.0232 ($4.5.10^{-6}$)	3.10^{-6} ($2.6.10^{-13}$)
SSIM	**Mean (variance)**	0.983 ($6.9.10^{-6}$)	0.932 ($3.76.10^{-6}$)	0.7913 ($6.64.10^{-4}$)	0.477 ($2.19.10^{-5}$)	0.188 ($2.02.10^{-5}$)	0.0536 ($2.41.10^{-6}$)	0.0123 ($8.04.10^{-7}$)	4.10^{-3} ($3.1.10^{-5}$)
MSSIM	**Mean (variance)**	0.48 ($3.6.10^{-4}$)	0.257 ($2.38.10^{-4}$)	0.124 ($1.2.10^{-3}$)	0.07 ($4.88.10^{-4}$)	0.035 ($1.3.10^{-4}$)	0.016 ($1.8.10^{-5}$)	0.0075 ($5.3.10^{-6}$)	0.002 ($4.6.10^{-6}$)
NQM (dB)	**Mean (variance)**	12.96 (0.59)	5.47 (0.153)	4.64 (0.012)	3.37 (0.04)	2.69 (0.025)	2.18 (0.05)	1.9 ($1.6.10^{-2}$)	$\cong$ 0 ($\cong$ 0)

Table 5.2: 5-level scale assessment of image quality.

Rate	Quality	Distortion
5	Excellent	Imperceptible
4	Good	Perceptible
3	Tolerable	Little distorted
2	Poor	Distorted
1	Bad	Very distorted

Single stimulus categorical rating: This type of evaluation assesses the quality of one image at a time. The test images are displayed one by one on a screen for a period with a lag time between the display of two images. This time allows the observer to note the quality of the image on a scale of five levels, such that: excellent, good, fair, poor or bad. Other scales of assessments can of course be used [171]. All test images are displayed randomly.

Double stimulus categorical rating: This method allows measuring the quality of an image according to another version; its original version or another one. For example, the reference image is first presented followed by a gray screen and in second, its degraded version is displayed followed by a gray screen. Notice that the gray screen is displayed so as not to affect the quality of the analysis. IN particular, the user will not be influenced by the "image of difference". It should be noted that the display duration must be the same during each stage. Then, the observer should rate the quality or the level of perception of the degradation. In this case, different rating scales can be used. An example of a 5-level rating scale [164] is given in Table 5.2.

Comparative methods: These methods look for quantifying the visual differences between images without having any information on their quality. Among these methods, one can find:

- *Ordering by force-choice pair-wise comparison* - this method consists in displaying a pair of images simultaneously [172]. Observers are asked to compare images and to determine the one of better quality, even if both images appear identical. Moreover, there is no time limit to make a decision.
- *Pair-wise similarity judgments* - like the force-choice pair-wise methods, images are presented in pairs. However, observers are asked to choose the image of a better quality and to quantify the degree of difference between images on an "*a priori*" fixed scale.

5.3.2.2 The adopted subjective pre-study

As previously mentioned, the objective of this subjective pre-study is to determine the embedding strengths β that introduce a visible distortion in lung thoracic CT images. It is important to notice that this pre-study has been performed by non-expert assessors/observers. As consequence, only the visual quality of the image will be evaluated. Indeed, the image diagnostic quality can be only assessed by experts. Consequently, the diagnostic quality of lung thoracic CT images will be evaluated by radiologists after getting the results of this pre-study.

Pre-study model and questionnaire

Due to this pre-study the purpose, we opted for an evaluation method based on the "ordering by force-choice pair-wise comparison" protocol in order to assess the quality of the watermarked images according to different embedding strengths. That is why, our basic idea is to present to the observers on one screen a series of pairs of original/watermarked images according to different embedding strengths. More clearly, each watermarked image is displayed alongside its corresponding original version. The observers have then to visually compare the original image to its watermarked versions.

Herein, based on the "ordering by force-choice pair-wise comparison" protocol, for each pair of original/watermarked images, the observer is asked to answer the question:

Table 5.3: Lung thoracic CT images selected for the test.

Image Nb.	Scanner manufacturer	Quality	Pathology	Nb. of nodules $\geq$ 3mm
1	GE MEDICAL SYSTEM	++	absent	0
2	TOSHIBA	+++	obvious	7
3	GE MEDICAL SYSTEM	++	absent	0
4	TOSHIBA	+++	obvious	1
5	TOSHIBA	++	obvious	1
6	TOSHIBA	+++	obvious	2
7	TOSHIBA	+++	obvious	4
8	TOSHIBA	++	obvious	6
9	TOSHIBA	++	obvious	2
10	TOSHIBA	+++	obvious	4
11	TOSHIBA	+++	absent	0
12	TOSHIBA	+++	obvious	2
13	TOSHIBA	++	obvious	1
14	TOSHIBA	+++	obvious	1
15	TOSHIBA	++	obvious	3
16	GE MEDICAL SYSTEM	++	absent	0
17	TOSHIBA	++	obvious	3
18	TOSHIBA	+++	obvious	1
19	GE MEDICAL SYSTEM	++	absent	0
20	TOSHIBA	+++	obvious	6

"Do you notice any difference between the two images?" with as expected answer: "yes" or "no". We do not wish to have an ambiguous answer.

Experimental design

We describe here the implementation of this pre-study. It has been established in such a way that the duration of a test remains less than 45 minutes per observer. This "time" factor has not been mentioned before, but it is important for several reasons: it is directly related to the attention or the fatigue of the observers on the one hand and to their availability, on the other hand. Other points were also considered in the development of this pre-protocol. We review them in the following.

Image test set - A test database has been created from the public database [170]. This test base consists of 20 16-bit depth images of lung scanners, issued from 20 different patients. These images have been acquired according to standard clinical procedures and issued from two different acquisition systems: TOSHIBA and General-Electric (GE) MEDICAL SYSTEM.

These images were chosen by a PhD student, having received no training in medical imaging, according to several criteria:

- Image visual quality (+: poor quality, ++: acceptable quality, and +++: good quality);
- The presence or not of a pathology (i.e. presence of nodule(s)$\geq$ 3mm;
- The number of nodules of size $\geq$ 3mm.

More details are given in Table 5.3.

Analysis biases

Several biases can affect the statistical analysis of results. Herein, in order to avoid and limit such biases, the most obvious ones are discussed.

- *Bias related to the test environment*: This bias can be related to several environmental conditions that we distinguish in two classes:
  - Viewing conditions - the room lighting as well as the distance between the screen and the observer plays an important role in the perception of a stimuli. For example, high lighting can dazzle the observer and thus alter his perception. Similarly, the light color in the room can influence the perception of certain shades of gray in the test images.
  - The screen - its calibration should not be modified during the test; otherwise the displayed colors may differ from the original stimulus. Zooming and windowing parameters' modification are also not allowed.

  To minimize the effects of environmental bias, the tests should be performed under the same conditions for all observers.

- *Bias related to the observer*: in our case, the observers are of distinct age and experience. Different other factors can influence their answers:
  - Psychological factors - they can be at the origin of a difference of perception. The interpretation of an image is closely related to the degree of concentration of the observer. Also, access to the image without any prior information on the watermarking strength can influence the interpretation of the image.
  - Understanding of the study objective - the watermarking concept can be abstract for people outside the domain. An observer may thus imagine distortions even if the images have not been watermarked or these distortions are not visible.

  To limit these biases, the test should start with an introductory session allowing the observer to better understand the objective of purpose of the test and the protocol to be followed.

- *Bias related to image test set*: For each original image, its watermarked versions are presented in a random manner. More clearly, watermarked images are not presented according to the embedding strength (sorted in ascending or descending order). Also, watermarked images

Table 5.4: The observers who have done the tests

Observer	O_1	O_2	O_3	O_4	O_5	O_6
Function	R&D director	Ph.D student	Engineer	Ph.D student	Trainee	Ph.D student
Degree of expertise (years)	>10	1	<1	>3	0	>3

related to the same image are not displayed in a successive way. More clearly, we do not display the same image when moving from one image to another in order to not to avoid the bias of remembering a specific condition.

The number of the original images used in this test may result in low representativeness of the results. However, as the test is based on a comparison between an original image and its 5 watermarked versions, increasing the number of watermarked images for an original image with a finer discrimination of the embedding strengths' space require a high and sustained concentration of observers and could impair the test quality.

Test environment - In order to avoid drifts during the statistical analysis of results, this subjective pre-study took place in SePEMED laboratory, under similar conditions to clinical practices. Images were presented using *MedMammo 4.0.0 64-bit* software from the Medecom Company on a *Dell* diagnostic workstation with a *NEC* 23" *LSD - MultiSync P232W* screen. Note that The contrast, zooming and windowing parameters are correctly calibrated in advance according to the preset values of the original image interpretation (for example, windowing factors of the used image dataset are of (Window Level, Window Width) = (-600, 1600), zoom = 100%) and have not to be modified during the experiments. Furthermore, watermarked images are always displayed to the right of their original versions.

Watermarked image dataset and display - The adopted watermarking method, i.e. the LSB substitution modulation, depends only on a single parameter, which is the embedding strength β (see Section 5.2). In this pre-study, we use the values retained from the objective pre-study, seen previously, that is to say for β values = $\{2, 3, 4, 5, 6\}$. 20 images are considered and watermarked with 5 strengths, leading to a set of 100 pairs of watermarked/original images, a number of the overall data used in the experiments sufficient enough for a statistical evaluation (minimum of 30 according to [173] and [166]). Each observer will be asked to compare an original image to one of its watermarked versions, where both images are displayed side by side on the same screen, and to answer the question given above. Watermarked images with distinct embedding strengths β will be randomly presented to the observer. A random ordering of the watermarked images brings more robustness to the results of this pre-study.

The observers - The tests were carried out by 6 IMT Atlantique and MEDECOM staffs as given in Table 5.4. Most of the observers have acquired experience in medical imaging.

Information to the observers - Before running the test, observers have received the same information about the objective of the test they would achieve. The idea is to ensure that all the readers receive the same instruction. Here is an example of the test provided for the test:

"*Image watermarking is a protection technique that consists of inserting a message (e.g. security attributes, digital signature) by modifying the pixels' gray values of the image. Due to these modifications, distortions are introduced into the watermarked images. We then want to study the impact of these distortions on the radiologist's diagnosis (i.e. diagnostic quality). The test consists of comparing watermarked images of lung scanners (displayed on the right) to their original versions (displayed on the left). Each original image has 5 watermarked versions whose embedding strength (i.e. the level of distortion) varies. The observer has thus to identify images watermarked with a visible distortion compared to their original image. Note that it is not allowed to modify the display parameters (zoom and windowing).*"

Table 5.5: Duration of test for each observer.

Observer	Duration of test (minutes)
O_1	35
O_2	35
O_3	40
O_4	29
O_5	29
O_6	34

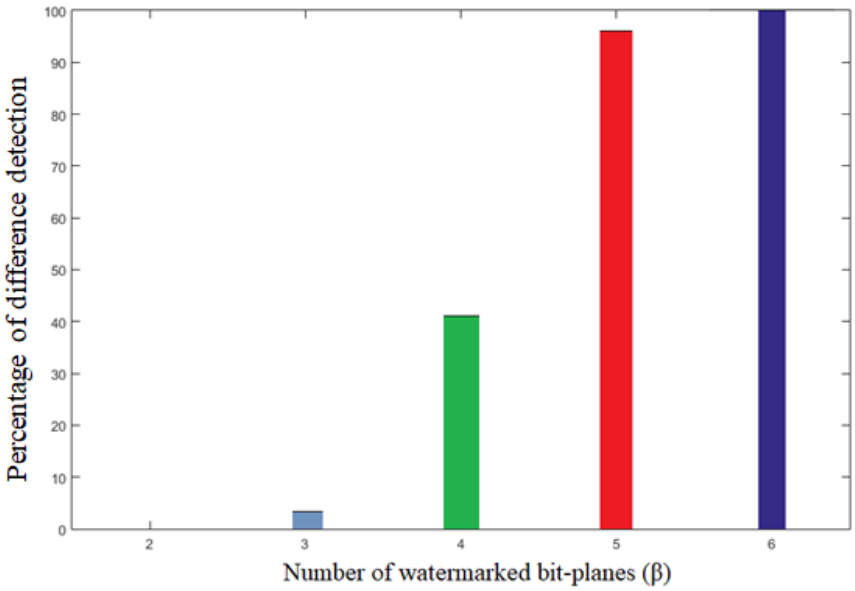


Figure 5.2: the percentage of observers who perceive a difference between one or more watermarked images and their original image according to the embedding strength β.

5.3.2.3 Test results

The following analysis is based on direct interpretation of test results as well as on some statistical data analysis methods.

Let us recall that the objective of this subjective pre-study is the identification of the values of the embedding strength (i.e. watermarking parameters) from which the watermark - or in other words the distortion caused by the watermark - becomes visible or is perceived by the observers. We will use these results to select the embedding strengths that will be subsequently used in the subjective study with the radiologists.

Observers' answers

Six observers participated in this test. Each of them was asked to compare an original image to one of its watermarked versions chosen randomly and to indicate if he/she sees a difference between them: '1' if so, '0' otherwise. Table 1 in the appendix provides the responses of the observers. Remember that for an original image, we have a sequence of five watermarked images. Therefore, for twenty original images, we have one hundred separate pairs. According to the duration of tests with the observers, as given in Table 5.5, the average duration of the test is of about 34 minutes/observer.

Watermark visibility

We are interested in determining the watermark visibility level according to the watermarking

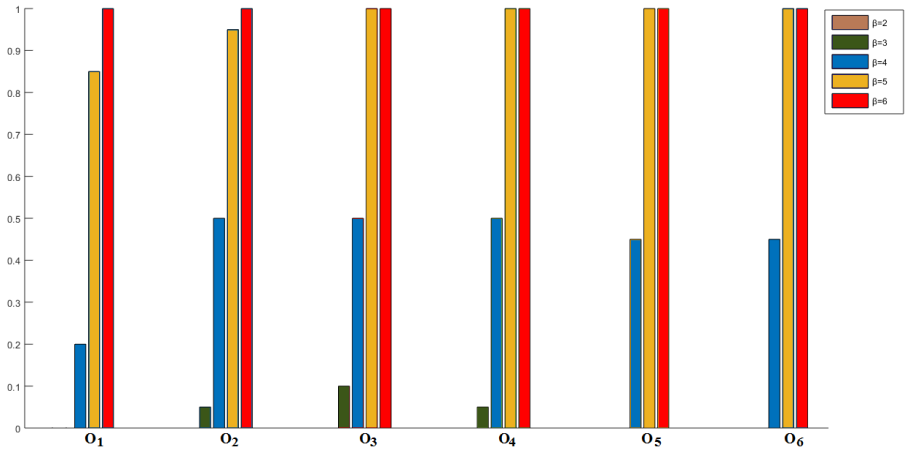


Figure 5.3: Watermark perception rate for each observer according to the embedding strength β.

Table 5.6: Pearson correlation matrix between the observers.

Variables	O_1	O_2	O_3	O_4	O_5	O_6
O1	1	0.7930	0.8009	0.8098	0.7523	0.8098
O_2	0.7930	1	0.8407	0.8202	0.84	0.9002
O_3	0.8009	0.8407	1	0.7816	0.8407	0.8617
O_4	0.8098	0.8202	0.7816	1	0.8202	0.8399
O_5	0.7523	0.84	0.8407	0.8202	1	0.9002
O_6	0.8098	0.9002	0.8617	0.8399	0.9002	1

parameter β. As mentioned above, the image display, i.e zoom, windowing, for the original images and their watermarked versions cannot be modified by the observers.

We give in Figure 5.2, the percentage of observers who perceive a difference between one or more watermarked images and their original image according to the embedding strength. This percentage increases as expected with the level of watermarking strength. The perception rate of the mark for the level of distortion $\beta = 3$ is very low (rate ≤ 0.1) and for $\beta = 2$ is null.

In Figure 5.3, the average rate of detection for each embedding strength level on all the images is plotted for each observer. As it can be see, all the observers can notice a difference between the original and its watermarked version at level β=6. As it can be also noticed, the observers O_3, O_4, O_5, O_6 can detect the distortion at level β=5, for 100% of images, O_1 and O_2 differentiate between 85% and 95% of images watermarked with β=5 from their original versions, respectively. According to images watermarked with β=4, as it can be seen, observers hold scattered opinions. Sometimes, they can detect the watermark and sometimes they cannot. O_2, O_3 and O_4 detect the watermark in 50 % of watermarked images at this level, and O_5, O_6 in 40% of them. O_1 perceives a difference in only 20% of images watermarked at level β=4. Note that O_1 is the most experienced of our observers (an engineer with over 13 years of experience). One can thus think that the experiment is a discriminating element between observers. Table 5.6 give the correlation coefficients of PEARSON between all the observers. Basically, two observers whose answers are close should have a correlation factor close to 1. As it can be noticed, the answers of all observers are close and correlated between each other, especially between O_2 and O_6 and between O_5 and O_6. O_1 and O_5 are the less correlated than others. As already mentioned in Table 5.4, O_1 and O_5 correspond to the most and least experienced observers, respectively. This can make of the level of experience a discriminating factor. Also, it is likely that the discriminating factor is a level of

Table 5.7: Threshold values of different quality measures.

Quality measure	Threshold value
PSNR	77 dB
UQI	0.8
SSIM	≈ 1
MSSIM	0.98
NQM	50

visual acuity. This hypothesis remains to be verified.

On the basis of these results, it seems relevant to test with radiologists the insertion strengths $\beta = \{4, 5, 6\}$, which are in the transition zone "the watermark is invisible, the watermark is visible" for non-specialist observers. We selected embedding strengths that make a watermark visible in the image in order to determine whether a low image visual quality has some impact on the image diagnostic quality.

5.3.3 Link between objective and subjective results of the pre-study

Even though the subjective study with the radiologists has been postponed to 2020, due to the unavailability of radiologists, based on the previous results, we think it is possible to establish a link between objective measures and subjective test results. As for $\beta_1 = 4$ (i.e. four bit planes watermarked), observers sometimes see a difference between the watermarked and original images and for $\beta_3 = 6$, they always detect the watermark. If objective measurements are coherent whatever the test images, one can use these values so as to determine embedding strength threshold not to exceed.

We give in Table 5.8 the objective measures obtained on our image test dataset. As it can be seen, the image distortion evaluated in terms of PSNR; UQI; SSIM; MSSIM and NQM, and introduced by the substitution modulation of LSBs at level $\beta_1 = 4$ is very close to or lower than the distortion introduced by lossy JPEG2000 compression as previously demonstrated (See Section 5.3.1). As a consequence, one can conclude that, in case of LSB watermarking modulation, the measurements of the distortion caused by watermarking up to the four least significant bit planes can be considered as threshold values not to be exceeded in order to better preserve the visual image quality. These threshold measures are given in Table 5.7.

To go beyond, we propose to see whether the distortions induced by our JWC and JWEC (presented in Chapter 2 and 3, respectively) schemes remain below these threshold values. As it can be seen in Table 5.8, these distortions respect the threshold values given in Table 5.7 and are so much lower that the ones induced by lossy JPEG2000 with a compression rate = 15:1 and by image watermarking up to level $\beta_1 = 4$.

5.4 Quality Assessment Protocol of the Subjective Study

The objectif of this study is to identify the embedding strength that can be used without compromising the "diagnostic value" of the image. Due to the modesty of our means, this study will at most allow us to identify the threshold value of the watermarking parameter β from which radiologists detect a difference between an image and its watermarked version in the case of Least Bit substitution watermarking modulation applied on CT medical images. As previously said, one use-case has been identified by CHUGA (CHU Grenoble-Alpes) radiologists as particularly relevant for a study of this type: the detection of pulmonary nodules in lung thoracic CT scans. Lung cancer is the most common type of cancer in the world (1.8 million new cases in 2012), and the most deadly (1.59 million deaths) [174] because it is most often discovered at an advanced stage. The lesions in chest scanner slices are of the order of a few millimeters. Their detection and characterization requires a thorough reading of the images that can be impacted by a degradation of the image quality or changes of texture.

Table 5.8: Image quality measurements obtained for LSB substitution watermarking modulation according to the selected embedding strength β values, lossy JPEG 2000 with a compression ratio of 15, our joint watermarking-compression (JWC) and joint watermarking-encryption-compression (JWEC) schemes.

Quality measure		JPEG2K Ratio=15	$\beta = 1$	$\beta = 2$	$\beta = 3$	$\beta = 4$ β_1	$\beta = 5$ β_2	$\beta = 6$ β_3	JWC	JWEC
PSNR (dB)	**Mean (variance)**	79.34 (3.215)	99.34 ($4.4.10^{-5}$)	90.63 ($8.4.10^{-5}$)	83.54 ($8.57.10^{-5}$)	77.03 ($3.07.10^{-4}$)	70.76 ($4.6.10^{-3}$)	64.75 ($3.3.10^{-3}$)	95 ($7.4.10^{-6}$)	86 ($1.9.10^{-5}$)
UQI	**Mean (variance)**	0.79 (3.10^{-5})	0.998 ($3.11.10^{-6}$)	0.988 ($1.18.10^{-4}$)	0.95 ($1.2.10^{-3}$)	0.93 ($2.6.10^{-5}$)	0.86 ($5.7.10^{-4}$)	0.77 ($1.8.10^{-3}$)	0.92 ($3.1.10^{-4}$)	0.85 (5.10^{-4})
SSIM	**Mean (variance)**	1 ($\cong$ 0)	1 ($\cong$ 0)	1 ($\cong$ 0)	1 ($\cong$ 0)	1 ($\cong$ 0)	0.9999 (3.10^{-6})	0.9997 (10^{-5})	0.9999 ($1.6.10^{-7}$)	0.99 (10^{-5})
MSSIM	**Mean (variance)**	0.98 ($2.9.10^{-3}$)	0.9998 ($\cong 0$)	0.998 ($\cong$ 0)	0.993 ($1.6.10^{-6}$)	0.98 ($8.2.10^{-5}$)	0.93 ($6.9.10^{-5}$)	0.854 ($2.5.10^{-4}$)	0.99 ($1.6.10^{-5}$)	0.98 3.10^{-7}
NQM (dB)	**Mean (variance)**	51.75 ($1.9.10^{-2}$)	67.41 (2.69)	60.26 (0.259)	56.91 (0.26)	50.53 (0.15)	43.8 (0.225)	36.68 ($1.1.10^{-2}$)	58.22 ($7.4.10^{-2}$)	52.64 (5.10^{-2})

Table 5.9: Distribution of the image test set for each radiologist

	Batch 1 (64 « mini-examinations »)	Batch 2 (64 « mini-examinations »)	Batch 3 (64 « mini-examinations »)	Batch 4 (64 « mini-examinations »)
Radiologist A	β_0	β_1	β_2	β_3
Radiologist B	β_3	β_0	β_1	β_2
Radiologist C	β_2	β_3	β_0	β_1
Radiologist D	β_1	β_2	β_3	β_0

5.4.1 Operational implementation

The definition of such a subjective protocol depends on several points, such as the number of images to be watermarked according to the previously selected parameters, the test environment, the choice of questions to ask the radiologists and, the observers (in this case the radiologists).

5.4.1.1 Image test set

A test database has been created from a database already acquired for the npUBD study (i.e. a study done in CHUGA). This test image set base consists of 64 16-bit depth images of lung scanners, issued from 50 different patients. A ratio of approximately 50% of "normal mini-examinations" (absence of nodules) and 50% of "abnormal mini-examinations" (presence of nodule $\geq$ 4mm will be respected.

5.4.1.2 Image watermarking

In this study, as in the above, the LSB modulation is considered parameterized by the embedding strength values retained from the subjective pre-study, seen previously, that is to say for β values $= \{4, 5, 6\}$. Notice that $\beta = 0$ refers to a non-watermarked image. The adopted watermarking strengths are thus $\beta = \{\beta_0, \beta_1, \beta_2, \beta_3\} = \{0, 4, 5, 6\}$.

For each of 4 radiologists who should participate to the test, a «mini-examination» database of 64 images is created, where:

- 64/4 non-watermarked (β_0) «mini-examinations»,
- 64/4 « mini-examinations » watermarked with β_1 ,
- 64/4 « mini-examinations » watermarked with β_2,
- 64/4 « mini-examinations » watermarked with β_3.

Each image (watermarked or not) of the mini-examination is only displayed once for each reader (see Table 5.9), ensuring thus the reading of the entire watermarked image test set.

5.4.1.3 Test environment

The tests will be done by respecting the following experimental conditions:

- The used screen will be one of those used for diagnosis in clinical routine,
- The contrast and luminance parameters will be correctly calibrated beforehand,
- The experimental conditions have to be identical for each expert reader (e.g. brightness of the room, distance between the reader and the screen).

5.4.1.4 Test

To avoid any reading bias, the "mini-examination" database will be presented to the radiologist randomly and blindly, that is to say that the radiologist does not know neither the embedding strength applied to the image nor its original diagnosis.

Experts will hence be asked to answer the following answers:

a) In this image, do you identify the presence of nodule(s) $\geq$ 4mm and not calcified?

   The answer has to be: yes, no or I do not know.

   If yes:

   i) You have to determine the number of nodules in the image and to localize them.

   ii) You have to identify the characteristics of each identified nodule.

b) How do you subjectively evaluate the quality of the image?

   Only one answer can be chosen from the following ones:

   - Excellent image quality: distinct anatomical details, absent or minimal noise.
   - Good visual image quality without altering the diagnostic quality: clear anatomical details, moderate increase in noise but not affecting the diagnosis.
   - Not diagnosable.

Unfortunately, due to time constraints, the tests with the radiologists have not yet been started. Nevertheless, we can identify some statistical data analysis methods that can be used to interpret the answers of radiologists for these questions. For instance, diagnostic accuracy (Question a, a.i and, b) can be assessed by comparing rater Sensitivity [175] (e.g. in terms of number of nodules in the image) and the Specificity (proportion of normal images - images without nodules - correctly identified) across the image test set watermarked with different embedding strengths β. According to the morphological feature of the nodules (Question a.ii), the concordance of the variable "density of the nodule" can be evaluated using the Cohen's Kappa coefficient [176].

5.5 Conclusion

Watermarking does not leave original medical image unimpaired. It is therefore necessary to evaluate such degradations. Our objectives were thus twofold. The first was to verify that, for lung thoracic CT images, a loss of information related to watermarking is tolerable. The second objective was to identify the watermarking parameters that can be used without any signification reduction of the visual and diagnostic quality of the image. Notice that assessing image quality through subjective studies for large variety of data present in medical imaging applications is time-consuming, expensive and difficult to set up. That is why, few subjective quality assessment methods of medical images have been done and most of them.

This study was conducted on a database of thoracic Computed Tomography (CT) scans for lung cancer diagnosis purposes that consist in the detection and interpretation of nodules within the CT images. In such a context, it may be possible that the watermarking masks nodules, introduces, or even changes their sizes. Before performing the experiment with radiologists, two pre-studies were conducted. The first one objectively compared watermarking distortions with those introduced by JPEG2000 lossy compression with the compression ratio recommended by the Canadian Association of Radiologists (CAR) [27], using common image quality metrics like PSNR, SSIM, and so on. The second pre-study was conducted with MEDECOM and IMT Atlantique staffs to subjectively assess the visual quality of watermarked images. The analysis of the results of these pre-studies allowed us to choose the ranges of values of watermarking parameters that will be used in the study with the radiologists.

Unfortunately, the subjective study with the experts has not yet been started due to time constraints. Nevertheless, according to our pre-studies, it is possible to establish a link in-between subjective and objectives studies. It appears that, for instance, when the PSNR is greater than 77 dB the watermarked CT image is of good visual quality. However, these results needs to be confirmed.

Conclusion

Medical images play an important role in diagnosis, patient follow-up and screening. This is why they are more and more shared in-between health professionals (e.g. telemedicine applications, cloud-based medical imaging applications) and also re-used and combined with machine learning techniques so as to develop tools for diagnosis aid support. However, as they are manipulated in highly open environments in which different users access information, security issues are increased due to the sensitive nature of medical images. As a consequence, healthcare organizations implement stringent policies and procedures to ensure data security so as to be used trustworthy in daily medical practice. Image security can thus expressed in terms of confidentiality, reliability and traceability.

We have seen that developing new security mechanism for the protection of medical images needs to take into account the specificity of the domain, in particular the Digital Imaging and Communications in Medicine (DICOM) standard and the fact that medical image data are usually stored and exchanged under a compressed form, due to the very large volume they represent (an hospital can generate at least 27,000 Terabytes per year [17]) DICOM has been introduced in order to facilitate the transmission handling of medical images. It specifies which lossless and lossy image compression algorithms can be used. JPEG-LS and JPEG are part of them. DICOM also takes into account different security aspects in its part 15. It recommends the use of the triple DES or AES cryptosystems so as to ensure image confidentiality while image integrity can be provided with the help of the DSA digital signature. Let us recall that the kind of protection these security mechanisms offer is "*a priori*", in the sense that once an image is decrypted or its digital signature is lost or deleted, it is no longer protected. Here comes the interest of an "*a posteriori*" protection, a kind of protection watermarking ensures due to the fact that it leaves the user accessing the data while maintaining them protected by the message.

From this standpoint, we have showed that it is desirable to develop solutions that can give access to watermarking-based security services (e.g. integrity and authenticity control) from the compressed domain, i.e. from the image compressed bitstream, and also from the encrypted domain.

In a first moment, we have studied the combination of watermarking and compression in order to be able to control image security without having to decompress it, even partially. In this context, we have proposed new joint watermarking-compression scheme [6, 7]. The system we propose combines, in a single operation, bit-substitution watermarking modulation with JPEG-LS,in the first scheme, and with JPEG in the second one. It is important to notice that with our schemes, it is possible to decompress the image with the common JPEG-LS or JPEG algorithms. More clearly, decompression as well as message extraction processes are conducted independently without having to be modified or adapted. With such an ability, our scheme is DICOM compliant. Furthermore, due to the fragile watermark, this solution makes possible to control image reliability (i.e. authenticity and integrity) directly from compressed domain. Note also that the proposed scheme saves the computational complexity as it does not require decompression to verify image reliability in the compressed domain. Future works will focus on improving the robustness of the watermark.

As stated above, there is a great interest to combine watermarking with encryption so as to simultaneously achieve an *a priori/ a posteriori* protection. We thus work on two crypto-watermarking techniques with as main objective to provide watermarking-based security services from encrypted data. The first one we proposed takes into account the fact that medical images are most of the time stored in a compressed form. It jointly combines watermarking, encryption and compression at once [8, 177]. Its originality is twofold. First, it allows accessing to watermarking-based security services from both encrypted and compressed image bitstreams without having to

decrypt or to decompress them, even partially. Second, it combines bit-substitution watermarking with JPEG-LS and the AES block cipher algorithm in its cipher block chaining (CBC) mode , in a single operation. It is important to notice that with our scheme, it is possible to decipher and decompress the image with the common AES and JPEG-LS algorithms. More clearly, decryption, decompression as well as message extraction processes are conducted independently without having to be modified or adapted. With such a capability, our scheme is DICOM compliant. Furthermore, this solution makes possible to control medical image reliability (i.e. authenticity and integrity) directly from both encrypted and compressed domains. Note also that the proposed scheme saves the computational complexity as it does not require decryption and decompression to verify the image reliability in the encrypted domain and the compressed one, respectively. In this work, we have also theoretically demonstrated the performance of the above scheme in terms of watermarking capacity and image quality distortion. These theoretical results have been validated experimentally on a real medical image data set and allow us to precisely predict the performance of our scheme. Moreover, they prove the suitability of the proposed scheme to the applications it is conceived for. Future works will focus on the possibility to embed a message in the encrypted-compressed image while being able to extract it from both encrypted and compressed domains.

The second crypto-watermarking scheme we propose consists in losslessly watermark an encrypted image for the purpose of verifying the reliability of an image in both encrypted and clear domains. The originality of the proposed system stands on the capability of embedding a message in the encrypted domain and being able to extract it from both encrypted and clear domains and to recover the original image. This is possible by means of a "pre-watermark" embedded before image encryption using a new robust reversible histogram shifting watermarking modulation with error-correction abilities. The message is then embedded in the encrypted image using a common watermarking technique. It is the impact of this message insertion process onto the "pre-watermark" that gives us access to the message in the clear domain, i.e. after the decryption process. Due to the robustness of the embedded pre-watermark, the original image can be thus exactly recovered. Message embedding/extraction processes are completely independent from encryption/decryption. The feasibility of our approach is demonstrated considering the Trivium stream cipher and a robust histogram shifting modulation that we developed. Experiments conducted on some natural images and medical images confirm the efficiency of our approach to make available a message in both clear and encrypted domains and to correctly recover the original image. However, the proposed approach is not DICOM-compliant because of the use of a stream cipher algorithm. Hence, future works will focus on making this scheme compliant to DICOM standard by using a block-cipher algorithm (e.g. AES, 3-DES).

Finally, we proposed to study the impact of lossy watermarking on lung thoracic CT images, in order to identify watermarking parameters or strengths to preserve the image quality. To do so, we have developed objective and subjective validation studies in cooperation with the CIC-IT, imaging pole and the University Clinic of Radiology and Imaging (Clinique Universitaire de Radiologie et Imagerie Médicale (CURIM)) of CHU Grenoble-Alpes, and MEDECOM so as to validate our distortion constraint hypothesis. Even though we were not able to make the test with radiologists due to time constraints, objective and subjective pre-studies was conducted with MEDECOM and IMT Atlantique staffs to assess the visual quality of watermarked images and to identify the watermarking parameters most likely to introduce visible distortion in the image. Herein, only the visual quality of the image was evaluated and not the diagnostic quality. This latter will be evaluated by radiologists by taking into account the results of the pre-study in order to limit the test duration with experts. In the continuity of this work, the impact of watermarking of medical images on automatic analysis methods can be studied as the modification of these images can have side effects on the extracted parameters.

Even though this Ph.D. thesis provides some contributions to medical image watermarking, there are still several open issues. In future works, we will focus on the better preservation of the image quality and improving the robustness of the watermark while reducing the algorithm complexity.

Résumé en français

L'imagerie médicale joue un rôle important et vital dans l'aide du diagnostic et la prise de décision. À ce propos, plusieurs techniques d'acquisition sont utilisées: Computer Tomographie (CT), échographie, Imagerie par Résonance Magnétique (IRM). D'autre part, l'incursion des nouvelles technologies de l'information et de la communication a entrainé d'importants développements dans le domaine de la santé. Ces technologies ont des différents impacts sur la pratique professionnelle, la gestion de l'information médicale et la prise en charge des patients. En particulier, l'échange numérique d'images médicales dans différentes applications de télémédicine (ex. téléradiologie, télédiagnostic, téléexpertise) permet de simplifier la prise de décision multidisciplinaire en faisant gagner du temps aux acteurs impliqués dans la chaîne de soin et d'améliorer la qualité des soins.

Cependant, avec l'évolution numérique du secteur, les données médicales sont de plus en plus exposées aux risques de sécurité. Du fait que ces données jouent un rôle important pour la santé d'un patient; son altération peut être la cause directe ou indirecte d'une atteinte grave à la santé d'un patient (ex. Selon le *Dr.* Sung Choi, chercheur à l'université de Vanderbilt, plus de 2100 décès de patients par an sont causés par des violations de données), le traitement de données médicales s'intègre à un cadre législatif et déontologique. Par exemple, au regard de la sensibilité des ces informations, le Règlement Général sur la Protection des Données (RGPD) interdit tout traitement de ces données, sans le consentement exprès des personnes concernées (i.e. patients) qui doivent obligatoirement être informée des buts et objectifs du traitement. Ce règlement interdit également que ce traitement se fasse en dehors des cas légalement prévus. 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.

Pour pouvoir répondre à ces besoins et assurer la sécurité des données médicales, il faut d'abord connaître les menaces et les risques rencontrés afin d'être capable de les éviter. Si l'on se réfère aux standards pour le déploiement de politiques de sécurité, comme l'ISO 27799 dédiée aux données de santé en particulier, les risques pour l'information médicale peuvent être classés en trois catégories:

- Les accidents (ex. les pannes matérielles, les phénomènes naturels).
- Les erreurs (ex. erreur lors de la saisie de l'information, des transferts).
- Les malveillances (ex. fraudes, détournement de l'information, intrusion dans le système).

Plusieurs solutions informatiques ont été proposées pour assurer la sécurité comme : le contrôle d'accès, le chiffrement, la signature numérique. Mais, elles restent insuffisantes. Prenons le cas par exemple du chiffrement d'images (voir Figure 5.4), cette opération a pour but de rendre le contenu d'une image inaccessible (confidentialité), sauf pour les détenteurs de la clé de déchiffrement. Il s'agit ainsi d'un mécanisme de protection de type que l'on peut qualifier de *"a priori"* car une fois déchiffrée, l'image n'est plus protégée. Le tatouage a été proposé comme un mécanisme complémentaire à ces solutions *"a priori"*. Il peut contribuer à une protection *continue* et *"a posteriori"*.

Dans son principe (voir Figure 5.5), le tatouage laisse l'accès à l'information tout en la maintenant protégée par une "marque" imperceptible, un message qui porte une information de protection (ex. des attributs de sécurité – signatures numériques, code d'authenticité). Dans le cas des images, l'insertion d'un message se fait par une distorsion contrôlée des niveaux de gris de celles-ci. Ces travaux de thèse portent sur la combinaison de techniques de tatouage avec des mécanismes de chiffrement de manière à développer de nouvelles solutions permettant d'assurer une protection à la fois *a priori* et *a posteriori*, et donc continue de données d'imagerie médicales partagées et échangées.

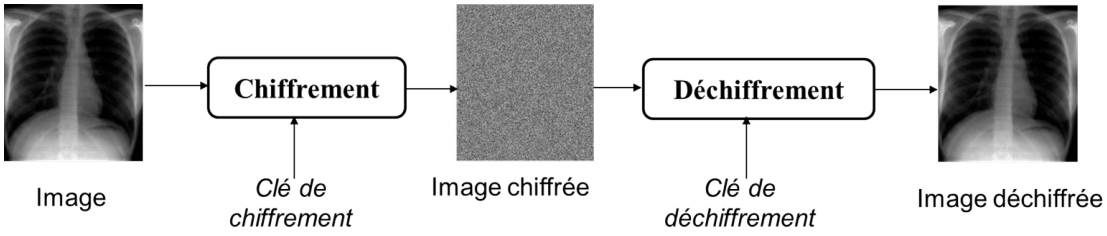


Figure 5.4: Schéma général du chiffrement d'images.

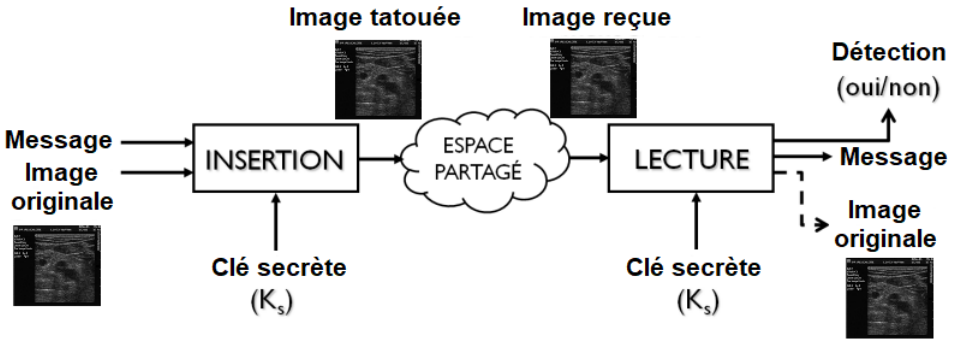


Figure 5.5: Étapes principales d'une chaine de tatouage classique. L'image tatouée est partagée (e.g. via l'Internet) et elle peut être manipulée entre l'insertion et la lecture. À la lecture, dans le cas du tatouage réversible (lossless), l'image originale peut être complètement récupérée.

Le déploiement de ces mécanismes de sécurité dans le domaine de la santé doit prendre en compte les spécificités de ce domaine. En particulier les images médicales constituent de grands volumes de données (un hôpital génère 27.000 Téraoctets de données par an). Afin d'optimiser les coûts de stockage et de transmission, ces images sont de plus en plus encodées sous formee compressées avec ou sans pertes. Avec la compression sans pertes, les données originales sont exactement récupérées à partir du flux compressé. Les méthodes de compression avec pertes apportent une distorsion aux images reconstruites.

Une autre contrainte forte que nous avons considérée dans ces travaux concerne l'interopérabilité des solutions proposées avec les standards de santé. Pour profiter pleinement du tatouage, il ne doit pas interférer avec ces derniers. Il doit être complètement transparent. Puisque nous travaillons sur des images médicales, nous avons pris en compte autant que possible le standard DICOM, qui stipule comment stocker et échanger de telles images. Il conviendra donc d'une part de considérer les algorithmes de chiffrement et de compression qu'il utilise et, d'autre part, de s'assurer en particulier que si un message est accessible dans l'image chiffrée et/ou compressée, il n'en empêche pas le déchiffrement ou la décompression par un système non-compatible en termes de tatouage.

Dans ce contexte, il est pertinent de pouvoir donner accès à des services de sécurité fondés sur le tatouage (ex. contrôle d'intégrité et d'authenticité) directement à partir du flux binaire compressé de l'image. Mais si on revient sur l'état de l'art, on trouve différentes méthodes qui combinent le tatouage et la compression. Les 3 grandes classes de méthodes, données dans Figure 5.6 : Elles appliquent un tatouage suivi d'une compression ou une compression puis tatouage, voire un tatouage/compression conjoint, où les deux opérations sont réalisées simultanément. Ces méthodes ne donnent accès à la marque qu'à partir d'une image totalement ou partiellement décompressée, mais pas directement dans le domaine compressé. D'où l'idée de combiner le tatouage et la compression, de manière à pouvoir accéder aux différents services de sécurité à partir du flux binaire compressé de l'image, sans le décompresser. Pour ce faire, différentes contraintes doivent être d'abord considérées. Travaillant sur des images médicales, une première est de préserver la qualité de l'image pour le diagnostic : le tatouage ne doit pas interférer avec l'interprétation de l'image. Une seconde contrainte est de déterminer où et comment le flux binaire peut être modifié sans en changer la sémantique. Enfin, il faut pouvoir extraire le message à partir du flux binaire compressé sans le décompresser, même partiellement. Plus clairement, le message doit être accessible sans avoir à décompresser l'image, même partiellement. Ceci a fait l'objet du premier Chapitre,

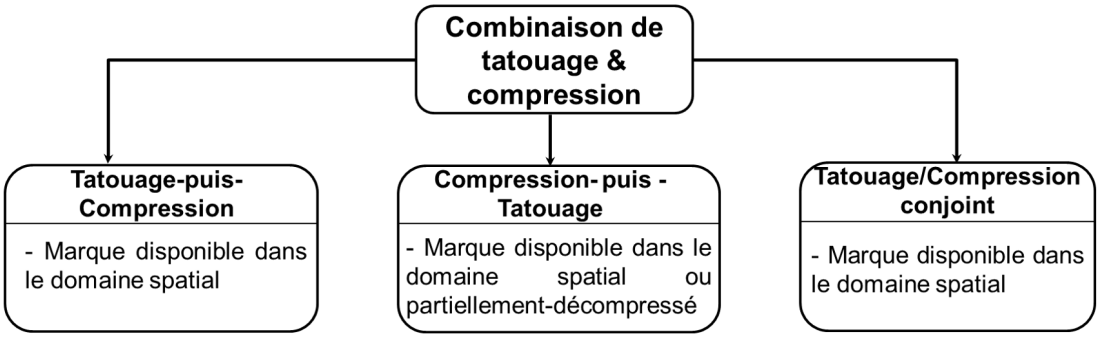


Figure 5.6: Classes de méthodes combinant le tatouage et la compression.

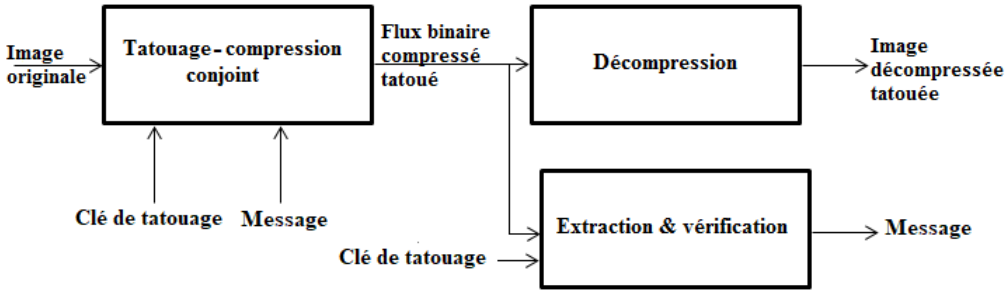


Figure 5.7: Architecture de la méthode "Tatouage-Compression Conjoint".

où nous avons proposons un schéma de tatouage-compression conjoint qui permet d'accéder à un message à partir du flux binaire compressé de l'image, sans avoir à le décompresser, même partiellement (voir Figure 5.7). Le message accessible dans le flux peut être utilisé pour vérifier l'integrité et l'authenticité d'une image sans la décompresser. Nous avons proposé une solution dont les principes restent les mêmes qu'elle soit appliquée à des images compressées JPEG-LS ou JPEG. L'architecture du système que nous proposons est donnée en Figure . Il exploite un tatouage fondé sur la substitution de bits appliqué lors de la compression JPEG-LS ou JPEG d'une image. Pour pouvoir accèder au message dans le domaine compressé, nous avons fixé un "mot-code" derrière lequel, à une position bien déterminée, un bit de message sera inséré chaque fois rencontré lors de la compression de l'image. Lors de la phase de vérification, le lecteur n'aura qu'à identifier les "mots-codes" dans le flux binaire pour retrouver les bits du message. Cette solution introduit une très faible dégradation de l'image et offre une capacité d'insertion suffisante pour aller au-délà du contrôle de la fiabilité de l'image. Aussi, la méthode proposée à été développée pour être compatible avec le standard DICOM.

Dans la continuité de ces travaux, nous nous sommes intéressés à vérifier la fiabilité des images compressées et chiffrées tout en maintenant leur confidentialité. Dans la littérature, on trouve plusieurs méthodes qui combinent le tatouage, le chiffrement et la compression de différentes manières mais qui combinent rarement tous les trois ensemble. Les méthodes qui le font ne donnent accès à un message qu'à partir d'un seul domaine et ne sont compatible avec le standard DICOM. Nous proposons ainsi une solution qui permet d'accéder à un message dans le domaine chiffré sans avoir à parser le flux binaire.

Son principe général (voir Figure 5.9) est fondé sur l'insertion de deux messages contenant des attributs de sécurité dans l'image durant la compression et le chiffrement de cette dernière. Chacun d'eux n'est accessible que dans un seul domaine : le domaine compressé ou le domaine chiffré, sans avoir à décompresser ou à déchiffrer l'image, même partiellement. Notre système comporte un processus de protection et un processus de vérification. Durant la phase de protection (Figure 5.9. a), une image I est compressée, chiffrée et tatouée simultanément avec deux messages m_c et m_e accessibles respectivement dans le domaine compressé et le domaine chiffré. L'insertion du message accessible à partir du flux binaire compressé de l'image est basée sur le schéma décrit précédemment, c'est à dire les bits du message sont insérés, durant la compression JPEG-LS, après un "mot-code" pré-défini pour pouvoir l'identifier dans le flux binaire compressé de l'image sans avoir à la décompresser. Pour donner l'accés à un deuxième message à partir de l'image chiffrée,

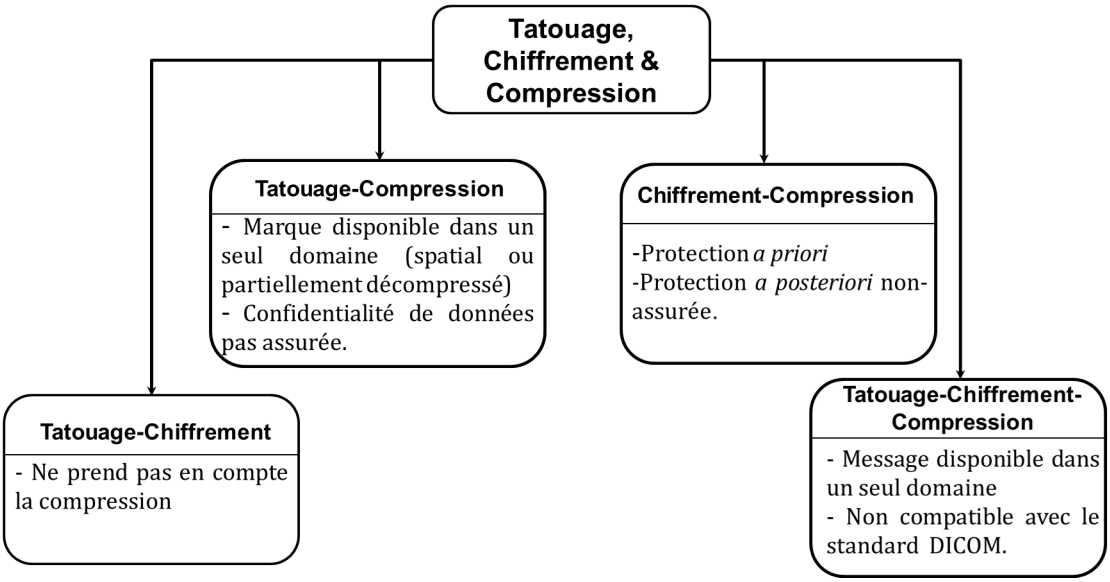


Figure 5.8: Classes de méthodes combinant le tatouage, le chiffrement et la compression.

notre idée est d'introduire une distorsion dans le flux binaire compressé de l'image afin d'extraire m_e ddans le domaine AES chiffré en utilisant une fonction d'extraction f_e, et sans déchiffrer les données. Plus clairement, comme le chifferement AES prend en entrée un bloc de 128 bits, notre solution consiste à sélectionner secrètement un bit de chaque bloc tatoué-compressé-chiffré et vérifier s'il correspond au bit à insérer de m_e: Si ce bit sélectionné secrètement est égal au bit du message, le bloc est alors considéré comme tatoué et le processus continue avec le bloc suivant. Sinon (c-à-d, si le bit secrètement sélectionné n'est pas égal au bit du message), le bloc tatoué-compressé-chiffré est déchiffré, tatoué et rechiffré jusqu'à ce que ce bit sélectionné dans le bloc chiffré par AES prenne la valeur du bit du message m_e. Lors de la phase de vérification (Figure 5.9, b), l'extraction des messages est effectuée indépendamment dans les deux domaines compressé et chiffré en utilisant la clé secrète de tatouage correspondante, K_{wc} et K_{we} respectivement. Chaque message est utilisé par la suite pour vérifier la fiabilité de l'image dans le domaine correspondant. Pour que notre schéma soit compatible avec le standard DICOM, l'algorithme de chiffrement AES en mode CBC et la compression JPEG-LS ont été utilisé. Ainsi, même si un système n'est pas compatible avec le tatouage, il peut déchiffrer l'image et y accéder, s'il connait bien sûr la clé de chiffrement de l'AES.

Après, nous nous sommes focalisés sur le tatouage réversible, qui permet d'enlever la marque et de réconstuire l'image 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 l'image. Ce fut l'objet du quatrième chapitre. Nous proposons une approche qui permet de tatouer une image chiffrée et donner accès à un message (e.g. des attributs de sécurité ou des métadonnées) que l'image soit chiffrée ou non. Mais si on revient sur l'état de l'art, il existe différentes méthodes qui tatouent des données chiffrées et qui donnent accès à un message dans les deux domaines spatial et chiffré (voir Figure 5.10). Ces méthodes font généralement du tatouage-chiffrement commutatif. Cependant, elles sont soit basées sur un chiffrement homomorphe qui est très complexe en terme de calcul, soit elles ne chiffrent qu'une partie des données à transmettre. Une partie de ces données apparaît alors en clair lors de la transmission et peut être utilisée par un pirate dans une attaque. L'autre classe de méthodes est le tatouage par pré-marquage. Son principe repose sur le fait d'insérer une pré-marque dans le domaine spatial, qui sera perturbée par l'insertion d'un message dans le domaine chiffré. Après, c'est la différence entre la prémarque originale et celle extraite, qui va permettre de décoder le message dans le domaine spatial. Le problème ici c'est que ces schémas ne sont pas réversible. Et donc pour y arriver, on a proposé une méthode qui permet d'insérer une pré-marque réversible, et robuste à l'insertion dans le domaine chiffré, pour pouvoir reconstruire l'image originale.

Comme montré dans Figure 5.11, sa principale originalité réside dans l'insertion, avant l'étape de chiffrement, d'une "pré-marque" M_s réversible et robuste à l'insertion d'un message M_e dans

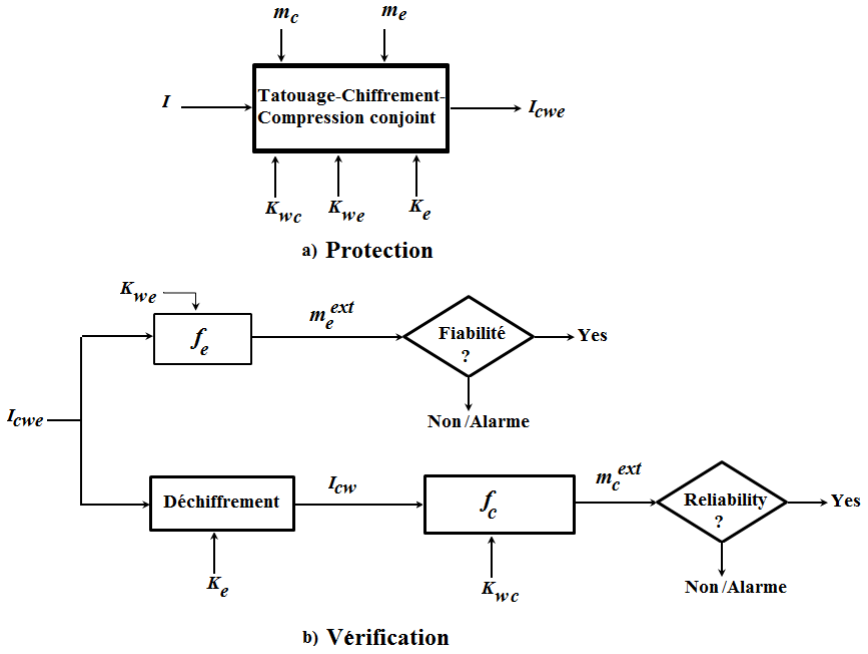


Figure 5.9: Architecture de "Tatouage-Chiffrement-Compression Conjoint".

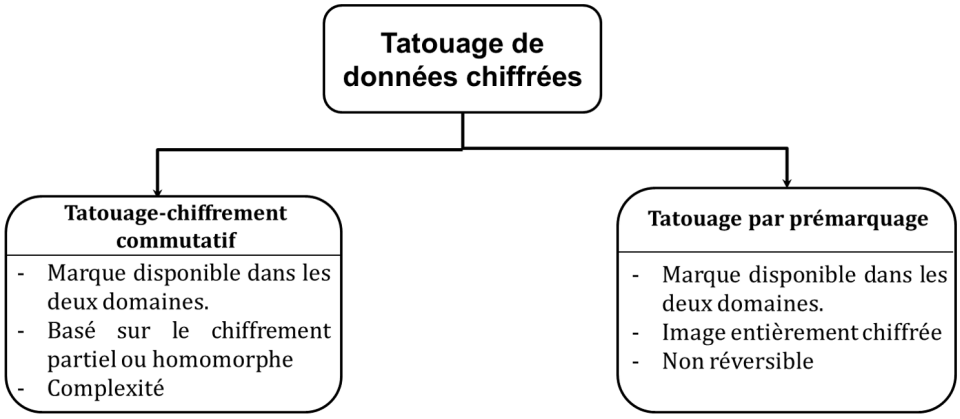


Figure 5.10: Classes de méthodes "Tatouage de données chiffrées".

le domaine chiffré. C'est la perturbation de cette pré-marque, lors du processus de marquage de l'image chiffrée, qui va donner ainsi accès au message M_e dans le domaine spatial. Pour ce faire, nous proposons une version robuste de la modulation de décalage d'histogramme (Histogram Shifting en anglais) pour l'insertion de la pré-marque M_s. Un code correcteur d'erreurs a été utilisé pour coder cette pré-marque pour pouvoir la reconstituer même si des erreurs, dues à l'insertion de M_e dans le domaine chiffré, l'ont altéré. L'insertion de M_e dans le domaine chiffré a été faite par substitution de bits de poids faible. Plus clairement, après avoir diviser l'image chiffrée en des blocs non-superposés et de taille bien définie, un bit de M_e est inséré dans chaque bloc à une position bien déterminée. L'extraction de ce message à partir de l'image chiffrée se fait en suivant les mêmes étapes qu'à l'insertion dans le domaine chiffré. Pour extraire la pré-marque M_s et reconstruire l'image originale à partir de sa version tatouée, il suffit alors d'appliquer le décalage d'histograme inverse tout en utilisant le code correcteur d'erreur pour récupérer correctement les bits de M_s et ainsi l'image originale. Comme dit précedemment, sur la base des erreurs détectées au moment d'extraction de la pré-marque, M_e - le message inséré dans le domaine chiffré - peut être ainsi extrait dans le domaine clair/déchiffré.

Finalement, dans le but de valider nos hypothèses en termes de distorsion, nous nous sommes intéressés à mesurer l'impact du processus de tatouage sur la qualité du diagnostic et donc sur l'interprétation des images tatouées. Pour ce faire, nous avons mis en œuvre un protocole d'étude

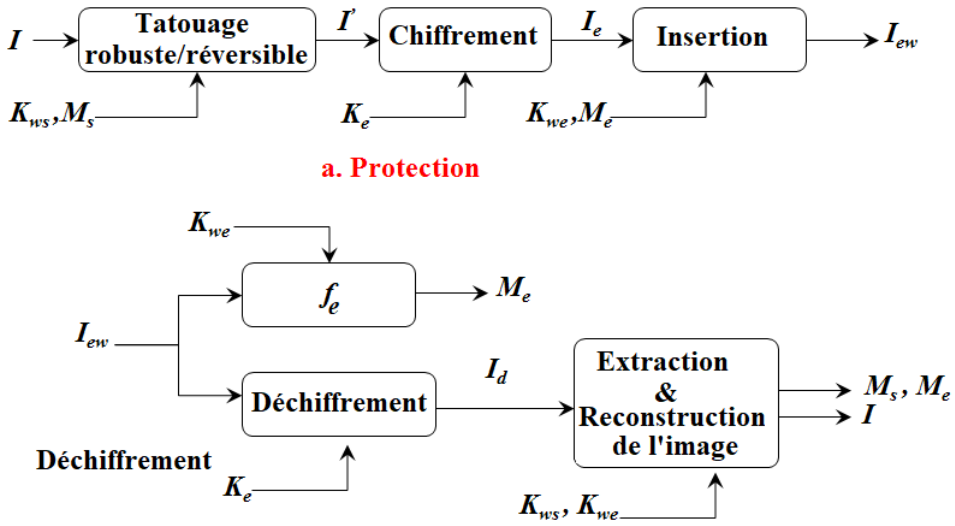


Figure 5.11: Architecture générale de "Tatouage réversible et robuste d'images chiffrées".

Figure 5.12: Image scanner thoracique plumonaire.

subjective. En fait, l'évaluation subjective est considérée comme le moyen le plus fiable pour juger la qualité d'une image, du fait qu'elle fait intervenir l'utilisateur final, c'est à dire les médecins, pour évaluer la qualité diagnostique d'images médicales tatouées. Un premier objectif de cette étude est de confirmer que pour les images scanners, une perte d'information liée au tatouage est tolérable et un deuxième objectif qui consiste à déterminer, pour un algorithme de tatouage, les valeurs seuils qui peuvent être utilisées sans aucune réduction significative de la qualité diagnostique de l'image. A partir de ces résultats, il sera possible d'établir un lien entre les mesures subjectives et objectives de la qualité d'image, pour nous permettre après d'ajuster les paramètres de tatouage de manière automatique.

Pour des raisons pratiques et pour réduire la complexité d'une telle étude, celle-ci est limitée aux images scanners, et plus particulièrement aux images thoraciques pulmonaires (Figure 5.12), à des fins d'identification de nodules.

Ce choix repose sur le fait que ces nodules sont de l'ordre de quelques millimètres, et leur détection requière une lecture approfondie des images qui peuvent être impactée par une dégradation de la qualité de l'image. Aussi, Les images ont déjà été collectées pour une autre étude. Ceci nous fait gagner un temps précieux vu que les données sont déjà anonymisées, et annotées.

Quant à l'algorithme de tatouage, la modulation par substitution de bits de poids faibles a été

utilisée. Cette modulation a été choisie car elle correspond au schéma de tatouage le plus basique et à partir de laquelle, on peut déterminer les paramètres des autres modulations

Afin de réduire davantage le nombre de tests et de limiter l'étude à un petit nombre de radiologues, ce travail s'est déroulé en plusieurs étapes. Dans un premier temps, nous avons réalisé deux pré-études qui visent à réduire la plage de valeurs des paramètres de tatouage. La première pré-étude vise à identifier les paramètres "maximaux" de tatouage à ne pas dépasser, sur la base de mesures objectives en comparaison avec les résultats de l'étude de l'association canadienne des radiologistes CAR (Canadian Association of Radiologists) sur la compression avec perte des images de scanner thoraciques. La seconde pré-étude consiste en une expérimentation subjective où les collaborateurs d'IMT Atlantique et de MEDECOM - personnes non expertes - sont invités à évaluer la qualité visuelle des images tatouées et à identifier un sous-ensemble de «valeurs seuils» de paramètres de tatouage les plus susceptibles d'introduire une déformation visible dans l'image dans le but de limiter les tests avec les radiologues. La dernière étape correspond donc à l'étude subjective avec des experts afin d'évaluer à la fois la qualité diagnostique et visuelle des images de scanner thoraciques pulmonaires tatouées. Malheureusement, cette étape n'a toujours pas commencé en raison de l'indisponibilité des radiologues et des contraintes de temps.

Bibliography

[1] 2019 annual breach barometer report. https://www.protenus.com/2019-breach-barometer. Accessed: 22-07-2019.

[2] How does the general data protection regulation (gdpr) affect gps? https://www.gponline.com/does-general-data-protection-regulation-gdpr-affect-gps/article/1460998. Accessed: April 4, 2018.

[3] Ross Fraser. Iso 27799: Security management in health using iso/iec 17799. In *Canadian Institute for Health Information (CIHI) Partnership Conference. June 2006*, 2006.

[4] Gouenou Coatrieux, Laurent Lecornu, Bulent Sankur, and Ch Roux. A review of image watermarking applications in healthcare. In *2006 International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 4691–4694. IEEE, 2006.

[5] Jorge Miguel Silva, Godinho T Marques, David Silva, and Carlos Costa. Web validation service for ensuring adherence to the dicom standard. *Studies in health technology and informatics*, 235:38–42, 2017.

[6] Sahar Haddad, Gouenou Coatrieux, Michel Cozic, and Dalel Bouslimi. Joint watermarking and lossless jpeg-ls compression for medical image security. *Irbm*, 38(4):198–206, 2017.

[7] Sahar Haddad, Gouenou Coatrieux, Michel Cozic, and Dalel Bouslimi. Joint watermarking and lossless jpeg-ls compression for medical image security. In *Proceedings of the International Conference on Watermarking and Image Processing*, pages 16–21. ACM, 2017.

[8] Sahar Haddad, Gouenou Coatrieux, and Michel Cozic. A new joint watermarking-encryption-jpeg-ls compression method for a priori & a posteriori image protection. In *2018 25th IEEE International Conference on Image Processing (ICIP)*, pages 1688–1692. IEEE, 2018.

[9] The digital universe: Driving data growth in healthcare. http://www.emc.com/analyst-report/digital-universe-healthcare-vertical-report-ar.pdf. Accessed: 31-07-2019.

[10] Elizabeth A Krupinski, Khan Siddiqui, Eliot Siegel, Rasu Shrestha, Edward Grant, Hans Roehrig, and Jiahua Fan. Influence of 8-bit vs. 11-bit digital displays on observer performance and visual search: A multi-center evaluation. *Journal of the Society for Information Display*, 15(6):385–390, 2007.

[11] Robert E Cooke Jr, Michael G Gaeta, Dean M Kaufman, and John G Henrici. Picture archiving and communication system, June 3 2003. US Patent 6,574,629.

[12] Clemens Scott Kruse, Priyanka Karem, Kelli Shifflett, Lokesh Vegi, Karuna Ravi, and Matthew Brooks. Evaluating barriers to adopting telemedicine worldwide: A systematic review. *Journal of telemedicine and telecare*, 24(1):4–12, 2018.

[13] Jeremy M Kahn et al. Virtual visits—confronting the challenges of telemedicine. *N Engl J Med*, 372(18):1684–1685, 2015.

[14] Jessica Kim Cohen. 10 ehr implementations with the biggest price tags in 2017. https://www.beckershospitalreview.com/ehrs/10-ehr-implementations-with-the-biggest-price-tags-in-2017.html. On Line: December 18th, 2017.

[15] Lokesh S Ramamoorthi and Saurabh Shukla. Cloud based telemedicine in neurology clinics: A new horizon. In *2018 9th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*, pages 934–938. IEEE, 2018.

[16] Rishi Saripalle, Christopher Runyan, and Mitchell Russell. Using hl7 fhir to achieve interoperability in patient health record. *Journal of biomedical informatics*, 94:103188, 2019.

[17] What's next for the health-care data center? http://www.datacenterjournal.com/whats-healthcare-data-center/. Accessed: 06-04-2015.

[18] Sachin P Nanavati and Prasanta K Panigrahi. Wavelets: applications to image compression-i. *Resonance*, 10(2):52–61, 2005.

[19] Xie Kai, Yang Jie, Zhu Yue Min, and Li Xiao Liang. Hvs-based medical image compression. *European journal of radiology*, 55(1):139–145, 2005.

[20] Lingyun Xiang, Yan Li, Wei Hao, Peng Yang, and Xiaobo Shen. Reversible natural language watermarking using synonym substitution and arithmetic coding. *Comput., Mater. Continua*, 55(3):541–559, 2018.

[21] Abhijit Jas, Jayabrata Ghosh-Dastidar, Mom-Eng Ng, and Nur A Touba. An efficient test vector compression scheme using selective huffman coding. *IEEE transactions on computer-aided design of integrated circuits and systems*, 22(6):797–806, 2003.

[22] Walter D Leon-Salas. Encoding compressive sensing measurements with golomb-rice codes. In *2015 IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 2177–2180. IEEE, 2015.

[23] Bogdan Rusyn, Oleksiy Lutsyk, Yuriy Lysak, Adolf Lukenyuk, and Lubomyk Pohreliuk. Lossless image compression in the remote sensing applications. In *2016 IEEE First International Conference on Data Stream Mining & Processing (DSMP)*, pages 195–198. IEEE, 2016.

[24] Alexander Zemliachenko, Vladimir Lukin, Nikolay Ponomarenko, Karen Egiazarian, and Jaakko Astola. Still image/video frame lossy compression providing a desired visual quality. *Multidimensional Systems and Signal Processing*, 27(3):697–718, 2016.

[25] Feng Liu, Miguel Hernandez-Cabronero, Victor Sanchez, Michael Marcellin, and Ali Bilgin. The current role of image compression standards in medical imaging. *Information*, 8(4):131, 2017.

[26] Acr–aapm–siim technical standard for electronic practice of medical imaging. https://www.acr.org/-/media/ACR/Files/Practice-Parameters/Elec-Practice-MedImag.pdf. Revised: 2017.

[27] David Koff, Peter Bak, Paul Brownrigg, Danoush Hosseinzadeh, April Khademi, Alex Kiss, Luigi Lepanto, Tracy Michalak, Harry Shulman, and Andrew Volkening. Pan-canadian evaluation of irreversible compression ratios ("lossy" compression) for development of national guidelines. *Journal of digital imaging*, 22(6):569, 2009.

[28] François-André Allaert, Liliane Dusserre, and B Leclercq. La sécurité des systèmes d'information médicohospitaliers. *Informatique et Santé*, 9:149–157, 1997.

[29] Menaces informatiques et pratiques de sécurité en france (2018). http://s3-eu-west-1.amazonaws.com/static.hospimedia.fr/documents/196371/3440/CLUSIF-MIPS2018.pdf?1530885161. Accessed: 07-2019.

[30] Patients stealing $52m worth of items from hospitals, says survey. https://www.healthleadersmedia.com/clinical-care/patients-stealing-52m-worth-items-hospitals-says-survey. Accessed: FEBRUARY 10, 2010.

[31] Healthcare cybersecurity and the human factor: Using risk-based authentication that considers behavioral factors. https://blog.identityautomation.com/healthcare-cybersecurity-and-the-human-factor. Accessed: August 14, 2018.

[32] 58% of all healthcare breaches are initiated by insiders. https://www.forbes.com/sites/louiscolumbus/2018/08/31/58-of-all-healthcare-breaches-are-initiated-by-insiders/#2d839a87601a. Accessed: August 31, 2018.

[33] Researcher: Hospital data breaches connected to patient deaths. https://digitalguardian.com/blog/researcher-hospital-data-breaches-connected-patient-deaths. Accessed: April 2, 2018.

[34] Steven H Miles. *The Hippocratic Oath and the ethics of medicine*. Oxford University Press, 2005.

[35] French code of medical ethics (edition of november 2013. https://www.conseil-national.medecin.fr/sites/default/files/external-package/edition/168yke7/code_de_deontologie_version_anglaise.pdf. Edited: 11-2013.

[36] Panocrim 2018 - des attaques au cœur des métiers - santé. https://clusif.fr/publications/panocrim-2018-des-attaques-au-coeur-des-metiers-sante/?visible=public. edited: 2018.

[37] George J Annas et al. Hipaa regulations-a new era of medical-record privacy? *New England Journal of Medicine*, 348(15):1486–1490, 2003.

[38] Félix Tréguer. Le droit pénal de la fraude informatique, nouvel ami des censeurs?. liberté d'expression (loi godfrain du 5 janvier 1988 et code pénal). *La Revue des droits de l'homme. Revue du Centre de recherches et d'études sur les droits fondamentaux*, 2015.

[39] John Mcdonald, Nouha Oualha, Armaud Puccetti, Artur Hecker, and Frederic Planchon. Application of ebios for the risk assessment of ict use in electrical distribution sub-stations. In *2013 IEEE Grenoble Conference*, pages 1–6. IEEE, 2013.

[40] Diomidis H Stamatis. *Failure mode and effect analysis: FMEA from theory to execution*. ASQ Quality press, 2003.

[41] Méthode Harmonisée d'Analyse de Risques. Mehari. *CLUSIF, France*, 2007.

[42] How risk assessment scores are calculated. https://https://jazz.net/help-dev/clm/index.jsp?topic=%2Fcom.ibm.rational.test.qm.doc%2Ftopics%2Fc_how_risk_is_calculated.html.

[43] David Ferraiolo, D Richard Kuhn, and Ramaswamy Chandramouli. *Role-based access control*. Artech House, 2003.

[44] Anas Abou El Kalam, R El Baida, Philippe Balbiani, Salem Benferhat, Frédéric Cuppens, Yves Deswarte, Alexandre Miege, Claire Saurel, and Gilles Trouessin. Organization based access control. In *Proceedings POLICY 2003. IEEE 4th International Workshop on Policies for Distributed Systems and Networks*, pages 120–131. IEEE, 2003.

[45] Chirag Langaliya and Rajanikanth Aluvalu. Enhancing cloud security through access control models: A survey. *International Journal of Computer Applications*, 112(7), 2015.

[46] Stephen P Kruger and Olgierd S Pieczul. Enabling granular discretionary access control for data stored in a cloud computing environment, March 24 2015. US Patent 8,990,950.

[47] Chia-Hui Liu, Yu-Fang Chung, Tzer-Shyong Chen, and Sheng-De Wang. The enhancement of security in healthcare information systems. *Journal of medical systems*, 36(3):1673–1688, 2012.

[48] Thomas B Slayton. Ransomware: The virus attacking the healthcare industry. *Journal of Legal Medicine*, 38(2):287–311, 2018.

[49] Gurpreet Singh. A study of encryption algorithms (rsa, des, 3des and aes) for information security. *International Journal of Computer Applications*, 67(19), 2013.

[50] Dag Arne Osvik, Joppe W Bos, Deian Stefan, and David Canright. Fast software aes encryption. In *International Workshop on Fast Software Encryption*, pages 75–93. Springer, 2010.

[51] Christophe De Canniere and Bart Preneel. Trivium. In *New Stream Cipher Designs*, pages 244–266. Springer, 2008.

[52] Allam Mousa and Ahmad Hamad. Evaluation of the rc4 algorithm for data encryption. *IJCSA*, 3(2):44–56, 2006.

[53] L'agrément d'hébergeur de données de santé d'orange healthcare. `https://healthcare.orange.com/wp-content/uploads/sites/14/2018/05/contenuhdsagrement_certificationv3.pdf`. Update: May, 2018.

[54] Dylan Yaga, Peter Mell, Nik Roby, and Karen Scarfone. Blockchain technology overview. *arXiv preprint arXiv:1906.11078*, 2019.

[55] Ingemar J Cox, Matthew L Miller, Jeffrey Adam Bloom, and Chris Honsinger. *Digital watermarking*, volume 53. Springer, 2002.

[56] Gouenou Coatrieux, Henri Maître, Bulent Sankur, Yann Rolland, and René Collorec. Relevance of watermarking in medical imaging. In *Proceedings 2000 IEEE EMBS International Conference on Information Technology Applications in Biomedicine. ITAB-ITIS 2000. Joint Meeting Third IEEE EMBS International Conference on Information Technol*, pages 250–255. IEEE, 2000.

[57] Wei Pan, Gouenou Coatrieux, Nora Cuppens-Boulahia, Frederic Cuppens, and Christian Roux. Medical image integrity control combining digital signature and lossless watermarking. In *Data privacy management and autonomous spontaneous security*, pages 153–162. Springer, 2009.

[58] Hui Huang, Gouenou Coatrieux, Huazhong Shu, Limin Luo, and Christian Roux. Blind integrity verification of medical images. *IEEE transactions on information technology in biomedicine*, 16(6):1122–1126, 2012.

[59] Wei Pan, Gouenou Coatrieux, Nora Cuppens-Boulahia, Frederic Cuppens, and Christian Roux. Watermarking to enforce medical image access and usage control policy. In *2010 Sixth International Conference on Signal-Image Technology and Internet Based Systems*, pages 251–260. IEEE, 2010.

[60] Rajendra Acharya, UC Niranjan, S Sitharama Iyengar, N Kannathal, and Lim Choo Min. Simultaneous storage of patient information with medical images in the frequency domain. *Computer methods and programs in biomedicine*, 76(1):13–19, 2004.

[61] Hong-yuan Chen and Yue-sheng Zhu. A robust watermarking algorithm based on qr factorization and dct using quantization index modulation technique. *Journal of Zhejiang University SCIENCE C*, 13(8):573–584, 2012.

[62] Alain Hore and Djemel Ziou. Image quality metrics: Psnr vs. ssim. In *2010 20th International Conference on Pattern Recognition*, pages 2366–2369. IEEE, 2010.

[63] Zhou Wang and Alan C Bovik. A universal image quality index. *IEEE signal processing letters*, 9(3):81–84, 2002.

[64] Xiyao Liu, Jieting Lou, Hui Fang, Yan Chen, Pingbo Ouyang, Yifan Wang, Beiji Zou, and Lei Wang. A novel robust reversible watermarking scheme for protecting authenticity and integrity of medical images. *IEEE Access*, 7:76580–76598, 2019.

[65] Osama M Alattar and Adnan M Alattar. Hierarchical watermark detector, February 20 2018. US Patent 9,898,792.

[66] Mehdi Rabizadeh, Maryam Amirmazlaghani, and Mahmoud Ahmadian-Attari. A new detector for contourlet domain multiplicative image watermarking using bessel k form distribution. *Journal of Visual Communication and Image Representation*, 40:324–334, 2016.

[67] Ran-Zan Wang, Chi-Fang Lin, and Ja-Chen Lin. Hiding data in images by optimal moderately-significant-bit replacement. *Electronics Letters*, 36(25):2069–2070, 2000.

[68] Brian Chen and Gregory W Wornell. Quantization index modulation: A class of provably good methods for digital watermarking and information embedding. *IEEE Transactions on Information Theory*, 47(4):1423–1443, 2001.

[69] Deepthi Anand and UC Niranjan. Watermarking medical images with patient information. In *Proceedings of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Vol. 20 Biomedical Engineering Towards the Year 2000 and Beyond (Cat. No. 98CH36286)*, volume 2, pages 703–706. IEEE, 1998.

[70] Hirak Kumar Maity and Santi Prasad Maity. Joint robust and reversible watermarking for medical images. *Procedia technology*, 6:275–282, 2012.

[71] Abhilasha Sharma, Amit Kumar Singh, and Satya Prakash Ghrera. Robust and secure multiple watermarking for medical images. *Wireless Personal Communications*, 92(4):1611–1624, 2017.

[72] Asokan Sivaprakash, Samuel Nadar Edward Rajan, and S Selvaperumal. A novel robust medical image watermarking employing firefly optimization for secured telemedicine. *Journal of Medical Imaging and Health Informatics*, 9(7):1373–1381, 2019.

[73] Hyeon-Uk Seo, Qun Wei, Seong-Geun Kwon, and Kyu-Ik Sohng. Medical image watermarking using bit threshold map based on just noticeable distortion in discrete cosine transform. *Technology and Health Care*, 25(S1):367–375, 2017.

[74] K Kalaivani. An efficient watermarking scheme for medical data security with the aid of neural network. *Brazilian Archives of Biology and Technology*, 59(SPE2), 2016.

[75] Preeti Bhinder, Kulbir Singh, and Neeru Jindal. Image-adaptive watermarking using maximum likelihood decoder for medical images. *Multimedia Tools and Applications*, 77(8):10303–10328, 2018.

[76] Mohamed Karasad, Dalel Bouslimi, Michel Cozic, and Gouenou Coatrieux. Watermarking of radiographic images based on quantum noise modulation. In *2016 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*, pages 9–12. IEEE, 2016.

[77] Ling Wang, Jianming Lu, Yeqiu Li, Takashi Yahagi, and Takahide Okamoto. Noise reduction using wavelet with application to medical x-ray image. In *2005 IEEE International Conference on Industrial Technology*, pages 33–38. IEEE, 2005.

[78] Frank Y Shih and Xin Zhong. High-capacity multiple regions of interest watermarking for medical images. *Information Sciences*, 367:648–659, 2016.

[79] Rohit Thanki, Surekha Borra, Vedvyas Dwivedi, and Komal Borisagar. A roni based visible watermarking approach for medical image authentication. *Journal of medical systems*, 41(9):143, 2017.

[80] Narendra K Pareek and Vinod Patidar. Medical image protection using genetic algorithm operations. *Soft Computing*, 20(2):763–772, 2016.

[81] Hui Liang Khor, Siau-Chuin Liew, and Jasni Mohd Zain. Region of interest-based tamper detection and lossless recovery watermarking scheme (roi-dr) on ultrasound medical images. *Journal of digital imaging*, 30(3):328–349, 2017.

[82] Ali Al-Haj, Ahmad Mohammad, et al. Crypto-watermarking of transmitted medical images. *Journal of digital imaging*, 30(1):26–38, 2017.

[83] Aleš Roček, Michal Javorník, Karel Slavíček, and Otto Dostál. Reversible watermarking in medical imaging with zero distortion in roi. In *2017 24th IEEE International Conference on Electronics, Circuits and Systems (ICECS)*, pages 356–359. IEEE, 2017.

[84] Osamah M Al-Qershi and Bee Ee Khoo. High capacity data hiding schemes for medical images based on difference expansion. *Journal of Systems and Software*, 84(1):105–112, 2011.

[85] Gouenou Coatrieux, Wei Pan, Nora Cuppens-Boulahia, Frédéric Cuppens, and Christian Roux. Reversible watermarking based on invariant image classification and dynamic histogram shifting. *IEEE Transactions on information forensics and security*, 8(1):111–120, 2012.

[86] R Rajkumar and A Vasuki. Reversible and robust image watermarking based on histogram shifting. *Cluster Computing*, 22(5):12313–12323, 2019.

[87] Xiang Yu, Xiang Wang, and Qingqi Pei. Reversible watermarking based on multi-dimensional prediction-error expansion. *Multimedia Tools and Applications*, 77(14):18085–18104, 2018.

[88] Eric Wu, Kevin Wu, David Cox, and William Lotter. Conditional infilling gans for data augmentation in mammogram classification. In *Image Analysis for Moving Organ, Breast, and Thoracic Images*, pages 98–106. Springer, 2018.

[89] Ferda Ernawan and Muhammad Nomani Kabir. A robust image watermarking technique with an optimal dct-psychovisual threshold. *IEEE Access*, 6:20464–20480, 2018.

[90] Nassiri Boujemaa, EL Yousef, Latif Rachid, Bsiss Mohammed Aziz, et al. Fragile watermarking of medical image for content authentication and security. *IJCSN-International Journal of Computer Science and Network*, 5(5), 2016.

[91] Samia Belkacem, Zohir Dibi, and Ahmed Bouridane. A masking model of hvs for image watermaking in the dct domain. In *2007 14th IEEE International Conference on Electronics, Circuits and Systems*, pages 330–334. IEEE, 2007.

[92] OuJun Lou, ShaoHua Li, ZhaoXia Liu, and ShuangTong Tang. A novel multi-bit watermarking algorithm based on hvs. In *2014 Sixth international symposium on parallel architectures, algorithms and programming*, pages 278–281. IEEE, 2014.

[93] Bijan G Mobasseri and Robert J Berger. A foundation for watermarking in compressed domain. *IEEE Signal Processing Letters*, 12(5):399–402, 2005.

[94] Jessica Fridrich, Miroslav Goljan, Qing Chen, and Vivek Pathak. Lossless data embedding with file size preservation. In *Security, Steganography, and Watermarking of Multimedia Contents VI*, volume 5306, pages 354–365. International Society for Optics and Photonics, 2004.

[95] A Venkata Subramanyam, Sabu Emmanuel, and Mohan S Kankanhalli. Robust watermarking of compressed and encrypted jpeg2000 images. *IEEE Transactions on Multimedia*, 14(3):703–716, 2011.

[96] Dalila Goudia, Marc Chaumont, William Puech, and Naima Hadj Said. Tatouage et compression conjoint dans jpeg2000 avec un algorithme de quantification codée par treillis (tcq). In *CORESA: COmpression et REprésentation des Signaux Audiovisuels*, 2010.

[97] Rohini Srivastava, Basant Kumar, Amit Kumar Singh, and Anand Mohan. Computationally efficient joint imperceptible image watermarking and jpeg compression: a green computing approach. *Multimedia Tools and Applications*, 77(13):16447–16459, 2018.

[98] Roberto Caldelli, Francesco Filippini, and Mauro Barni. Joint near-lossless compression and watermarking of still images for authentication and tamper localization. *Signal Processing: Image Communication*, 21(10):890–903, 2006.

[99] The Joint Photographic Experts Group (JPEG)/ FCD 14495. Lossless and near-lossless coding of continuous-tone still image jpeg-ls. *The International Standards Organization (ISO)/The International Telegraph and Telephone Consultative Committee (CCITT)*, July 1997.

[100] Marcelo J Weinberger, Gadiel Seroussi, and Guillermo Sapiro. The loco-i lossless image compression algorithm: Principles and standardization into jpeg-ls. *IEEE Transactions on Image processing*, 9(8):1309–1324, 2000.

[101] Gregory K Wallace. The jpeg still picture compression standard. *IEEE transactions on consumer electronics*, 38(1):xviii–xxxiv, 1992.

[102] Sherly Kk and Neethu Mathai. A modified framework for secure and robust blind data hiding in videos using chaotic encryption and forbidden zone concept. *International Journal of Scientific and Engineering Research*, Volume 4:Page 1, 08 2013.

[103] Dalel Bouslimi and Gouenou Coatrieux. A crypto-watermarking system for ensuring reliability control and traceability of medical images. *Signal Processing: Image Communication*, 47:160–169, 2016.

[104] P SUCKLING J. The mammographic image analysis society digital mammogram database. *Digital Mammo*, pages 375–386, 1994.

[105] Richard Dosselmann and Xue Dong Yang. A comprehensive assessment of the structural similarity index. *Signal, Image and Video Processing*, 5(1):81–91, 2011.

[106] Keshi Chen and Tenkasi V Ramabadran. Near-lossless compression of medical images through entropy-coded dpcm. *IEEE Transactions on Medical Imaging*, 13(3):538–548, 1994.

[107] Dalel Bouslimi, Gouenou Coatrieux, Michel Cozic, and Christian Roux. A joint encryption/watermarking system for verifying the reliability of medical images. *IEEE Transactions on Information Technology in Biomedicine*, 16(5):891–899, 2012.

[108] Yingliang He, Gaobo Yang, and Ningbo Zhu. A real-time dual watermarking algorithm of h. 264/avc video stream for video-on-demand service. *AEU-International Journal of Electronics and Communications*, 66(4):305–312, 2012.

[109] Xinpeng Zhang. Reversible data hiding in encrypted image. *IEEE signal processing letters*, 18(4):255–258, 2011.

[110] Shilpa P Metkar and Milind V Lichade. Digital image security improvement by integrating watermarking and encryption technique. In *2013 IEEE international conference on signal processing, computing and control (ISPCC)*, pages 1–6. IEEE, 2013.

[111] Zhenxing Qian, Xinpeng Zhang, Yanli Ren, and Guorui Feng. Block cipher based separable reversible data hiding in encrypted images. *Multimedia Tools and Applications*, 75(21):13749–13763, 2016.

[112] Xinpeng Zhang. Separable reversible data hiding in encrypted image. *IEEE transactions on information forensics and security*, 7(2):826–832, 2011.

[113] Xiaochun Cao, Ling Du, Xingxing Wei, Dan Meng, and Xiaojie Guo. High capacity reversible data hiding in encrypted images by patch-level sparse representation. *IEEE transactions on cybernetics*, 46(5):1132–1143, 2015.

[114] Roland Schmitz, Shujun Li, Christos Grecos, and Xinpeng Zhang. A new approach to commutative watermarking-encryption. In *IFIP International Conference on Communications and Multimedia Security*, pages 117–130. Springer, 2012.

[115] Dalel Bouslimi, Gouenou Coatrieux, Michel Cozic, and Christian Roux. Data hiding in encrypted images based on predefined watermark embedding before encryption process. *Signal Processing: Image Communication*, 47:263–270, 2016.

[116] Dawen Xu and Rangding Wang. Watermarking in h. 264/avc compressed domain using exp-golomb code words mapping. *Optical Engineering*, 50(9):097402, 2011.

[117] N Naveen and Ebin M Manuel. Tamper detection and authentication in jpeg 2000 images using chaotic watermarking scheme. In *2011 International Conference on Signal Processing, Communication, Computing and Networking Technologies*, pages 370–373. IEEE, 2011.

[118] Jing Long, Zhaoxia Yin, Jinpeng Lv, and Xinpeng Zhang. Rotation based reversible data hiding for jpeg images. *IETE Technical Review*, 33(6):607–614, 2016.

[119] Fangchao Wang and Sen Bai. Jpeg image encryption by shuffling dct coefficients in defined block. In *2013 International Conference on Computational and Information Sciences*, pages 60–63. IEEE, 2013.

[120] Yuanyuan Sun, Rudan Xu, Lina Chen, and Xiaopeng Hu. Image compression and encryption scheme using fractal dictionary and julia set. *IET Image Processing*, 9(3):173–183, 2015.

[121] BM Shreedhar, IL Vishal, and N Hemavathi. Image encryption-then-compression system via prediction error clustering and lossless encoding. *International Journal of Innovative Research in Information Security*, 4(2):33–39, 2015.

[122] Harmanpreet Kaur Aujla and Rajesh Sharma. Designing an efficient image encryption-then compression system with haar and daubechies wavelet. *International Journal of Computer Science and Information Technologies*, 5(6):7784–7788, 2014.

[123] Camit Hazay, Ashish Jagmohan, Demijan Klinc, Hugo M Krawczyk, and Tal Rabin. Compressing block-cipher encrypted data, January 13 2015. US Patent 8,934,630.

[124] Xiao-Jun Tong, Penghui Chen, and Miao Zhang. A joint image lossless compression and encryption method based on chaotic map. *Multimedia Tools and Applications*, 76(12):13995–14020, 2017.

[125] Ayman Alfalou, C Brosseau, and Nadine Abdallah. Simultaneous compression and encryption of color video images. *Optics Communications*, 338:371–379, 2015.

[126] Zhenxing Qian, Haisheng Xu, Xiangyang Luo, and Xinpeng Zhang. New framework of reversible data hiding in encrypted jpeg bitstreams. *IEEE Transactions on Circuits and Systems for Video Technology*, 29(2):351–362, 2018.

[127] Vinita Shadangi, Siddharth Kumar Choudhary, K Abhimanyu Kumar Patro, and Bubhudendra Acharya. Novel arnold scrambling based cbc-aes image encryption. *Int J Control Theory Appl*, 10(15):93–105, 2017.

[128] Unary coding. `https://vicente-gonzalez-ruiz.github.io/unary_coding//`. Accessed: 31-12-2018.

[129] Adriana Vasilache. Order adaptive golomb rice coding for high variability sources. In *2017 25th European Signal Processing Conference (EUSIPCO)*, pages 1789–1793. IEEE, 2017.

[130] Pitools medical. `https://accusoft.com/products/pictools-medical\/features/`.

[131] Lossless jpeg. `en.wikipedia.org/wiki/Lossless_JPEG`. edited: 19-04-2019.

[132] Mohammed Nazeh Abdul Wahid, Abdulrahman Ali, Babak Esparham, and Mohamed Marwan. A comparison of cryptographic algorithms: Des, 3des, aes, rsa and blowfish for guessing attacks prevention. *Journal Computer Science Applications and Information Technology*, 3:1–7, 2018.

[133] Gouenou Coatrieux, Catherine Quantin, Julien Montagner, Maniane Fassa, François-André Allaert, and Christian Roux. Watermarking medical images with anonymous patient identification to verify authenticity. In *MIE*, volume 136, pages 667–672. Citeseer, 2008.

[134] Yong Zhang, Xueqian Li, and Wengang Hou. A fast image encryption scheme based on aes. In *2017 2nd International Conference on Image, Vision and Computing (ICIVC)*, pages 624–628. IEEE, 2017.

[135] Xinmin Zhou, Weidong Zhao, Zhicheng Wang, and Li Pan. Security theory and attack analysis for text watermarking. In *2009 International Conference on E-Business and Information System Security*, pages 1–6. IEEE, 2009.

[136] Jun Tian. Reversible data embedding using a difference expansion. *IEEE transactions on circuits and systems for video technology*, 13(8):890–896, 2003.

[137] Chunlin Song, Jie Sang, and Sud Sudirman. A buyer-seller watermarking protocol for digital secondary market. *Multimedia Tools and Applications*, 77(1):225–249, 2018.

[138] S Priya and B Santhi. A novel visual medical image encryption for secure transmission of authenticated watermarked medical images. *Mobile Networks and Applications*, pages 1–8, 2019.

[139] Li Jiang. The identical operands commutative encryption and watermarking based on homomorphism. *Multimedia Tools and Applications*, 77(23):30575–30594, 2018.

[140] Vaibhav B Joshi and Mehul Raval. An improved commutative reversible watermarking and encryption for fingerprint image. *OSF Preprints*, 2019.

[141] Michela Cancellaro, Federica Battisti, Marco Carli, Giulia Boato, Francesco GB De Natale, and Alessandro Neri. A commutative digital image watermarking and encryption method in the tree structured haar transform domain. *Signal Processing: Image Communication*, 26(1):1–12, 2011.

[142] Liu Yao and Xue Shuai. Accelerate the paillier cryptosystem in cryptdb by chinese remainder theorem. In *2018 20th International Conference on Advanced Communication Technology (ICACT)*, pages 74–77. IEEE, 2018.

[143] Dalel Bouslimi, Gouenou Coatrieux, and Christian Roux. A joint encryption/watermarking algorithm for verifying the reliability of medical images: Application to echographic images. *Computer methods and programs in biomedicine*, 106(1):47–54, 2012.

[144] Hiba Abdel-Nabi and Ali Al-Haj. Efficient joint encryption and data hiding algorithm for medical images security. In *2017 8th international conference on information and communication systems (ICICS)*, pages 147–152. IEEE, 2017.

[145] Dalel Bouslimi, Gouenou Coatrieux, Michel Cozic, and Ch Roux. Combination of watermarking and joint watermarking-decryption for reliability control and traceability of medical images. In *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 4495–4498. IEEE, 2014.

[146] Mehmet Utku Celik, Aweke N Lemma, Stefan Katzenbeisser, and Michiel van der Veen. Lookup-table-based secure client-side embedding for spread-spectrum watermarks. *IEEE Transactions on Information Forensics and Security*, 3(3):475–487, 2008.

[147] Ross Anderson and Charalampos Manifavas. Chameleon—a new kind of stream cipher. In *International Workshop on Fast Software Encryption*, pages 107–113. Springer, 1997.

[148] Zhenxing Qian and Xinpeng Zhang. Reversible data hiding in encrypted images with distributed source encoding. *IEEE Transactions on Circuits and Systems for Video Technology*, 26(4):636–646, 2015.

[149] Shuli Zheng, Dandan Li, Donghui Hu, Dengpan Ye, Lina Wang, and Jinwei Wang. Lossless data hiding algorithm for encrypted images with high capacity. *Multimedia Tools and Applications*, 75(21):13765–13778, 2016.

[150] Jiantao Zhou, Weiwei Sun, Li Dong, Xianming Liu, Oscar C Au, and Yuan Yan Tang. Secure reversible image data hiding over encrypted domain via key modulation. *IEEE transactions on circuits and systems for video technology*, 26(3):441–452, 2015.

[151] Hao-Tian Wu, Yiu-ming Cheung, and Jiwu Huang. Reversible data hiding in paillier cryptosystem. *Journal of Visual Communication and Image Representation*, 40:765–771, 2016.

[152] Dalel Bouslimi, Gouenou Coatrieux, Michel Cozic, and Ch Roux. An a priori and a posteriori protection by means of data hiding of encrypted images: application to ultrasound images. In *The International Conference on Health Informatics*, pages 220–223. Springer, 2014.

[153] Dalel Bouslimi, Reda Bellafqira, and Gouenou Coatrieux. Data hiding in homomorphically encrypted medical images for verifying their reliability in both encrypted and spatial domains. In *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 2496–2499. IEEE, 2016.

[154] Zhaoxia Yin, Andrew Abel, Jin Tang, Xinpeng Zhang, and Bin Luo. Reversible data hiding in encrypted images based on multi-level encryption and block histogram modification. *Multimedia Tools and Applications*, 76(3):3899–3920, 2017.

[155] Xiaotian Wu and Wei Sun. High-capacity reversible data hiding in encrypted images by prediction error. *Signal processing*, 104:387–400, 2014.

[156] Yuanzhi Yao, Weiming Zhang, Hui Wang, Hang Zhou, and Nenghai Yu. Content-adaptive reversible visible watermarking in encrypted images. *Signal Processing*, 164:386–401, 2019.

[157] S Sriadhi, Robbi Rahim, and Ansari Saleh Ahmar. Rc4 algorithm visualization for cryptography education. In *Journal of Physics: Conference Series*, volume 1028, page 012057. IOP Publishing, 2018.

[158] Pim Tuyls, Henk DL Hollmann, Jack H Van Lint, and LMGM Tolhuizen. Xor-based visual cryptography schemes. *Designs, Codes and Cryptography*, 37(1):169–186, 2005.

[159] Anne Canteaut, Sergiu Carpov, Caroline Fontaine, Jacques Fournier, Benjamin Lac, Marıa Naya-Plasencia, Renaud Sirdey, and Assia Tria. End-to-end data security for iot: from a cloud of encryptions to encryption in the cloud. In *Proc. IEEE Conf.(Cesar)*, pages 1–21, 2017.

[160] Zhicheng Ni, Yun-Qing Shi, Nirwan Ansari, and Wei Su. Reversible data hiding. *IEEE Transactions on circuits and systems for video technology*, 16(3):354–362, 2006.

[161] Jasni M Zain, Abdul RM Fauzi, and Azian A Aziz. Clinical evaluation of watermarked medical images. In *2006 International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 5459–5462. IEEE, 2006.

[162] Mertcan Özdemir, Osman Eroğul, and Aytekin Ünlü. Investigation of ballistic gelatin based phantom models for computed tomography, x-ray and ultrasound imaging devices. In *2019 Medical Technologies Congress (TIPTEKNO)*, pages 1–4. IEEE, 2019.

[163] Nedia Nouri, Denis Abraham, Jean-Marie Moureaux, Michel Dufaut, Jacques Hubert, and Manuela Perez. Evaluation subjective de la qualité de vidéos encodées MPEG2 dans un contexte de télé-robotique chirurgicale. In *Sixième Conférence Internationale Francophone d'Automatique, CIFA 2010*, June 2010.

[164] Aladine Chetouani. *Vers un système d'évaluation de la qualité d'image multi-critères*. PhD thesis, Université Paris-Nord, 2010.

[165] Gouenou Coatrieux, Henri Maitre, and Bulent Sankur. Strict integrity control of biomedical images. In *Security and watermarking of multimedia contents III*, volume 4314, pages 229–240. International Society for Optics and Photonics, 2001.

[166] Christine Cavaro-Ménard, Patrick Le Callet, Dominique Barba, and Jean Yves Tanguy. Quality assessment of lossy compressed medical images, 2010.

[167] Olav Christianson, Joseph JS Chen, Zhitong Yang, Ganesh Saiprasad, Alden Dima, James J Filliben, Adele Peskin, Christopher Trimble, Eliot L Siegel, and Ehsan Samei. An improved index of image quality for task-based performance of ct iterative reconstruction across three commercial implementations. *Radiology*, 275(3):725–734, 2015.

[168] Soerjomataram I Siegel RL Torre LA Jemal A Bray F, Ferlay J. Global cancer statistics 2018: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. ca cancer j clin, in press. `https://www.wcrf.org/dietandcancer/cancer-trends/lung-cancer-statistics`. Accessed: 2018.

[169] Mizuho Nishio. Computer-aided diagnosis of lung nodules: Systems for estimation of lung cancer probability and false-positive reduction of lung nodule detection. *Lung Imaging and CADx*, page 107, 2019.

[170] G Samuel, Geoffrey McLennan, Luc Bidaut, Michael F McNitt-Gray, et al. Data from lidc-idri. *The Cancer Imaging Archive*, 2015.

[171] BT Series. Methodology for the subjective assessment of the quality of television pictures. *Recommendation ITU-R BT*, pages 500–13, 2012.

[172] Rafał K Mantiuk, Anna Tomaszewska, and Radosław Mantiuk. Comparison of four subjective methods for image quality assessment. In *Computer graphics forum*, volume 31, pages 2478–2491. Wiley Online Library, 2012.

[173] Olga Ferrer-Roca and Marcelo C Sosa-Iudicissa. *Handbook of telemedicine*, volume 54. IOS press, 1998.

[174] Agence internationale pour la recherche du cancer, globocan : Lung cancer estimated incidence, mortality and prevalence worldwide in 2012. `http://globocan.iarc.fr/Pages/fact_sheets_cancer.aspx`. Accessed: 2012.

[175] Will Kenton. Sensitivity analysis. `https://www.investopedia.com/terms/s/sensitivityanalysis.asp`. Updated: Sep 29, 2019.

[176] Mary McHugh. Interrater reliability: The kappa statistic. *Biochemia medica : časopis Hrvatskoga društva medicinskih biokemičara / HDMB*, 22:276–82, 10 2012.

[177] S. Haddad, G. Coatrieux, A. Moreau-Gaudry, and M. Cozic. Joint watermarking-encryption-jpeg-ls for medical image reliability control in encrypted and compressed domains. *IEEE Transactions on Information Forensics and Security*, pages 1–1, 2020.

UNIVERSITE BRETAGNE LOIRE / MATHSTIC

IMT Atlantique
Bretagne-Pays de la Loire
École Mines-Télécom

Titre : Protection d'images médicales chiffrées et/ou compressées par tatouage.

Mots clés : Imagerie médicale, Sécurité, Tatouage, Chiffrement, Compression JPEG-LS/JPEG.

Résumé : L'évolution des technologies du multimédia et des communications s'exprime dans le domaine de la santé par la mise à disposition de nouveaux moyens de partage et d'accès distant aux données de l'imagerie médicale. Dans un tel contexte, la question de la sécurité de ces données est particulièrement sensible, notamment en termes de confidentialité, fiabilité et traçabilité. Ces travaux de thèse ont trait à combinaison voire la fusion de tatouage et de chiffrement de données pour la protection des images médicales. Cependant, le déploiement de ces mécanismes de sécurité dans le domaine de la santé doit prendre en compte les spécificités de ce domaine. Notamment, du fait que les images médicales constituent de grands volumes de données, elles sont le plus souvent encodées sous forme compressée afin de réduire les coûts de transmission et de stockage. Une première partie de ces travaux a porté sur le tatouage-compression conjoint d'images médicales de manière à pouvoir vérifier l'intégrité et l'authenticité d'une image sans la décompresser, à l'aide d'un message tatoué que l'on peut extraire sans décoder, même partiellement, le flux binaire. Dans la continuité de ces travaux, nous nous sommes intéressés à vérifier la fiabilité des images compressées et chiffrées tout en maintenant leur confidentialité. Son principe général est fondé sur l'insertion de deux messages contenant des attributs de sécurité dans l'image durant la compression et le chiffrement de cette dernière. Chacun d'eux n'est accessible que dans un domaine : le domaine compressé ou le domaine chiffré, sans avoir à décompresser ou déchiffrer l'image, même partiellement. Ces schémas ont par ailleurs été développés pour être compatible avec le standard DICOM. Une deuxième partie de ces travaux de recherche a focalisé sur le tatouage réversible d'images médicales. Cette propriété garantit la reconstruction de l'image originale après avoir retiré la marque insérée. Nous avons développé un schéma de tatouage original permettant d'insérer une marque de manière réversible dans une image chiffrée, sur la base d'une nouvelle modulation de tatouage par décalage d'histogramme robuste. Ce message est accessible dans les deux domaines chiffré et spatial. Enfin, dans le but de valider les solutions proposées vis-à-vis de la distorsion introduite par le tatouage, nous avons mis en place un protocole d'évaluation « psychovisuelle » du tatouage.

Title : Protection of encrypted and/or compressed medical images by means of watermarking

Keywords : Medical imaging, Security, Watermarking, Encryption, JPEG-LS/JPEG Compression.

Abstract : The rapid growth of information and communication technologies have offered new possibilities to store, access and transfer medical images over networks. In this context, data leaks, robbery as well as data manipulations represent a real danger needing thus new protection solutions, in terms of data confidentiality, reliability control and data traceability. The work conducted during this Ph.D. thesis aims at the combination, or even the fusion of watermarking and data encryption for medical image security. Nevertheless, the deployment of these protection mechanisms in the healthcare domain should take into account its specificities. Particularly, as medical images constitute large volumes of data, they are usually encoded in lossy or lossless compressed form so as to minimize transmission and storage costs. The first part of this work focused on joint watermarking-compression of medical images so as to be able to verify image reliability from the compressed bitstream without decompressing it, even partially. In the continuity of this work, we were interested in verifying the reliability of compressed and encrypted images while maintaining their confidentiality. Its main principle is based on the insertion of two messages, containing security attributes, in the image during the image compression and encryption. Each of them is only accessible in one domain: the compressed domain or the encrypted domain, without having to decompress or decrypt the image, even partially. Note that these schemes have been developed to be compliant to the DICOM standard. A second part of these research activities focused on the reversible watermarking of encrypted medical images. The reversibility property guarantees the recovery of the original image after removing the inserted watermark. We have developed an original watermarking scheme allowing a reversible insertion of a message in an encrypted image, based on a new robust histogram shifting reversible watermarking modulation. This message is accessible in both encrypted and clear domains. Finally, in order to validate the distortion introduced by watermark embedding in the proposed solutions, we have implemented a "psychovisual" assessment protocol for medical image watermarking.