33rd Annual International Conference of the IEEE EMBS
Boston, Massachusetts USA, August 30 - September 3, 2011

A Joint Watermarking/Encryption Algorithm for Verifying Medical Image Integrity and authenticity in Both Encrypted and Spatial Domains

D. Bouslimi, G. Coatrieux, *Member, IEEE,* and Ch. Roux, *Fellow*, *IEEE*

Abstract— **In this paper, we propose a new joint watermarking/encryption algorithm for the purpose of verifying the reliability of medical images in both encrypted and spatial domains. It combines a substitutive watermarking algorithm, the quantization index modulation (QIM), with a block cipher algorithm, the Advanced Encryption Standard (AES), in CBC mode of operation. The proposed solution gives access to the outcomes of the image integrity and of its origins even though the image is stored encrypted. Experimental results achieved on 8 bits encoded Ultrasound images illustrate the overall performances of the proposed scheme. By making use of the AES block cipher in CBC mode, the proposed solution is compliant with or transparent to the DICOM standard.**

I. Introduction

THE rapid evolution of multimedia and communication technologies offers new means of sharing and remote access to patient data. Medical images naturally play important roles in applications like telesurgery, telediagnosis and so on. But at the same time, this ease of transmission and sharing increases security issues in terms of [1]:

- Confidentiality, which means that only authorized users can access to medical data.
- Availability, which guarantees access to medical information in the normal scheduled conditions of access and exercise.
- Reliability, which is based on: i) integrity - a proof that the information hasn't been altered or modified by unauthorized persons; ii) Authentication - a proof of the information origin and of its attachment to the correct patient. Reliable information can be used confidently by the physician.

In an information system, data confidentiality, integrity and non-repudiation services are usually achieved by cryptographic means. However, once decrypted or its digital signature deleted or lost, the information is no longer protected and it becomes hard to verify its integrity and its origins. From this point of view, these cryptographic means, especially encryption, rather appear as a priori protection mechanisms.

Watermarking has been proposed as a complementary mechanism to improve the security of medical images [2]. When it is applied to images, watermarking modifies or modulates the image pixels' gray level values in an imperceptible way, in order to encode or insert a message within this image. It allows us to intimately associate protection data with the information to be protected. For instance, it can be used to control the integrity and the authentication of an image by inserting a digital signature of this one. As defined, watermarking is an a posteriori control mechanism because the image content is still available and interpretable, while staying protected.

Different approaches have been proposed in order to benefit from the complementarity of these two mechanisms in terms of a priori/ a posteriori protection, essentially in the context of copyright protection. Technically, two categories of methods can be distinguished according to the way watermarking and encryption are merged:

- Joint Decryption/ Watermarking (JDW), where watermark embedding is conducted during the decryption process [3-6].
- Joint Encryption/ Watermarking (JEW), where watermarking and encryption processes are merged. In this case, the watermark can be extracted in the spatial domain, i.e. after the decryption process, or in the encrypted domain, or in both domains [7].

The method we propose in this paper belongs to the second category. It merges a block cipher algorithm, AES (Advanced Encryption Standard), and a substitutive watermarking algorithm, the Quantization Index Modulation (QIM) [8]. The objective of this operation is to give access to watermarked information (i.e., security attributes) in the encrypted and spatial domains for the purpose of verifying the reliability of images. We decided to consider the AES because its use is recommended by DICOM[1], the standard of reference in medical imaging.

The rest of this paper is organized as follows. We detail the proposed method in section II, and evaluate its performances in section III. Conclusions are given in section IV.

II. Joint Watermarking/Encryption Algorithm

A. System architecture and principle

As illustrated in figure 1, the proposed system relies on two main procedures: protection and verification. The first one (fig. 1a) conducts the watermarking and the encryption of an image *I,* jointly. It allows us to insert two messages, m_s and m_e, which will be available in the spatial and encrypted domains, respectively. These two messages contain security attributes that will assess the image reliability in each domain. Indeed, each message contains an authenticity codes

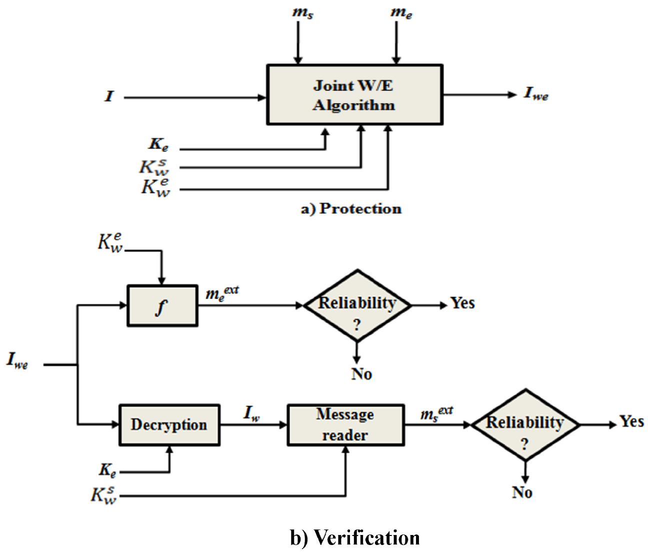
D. Bouslimi, G. Coatrieux and Ch. Roux are with the Institut Telecom; Telecom Bretagne; Unite INSERM 650 Latim, Technopole Brest-Iroise, CS 83818, 29238 Brest Cedex 3 France (e-mail: {dalel.bouslimi, gouenou.coatrieux, christian.roux}@telecom-bretagne.eu).

[1] Digital Imaging and Communications in Medicine - http://medical.nema.org/

978-1-4244-4122-8/11/$26.00 ©2011 IEEE

which identifies the image origin and an integrity proof. Integrity can be carried out by one digital signature calculated by making use of a secure hash function or a pseudo-random sequence. In both cases, integrity verification resides in detecting the presence of this digital signature or pseudo random sequence. As it can be seen in figure 1b, protection data is made available from the encrypted image or from the decrypted image for a subsequent verification stage. Thus, if watermarking and encryption are jointly conducted, watermark extraction and image decryption are two independent processes.

In the following sections, we present the watermarking modulation before describing how it is merged with the encryption process.



a) Protection

b) Verification

Fig. 1. General scheme of our system. I, I_{we}, I_w, K_e, K_w^s and K_w^e denote the original image, the watermarked encrypted image, the watermarked decrypted image, the encryption key and the watermarking keys, respectively. m_e and m_e^{ext} are the embedded and extracted messages in the encrypted domain, respectively. m_s and m_s^{ext} denote the embedded and extracted messages in the spatial domain, respectively. f is the extraction function of m_e in the encrypted domain.

B. QIM

Quantization Index Modulation (QIM), proposed by Chen and Wornell [8], relies on quantifying the components of one image according to a set of quantizers based on codebooks in order to insert a message. More clearly, to each message m_{si} issued from a finite set of possible messages $Ms=\{m_{si}\}_{i=1,\dots,p}$, the QIM associates elements of a codebook $C_{m_{si}}$ such as:

$$C_{m_{si}} \cap C_{m_{sj}} = \emptyset, i \neq j \quad (1)$$

Substituting one component of the image by its nearest element in the codebook $C_{m_{si}}$ thus allows the insertion of m_{si}. Let us consider one image component 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), each cell is associated to a codebook $C_{m_{si}}$, $i=1,\dots,p$. As a consequence, one message m_{si} has several representations in $\mathbb{N}^N$. 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. During the extraction step, the knowledge of the cell to which X_w belongs is enough to identify the embedded message.

C. Combination of encryption and watermarking

Our objective is to give access to two messages: m_{si}, the message available in the spatial domain and, m_{ej}, the message available in the encrypted domain. Similarly to m_{si}, m_{ej} is a message issued from a finite set of possible messages $Me = \{m_{ej}\}_{j=1,\dots,q}$.

In order to conduct jointly this double watermarking process and to avoid any interference between them, we propose to adapt the QIM described above according to the following principles.

Each codebook $C_{m_{si}}$ is decomposed into sub-codebooks $C_{m_{si}m_{ej}}$ such as

$$C_{m_{si}} = \bigcup_{j=1}^{q} C_{m_{si}m_{ej}}$$

$$C_{m_{si}m_{ej}} \cap C_{m_{si}m_{ek}} = \emptyset, j \neq k \ (2)$$

Thus m_{si} and m_{ej} are embedded simultaneously within a vector image $X \in \mathbb{N}^N$ by replacing X with X_w which corresponds to the nearest element of X in $C_{m_{si}m_{ej}}$. The watermarked X (X_w) is then given by:

$$X_w = \min_{j}(\|X - Y_{ij}\|),\ Y_{ij} \in C_{m_{si}m_{ej}}, j = 1, \dots, q$$

Making the message m_{ej} available in the encrypted domain depends on the sub-codebook construction a process intimately linked with the encryption algorithm. Considering an encryption algorithm E and its encryption key K_e, sub-codebooks $C_{m_{si}m_{ej}}$ are built so as to verify:

$$C_{m_{si}m_{ej}} = \{Y \in C_{m_{si}} / f(E(Y, K_e), K_w^e) = m_{ej}\} \ (3)$$

where f is the extraction function of m_{ej} in the encrypted domain. To sum up this process, m_e is made available in the encrypted domain by modulating pixel values in the spatial domain.

D. Implementation with AES

In this study, for sake of simplicity, we work with binary messages, $m_{si}=\{0,1\}$ and $m_{ej}=\{0,1\}$ and consequently with two codebooks C_0 ($m_{s1}=0$) and C_1 ($m_{s2}=1$), and four sub-codebooks $C_{m_{si}m_{ej}}$: C_{00} and C_{01} and C_{10} and C_{11} derived from C_0 and C_1 respectively.

In this work, we use the well known block cipher algorithm AES [9] in CBC mode of operation in order to be compliant with the DICOM standard. A simplified view of this algorithm is given in figure 2. The concept of mode of operation refers to the manner in which plaintext blocks (sequence of bytes) are treated at the encryption stage (resp. decryption stage). When the CBC mode is applied, the plaintext block is combined, through an XOR operation, with the previous ciphertext block before being encrypted with the AES.

In order to insert m_{ej} within the ciphertext, we distort the image pixels so that the value of f is equal to m_{ej}. The insertion of both m_{si} and m_{ej} within a byte block B involves replacing B with an element Y in $C_{m_{s_i}}$ that satisfies:

$$f(Y_e) = m_{ej} \quad \text{where} \quad Y_e = AES(Y, K_e) \ (4)$$

and to build the subsequent sub-codebooks $C_{m_{si}0}$ and $C_{m_{si}1}$. However, because of the AES in CBC mode, sub-codebooks have to be determined at each block. To reduce the complexity of our algorithm, it is preferable browse the elements of $C_{m_{si}}$ until to find the one that satisfies (4).

Here, the function f used to extract the message m_{ej} from an encrypted byte block B^e, is defined as follows:

$$f(B^e, K_w^e) = S_k, \text{ where } \quad S = SHA(B^e) \ (5)$$

where S_k corresponds to the k^{th} bit of S an cryptographic hash issued by the Secure Hash Algorithm. The choice of the rank k depends on the secret watermarking key K_w^e.

The SHA-1 is herein used in order to support the integrity checking of the image in the encrypted domain. Indeed, based on the fact that the "strength" of the SHA-1 is 80bits. This means that if one bit of Y_e changes, then there is half a chance that the value of f , i.e. S_k, commutes. Thus integrity verification in the encrypted domain will be based on computing the SHA of the encrypted block and on verifying the value of S_k that has been inserted. SHA depends on all encrypted block bits; any bit changes will give a different digital signature. If an unauthorized person modifies the encrypted image, there is a high probability to detect it is altered.

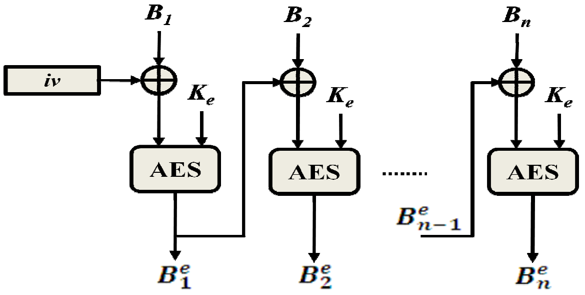


Fig. 2. AES Encryption in CBC mode. B_i and B_i^e denote the plaintext block and the encrypted block, respectively. iv is a random initialization vector.

In the implementation used in the sequel, $C_{m_{si}}$ is built as follows:

$$C_{m_{si}} = \{Y \in \mathbb{N}^N / \left\lfloor \frac{Y_k}{\Delta} \right\rfloor \ mod\ 2 = m_{si}\} \ (6)$$

where Y_k is the k^{th} byte or more precisely the k^{th} pixel of the block B to be encrypted. The choice of the rank k depends on the watermarking key K_w^s. $C_{m_{si}}$ is thus used to insert the message in the spatial domain.

The message available in the encrypted domain m_e is generated as follows:

$$m_e = \sigma(PGNR(K_w^e)||AC, K_w^e)$$

where || is the concatenation operator; $\sigma(.)$ denote a secret random permutation operator which depends on K_w^e and; *PGNR* (Pseudo-Random Number Generator) generates randomly a bit sequence, and; AC is the image authenticity code. The use of σ increases the message security. Even if an unauthorized person read the inserted message, he cannot identify surely the authenticity code bits and modify them without modifying the integrity proof.

The watermarking/encryption algorithm for a grayscale image I of 8 bit depth we propose can be defined in two steps:

1) I is splitted into non-overlapping blocks, $\{B_i\}_{i=1..U}$, of 16 pixels. For each block, n pixels are selected. This step depends on the secret watermarking key K_w^s. The six Most significant bits of each selected pixels and all bits of *non-selected* pixels are concatenated to form a bit sequence V. m_s contains the digital signature of this sequence and the image authenticity code AC

$$m_s = \sigma(SHA(V)||AC, K_w^s)$$

2) The QIM is applied to B_i using the sub-codebooks $C_{m_{si}m_{ei}}$ to insert one bit m_{si} of m_s, $m_s = \{0,1\}^U$ (U is the length of the message), and one bit of m_e, $m_e = \{0,1\}^U$. At the same time, the watermarked version of B_i (i.e. B_i^w) is encrypted through the AES. In this implementation, the difference between B_i and B_i^w stands in the two Least Significant Bit of selected pixels, difference on which QIM sub codebooks are built (see eq. 3).

As stated before, extraction can be conducted independently in both the encrypted and spatial domain. In the encrypted domain, the encrypted image (I_{we}) is decomposed in blocks of 16 pixels. Then, the function f is applied to each block to determine the message m_e. In the spatial domain, the message m_s is similarly extracted based on principles of the QIM. Each message is herein used to verify the image reliability in one domain. One main advantage of the proposed approach is that it is transparent to the DICOM standard. More clearly, if a system is not watermarking interoperable it can decrypt and access the image if it knows the AES encryption key.

III. Experimental results and discussion

To evaluate the performances of our approach, two indicators are considered:

- the peak signal to noise ratio (PSNR) computed between the original image I and its watermarked and deciphered version I_{wd} to measure the image distortion

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

$$MSE(I, I_{wd}) = \frac{1}{ML}\sum_{k=1}^{MxL}(I(k) - I_{wd}(k))^2$$

where M×L corresponds to the number of pixels of I.

- The entropy H of an image to measure the degree of security provided.

$$H(I) = -\sum_{i=0}^{G-1} Pr(\alpha_i)\, log_2(Pr(\alpha_i))$$

where $Pr(\alpha_i)$ is the occurrence probability of the gray level α_i in the image I and G is the total number of gray levels. This measure is applied to the original image and the jointly watermarked/ encrypted image I_{we}.

The proposed joint encryption/watermarking algorithm has been tested over 100 ultrasound images of 576×688 pixels.

Considering $\Delta=2$ in eq. 6, the capacity rate in each domain is 1/16bpp. As a consequence, the global capacity achieved in each domain is of 24768 bits, capacity sufficient enough for the insertion of one integrity proof and one authenticity code.

Obtained PSNR values are greater than 60dB, which is large enough to establish a good diagnosis. Indeed, it has been shown in [10] that some loss of information due to lossy JPEG compression can be tolerated without affecting image diagnostic quality for ultrasound images if image distortion is maintained in the range of 40 and 50dB. In order to respect such a recommendation, we have considered the highest PSNR value our system can provide. Anyway, a more complete study has to be conducted in order to evaluate how such a watermark may interfere with image interpretation.

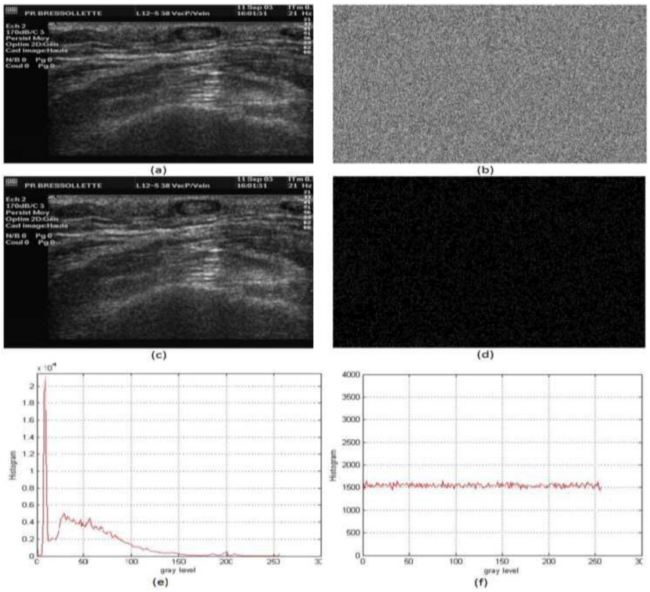


Fig. 3. ultrasound image samples used for experimentations. a) Original image; b) joint watermarked/ciphered image; c) deciphered watermarked image; d) difference between original and decrypted watermarked images; e) Histogram of the original image (a) with an entropy of 6.76 bits/pixel f) Histogram of the protected image (b) with an entropy of 8 bits/pixel.

Histograms of original images (see fig.3e) are significantly different from those of jointly watermarked/encrypted images (see fig.3f). Furthermore, as it can be seen in figure 3f, occurrence probability of each gray value is fairly equitably distributed leading to a very high entropy encrypted image (about 8 bits/pixel for all tested images), also entropy measurements on our jointly protected images are nearly equal to those obtained with AES only. Thus, the insertion of the messages does not impact general performances of the encryption algorithm.

If now we look at the complexity of our algorithm, experiments show that the joint watermarking /encryption algorithm needs, on average, twice the time necessary for encrypting an image with the AES only. Whence, this algorithm may not be suitable for real time transmission of image. Notice also that embedded watermarks are fragile. Messages will be lost after any image modifications. However, this is not an issue due to the fact that we focus on verifying data reliability. If data integrity is not valid, thus data reliability is lost.

IV. Conclusion

In this paper, we have proposed a new joint watermarking/encryption algorithm, which guarantees the a priori and a posteriori protection. It merges the QIM and a block cipher algorithm the AES in CBC mode. This makes it compliant with the DICOM standard and also gives access to two distinct messages in the spatial domain and in the encrypted domain, respectively. These two messages are used for verifying the image reliability even though it is encrypted. Experimental results show that the image distortion is very low. Future works will focus on: i) making our scheme more robust to attacks like lossy image compression (ex. JPEG) and; ii) reducing the complexity of our algorithm so as being able to operate in real-time.

Acknowledgment

The work presented in this paper is supported by the ANR through the funding of the ANR SELKIS project.

References

[1] G. Coatrieux, H. Maître, B. Sankur, Y. Rolland, R. Collorec, "Relevance of watermarking in medical imaging," *in proc. of Int. Conf. on IEEE EMBS* ITAB, USA, pp. 250-255, 2000.

[2] G. Coatrieux, C. Le Guillou, J.-M. Cauvin, C. Roux, "Reversible watermarking for knowledge digest embedding and reliability control in medical images," *IEEE Trans. Inf. Technol. Biomed.*, 2009 Mar., 13(2):158-165.

[3] R. Anderson, C. Manifavas, "Chameleon- A New Kind of Stream Cipher," *FSE '97*, vol. 1267, pp. 107–113, 1997.

[4] A. Adelsbach, U. Huber, A.S. Sadeghi, "Fingercasting–Joint Fingerprinting and Decryption of Broadcast Messags," *ACISP '06*, LNCS, vol. 4058, pp. 136–147, 2006.

[5] M. Celik, A.N. Lemma, S. Katzenbeisser, M. van der Veen, "Secure Embedding of Spread Spectrum Watermarks Using Look-up-Tables," *ICASSP '07*, vol. 2, pp. 153-156, 2007.

[6] L. Shiguo, L. Zhongxuan, R. Zhen, W. Haila, "Joint Fingerprint Embedding and Decryption for Video Distribution," *IEEE Int. Conf. on Multimedia and Expo,* pp.1523-1526, 2007.

[7] L. Shiguo, L. Zhongxuan, R. Zhen, W. Haila, "Commutative encryption and watermarking in video compression," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 17, n°6, pp.774-778, 2007.

[8] B. Chen, G.W. Wornell, "Quantization Index Modulation: A Class of Provably Good Methods for Digital watermarking and information embedding," *IEEE Trans. on Information Theory*, vol. 47, n°. 4, pp. 1423- 1443, 2001.

[9] http://buchholz.hs-bremen.de/aes/aes.htm

[10] K. Chen, T.V. Ramabadran, "Near-Lossless Compression of Medical Images Through Entropy Coded DPCM," *IEEE Trans. on Medical Imaging*, vol. 13, n°. 3, pp. 538-548, 1994.