Data Hiding in Homomorphically Encrypted Medical Images for Verifying Their Reliability in both Encrypted and Spatial Domains

D. Bouslimi, R. Bellafqira, G. Coatrieux, Senior Member, IEEE

Abstract— In this paper, we propose a new scheme of data hiding of encrypted images for the purpose of verifying the reliability of an image into both encrypted and spatial domains. This scheme couples the Quantization Index Modulation (QIM) and the Paillier cryptosystem. It relies on the insertion into the image, before its encryption, of a predefined watermark, a "pre-watermark". Message insertion (resp. extraction) is conducted into (resp. from) the encrypted image using a modified version of QIM. It is the impact of this insertion process onto the "pre-watermark" that gives access to the message in the spatial domain, i.e. after the image has been decrypted. With our scheme, encryption/decryption processes are completely independent from message embedding/extraction. One does not need to know the encryption/decryption key for hiding a message into the encrypted image. Experiments conducted on ultrasound medical images show that the image distortion is very low while offering a high capacity that can support different watermarking based security objectives.

I. INTRODUCTION

Cloud computing services become important solutions for the storage and continuous availability of data supplied by multiple sources. Due to the outsourcing of data and services, they are exposed to many threats that strongly security requirements in terms confidentiality, availability and reliability (i.e. integrity and authentication). Among available security mechanisms, encryption is commonly used so as to ensure medical data confidentiality. However, once decrypted, one piece of information is no longer protected and it becomes hard to verify its integrity and its origin. From this point of view, encryption appears as an "a priori" protection. Watermarking has been proposed as a complementary mechanism that can improve security of medical images. 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 some security attributes (i.e. the watermark) into it. As defined, such a protected image can be accessed while remaining protected by these hidden security attributes that can be used for example for verifying the image reliability (i.e., its integrity, its origins and its attachment to one patient). Thus, combining watermarking with encryption may allow us ensuring an a priori/ a posteriori protection at the same time. In practice, watermarking is usually conducted before encryption or during the encryption/decryption processes.

D. Bouslimi, R. Bellafqira, G. Coatrieux are with the Institut Mines - Telecom; Telecom Bretagne; Unite INSERM 1101 Latim, Technopole Brest-Iroise, CS 83818, 29238 Brest Cedex 3 France (e-mail: {dalel.bouslimi,reda.bellafqira,gouenou.coatrieux}@telecom-bretagne.eu).

However, in order to watermark outsoured data without endangering privacy and data confidentiality, different approaches have been proposed so as to embed a message directly into the encrypted image, essentially in the framework of copyright protection. Three categories of approaches can be distinguished according to the availability of the embedded message into the spatial domain (i.e. after decryption process) and/or in the encrypted domain:

- Message available in the spatial domain (MSD)- The scheme proposed in [2] exploits homomorphic encryption, which allows modifying an encrypted image for the embedding of a watermark.
- Message available in the encrypted domain (MED)- As example, in [3], the image is firstly divided into patches. Before encryption, some patches are replaced by patches computed from their sparse coefficients while the residual errors in-between patches are reversibly embedded into the rest of patches; leaving thus some free space by next is used for message embedding in the encrypted domain
- Message available in both encrypted and spatial domains (MSED)- most of these methods are based on partial encryption [4] or invariant encryption [5]. With those methods, only some parts of the host image are encrypted while the rest of it is watermarked. Recently, a novel concept, called VRBE (Vacating Room Before Encryption) has been proposed in [6]. Its principle is to reversibly watermark an image before encrypting it so as to leave some free space into the encrypted domain for message embedding. However, to make possible the retrieval of this free space into the encrypted domain, the image has to be reorganized before encryption. Moreover, the decryption process is modified so as to make possible message extraction in the spatial domain. As example, in [6] watermarkable positions in the encrypted image are placed at the beginning of the bit stream and, at the reception, watermarked positions are not decrypted.

In this paper, we propose a new scheme of MSED type, which principles allow the insertion of some security attributes into an encrypted image; attributes by next available in both encrypted and spatial domains for verifying the image reliability in both domains. Compared to other MSED methods, our approach entirely encrypts the image; and watermarking and encryption/decryption are independent. Indeed, message insertion and extraction processes (resp. encryption and decryption) do not require the knowledge of the encryption key (resp. watermarking key or other extra parameters).

The remainder of this paper is organized as follows. In Section II, we present QIM modulation and Paillier cryptosystem we used to implement our scheme. We describe the proposed system in section III and evaluate its

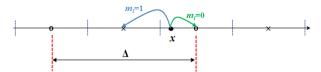


Figure 1. Insertion with QIM mono-dimensional of a binary message into a pixel value x. Symbols o and × denote cells' centers that encode 0 and 1, respectively.

performance through some experiments on ultrasound images in section IV.

II. CRYPTOGRAPHIC AND WATERMARKING PRIMITIVES

A. Watermarking primitive: QIM

Quantization Index Modulation [7] relies on quantifying some image components according to a set of quantizers based on codebooks in order to insert a message. More clearly, to one message m_i issued from a finite set of possible messages $M = \{m_i\}_{i=0,\dots,q}$, QIM associates the elements of a codebook C_{m_i} such as

$$C_{m_i} \cap C_{m_j} = \emptyset, i \neq j. \tag{1}$$

By substituting one component of the image by its nearest element in the codebook C_{m_i} , one allows the insertion of m_i . Let us consider one image component such as a pixel value $x \in \mathbb{N}$ and a binary message m_i , i.e. $m_i \in \{0,1\}$. In this case, two codebooks C_0 and C_1 are defined. They can be built up by uniformly quantizing the dynamic of x with a quantization step Δ , a QIM parameter, as illustrated in Fig. 1. In this example, intervals centered on crosses represent C_1 $(m_i = 1)$ whereas intervals centered on circles represent C_0 $(m_i = 0)$. Thus, x will be moved to the nearest cross or circle in order to encode m_i . Herein, the codebook C_{m_i} can be defined as

$$C_{m_i} = \left\{ c_{m_{i,k}} \right\} = \left\{ (k + m_i/2) \Delta, k \in \mathbb{N} \right\}$$
 and the watermarked version of x , i.e. x_w , is given by

 $x_w = Q_{\Delta}(x, m_i)$ where $Q_{\Delta}(x, m_i) = min_k (|x - c_{m_{i,k}}|)$ (3) Let us consider \hat{x} a possible attacked version of x_w . During the extraction step, the knowledge of the interval (or the codebook) to which belongs \hat{x} is enough to identify the embedded message and, if \widehat{m}_i is the extracted message from \hat{x} then

$$\widehat{m_l} = Q_{\Delta}^{ext}(\widehat{x}) \tag{4}$$

where Q_{Δ}^{ext} is the QIM extraction function defined as:

$$Q^{ext}_{\Delta}(\hat{x}) = arg \; \mathrm{min}_{m_i \in \{0,1\}} \, \mathrm{min}_{c_{m_i,k} \in C_{m_i}} \big| c_{m_i,k} - \hat{x} \big|$$

B. Homomorphic encryption algorithm: Paillier cryptosystem

In this work, we opted for the asymmetric Paillier cryptosystem because of its additive homomorphic property [8]. It stands on a public-private key pair (K_p, K_s) , such as:

$$K_p = pq$$
 and $K_s = (p-1)(q-1)$ (5) where p and q are two large prime integers. The Paillier encryption of a plaintext m into the ciphertext c using the public key K_p is given by:

$$c = E[m, r] = (1 + K_p)^m r^{K_p} \mod K_p^2$$
 (6)
where $r \in \mathbb{Z}_N^*$ is a random integer associated to m making the Paillier cryptosystem probabilistic or semantically

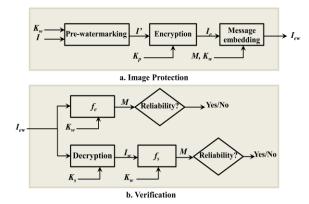


Figure 2. General architecture of our system. K_w and (K_p, K_s) are the watermarking key and the emetter or receiver public-private key pair. secure. Based on r, the encryption of the same plaintext message yields to different ciphertexts even though the encryption key is the same. The decryption of c using the private key K_s is as follows

$$m = (c^{K_S} - 1)/K_p \bmod K_p \tag{7}$$

If we consider two plaintexts m_1 and m_2 , the additive homomorphic property of Paillier cryptosystem ensures that

$$E[m_1, r_1] \times E[m_2, r_2] = E[m_1 + m_2, r_1 + r_2]$$

$$E[m_1, r_1]^{m_2} = E[m_1 m_2, r_1^{m_2}]$$
(8)

$$E[m_1, r_1]^{m_2} = E[m_1 m_2, r_1^{m_2}]$$
 (9)

III. DATA HIDING IN HOMOMORPHIC DOMAIN

A. System architecture and principle

The architecture of our system is depicted in figure 2. It relies on two main procedures: protection and verification. The first one (Fig. 2a) allows the embedding within an encrypted image of a message M that will be available in both encrypted and spatial domains (i.e. in the encrypted and decrypted images). Herein, M is a sequence of bits that encodes some security attributes assessing the image integrity and authenticity in both domains. In order to make possible such embedding, and especially the extraction of M after the image has been decrypted, our solution relies on a prewatermarking procedure. This one consists in the QIM embedding into the image I of a predefined watermark Wbefore the encryption process. Once I is pre-watermarked into I', I' is encrypted into I_e . M is then embedded into I_e leading to the watermarked encrypted image I_{ew} . It is the impact of the distortion induced by this watermarking operation in the encrypted domain onto W that will give access to M in the spatial domain. More clearly, due to the fact the embedding of M modifies the encrypted image I_e into I_{ew} , the resulting decrypted image I_w will then differ from I'. It is the differences between the original watermark W and the extracted one that allows the encoding of M into I_w . As illustrated in Fig. 2.b., the access to M in both encrypted (i.e. from I_{ew}) and spatial (i.e. from I_w) domains, relies on two extraction functions f_e and f_s , respectively. These functions are defined according to the watermarking modulations used in the encrypted and spatial domains. We will come back on this point in the next section. It is important to notice that, in our approach, message insertion/extraction and image encryption/decryption are considered as independent processes. Moreover, all of pre-watermark embedding, encryption and data hiding processes do not have to be conducted at the same time. In the following, we detail the different steps of our system and how they interact considering image protection and message extraction separately in application to 8 bit grayscale images.

B. Image Protection

This procedure is constituted of three main steps (see Fig. 2.a): image pre-watermarking, pre-watermarked image encryption, message embedding in the encrypted domain.

1) Pre-watermarking step

Based on the watermarking key K_w , a watermark W is first generated. From here on, W corresponds to a random sequence of p bits: $W = \{w_1, ..., w_i, ..., w_p\}, w_i \in \{0,1\}$. W is then embedded into the image I as follows:

i. using K_w , I is secretly splitted into N non-overlapping and distributed subsets of p pixels $\{I_i\}_{i=1,...N}$ with

$$I_i = \{I_i^1, \dots, I_i^j, \dots, I_i^p\}, p \ge 2 \text{ and } I_i^j \in \mathbb{N}$$
 where I_i^j corresponds to the j^{th} pixel of I_i .

ii. W is then embedded into each subset I_i by embedding one bit w_i within I_i^J using QIM and a quantization step Δ (see Eq. 3)

$$I_i^{j'} = Q_{\Delta}(I_i^j, w_j), \forall j \in \{1, ..., p\}$$
 (11)
The resulting pre-watermarked image I' is thus given by

$$I' = \{I'_i\}_{i=1,..,N} \text{ where } I'_i = \{I^{1'}_i, ... I^{j'}_i, ..., I^{p'}_i\}$$
 (12)

2) Encryption step

During this step, the previous pre-watermarked image I' is encrypted into I_e using the Paillier cryptosystem and the public key K_p : $I_e = \{I_{ie}\}_{i=1,..,N}$. Encrypted subsets $\{I_{ie}\}_{i=1,..,N}$ are obtained from the pre-watermarked pixel subsets $\{I_i'\}_{i=1,\dots,N}$ such as (see Eq 6)

$$I_{ie} = \{I_{ie}^1, \dots I_{ie}^j \dots, I_{ie}^p\} \text{ with } I_{ie}^j = E[I_i^{j\prime}, r_i^j]$$
 (13) where r_i^j is random and associated to $I_i^{j\prime}$ (see section II.B) 3) Message embedding step

This process enables an authorized user who knows the watermarking key K_w to embed into the encrypted image I_e a message M, which is a sequence of bits $\{b_i\}_{i=1,...N}$ uniformly distributed, i.e. $Pr(b_i=0)=Pr(b_i=1)=0.5$. Based on the knowledge of K_w , subsets of encrypted pixels $\{I_{ie}\}_{i=1,...,N}$ are retrieved. One bit of M is embedded per subset I_{ie} and will be available in both encrypted and spatial domains such as

$$b_i = f_e(I_{iew}) = f_s(I_{iw})$$
 (14)
where I_{iew} is the watermarked version of the subset I_{ie} ; I_{iw} is

the decrypted version of I_{iew} .

In this paper, one bit b_i of M is embedded in all the pixel I_{ie}^J of the encrypted subset I_{ie} using a modified version of QIM. More clearly, b_i is repeated p times into the subset I_{ie} . At the same time, and in order to make b_i available in the spatial domain, this process should guarantee that the prewatermark is or not modified. The rule we imposed to make this possible is the following one: if the pre-watermark W in the decrypted subset I_i^j is unchanged then a '0' was embedding, on the contrary '1' has been inserted. As a consequence, the embedding process in the encrypted domain should make sure that the pre-watermark bit w_i into I_i^j commutes if b_i ='1' or remains unchanged if b_i ='0'.

As illustrated in Fig. 1, to modify (resp. to not) the embedded bit w_j , the necessary distortion d_{wj} induced by the insertion process of b_i into I_{ie}^j must verify: $\frac{\Delta}{4} < \left| d_{wj} \right| < \frac{3\Delta}{4}$ $(resp. |d_{wj}| < \frac{\Delta}{4})$. To guarantee such degree of distortion, we exploit the additive homomorphic property of Paillier cryptosystem. It allows modulating I_{ie}^{j} for both the embedding of b_i and the introduction of a distortion d_{wj} in the spatial domain. Basically, the watermarked version of I_{ie}^{J} , I_{iew}^{j} is obtained by multiplying I_{ie}^{j} with an encrypted version of d_{wi} such as:

$$I_{iew}^{j} = I_{ie}^{j} \times E[d_{wj}, r_{k}] = E\left[I_{i}^{j'}, r_{i}^{j'}\right] \times E[d_{wj}, r_{k}]$$

$$= E\left[I_{i}^{j'} + d_{wj}, r_{i}^{j} + r_{k}\right] \qquad (15)$$
where r_{k} is a random integer that satisfies
$$Q_{\Delta}^{ext}(I_{ie}^{j} \times E[d_{wj}, r_{k}]) = b_{i} \qquad (16)$$
and d_{wj} verifies

$$Q_{\Delta}^{ext} \left(I_{ie}^{J} \times E \left[d_{wj}, r_{k} \right] \right) = b_{i} \tag{16}$$

$$\begin{cases} \frac{\Delta}{4} < |d_{wj}| < \frac{3\Delta}{4} & \text{if } b_i = 1\\ |d_{wj}| < \frac{\Delta}{4} & \text{if } b_i = 0 \end{cases}$$

$$(17)$$

As it can be seen, by choosing an appropriate random value r_k one can insert b_i into an encrypted pixel with QIM (eq. 16); and at the same time, we induce the desired distortion into the spatial domain (eq. 17). Notice that in our implementation, we work with $d_{wj} = \frac{b_i \Delta}{2}$. By doing so and considering D the decryption function of the Paillier cryptosystem, the decryption version of I_{iew}^{j} , I_{iw}^{j} , is equal to

$$I_{iw}^{j} = D\left[I_{iew}^{j}\right] = D\left[E\left[I_{i}^{j'} + \frac{b_{i}\Delta}{2}, r_{i}^{j} + r_{k}\right]\right] = I_{i}^{j'} + \frac{b_{i}\Delta}{2}$$
 (18) As a consequence, the insertion of $b_{i} = '0'$ into I_{ie} doesn't modify I_{i}' (i.e. $I_{iw}^{j} = I_{i}^{j'}$). On the contrary, embedding $b_{i} = '1'$ into I_{ie} implicitly modifies the embedded value w_{i} due to the fact that $I_{iw}^{j} = I_{i}^{j'} + \frac{\Delta}{2}$.

To sum-up, the extraction functions f_e we use to extract b_i from one encrypted watermarked pixel subset I_{iew} is such as

one encrypted watermarked pixel subset
$$I_{iew}$$
 is such as
$$b_i = f_e(I_{iew}) = \begin{cases} 1 & if \sum_{j=1}^p Q_{\Delta}^{ext}(I_{iew}^j) > \left\lfloor \frac{p}{2} \right\rfloor \\ 0 & else \end{cases}$$
le, in the spatial domain, the extraction function f_s is

While, in the spatial domain, the extraction function
$$f_s$$
 is
$$b_i = f_s(I_{iw}) = \begin{cases} 0 & \text{if } Q_{\Delta}^{ext}(I_{iw}^j) = w_j \ \forall j \in \{1, ..., p\} \\ 1 & \text{otherwise} \end{cases}$$
(20)

C. Message extraction step

As said previously, the message M can be extracted in both encrypted and spatial domains. In the encrypted domain, according to K_w , the encrypted image (I_{ew}) is firstly splitted into subsets of p encrypted pixels $\{I_{iew}\}$. Then, f_e (Eq. 19) is used to extract from each of them one bit of the message M. Once the image is decrypted using K_s , M is simply read with the help of f_s (see eq. 20) applied to the pixel subsets $\{I_{iw}\}\$ of the decrypted image (I_w) .

IV. EXPERIMENTAL RESULTS AND DISCUSSION

Experiments were conducted on 100 8 bit depth ultrasound images of 576×688 pixels. Two indicators are considered to evaluate the performance of our system: capacity and watermark imperceptibility.

Capacity: Because one bit of message is embedded per pixel subset, the capacity one can insert into an image depends on the dimensions of the pixel subsets and of the image. Indeed, the achieved capacity rate is equal to 1/p bpp (bit per pixel). Working with p=1 leads to a capacity of 1bpp or equivalently to a message of about 396 Kbits. This capacity is large enough for the insertion of some security attributes assessing the image reliability. For instance, M may contain an authenticity code (e.g. about 1000 bits by combining the French National Identifier with the Unique Identifier of DICOM the standard for medical images [9]), and an integrity proof which can be a secret pseudorandom binary sequence [10]. The integrity of the encrypted or decrypted image can thus be checked based on verifying the presence of this sequence within the image. Moreover, one can better enhance the robustness of the embedded message in the encrypted domain by working with p>2; allowing repeating the message at least 3 times.

Distortion: As our algorithm introduces in average the same image distortion in each pixel block, we decided to use the Peak Signal to Noise Ratio (PSNR) to measure the distortion between the image I and its watermarked and deciphered version I_w . The lower bound of PSNR can be theoretically determined according to Δ . Let us assume that the pixels of the image (i.e. I) are uniformly distributed over the cells of QIM codebooks (see Fig.1). This means that the probability that one subset pixel I_i^j belongs to the cell that encodes '0' (resp. '1') is 0.5. Since W is a binary sequence uniformly distributed, the probability that the pixel I_i^J belongs to the cell that encodes w_i is 0.5. As can be seen in Fig 1, the maximum distortions one may introduce to embed w_i into I_i^j by moving it to the center of the nearest cell that encodes w_i are $\Delta/4$ and $\Delta/2$ in the cases where I_i^j "naturally" encodes w_i or not, respectively. Knowing that, we can consider that the maximum distortion induced by the insertion process of w_i into I_i^j is thus $d_{ms} = \frac{1}{2} \left(\frac{\Delta}{4} + \frac{\Delta}{2} \right) = \frac{3}{8} \Delta$. After the insertion of M into the encrypted image, only subsets that encode $b_i = '1'$ are modified (see eq.18). Due to fact that the distortion induced by the insertion of $b_i = '1'$ into $I_i^{j'}$ is $\Delta/2$ and that the probability that $b_i = '1'$ is 0.5, the distortion induced by the insertion of b_i into I_{ie}^j is then $d_{me} = \frac{\Delta}{4}$. Therefore, the maximum distortion induced per pixel by our data hiding scheme is $d_m = d_{ms} + d_{me} = \frac{5}{8} \Delta$. Consequently, the PSNR lower-bound can be refined as

$$PSNR_{Pail}(I, I_{wd}) \ge 20Log_{10}\left(\frac{408}{\Delta}\right) \tag{21}$$

We give in Fig.3 the variation of this limit for different values of Δ . In practice, achieved PSNR values are much greater (see Fig.3). They are about 51.15 dB, 44.2dB and 37.8dB for Δ = {2,4,8}, respectively. This can be explained by the fact that, in practice, the pixels of the image source are not uniformly distributed over the cells of QIM codebooks. With our system, an information loss thus occurs. But, this loss remains small in particular when Δ <5. In healthcare, Chen *et al.* reported in [11] that some loss can be tolerated until the PSNR stays in the range of 40 and 50dB.

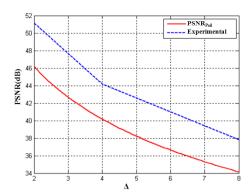


Figure 3. Lower theoretical PSNR bound ($PSNR_{Pail}$) and obtained experimental PSNR values for different values of Δ .

V.CONCLUSION

In this work, we have proposed a new data hiding scheme of encrypted images that allows accessing a message in both the encrypted and spatial domains. This message can be used for verifying the image reliability even though it is encrypted. Its originality stands on the use of a prewatermark which makes the insertion /extraction processes independent of the encryption/decryption processes, and vice versa. We have also provided an implementation of our scheme based on QIM modulation and Paillier cryptosystem. It provides an important capacity rate while minimizing image distortion. Future works will focus on making this implementation more robust to attacks like lossy image compression (ex. JPEG) and on enhancing the quality of the watermarked images.

REFERENCES

- D. Bouslimi, et al., "A telemedicine protocol based on watermarking evidence for identification of liabilities in case of litigation", Healthcom (2012), 506-509.
- [2] N. Memon, P. Wong, "A buyer-seller watermarking protocol", *IEEE Transactions on Image Processing*, (2001) 643–649.
- [3] X. Cao, et al., "High capacity reversible data hiding in encrypted images by patch-level sparse representation", IEEE Transactions on Cybernetics, 2015.
- [4] D. Xiao, S. Chen, "Separable data hiding in encrypted image based on compressive sensing", *Electronics Letters*, 50 (8) (2014) 598–600.
- [5] R. Schmitz, et al., "Towards more robust commutative watermarkingencryption of images", *IEEE International Symposium on Multimedia* (ISM) (2013) 283–286.
- [6] K. Ma, et al, "Reversible data hiding in encrypted images by reserving room before encryption", IEEE Transactions on Information Forensics and Security, 8 (3) (2013) 553–562.
- [7] 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*, 47(4) (2001), 1423-1443
- [8] P. Paillier, "Public-Key Cryptosystems Based on Composite Degree Residuosity classes", *Proc Eurocrypt*, 1592 (1999) 223-238.
- [9] W. Pan, et al., "Medical image integrity control combining digital signature and lossless watermarking", In: Data privacy management and autonomous spontaneous security. Springer Berlin Heidelberg, (2010), 153-162.
- [10] D. Bouslimi, et al., "A joint encryption/watermarking system for verifying the reliability of medical images", IEEE Transactions on Information Technology in Biomedicine 16 (2012) 891–899.
- [11] K. Chen, T.V. Ramabadran, "Near-lossless compression of medical images through entropy coded DPCM", *IEEE Transactions on Medical Imaging*, 1994.