Signal Processing: Image Communication 47 (2016) 263–270

Contents lists available at ScienceDirect

Signal Processing: *Image Communication*

journal homepage: www.elsevier.com/locate/image

Data hiding in encrypted images based on predefined watermark embedding before encryption process

Dalel Bouslimi [a,b,*], Gouenou Coatrieux [a,b], Michel Cozic [c], Christian Roux [a,b]

[a] *Telecom Bretagne, Département Image et Traitement de l'Information, Brest 29238, France*
[b] *INSERM U1101, Laboratoire de Traitement de l'Information Médicale, Brest 29238, France*
[c] *MEDECOM, Plougastel Daoulas 29470, France*

ARTICLE INFO

Article history:
Received 10 February 2015
Received in revised form
23 June 2016
Accepted 23 June 2016
Available online 30 June 2016

Keywords:
Data hiding
Image security
Encryption
Stream cipher

ABSTRACT

In this paper, we propose a novel approach which allows embedding a message into an encrypted image; a message that can be accessed whether the image is encrypted or not. To do so, the proposed solution relies on the insertion into the image of a predefined watermark, a "pre-watermark", before the encryption process. Message insertion (resp. extraction) is then commonly conducted into (resp. from) the encrypted image. It is the impact of this data hiding process onto the "pre-watermark" that gives us access to the message into the spatial domain, i.e. after the decryption process. By doing so, the watermark processing process is independent of the knowledge of the encryption key and one only just has to know the watermarking key so as to embed the message and extract it from the encrypted or decrypted image. Reciprocally, encryption/decryption processes are completely independent from message embedding/extraction. We illustrate the feasibility of our approach considering the RC4 stream cipher algorithm and the Least Significant Bit substitution watermarking modulation. Experiments conducted on natural test images and ultrasound medical images demonstrate the general capabilities of our system to securely make available a message in both spatial and encrypted domains while minimizing image distortion. We further discuss the use of different encryption and watermarking algorithms and we illustrate how our system can be used for ensuring image integrity and authenticity control.

1. Introduction

Nowadays, research on processing data in an encrypted form gains in interest. It allows for example working on externalized data without endangering privacy and data confidentiality [1, 2]. Among possible encrypted data processing, watermarking offers the opportunity for better data protection. When applied to images, this latter modifies or modulates image pixels' gray values in an imperceptible way in order to encode or insert a message, the content of which can be some security attributes (e.g. digital signature and users' identifiers). As defined, watermarking is an "*a posteriori*" protection mechanism due to the fact it leaves access to the image while maintaining it protected. It completes encryption an "*a priori*" protection mechanism, as encrypted data need to be deciphered before being accessed. Thus, combining watermarking or more generally data hiding with encryption can provide an *a priori/ a posteriori* protection at the same time [3].

Different approaches combining both technologies have been already proposed, essentially in the copyright protection framework. From a technical point of view, we can distinguish three categories of methods depending on the availability of the embedded message into the spatial domain (i.e. when the image is decrypted) and/or in the encrypted domain:

- *Message available in the spatial domain* (*MASD*) [4–8] – In [4], Memon et al. exploit homomorphic encryption [9]. Based on the homomorphic property, an operation in the encrypted domain on the cipher text (e.g. '×', '+'…) is translated into another operation on the plain text (e.g. '+','×'…). It is then possible to modify an encrypted image and consequently insert a watermark into the spatial domain. The approach described in [5] is another strategy where a bit is embedded in each encrypted image block by flipping the three least significant bit (LSB) planes of half the pixels in this block. The recipient has first to decrypt the image so as to extract the message with the help of correlation measurement in between pixels in each decrypted block. This approach was improved by Hong et al. by exploiting the pixel correlations between neighboring blocks so as to

[*] Corresponding author at: Telecom Bretagne, Département Image et Traitement de l'Information, Brest 29238, France.
E-mail address: Dalel.Bouslimi@telecom-bretagne.eu (D. Bouslimi).

http://dx.doi.org/10.1016/j.image.2016.06.012
0923-5965/© 2016 Elsevier B.V. All rights reserved.

decrease the message extraction error rate [7]. In [8], Qian et al. nearly reversibly embed a message into an encrypted JPEG bitstream. By using the encryption and watermarking keys, the embedded message is extracted and the original bitstream is approximately recovered by analyzing the blocking artifacts caused by message embedding.

- *Message available in the encrypted domain* (*MAED*) [10–13] – In [10], Puech et al. insert a message into the encrypted image by substituting one bit of a block with one of the message to be embedded. At the recipient side, each pixel block of the image is decrypted twice testing the two possible values of the original substituted bit (i.e. 0 or 1). The decrypted block with the minimum standard deviation is the original block. In such a scheme, the message is removed from the image during the decryption process. In [11], Zhang applies a lossless compression onto the LSBs of an image encrypted with a stream cipher algorithm. The space gain is then used for message insertion but as in the previous scheme, inserted data are only available in the encrypted domain and have to be removed before image decryption. Recently, Cao et al. propose a reversible scheme, where the original image are firstly divided into patches [12]. 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 which is used for message embedding in the encrypted domain.
- *Message available in both encrypted and spatial domains* (*MASED*) – These techniques are usually based on partial encryption [14–16] or invariant encryption [17, 18]. With these solutions, only some parts of the host image are encrypted, while the rest of it is watermarked [19, 20]. In [17] the image is first encrypted by permuting pixels' positions randomly without changing their gray values, which are modulated for message embedding. Embedded data are thus available in both encrypted and spatial domains. Due to the fact that encryption and watermarking are conducted on different host parts, one may consider its confidentiality is not fully ensured while making possible attacks based on this information leakage. A novel concept, called VRBE[1] has been proposed in [21, 22]. 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. For example, in [21], the LSBs of some image pixels (noted A) and their positions are firstly reversibly embedded into the rest of the image pixels to vacate some free space. The pixels of A are then rearranged on the top of the image, which is next RC4 encrypted. Being at the beginning of the encrypted bitstream, vacated LSBs can be easily identified and modified for message embedding. If the message can be extracted from the encrypted image, the RC4 decryption process has been modified so as to make possible the message extraction in the clear domain. Indeed, it decrypts the image except the LSBs that have been watermarked. Once this decryption process is achieved, the system has to reorganize the pixels so as to finally obtain the watermarked decrypted image. It appears thus that decryption and watermarking processes are not independent: decryption process requires the knowledge of the positions of watermarked LSBs and some other insertion parameters. Ref. [22] follows the same idea as [21] working with reversible histogram shifting modulation in application to prediction errors instead of pixel gray values. Again, image decryption requires the knowledge of encryption and data hiding keys and has to reorganize the image bitstream.

Beyond, most of the above methods, and especially those belonging to MASD and MAED classes, are limited to some practical scenarios. Let us consider the general framework of an image exchange between two users, the transmitter (*T*) and the receiver (*R*), who have various security needs: confidentiality, integrity, authenticity (i.e. verifying the origin of an image). Herein, confidentiality is ensured with the help of data encryption and other security services by means of watermarking. This communication may not be direct and some intermediaries (e.g. network nodes) may also contribute to data transmission. In this framework, different ways to merge watermarking/encryption can be considered. In a first time, if *T* stores images in an encrypted form, he may not want to decrypt and watermark them with appropriate security attributes (e.g. digital signature, transmitter's name and so on) before re-encrypting and sending them. Beyond the fact such a process is time consuming, it also introduce a confidentiality protection discontinuity as at one time data appear in a clear form. Being able to directly watermark encrypted data while making the embedded message available in the decrypted image is herein of interest. In a second time, *T* may also want that intermediaries have access to some image security attributes while this one is encrypted for verifying data integrity/authenticity as example. *T* must therefore be able to insert within an encrypted image a message that can be extracted in the spatial or encrypted domains, i.e. whether the image is encrypted or not. It is easy to see that MASD and MAED methods are not adequate because embedded data are available only in one domain. On the contrary, MASED methods can be used. However, due to the fact they are usually based on partial or invariant encryption of host data; they do not offer a full confidentiality protection. On their side, VRBE methods [21, 22] require reorganizing image bitstream before encryption and after decryption so as to reconstruct the image. With these methods, encryption and decryption processes are not independent leading to proprietary systems.

In this paper, we propose a new approach of MASED type, the principles of which allow the insertion of a message into encrypted data; a message by next available in both encrypted and spatial domains. More clearly, the same message can be extracted before and after image decryption. Compared to the MASED methods, the contributions of our approach are as follows:

1. Unlike methods based on partial and invariant encryption, the entire image is encrypted with our proposal.
2. Unlike VRBE methods, watermarking and encryption are independent. Message insertion and extraction processes (resp. encryption and decryption) do not require the knowledge of the encryption key (resp. watermarking key and extra parameters). Moreover, it does not require the reorganization of the data bitstream before its protection. More clearly, watermarking is completely transparent to the decryption process and if a system is not watermarking interoperable, it will be able to decrypt and access the image by simply applying the decryption process based on the encryption key.

The rest of this paper is organized as follows. In Section 2, we give the general principle of our approach and illustrate one possible implementation based on the LSB substitution watermarking modulation and the Rivest Cipher 4 (RC4) stream cipher algorithm. We provide some experimental results in Section 3 as well as some capacity/distortion bounds of such an implementation. Section 4 is devoted to a general discussion about the security of our proposal, its implementation with various encryption and watermarking algorithms as well as its deployment for verifying the integrity and authenticity of an image.

[1] Vacating room before encryption.