Signal Processing 143 (2018) 336–345

Contents lists available at ScienceDirect

Signal Processing

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

On the security of a stream cipher in reversible data hiding schemes operating in the encrypted domain

Fouad Khelifi

Department of Computer and Information Sciences, Northumbria University, NE2 1XE, UK

ARTICLE INFO

Article history:
Received 18 April 2017
Revised 14 August 2017
Accepted 19 September 2017
Available online 21 September 2017

Keywords:
Reversible data hiding
Encryption
Security
Inter-pixel redundancy

ABSTRACT

Reversible data hiding in encrypted images has recently emerged as an effective approach to embed and extract a message in the encrypted domain and losslessly recover the host data while maintaining its confidentiality through encryption. That is, the data hider can embed and extract additional data without knowing the image. This approach can be used in cloud applications where the service provider, i.e., the data hider, is not authorized to access the visual content of the host data for security and privacy purposes. Most existing techniques that have been reported in the literature apply a bit-wise encryption method, also known as the stream cipher, prior to data hiding. However, because of the spatial redundancy that characterizes natural images, the security of such an encryption could be compromised. This work is the first one that analyzes reversible data hiding in encrypted images from a security perspective. It proposes a Ciphertext-Only Attack (COA) and highlights the weakness of current state-of-the-art data hiding systems in the encrypted domain. We particularly show how the data hider can break the security of the encryption system and consequently discloses the visual content of encrypted images. Finally, possible solutions to combat COA with existing systems are discussed.

1. Introduction

Reversible Data Hiding (RDH) was first introduced in the pixel domain with the aim of embedding as much data as possible in the host image provided that the distortion incurred is visually acceptable [1–3]. Reversibility suggests that the decoder can recover the original host image losslessly. The idea of RDH in encrypted images is inspired by existing works on information processing in the encrypted domain for cloud computing and privacy-preserving applications such as data compression [4,5], watermarking [6] and signal representation and transforms [7,8]. The idea has also been extended in [9] for RDH in encrypted compressed bitstream. As the name suggests, RDH in the encrypted domain consists of embedding and extracting the message in the encrypted domain [10,11]. This ensures the confidentiality of the host image through encryption since the data hider can only embed and extract additional data without access to the visual content of the image. Such systems can be used in applications where the content is protected from the service provider (i.e. the data hider) while being shared with authorized users.

RDH techniques in the encrypted domain can be cast in two classes based on whether the embedding room is created before or after encryption [12]. The very first attempts in the literature create space after encryption [10,11,13]. This approach has also been broadly adopted in recent years [14–23]. In the other category of RDH in encrypted images, the embedding space is created before encryption [12,24–26]. Because this approach attempts to create room in the spatial domain, inter-pixel redundancy can be exploited in an efficient way for higher embedding capacity. It is also worth mentioning an interesting feature in some systems, called separability, which has been introduced in [14] for potential use in areas where the extraction of additional data needs to be performed *separately* from image recovery [12,14,15,17,22,24,26].

With the exception of a single attempt, reported recently in [21], all existing data hiding techniques in the encrypted domain have been assessed on the basis of two conflicting measures, i.e., the capacity of data hiding and the quality of marked images after decryption. Thus, the security aspect has been absent in existing works on RDH in encrypted images. Nonetheless, the security of such systems is paramount because the primary aim of encryption is to protect the image content from the data hider. The aforementioned paper that has touched on the security of a RDH system in encrypted images considered the Watermarked Only Attack (WOA) in which the attacker has access only to the encrypted marked image [21]. However, the problem where the attacker has access to the encrypted image before data hiding was never addressed. Note that this attack, which is often referred to as the ciphertext-only attack, is very likely to occur since the data hider can act as the

E-mail address: fouad.khelifi@northumbria.ac.uk

https://doi.org/10.1016/j.sigpro.2017.09.020
0165-1684/

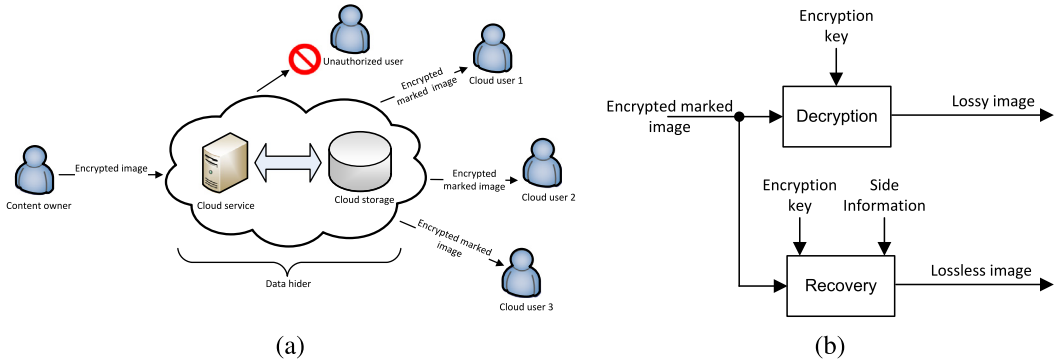

Fig. 1. Application of RDH in encrypted images. (a) Secret image sharing through the cloud using RDH in encrypted images. (b) Reconstruction of the original image by the cloud user.

malicious attacker. It is worth noting that the ciphertext-only attack has been used in [27] to demonstrate the weakness of a selective image scrambling scheme. It has also been successful in breaking a number of permutation-only image encryption systems [28,29] though its performance remains clearly lower than that of the known-plaintext attack and the chosen-plaintext attack [30–32]. In this context, it is also worth mentioning other attempts that break the security of image-based systems when the key is re-used multiple times [27,32–35].

This paper analyzes the security of a stream cipher that has been widely applied in current state-of-the-art RDH techniques operating in the encrypted domain. From the aforementioned techniques, the stream cipher was particularly used in [11–18,21–23,25,26]. Because the primary aim of encryption in such RDH schemes is to protect the content of images from the data hider, the cipher security is of paramount significance for privacy-preserving applications. Compared with other image encryption techniques, the stream cipher is faster and provides more flexibility for the data hider to create room and embed additional data. However, because of the spatial redundancy that characterizes natural images, the security of such a cipher could be compromised. This paper proposes a ciphertext-only attack and highlights the weakness of current state-of-the-art data hiding systems in the encrypted domain. A new algorithm is proposed to estimate the stream cipher key and break the security of the encryption system. As a result, the estimated key enables the data hider to disclose the visual content of any encrypted image. The novelty of this work is twofold: (i) A cryptanalysis on the stream cipher in RDH systems is provided. Unlike existing work, this paper evaluates RDH systems in the encrypted domain from a *security* point of view where the data hider acts as the attacker to disclose the image content. (ii) A new ciphertext-only attack exploiting inter-pixel redundancy is proposed. The technique can efficiently estimate the stream cipher key and consequently disclose the visual content of stream cipher encrypted images.

The rest of the paper is structured as follows. Section 2 presents a formulation of the security problem with current schemes of reversible data hiding in encrypted images. In Section 3, the proposed technique for breaking the stream cipher key is described. Section 4 demonstrates the performance of the proposed technique through an experimental analysis on natural images. Section 5 discusses possible solutions to tackle the security problem with existing RDH in encrypted images. Finally, Section 6 summarizes the contributions and concludes the paper.

2. Problem formulation

Practical scenarios of RDH in encrypted images have been suggested in [11,12,18,22,25] for secret image sharing and secure cloud computing. As illustrated by Fig. 1(a), sensitive visual data such as biometrics (face, fingerprint, iris, etc.) and medical images need to be encrypted before they are sent to the cloud for storage. This ensures the confidentiality of the data since only authorized cloud users can disclose the content of images via decryption and image recovery. On the other hand, the Cloud Service Provider (CSP) may embed some meta data about the owner and the date of upload in encrypted images to facilitate their management in the database (i.e. storage, search, authentication, and retrieval). As a result, the CSP acts as a data hider by embedding meta data in encrypted images and, upon request, serves authorized users with encrypted marked images. Therefore, the design suggests that the authorized users must be able to reconstruct the original image from the encrypted marked version. In the literature, the reported RDH schemes that operate in the encrypted domain normally offer both lossy and lossless image reconstruction depending on the key and information available at the receiver side [11–18,20,21,24–26]. In fact, if the encrypted marked image is directly decrypted, the reconstructed image will be similar to the original but with some loss of information due to the data embedding process. Existing research works consider two conflicting design requirements in this case. (i) The reconstruction quality, often measured by PSNR, on one hand, and (ii) The data hiding capacity, measured by the number of bits embedded, on the other hand. In addition to the encryption key, if the receiver also has access to some side information about the data hiding process (i.e., method used and location of bits altered) he can recover the image *losslessly*. This is shown in Fig. 1(b). In short, secure and efficient data storage in the cloud involves two separate systems operating in cascade, i.e., the encryption system, used by the content owner, on one hand and the data hiding system, used by the CSP, on the other hand. The embedding room could be created before or after encryption (see Fig. 2). However, as pointed out earlier, none of the aforementioned papers considered the robustness of RDH in encrypted images thoroughly from a *security* perspective. Obviously, the encryption process is of crucial importance here because the data hider is not meant to access the image content. This paper is the first one to analyze RDH schemes in the encrypted domain from a *security* point of view where image content protection from the CSP is considered.

In most state-of-the-art techniques [11–18,21,25,26], the stream cipher is adopted to protect the image content. Denote by X^k an image to be encrypted and transmitted by the content owner where $k = 1, \cdots, \ell$ and ℓ is the number of received images. Each pixel is supposedly encoded with 8 bits. The representation bit for a given pixel $X^k(i, j)$ at (i, j) can be given as

$$b_q^k(i,j) = \left\lfloor \frac{X^k(i,j)}{2^q} \right\rfloor \mod 2\,, \quad q = 0, 1, \cdots, 7. \tag{1}$$