REVERSIBLE DATA HIDING IN ENCRYPTED IMAGES BASED ON RESERVING ROOM AFTER ENCRYPTION AND MULTIPLE PREDICTORS

Ioan Catalin Dragoi and Dinu Coltuc

Electrical Engineering Dept.
Valahia University of Targoviste, Romania
Email: {catalin.dragoi, dinu.coltuc}@valahia.ro

ABSTRACT

This paper proposes a new vacating room after encryption reversible data hiding scheme. Hidden data is embedded into the encrypted host image by bit-flipping a preselected bit-plane of a randomly formed pixel group. The major novelty of the paper is the use of multiple predictors in an adaptive procedure for detecting between original and modified pixels. Four predictors are used on a context of four neighbors, namely the average of the four pixels, a weighted average based on local gradients, the median and the midpoint. Experimental results are provided. Compared with the state-of-the-art reserving room after encryption schemes, the proposed approach provides higher embedding bit-rates at lower distortion.

Index Terms— reversible data hiding, vacating room after encryption, prediction, bit-flipping

1. INTRODUCTION

Reversible data hiding methods in encrypted images (RDH-EI) are generally divided into two classes: reserving room before encryption (RRBE) and vacating room after encryption (VRAE) [1]. RRBE methods, like [2] and [3], use a preprocessing stage before encryption together with a proprietary encryption method. The VRAE methods use a standard encryption schemes and there is no secret key sharing between the data-owner and the data-hider. This makes VRAE a challenging research area. The first VRAE scheme was introduced in [4], this approach was further refined in [5, 6, 7, 8]. Also note that RRBE methods directly exploit the correlation found in the original image, outperforming their VRAE counterparts in both capacity and embedding distortions. But their need for preprocessing and proprietary encryption makes them much more limited in terms of applicability.

As far as we know, the VRAE RDH schemes proposed so far use a single criterion to distinguish between original and modified pixels. This criterion is either an equation like, for instance, the different versions of block based neighboring pixel differences found in [4, 5, 6] or the prediction error for randomly selected groups used in [7, 8].

This paper proposes a new vacating room after encryption RDH scheme. The main novelty of the paper is the use of multiple predictors with randomly selected groups in order to better distinguish between original and modified pixels. Each group contains either only modified or only original pixels. A predictor is considered reliable for a group of pixels, if it provides a clear majority of modified/original pixels in the group and unreliable if not. The decision on original/modified is taken adaptively, depending on the reliability/unreliability of each predictor.

The outline of the paper is as follows. The two staged reserving room after encryption RDH framework of [8] is discussed in Section 2. The proposed multiple predictor based data extraction scheme is introduced in Section 3 and its performance is evaluated in Section 4. The conclusions are drawn in Section 5.

2. PREDICTION BASED RDH

The proposed scheme uses the two staged RDH framework recently proposed in [8] as an improvement of [7]. The hidden data is embedded into encrypted images generated with the standard XOR based stream-cipher encryption.

An encrypted image C is obtained based on the original image I and r, a pseudorandom bitstream sequence generated by the encryption key:

$$C_t = I_t \oplus r_t \tag{1}$$

where $\oplus$ is the exclusive-or operator. Equation (1) is used on each $t \in \{1, 2...8\}$ bit-plane of the image.

The data-hider selects a t bit plane and splits the encrypted image into three distinct sets (Fig. 1). Set U is kept unchanged. It is used at the decoding stage for predicting the watermarked pixels belonging to the other two sets.

Set A is the first to be embedded. The pixels belonging to A are randomly distributed into groups based on a data hiding

This work was supported by UEFISCDI Romania, in the frame of the PNIII-P4-IDPCE-2016-0339 and PN-III-P1-1.1-PD-2016-1666 Grants.

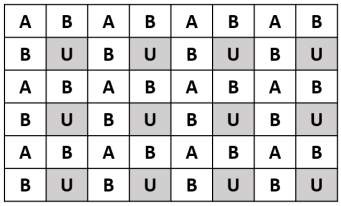
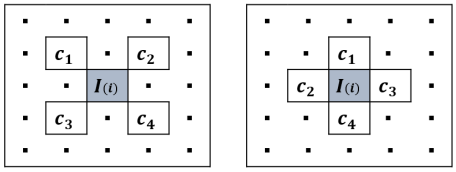


Fig. 1. Pixel distribution for [8] and for the proposed scheme.



(a) set A (b) set B

Fig. 2. The prediction context for set A and set B.

key. The first groups have always a fixed size and are used for storing the parameters of the embedding (watermark size, other group size). The remaining groups are used for data embedding. The embedding depends on the selected version of the RDH framework (joint or separate) and is similar for all the groups.

Let n be the size of a group and let $b \in \{0, 1\}$ be a hidden data bit. The joint version flips the t bit of each pixel in a group for $b = 1$ or keeps the group unchanged for $b = 0$:

$$C'_t(i) = \begin{cases} \sim C_t(i), & \text{if } b = 1, \\ C_t(i), & \text{if } b = 0. \end{cases} \tag{2}$$

where $\sim$ is the not operator and $i \in \{1, 2, ...n\}$. This equation allows both hidden data extraction and image restoration after image decryption.

The separate version first computes the parity on the t bit plane for the selected group:

$$s = C_t(1) \oplus C_t(2) \oplus ... \oplus C_t(n) \tag{3}$$

Note that this version of the RDH scheme requires an odd number n. Flipping the t bit values of all the pixels in the group will always flip the parity. Thus, b replaces the parity value:

$$C'_t(i) = \begin{cases} \sim C_t(i), & \text{if } s \neq b, \\ C_t(i), & \text{if } s = b. \end{cases} \tag{4}$$

where $i \in \{1, 2, \ldots, n\}$. This equation allows embedded data reading directly from the encrypted image. The host image is restored from its decrypted watermarked version.

If set A provides the necessary capacity for watermark embedding, the procedure stops. If not, the embedding continues for set B.

The watermarked image is obtained by decrypting each bit plane of the watermarked encrypted image:

$$I'_t = C'_t \oplus r_t \tag{5}$$

The decrypted image is then split into the U, A and B sets The joint version of the RDH framework checks each t bit plane for the embedding parameters, starting from the most significant bit planes, by determining if the corresponding bits were flipped (a similar operation is performed on the encrypted image for the separate version by reading the parity values). I'' is generated by flipping the t bits of a group:

$$I''(i) = \sim I'(i) \tag{6}$$

where $i \in \{1, 2, ...n\}$. A predicted value $\hat{I}(i)$ is generated for each pixel in the group using a predictor (the median value further discussed in the next section) on a prediction context formed by pixels of set U (Fig. 2.a). The bit plane of a group is considered flipped if, on average, the $I''(i)$ value is closer to $\hat{I}(i)$ then $I'(i)$ is to $\hat{I}(i)$, otherwise the pixels in the group are considered original:

$$I(1...n) = \begin{cases} I''(1...n), & \text{if } \sum_{i=1}^{n} \left|I'(i) - \hat{I}(i)\right| > \sum_{i=1}^{n} \left|I''(i) - \hat{I}(i)\right| \\ I'(1...n), & \text{if } \sum_{i=1}^{n} \left|I'(i) - \hat{I}(i)\right| \leq \sum_{i=1}^{n} \left|I''(i) - \hat{I}(i)\right| \end{cases} \tag{7}$$

where $|x| = \begin{cases} x, & \text{if } x \geq 0 \\ -x, & \text{if } x < 0 \end{cases}$. For the joint method, the hidden bit is extracted using:

$$b = \begin{cases} 1, & \text{if } I(1...n) = I''(1...n) \\ 0, & \text{if } I(1...n) = I'(1...n) \end{cases} \tag{8}$$

After set A is restored, the process is repeated for set B. A prediction context formed of pixels from U and A is used to predict the pixels in B (Fig. 2.b).

Note that both [7] and [8], like all reserving room after encryption schemes that rely on a standard encryption algorithm, can have decoding errors. Some bit flipped versions of pixels are closer to the corresponding predicted value than the original pixels. This is more common in textured regions. While this problem is unavoidable, its effects are mitigated by selecting the appropriate values for t and n (based on testing).

3. DATA EXTRACTION BASED ON MULTIPLE PREDICTORS

The proposed scheme uses four distinct predictors to evaluate if the t bit plane of the n pixel group was modified. The prediction context comprises the four diagonal neighbors for set A (Fig. 2.a) and the four closest horizontal/vertical neighbors for set B (Fig. 2.b). The predictors, selected by intensive testing, are:

- the average of the four neighbors:

$$\hat{I}_1 = \frac{c_1 + c_2 + c_3 + c_4}{4}; \tag{9}$$

- a weighted average based on local gradients:

$$\hat{I}_2 = \frac{(D_a + 1)\frac{c_1+c_4}{2} + (D_b + 1)\frac{c_2+c_3}{2}}{D_a + D_b + 2}; \tag{10}$$

where $D_a = |c_2 - c_3|$ and $D_b = |c_1 - c_4|$;

- the median:

$$\hat{I}_3 = \frac{c(2) + c(3)}{2} \tag{11}$$

where $c(1) \leq c(2) \leq c(3) \leq c(4)$;

- the midpoint (the average of the min and max values):

$$\hat{I}_4 = \frac{c(1) + c(4)}{2}. \tag{12}$$

The modified/original evaluation for each $I'(i)$ pixel in a group is performed for predictor k, $k = 1, \ldots, 4$:

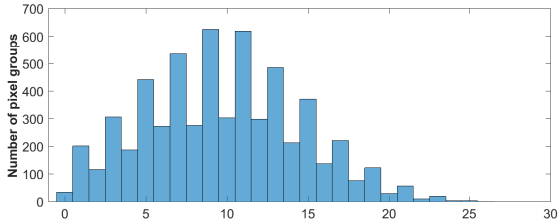
$$e_k(i) = \begin{cases} -1, & \text{if } \left|I'(i) - \hat{I}(i)_k\right| < \left|I''(i) - \hat{I}(i)_k\right| \\ 0, & \text{if } \left|I'(i) - \hat{I}(i)_k\right| = \left|I''(i) - \hat{I}(i)_k\right| \\ 1, & \text{if } \left|I'(i) - \hat{I}(i)_k\right| > \left|I''(i) - \hat{I}(i)_k\right| \end{cases} \tag{13}$$

A predictor is considered reliable for the current pixel group if $|\overline{e}_k| > n/5$, where $\overline{e}_k = \sum_{i=1}^{n} e_k(i)$. The sum of the $\overline{e}_k$ values of the reliable predictors is considered and the original value of the group is determined as:

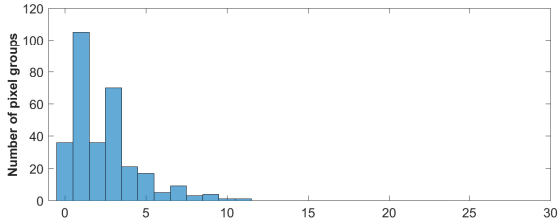
$$I(1...n) = \begin{cases} I''(1...n), & \text{if } \sum \overline{e}_k \leq 0 \\ I'(1...n), & \text{if } \sum \overline{e}_k > 0 \end{cases} \tag{14}$$

If no reliable predictors are found for the current group, equation (14) is computed by using the $\overline{e}_k$ values from all the selected predictors.

Before going any further, the selection of $n/5$ as the threshold value for $|\overline{e}_k|$ must be discussed. $|\overline{e}_k|$ represents the absolute difference between pixels where I' is considered the original value and those where I'' is chosen as the original. As can be seen from the example provided in Fig. 3, only the first half of the $|\overline{e}_k|$ domain is at risk of producing decoding errors. Meanwhile, these values also contain a significant number of correctly recovered groups. For $n = 31$, i.e. $n/5 \approx 6$, there is a good compromise between maintaining the predictor for groups that are properly decoded by it and attempting a new prediction for groups that are more likely to produce a decoding error. In the provided example, only 11 out of 284 incorrect blocks have a $|\overline{e}_k|$ value larger than 6. For this example, the use of multiple predictors has reduced the number of incorrectly decoded groups to around 245.



(a) $|\overline{e}_k|$ for correctly decoded groups



(b) $|\overline{e}_k|$ for incorrectly decoded groups

Fig. 3. Group distribution based on the selection value $|\overline{e}_k|$ for the median predictor on the test image *Mandrill*, $n = 31$, $t = 4$.

Fig. 4. The test images: 8 classic images and the Kodak set.

As a side note, the even values of $|\overline{e}_k|$ in Figure 3 have less groups because n is odd. Only groups with one or more pixels with $\left|I'(i) - \hat{I}(i)_k\right| = \left|I''(i) - \hat{I}(i)_k\right|$ (from equation (13)) can produce such values.

Another improvement with respect to [8] is the use of a different n value based on the current pixel set. The diagonal neighbors used for predicting the pixels in A are less reliable than the horizontal/vertical neighbors used in B, therefore a slightly larger value for n is needed for set A in order to compensate for the weaker prediction.

4. EXPERIMENTAL RESULTS

In this section, the proposed scheme is compared with other recent reserving room after encryption RDH schemes, namely the ones introduced in [7] and [8]. The performance of the three RDH schemes is evaluated on the 32 image set shown in Fig. 4 (8 standard 512×512 test images and the graylevel ver-

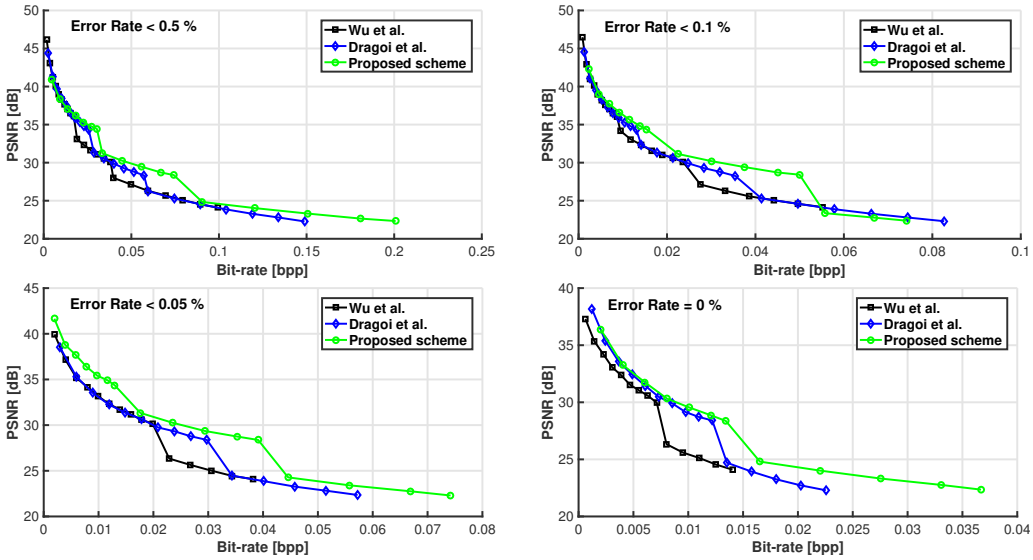


Fig. 5. Average PSNR versus Bit-rate results on the 32 image set for different error rate thresholds.

sions of the 768×512 Kodak set images). The Peak Signal-to-Noise Ratio (PSNR) between the original imagine and its decrypted watermarked version is used to evaluate the distortions introduced by RDH (distortion that are removed by extracting the hidden data).

We mention that each test was repeated 20 times in order to account for the variation in performance caused by the random group selection. The results are the averages of those tests and were obtained by varying the embedded bit plane t from 3 to 6, the group size n from 3, 4 and 5, after which it was incremented by 4 up to 61 pixels. The proposed scheme uses the same values for n in set B and for set A, n was either increased by 4, 8, 16 compared to B or was kept unchanged.

As can be seen from Fig. 5, the proposed scheme offers a noticeable increase in PSNR with respect to the other two RDH schemes. A gain in bit-rate is also observed for three of the four error rate thresholds: 0.05 bpp with an error under 0.5% and 0.015 bpp for both errors under 0.005% and decoding without errors.

5. CONCLUSIONS

An original vacating room after encryption reversible data hiding scheme has been proposed. The primary feature of the proposed scheme is a data extraction stage that is based on multiple predictors. Four predictors were selected for this purpose: the average of the four pixels, a weighted average based on local gradients, the median and the midpoint. The proposed approach outperforms several state-of-art vacating room after encryption reversible data hiding scheme.

6. REFERENCES

[1] Y. Q. Shi, X. Li, X. Zhang, H. T. Wu and Ma, "Reversible data hiding: advances in the past two decades", *IEEE Access*, 4, pp. 3210-3237, 2017.

[2] K. Ma, W. Zhang, X. Zhao, N. Yu, and F. Li, "Reversible Data Hiding in Encrypted Images by Reserving Room Before Encryption", *IEEE Trans. Inf. Forensics Security*, vol. 8, pp. 553–568, 2013.

[3] X. Cao, L. Du, X. Wei, D. Meng, and X. Guo, "High capacity reversible data hiding in encrypted images by patch-level sparse representation", *IEEE Trans. Cybernetics*, vol. 46, pp. 1132–1143, 2016.

[4] X. Zhang, "Reversible data hiding in encrypted images", *IEEE Signal Process. Lett.*, vol. 18, pp. 255–258, 2011.

[5] W. Hong, T. Chen, and H.Wu, "An improved reversible data hiding in encrypted images using side match", *IEEE Signal Process. Lett.*, vol. 19, pp. 199–202, 2012.

[6] Y.-S. Kim, K. Kang and D.-W. Lim, "New Reversible Data Hiding Scheme for Encrypted Images using Lattices", *Appl. Math.*, vol. 9, pp. 2627–2636, 2015.

[7] X. Wu and W. Sun, "High-capacity reversible data hiding in encrypted images by prediction error", *Signal Processing*, pp. 387–400, 2014.

[8] I.-C Dragoi, H.-G. Coanda and D. Coltuc, "Improved Reversible Data Hiding in Encrypted Images Based on Reserving Room After Encryption and Pixel Prediction", *in Proc. 25th Eur. Conf. Signal. Process. (EUSIPCO)*, pp. 2250–2254, 2017.