Scramble and Transform: An Image Data Hiding Technique

Mohammad Rezaei
Security Analysis Laboratory
SALAB
Tehran, Iran
rezaei@salab.ir

Saeed Montazeri Moghaddam
Security Analysis Laboratory
SALAB
Tehran, Iran
montazeri@salab.ir

Abstract—**Security is the most important concern in steganography which means hiding from visual or statistical attacks. Current steganography methods usually embed the secret data in the spatial or transform domains. Modern steganalysis methods extract features from different possible domains in which the data is hidden. To provide a high secure steganography method, this paper proposes to hide data in a secret domain. The secret domain is achieved by scrambling the image in the spatial domain and then applying integer wavelet transform. The message bits are embedded in the LSBs of randomly selected wavelet coefficients. Applying inverse wavelet transform and unscrambling, the stego image in the spatial domain is obtained. Experimental results were carried out on BOSSbase 1.01 database, and the results are compared against LSB matching (LSBM),** ***highly undetectable stego*** **(HUGO) and** ***spatial embedding of universal wavelet residual distortion*** **(S-UNIWARD) steganographic methods.**

Keywords—**Data hiding, steganography, integer wavelet transform, image scrambling, transform domain embedding**

I. Introduction

Steganography is concerned with hiding secret message into digital media without raising any suspicion. It is, however, a challenging task due to improvements in *steganalysis* which aims to detect the presence of hidden data [1]. Digital image is the most common carrier used for steganography [2]. Most of the steganographic techniques use the spatial domain or transform domain, such as *discrete cosine transform* (DCT) or *wavelet transform* (WT), to embed the message. In contrast, steganalysis methods investigate abnormal statistical artifacts resulted from data hiding in these domains. For example, numerous data hiding methods proposed for *joint photographic experts group* (JPEG) images embed the message by manipulating DCT coefficients. In this case, the JPEG steganalysis methods that extract features from DCT domain are more successful than the methods which extract features from spatial or wavelet domains. A solution is to embed the message in an unknown domain to steganalyzer.

Steganalysis methods are broadly classified into two groups: *specific* and *universal*. A specific or targeted method is designed to detect a known steganographic method while a universal or blind steganalysis method is based on extracting a large set of features and training a classifier in order to detect different types of embedding methods [2], [3].

One of the well-known specific methods is *sample pair analysis* (SPA) [4]. Some sensitive statistical properties to LSB embedding such as local correlation are extracted from two adjacent pixels called *sample pair*. Looking into the features of this method and other specific methods reveals that they provide accurate results only if they extract the statistics from domain in which the data is embedded.

One of the first blind methods was proposed by Farid, where 72 features are extracted from the wavelet transform of the image [5], [6]. This method provides less accurate results when the message is embedded in another domain, for example, in DCT coefficients of a JPEG image. In the case of JPEG images, several methods have been proposed that extract features directly from the DCT domain. They provide highly accurate results for a wide spectrum of steganographic methods [7]–[9]. *Subtractive pixel adjacency matrix* (SPAM) [10] is the state-of-art steganalysis method, which is widely used in the spatial domain. This method first models the differences between adjacent pixels using first-order and second-order Markov chains [11]. Then, subsets of transition probability matrices are used as features for a classifier implemented by support vector machines (SVM). According to [10], 686 second-order SPAM features provide highly accurate results.

The steganographic algorithm YASS produces stego JPEG files but embeds the data in a different domain than DCT [12]. It hides the data in the blocks of the image in the spatial domain, where the positions of the blocks are controlled by a secret key. The image is partitioned into 10x10 blocks and an 8x8 block is then randomly selected from each larger block to be used for data embedding. After hiding the data in the DCT domain, the image is decoded back to the spatial domain and converted to a JPEG file. This embedding procedure makes it difficult to identify the statistical properties that could lead to detect YASS [13]. Although YASS introduces more changes to the image than many other steganography methods, the steganalytic methods until 2007 were not able to detect it [13].

In this paper, we propose a novel steganography method by hiding the data in a different domain in order to defeat the steganalysis methods that extract features from the spatial domain or transform domain of the image. The image is first scrambled, and then, integer wavelet transform is applied to

978-1-7281-1508-5/19/$31.00 ©2019 IEEE

embed the message in the LSBs of the wavelet coefficients. After embedding, the image is decoded back to the spatial domain and unscrambled.

II. The Proposed Approach: Scramble and Transform

In this section, we first describe the proposed data embedding procedure and then analyze the main components of the method.

A. Embedding and Extracting the Message

The proposed scheme, as shown in Figure 1 (top), first scrambles the image in the spatial domain and then embeds the encrypted message in the transform domain. The image is scrambled based on a secret key shared between the sender and the receiver. Secret key initiates a function to produce a pseudo-random permutation of the integers from 1 to the number of pixels in the image, inclusively. Integer-to-integer wavelet transform is applied to the scrambled image.

The wavelet transforms which use filters with limited length can be computed using lifting-based scheme, which consists of three steps: split (*Lazy* wavelet transform), primal and dual lifting, and scaling [14]. The Lazy wavelet is a trivial wavelet which splits the original 1-D signal into odd and even indexed samples. A reversible integer-to-integer wavelet transform can be obtained by combining the lifting results with rounding off operation. This type of wavelet transform approximates *Haar* wavelet as given in (1).

$$\begin{aligned} h_i &= x_{2i+1} - x_{2i}, \\ l_i &= x_{2i} - \lfloor \frac{h_i}{2} \rfloor. \end{aligned} \tag{1}$$

where $\boldsymbol{x}$=(x_1,x_2,...,x_N) is the input integer vector with length N, and $\boldsymbol{l}$=(l_1,l_2,...,$l_{(N/2)}$) and $\boldsymbol{h}$=(h_1,h_2,...,$h_{(N/2)}$) are the resulting low-frequency (L) and high-frequency (H) wavelet subbands, respectively, and $\lfloor . \rfloor$ is rounding function which returns the largest integer not greater than its given value. To construct integer wavelet for an image, (1) is first applied to each row of the image. The resulting matrix with the same size as the image contains low-frequency (L) coefficients on the left and high-frequency coefficients on the right. Equation (1) is then applied to each column of the matrix to produce four subbands LL, LH, HL, HH. In this study, we use the Haar integer-to-integer wavelet transform.

The three wavelet sub-bands including LH, HL, and HH are chosen for data embedding. The message bits are embedded in the LSBs of the wavelet coefficients. To hide *n* message bits, *n* coefficients are selected randomly based on another secret key. However, the same key for image scrambling can be used. By applying inverse wavelet transform, the blocks are obtained in the spatial domain. The scrambled image containing hidden data is reconstructed from the blocks, and finally unscrambling is performed to obtain the intelligible stego image. To further clarify the embedding procedure, we provide a pseudo-code in Algorithm 1.

```
*Declarations*
*cover image: C
*stego image: S
*key: K
*a coefficient in wavelet Sub-bands: p
*desired payload: R
*length of R: L

*Embedding starts*
C_scrambled = Scramble C with K
[LL, LH, HL, HH] = Integer wavelet transform of
   C_scrambled
i = 1
WHILE i < L   *the payload is fully embedded*
    Choose an arbitrary p from one of LH, HL, or
    HH sub-bands
    Embed one bit of message, R(i), in p
    i = i + 1
S_scrambled = Inverse integer wavelet of (LL, LH,
   HL, HH)
S = Descramble S_scrambled using K
*Embedding Ends*
```

Algorithm 1: Pseudo-code of the proposed method

The procedure of extracting the message starts with scrambling the stego image in the same way as the embedding procedure, see Figure 1 (bottom). Integer wavelet transform is applied to the scrambled image, and the message bits are extracted from the selected wavelet coefficients specified by the same secret key used for embedding.

B. Analysis of the Proposed Method

In this section, the main components of the proposed steganography approach including image scrambling, integer wavelet transform, and embedding method are analyzed.

By scrambling the image in the spatial domain and applying wavelet transform, a secret transform domain to the steganalyzer is achieved. This technique increases the security and resistance against steganalysis methods. In the proposed approach, we scramble the image by pixel-level permutation based on a secret key, however, various scrambling methods can be used. For example, 2x2 or 4x4 blocks of the image could be scrambled, or scrambling techniques such as *Arnold* [15] used in image cryptography may be employed. There are a few works in the data hiding literature which mention image scrambling. However, some of them use scrambling just as an alternative for random selection of the pixels or coefficients in the transform domain [16]. Some others aim at hiding a scrambled image inside a cover image [17], [18]. In [19], a

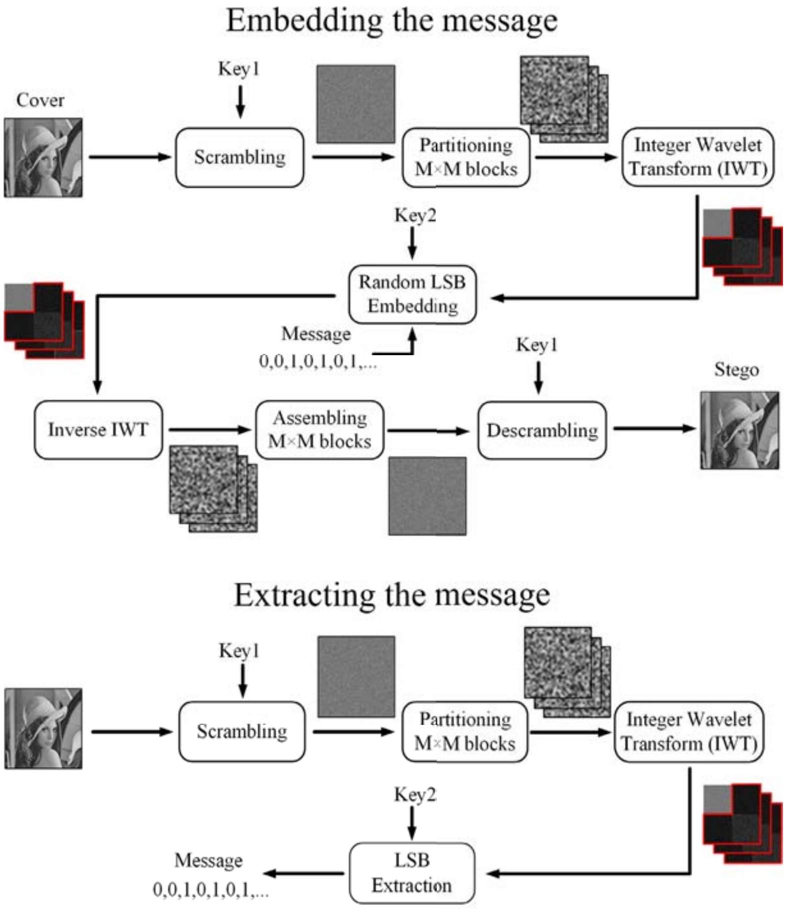

Fig. 1: The proposed data hiding procedure.

complicated approach based on scrambling in the DCT domain is proposed. However, it significantly degrades the image. The authors suggest to modify the JPEG quantization table to alleviate this problem. However, an abnormal quantization table can itself be an indication of steganography.

The secret data embedded in the transform domain can be extracted correctly only if a lossless transform is used. The reason is that inverse transform and unscrambling is applied after embedding to get the original image form. To extract the message, scrambling and transforming are needed to get the coefficients in which the data is hidden. However, if a lossy transform is used, different coefficients may be produced. Therefore, we employ the integer wavelet transform which is a lossless transform [20]–[22]. The transform can be applied to the entire image or non-overlapping blocks of the image. Since embedding in the low-frequency wavelet subband (LL) significantly affects the image quality, the data is embedded in the three subbands LH, HL, and HH. Changing wavelet coefficients in the data embedding process may cause underflow or overflow in the pixel values in the spatial domain (e.g. the values lower than 0 or larger than 255). This range violation may result in losing some parts of the embedded message. To overcome this problem, we utilized the cover adjustment method described in [23]. It is noteworthy that lossy transform such as DCT may be used but since some of the message bits are not correctly extracted, error correcting codes should be used. However, for the same payload, using lossy transform comparing to lossless transform introduces more changes leading to more image distortion.

Any embedding method may be used to hide the message in the wavelet coefficients. We use simple LSB replacement where the bits of the message are embedded in randomly selected coefficients of the subbands LH, HL, and HH.

III. Experiments

The proposed method is evaluated using three steganalysis methods including SPA [4], Farid [5], and SPAM [10], where the first one is targeted and the two others are universal steganalysis methods. SPA searches for statistical distortions in the image caused by LSB embedding. Farid's method extracts features from the wavelet domain. As we hide the data in the wavelet coefficients of the image, this method has been chosen. Embedding in the wavelet domain will affect the pixel values in the spatial domain. Therefore, SPAM, which is designed to assess the short-range dependences in the spatial domain, is also used for evaluating the proposed method.

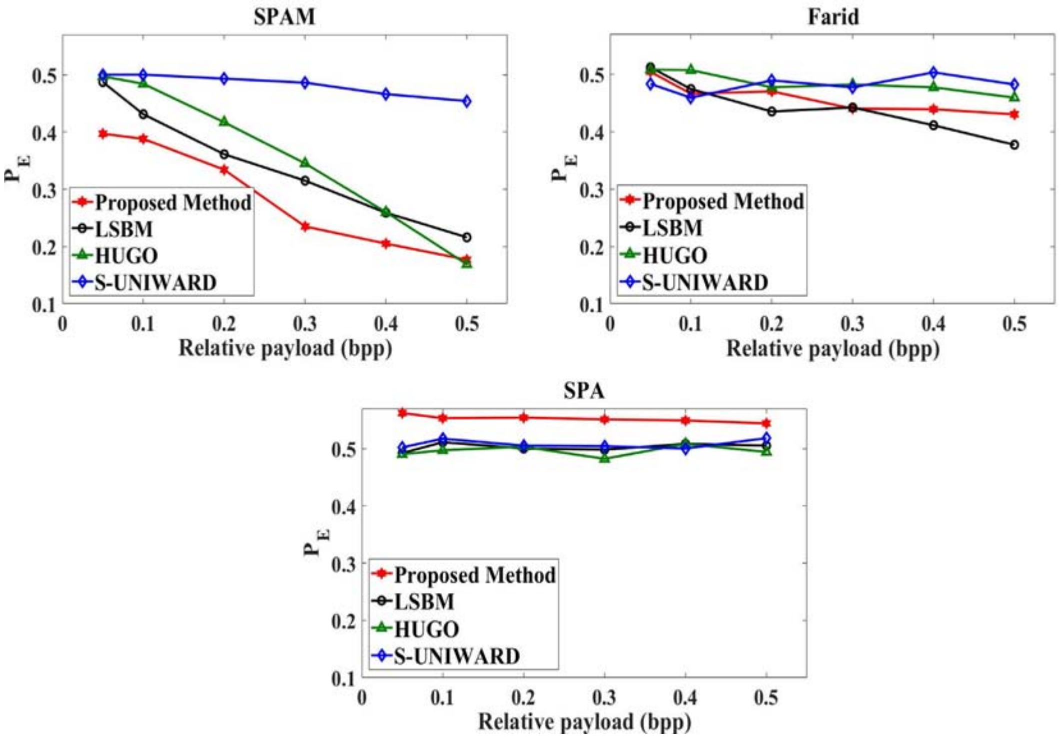

Fig. 2: Comparison of the proposed method with the steganography methods LSBM, HUGO, and S-UNIWARD based on the three steganalysis methods: SPA, Farid, and SPAM.

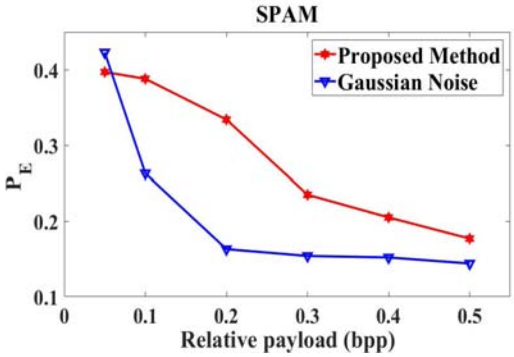

Fig. 3: SPAM results on non-stego images with Gaussian noise and stego images produced by the proposed method.

We provide comparisons with three well-known steganography methods: LSBM [24], HUGO [25], and S-UNIWARD [26]. LSBM is supposed to hide the stego data in the pixels as an additive noise which is randomly scattered in the image [24]. HUGO and S-UNIWARD are content-adaptive methods that hide stego data in the complex regions [27]. These methods use the *syndrome-trellis codes* (STCs) to select embedding locations which cause minimum distortion to the image. Following [25], [26], we set the constraint height of STCs $h = 10$ and $h = 12$ for HUGO and S-UNIWARD, respectively.

The experiments are conducted on the BOSSbase 1.01, including 10000 grayscale images of the fixed size 512×512, which were taken by seven digital cameras [28]. Given this dataset as input cover images, the three steganography methods were applied with six relative payloads (0.05, 0.1, 0.2, 0.3, 0.4, and 0.5 bpc), resulting in 180000 stego images.

The detection error of the steganalysis methods is measured by the *minimal average decision error* [25]:

$$P_E = min(\frac{P_{FP} + P_{FN}}{2}) \tag{2}$$

where P_{FN} and P_{FP} are the probability of false negative and the probability of false positive, respectively. Figure 2 shows the detection error of the steganalysis methods as the embedding rate increases from 0.05 to 0.5. SPA and Farid's method provide weak results for all steganography methods including the proposed approach and do not perform much better than a random guess. SPAM, however, as a more powerful steganalysis method, shows a much better performance for all the steganography methods except S-UNIWARD. It cannot detect S-UNIWARD even for the payloads as large as 0.3 bpp. The reason is that S-UNIWARD was designed to resist SPAM method. The detection error of SPAM, as can be seen in Figure 2, is linearly decreased as the embedding rate increases. The proposed method seems more detectable than LSBM and HUGO. The reason is that embedding data in the wavelet domain causes more changes in the spatial domain. Changing a wavelet coefficient can produce changes in several pixels in the spatial domain. However, this low detection rate of the proposed method cannot be considered as the strength of SPAM.

The point is that SPAM detects the stego images as well as naturally noisy images. To prove this, we performed another experiment. We added Gaussian noise in six different levels to the cover images. The amount of noise was set so that the same quality as stego images was obtained. For example, the noise level 2 has the same quality (measured by PSNR) as a stego image with payload 0.1. Gaussian noise is a common photographic noise which almost exists in every photo captured by a camera sensors [29]. The results in Figure 3 show that the detection error of SPAM for noisy images is less than the error for the proposed method. Embedding data in the wavelet domain of the scrambled image introduces some changes in the spatial domain, which are noise-like. No specific patterns on image statistics are built in the spatial domain, which indicates the stego data.

IV. Conclusions

This paper introduces a new steganography method which hides secret data in the wavelet transform of the scrambled image. After the embedding process, inverse wavelet transform and unscrambling are applied to provide the stego image in the spatial domain. The main goal of the proposed method is to embed data in an unknown domain to steganalyzers. The motivation is that the steganalysis methods are more successful when they extract features from the same domain that the data is hidden.

Our experiments show that the presented method is secure against SPA and Farid steganalysis methods, but SPAM detects our steganography method. However, it also classifies non-stego noisy images as stego. Therefore, SPAM seems sensitive to any changes to the image originated either from natural noise or from steganography.

References

[1] B. Li, M. Wang, X. Li, S. Tan, and J. Huang, “A strategy of clustering modification directions in spatial image steganography,” IEEE Trans. Information Forensics and Security, vol. 10, pp. 1905-1917, 2015.

[2] J. Fridrich, Steganography in Digital Media: Principles, Algorithms, and Applications, Cambridge University Press, 2009.

[3] B. Li, J. He, J. Huang, and Y. Q. Shi, “A survey on image steganography and steganalysis,” J. Information Hiding and Multimedia Signal Processing, vol. 2, pp. 142-172, 2011.

[4] S. Dumitrescu, X. Wu, and Z. Wang, “Detection of LSB steganography via sample pair analysis,” IEEE Trans. Signal Processing, vol. 51, pp. 1995-2007, 2003.

[5] H. Farid, “Detecting steganographic messages in digital images,” Technical Report TR2001-412, Deparment of Computer Science, Dartmouth College, 2001.

[6] H. Farid, “Detecting hidden messages using higher-order statistical models,” Proc. Int. Conf. on Image Processing, pp. 905-908, 2002.

[7] J. Fridrich, “Feature-based steganalysis for JPEG images and its implications for future design of steganographic schemes,” Int. Workshop on Information Hiding, pp. 67-81, 2004.

[8] T. Pevny and J. Fridrich, “Merging Markov and DCT features for multiclass JPEG steganalysis,” Electronic Imaging, vol. 6505, pp. 3 1-3 14, 2007.

[9] N. Provos and P. Honeyman, “Hide and seek: an introduction to steganography,” IEEE Security & Privacy, vol. 99, pp. 32-44, 2003.

[10] T. Pevny, P. Bas, and J. Fridrich, “Steganalysis by subtractive pixel adjacency matrix,” IEEE Trans. Information Forensics and Security, vol. 5, pp. 215-224, 2010.

[11] Y. Q. Shi, C. Chen, and W. Chen, “A Markov process based approach to effective attacking JPEG steganography,” Int. Workshop on Information Hiding, pp. 249-264, 2006.

[12] K. Solanki, A. Sarkar, and B. Manjunath, “YASS: Yet another steganographic scheme that resists blind steganalysis,” Int. Workshop on Information Hiding, pp. 16-31, 2007.

[13] J. Kodovský, T. Pevný, and J. Fridrich, “Modern steganalysis can detect YASS,” IS&T/SPIE Electronic Imaging, vol. 7541, pp. 02-01-02-11, 2010.

[14] A. R. Calderbank, I. Daubechies, W. Sweldens, and B. L. Yeo, “Wavelet transforms that map integers to integers,” Applied and Computational Harmonics Analysis, vol. 5, no. 3, pp. 332-369, 1998.

[15] W. Ding, “Digital image scrambling technology based on Arnold transformation,” J. Computer Aided Design and Computer Graphics, vol. 13, pp. 338-341, 2001.

[16] X. Song, F. Liu, Z. Zhang, C. Yang, X. Luo, and L. Chen, “2D Gabor filters-based steganalysis of content-adaptive JPEG steganography,” Multimedia Tools and Applications, vol. 76, no. 24, pp. 26391-26419, 2017.

[17] S. Khan, T. Khan, M. Ismail, M. H. Zafar, R. Ashraf, and N. Ahmad, “5LSB steganogaraphyusing monotonie RGB color image as cover medium,” 6th Int. Conf. Innovative Computing Technology (INTECH), pp. 275-278, 2016.

[18] O. Erdem and M. Turan, “A case study for automatic detection of steganographic images in network traffic,” 10th Int. Conf. Electrical and Electronics Engineering (ELECO), pp. 885-889, Bursa, 2017.

[19] K. Wong and K. Tanaka, “DCT based scalable scrambling method with reversible data hiding functionality,” 4th Int. Symposium on Communications, Control and Signal Processing (ISCCSP), pp. 1-4, 2010.

[20] A. Calderbank, I. Daubechies, W. Sweldens, and B.-L. Yeo, “Wavelet transforms that map integers to integers,” Applied and Computational Harmonic Analysis, vol. 5, pp. 332-369, 1998.

[21] I. Daubechies and W. Sweldens, “Factoring wavelet transforms into lifting steps,” J. Fourier Analysis and Applications, vol. 4, pp. 247-269, 1998.

[22] W. Sweldens, “The lifting scheme: A construction of second generation wavelets,” SIAM J. Mathematical Analysis, vol. 29, pp. 511-546, 1998.

[23] M. Tolba, M. Ghonemy, I. Taha, and A. Khalifa, “Using integer wavelet transforms in colored image steganography,” Int. J. Intelligent Cooperative Information Systems, vol. 4, pp. 230-235, 2004.

[24] T. Sharp, “An implementation of key-based digital signal steganography,” Int. Workshop on Information Hiding, 2001, pp. 13-26.

[25] T. Pevny, T. Filler, and P. Bas, “Using high-dimensional image models to perform highly undetectable steganography,” Int. Workshop on Information Hiding, pp. 161-177, 2010.

[26] V. Holub, J. Fridrich, and T. Denemark, “Universal distortion function for steganography in an arbitrary domain,” EURASIP J. Information Security, vol. 2014, no. 1, pp. 1-13, 2014.

[27] T. Denemark, V. Sedighi, V. Holub, R. Cogranne, and J. Fridrich, “Selection-channel-aware rich model for steganalysis of digital images,” IEEE Int. Workshop on Information Forensics and Security (WIFS), pp. 48-53, 2014.

[28] P. Bas, T. Filler, and T. Pevný, “Break our steganographic system: the ins and outs of organizing BOSS,” Int. Workshop on Information Hiding, pp. 59-70, 2011.

[29] V. Sedighi, R. Cogranne and J. Fridrich, “Content-adaptive steganography by minimizing statistical detectability” IEEE Trans. Information Forensics and Security, vol. 11, pp. 221-234, 2016.