Reversible Data Hiding in Encrypted Image with Distributed Source Encoding

Zhenxing Qian, Xinpeng Zhang

Abstract—This paper proposes a novel scheme of reversible data hiding (RDH) in encrypted images using distributed source coding (DSC). After the original image is encrypted by the content owner using a stream cipher, the data-hider compresses a series of selected bits taken from the encrypted image to make room for the secret data. The selected bit series is Slepian-Wolf encoded using low density parity check (LDPC) codes. On the receiver side, the secret bits can be extracted if the image receiver has the embedding key only. In case the receiver has the encryption key only, he/she can recover the original image approximately with high quality using an image estimation algorithm. If the receiver has both the embedding and encryption keys, he/she can extract the secret data and perfectly recover the original image using the distributed source decoding. The proposed method outperforms previously published ones.

Index Terms—Reversible data hiding, image encryption, image recovery

I. INTRODUCTION

Information processing in the encrypted domain has attracted considerable research interests in recent years [1]. In many applications such as cloud computing and delegated calculation, the content owner needs to transmit data to a remote server for further processing. In some cases, the content owner may not trust the service supplier, and needs to encrypt the data before uploading. Thus, the service provider must be able to do the processing in the encrypted domain. Some works have been done for data processing in an encrypted domain, for example, compressing encrypted images [2]-[4], adding a watermark into the encrypted image [5][6], and reversibly hiding data into the encrypted image [7]-[13].

Unlike robust watermarking, reversible data hiding emphasizes perfect image reconstruction and data extraction, but not the robustness against malicious attacks [14]. Many RDH methods for plaintext images have been proposed [15-19], for example, a common framework of redundancy compression [14], difference expansion (DE) [15] and histogram shifting (HS) [16] approaches. However, these are not applicable to encrypted images since the redundancy in the original image cannot be used directly after image encryption.

As a new trend, reversible data hiding in encrypted images allows the service provider to embed additional messages, e.g., image metadata, labels, notations or authentication information, into the encrypted images without accessing the original contents. The original image is required to be perfectly recovered and the hidden message completely extracted on the receiving side. Reversible data hiding in encrypted images is desirable. For example, in medical applications, a patient does not allow his/her medical images to be revealed to any outsiders, while the database administrator may need to embed medical records or the patient’s information into the encrypted images. On the other hand, the original medical image for diagnosis must be recovered without error after decryption and retrieval of the hidden message. The emerging methods [7]-[13] on reversible data hiding in encrypted images are reviewed in Section II.

This paper aims to enhance embedding payload in encrypted images. We propose a separable reversible data hiding method for encrypted images using Slepian-Wolf source encoding [21]. The idea is inspired by the DSC [22-24], in which we encode the selected bits taken from the stream-ciphered image using LDPC codes [25] into syndrome bits to make spare room to accommodate the secret data. With two different keys, the proposed method is separable. The hidden data can be completely extracted using the embedding key, and the original image can be approximately reconstructed with high quality using the encryption key. With both keys available, the hidden data can be completely extracted, and the original image perfectly recovered with the aid of some estimated side information. The proposed method achieves a high embedding payload and good image reconstruction quality, and avoids the operations of room-preserving by the sender.

The rest of the paper is organized as follows. Previous works of RDH in encrypted images are surveyed in Section II. The proposed system is described in Section III. Section IV presents the procedures of image encryption and data embedding. Data extraction and image recovery are elaborated in Section V. Section VI presents the experimental results and Section VII discusses the proposed method. Section VIII concludes the paper.

II. PREVIOUS WORKS

In this section, the state-of-the-art RDH techniques of embedding secret message in encrypted images are reviewed. RDH for encrypted image are usually designed for the applications in which the data-hider and the image owner are not the same party. The data-hider cannot access the image content, and the secret message is held by the data-hider. Thus,
encryption is done by the sender, hiding by the data-hider, and data extraction and/or image reconstruction by the receiver. For the ease of discussion, explanations of some frequently used terms are listed in Table I.

<table>
<thead>
<tr>
<th>Terms</th>
<th>Explanations</th>
</tr>
</thead>
<tbody>
<tr>
<td>original image</td>
<td>natural image in plaintext form</td>
</tr>
<tr>
<td>encrypted image</td>
<td>image obtained by encrypting the original image</td>
</tr>
<tr>
<td>additional bits</td>
<td>secret message to be embedded into the encrypted image</td>
</tr>
<tr>
<td>marked encrypted image</td>
<td>encrypted image containing additional data bits</td>
</tr>
<tr>
<td>approximate image</td>
<td>the reconstructed image close to the original image</td>
</tr>
<tr>
<td>recovered image</td>
<td>perfectly restored image that is identical to the original image</td>
</tr>
<tr>
<td>sender</td>
<td>owner of the original image who encrypts the original image and uploads the encrypted image to the server</td>
</tr>
<tr>
<td>data-hider</td>
<td>service provider who embeds additional message into the encrypted image</td>
</tr>
<tr>
<td>receiver</td>
<td>one who receives the marked encrypted image, and performs data extraction and/or image reconstruction</td>
</tr>
</tbody>
</table>

Existing RDH methods for encrypted images can be classified into two categories: “vacating room after encryption (VRAE)” and “vacating room before encryption (VRBE)” [11].

In VRAE, the original image is encrypted directly by the sender, and the data-hider embeds the additional bits by modifying some bits of the encrypted data. The idea was first proposed by Puech et al. [7], in which the owner encrypts the original image by Advanced Encryption Standard (AES), and the data-hider embeds one bit in each block containing n pixels, meaning that the embedding rate is 1/n bit-per-pixel (bpp). On the receiver side, data extraction and image recovery are realized by analyzing the local standard deviation during decryption of the marked encrypted image. This method requires that image decryption and data extraction operations must be done jointly. In other words, extraction and decryption are inseparable.

With a different idea, Zhang proposed a practical RDH method for encrypted images in [8], in which the data-hider divides the encrypted image into blocks and embeds one bit into each block by flipping three least significant bits (LSB) of half the pixels in the block. On the receiver side, the marked encrypted image is decrypted to an approximate image. The receiver flips the three LSBs of pixels to form a new block. Due to spatial correlation in natural images, original block is presumed to be much smoother than interfered block. Thus the embedded bits can be extracted and the original image recovered jointly. Embedding rate of this method depends on the block size. If an inappropriate block size is chosen, errors may occur during data extraction and image recovery. This method was improved in [9] by exploiting spatial correlation between neighboring blocks and using a side-match algorithm to achieve a better embedding payload with much lower error rates in image recovery. Both methods [8] and [9] are feasible based on the spatial correlation in natural images. Data extraction, however, are inseparable.

To overcome the drawback of inseparability in [7]-[9], a separable RDH scheme was proposed for encrypted images in [10]. The data-hider pseudo-randomly permutes and divides the encrypted image into groups with size of L. The P LSB-planes of each group are compressed with a matrix G sized (P–L)×P×L to generate corresponding vectors. Thus, S bits are available for data embedding. On the receiver side, a total of (8–P) most significant bits (MSB) of pixels are obtained by decryption. The receiver then estimates the P LSBs by the MSBs of neighboring pixels. By comparing the estimated bits with the vectors in the cost Ω corresponding to the extracted vectors, the receiver can recover the original bits of the P LSBs. Because the additional bits are embedded in LSBs of the encrypted images, which can be extracted directly before image recovery, data extraction and image recovery are therefore separable. Besides, this method achieves a better embedding rate than [8] and [9]. Another separable method was proposed in [12], in which the data-hider embeds additional bits by a histogram shifting and n-ary data hiding scheme, greatly improving the embedding payload as compared to [8]-[10]. However, as the original image is encrypted with pixel permutation and affine transformation, leakage of image histogram is inevitable under exhaustive attack.

In the VRBE, the original images are processed by the owner before encryption to create spare space for data embedding, and the secret data are embedded into specified positions by the data-hider. For example, the method in [11] creates embedding room in the plaintext image by embedding LSBs of certain pixels into other pixels using a traditional RDH method. The pre-processed image is then encrypted by the owner to generate an encrypted image. Thus, positions of these vacated LSBs in the encrypted image can be used by the data-hider, and a large payload up to 0.5 bpp, can be achieved. With a similar idea, another method based on an estimation technique was proposed in [13], in which a large portion of pixels is used to estimate the rest before encryption, and the final version of encrypted image is formulated by concatenating the encrypted estimating errors and a large group of encrypted pixels. Additional bits can be embedded into the encrypted image by modifying the estimating errors. With this method, PSNR of the approximate image reconstructed by the receiver is higher than previous methods.

Both [11] and [13] are separable RDH methods with good embedding rates and reconstruction capability, but require an additional RDH operation by the sender before image encryption. That means the issue of RDH in encrypted images is actually transformed into a traditional RDH in plaintext images.

In summary, methods in both VRAE and VRBE categories are effective for RDH in encrypted images. However, there are some limitations. In VRAE RDH methods for encrypted
images, estimation technique is necessary for the receiver, because no prior information of the original content is available except that he/she knows that the cover is a natural image. In traditional methods, LSB planes of the encrypted image are modified to accommodate the additional message, and the image recovery is based on estimating the original LSB planes with an assessment criterion, such as the “fluctuation function” in [8]-[10]. Because these estimations are not accurate enough, the recovery is only suitable for the case when a small amount of additional bits were embedded.

Although VRBE can achieve a higher payload, it requires that the sender must perform an extra RDH before image encryption. This may be impractical, because the sender has no idea of the forthcoming data hiding by the data-hider, or he/she has no computational capability of the traditional RDH. On the other hand, in case the sender can reserve room for embedding by reversibly hiding redundant bits into the original plain image, all embedding tasks can also be done on the sender side and then the data-hider becomes redundant.

In view of these problems, we propose a method to achieve high embedding payload by combining the MSB estimation with DSC. As estimating MSB is much more accurate than estimating LSB planes, the original data of the MSB plane can be recovered by DSC decoding with an acceptable decoding error probability. In other words, large embedding capacity can be achieved by this kind of combination.

III. SYSTEM DESCRIPTION

The proposed system is sketched in Fig. 1, which consists three phases: image encryption, data embedding, and data extraction/image recovery. In phase I, the sender encrypts the original image into an encrypted image using a stream cipher and an encryption key. In phase II, the data-hider selects and compresses some MSB of the secret image using LDPC codes to generate a spare space, and embeds additional bits into the encrypted image using an embedding key. In phase III, the receiver extracts the secret bits using the embedding key. If he/she has the encryption key, the original image can be approximately reconstructed via image decryption and estimation. When both the encryption and embedding keys are available, the receiver can extract the compressed bits, and implement the distributed source decoding using the estimated image as side information to perfectly recover the original image.

IV. IMAGE ENCRYPTION AND DATA EMBEDDING

A. Image Encryption

Without loss of generality, we assume the original image \( O \) is a grayscale image with all pixel values falling into \([0, 255]\), and the image size is \( M \times N \) where both \( M \) and \( N \) are power of 2. First, the image owner turns the original image into plain bits by decomposing each pixel into 8 bits using

\[
b_{i,j,u} = \left[ O_{i,j}/2^u \right] \mod 2, \quad u = 0, 1, 2, \ldots, 7
\]

(1)

where \( O_{i,j} \) are pixels of the original image, \( 1 \leq i \leq M, 1 \leq j \leq N \).

The owner then chooses an encryption key \( K_{\text{ENC}} \) to generate pseudo-random bits using a stream cipher function (e.g., RC4 or SEAL), and encrypts the bitstream of the original image by

\[
e_{i,j,u} = b_{i,j,u} \oplus K_{i,j,u}, \quad u = 0, 1, 2, \ldots, 7
\]

(2)

where \( k_{i,j,u} \) are the key stream bits, \( e_{i,j,u} \) the generated cipher text, and \( \oplus \) denotes exclusive OR.

Accordingly, the encrypted image \( E \) can be constructed by

\[
E_{i,j} = \sum_{u=0}^{7} e_{i,j,u} \cdot 2^u
\]

(3)

\( E_{i,j} \) are pixel values of the encrypted image, \( 1 \leq i \leq M, 1 \leq j \leq N \). Note that the stream cipher in (2) only scrambles pixel values but does not shuffle pixel locations.

B. Data Embedding

After image encryption, the content owner sends the encrypted image to the data-hider. To embed additional data into the image, the data-hider first decomposes the encrypted image \( E \) into four sub-images \( E^{(1)}, E^{(2)}, E^{(3)} \) and \( E^{(4)} \), each sized \( M/2 \times N/2 \).

\[
\begin{align*}
E^{(1)}(i, j) &= E(2i - 1, 2j - 1) \\
E^{(2)}(i, j) &= E(2i - 1, 2j) \quad i = 1, \ldots, M/2 \\
E^{(3)}(i, j) &= E(2i, 2j - 1) \quad j = 1, \ldots, N/2 \\
E^{(4)}(i, j) &= E(2i, 2j)
\end{align*}
\]

(4)

\( E^{(k)}(i,j) (k=1,\ldots,4) \) are pixels of the sub-images. Fig. 2 shows an example, in which a 4x4 image is down-sampled to four 2x2 sub-images.

![Fig. 2 An example of image decomposing](image-url)
where the arithmetic is modulo-2, $k$ the group index, and $S$ the resulting syndrome.

According to Eq. (5), groups $[C(k,1), C(k,2), \ldots, C(k,n)]$ are compressed into $[S(k,1), S(k,2), \ldots, S(k,r)]$. Thus, $n-r$ bits are vacated for data hiding. Assuming there are $K(n-r)$ additional bits to be embedded, the data-hider encrypts these data using a stream cipher algorithm such as RC4 or SEAL with another key $K_{SE}$. Divide the encrypted additional bits into $K$ segments and append each segment to $[S(k,1), S(k,2), \ldots, S(k,r)]$ to create a composition $[C'(k,1), C'(k,1), \ldots, C'(k,n)]$, $[C(k,1), C(k,2), \ldots, C(k,n)]$ are then replaced with $[C'(k,1), C'(k,1), \ldots, C'(k,n)]$ and put into the original MSB positions after inverse shuffling using the shuffle key $K_{SS}$. This way, a marked encrypted image is generated.

The keys $K_{SE}$, $K_{SS}$, and $K_{SC}$ constitute the embedding key $K_{E} = (K_{SE},K_{SS},K_{SC})$. Through a trusted channel, the embedding key $K_{E}$ and parameters $P_{k} = (L, n, r)$ are transmitted to the receiver.

### C. Virtual Channel and Embedding Rate

As the proposed method is based on DSC, we only use the LDPC matrix $H$ to generate the syndrome bits that compresses $n$ bits into $r$ bits. Here the ratio $\beta$ should be determined by correlation statistics between the source and the side information. DSC is depicted in Fig. 3, in which $X$ is the source to be encoded, $Y$ is the side information for decoding, and a virtual channel between $X$ and $Y$ is assumed [26].

![Fig. 3 Distributed source coding and decoding](image)

According to the Slepian-Wolf theorem of distributed source coding [22], $X$ can be compressed with the rate

$$R \geq H(X | Y)$$

(6)

where $H(x) = x \cdot \log_2(x) - (1-x) \cdot \log_2(1-x)$ is the entropy function.

Suppose we hide data into the spare room generated by compressing the source $X$ with an ideal Slepian-Wolf encoding. Thus the embedding capacity is

$$C_{SW} \leq H(X) - R = H(X) - H(X | Y)$$

(7)

For statistically independent and identically distributed (i.i.d.) binary source $X$ and $Y$, the embedding capacity is

$$C_{EB} \leq 1 - H(q)$$

(8)

where $q$ is the crossover probability of the virtual channel.

In the proposed method, $L$ bits are selected from $MN$ pixels as the embedding cover. Although the selected bits may be highly correlated, correlation is weakened after pseudo-random shuffling and segmentation. These bits can be assumed i.i.d., and the virtual channel can be treated as a binary symmetric channel (BSC) with error probability $q$. Thus, the upper bound of embedding rate in bits-per-pixel (bpp) is

$$R_{c} = \frac{L}{MN} \cdot \frac{C_{EB}}{3} = \frac{3}{4} \alpha [1 - H(q)]$$

(9)

With the LDPC algorithm for Slepian-Wolf compression, the actual embedding rate is less than the bound $R_{c}$. During data hiding, a total of $(n-r)K$ bits are reserved in all groups to accommodate the secret data. Thus the actual embedding rate between the amount of secret bits and the total number of pixels in bpp is

$$R_{e} = \frac{(n-r)K}{MN} \approx (1-\beta) L = \frac{3}{4} \alpha (1-\beta)$$

(10)

where $\beta = \frac{r}{n}$.

From the value of $q$, the ratio $\beta$ can be determined, and the LDPC matrix $H$ can be constructed in several different forms. The data-hider arbitrarily chooses a parity-check matrix $H$ corresponding to a regular or irregular LDPC code by setting the numbers of variable nodes and the check nodes. Algorithms have been proposed for the matrix construction, for example, matrices used in Gallager codes, MacKay codes, and finite geometry codes. Details of the LDPC matrix $H$ can be found in [23]. Since the data-hider has no knowledge of the virtual channel and there is no feedback channel between the data-hider and the receiver, a precise $q$ cannot be determined for every image. Therefore an empirical value of $q$ suitable for most natural images can be specified, which will be discussed in the experiment section.

### V. DATA EXTRACTION AND IMAGE RECOVERY

On the receiver end, with the marked encrypted image, the hidden data can be extracted using the embedding key, and the original image can be approximately reconstructed using the encryption key, or losslessly recovered using both of the keys. Three cases are analyzed below in Subsections A, B and C respectively, in which the receiver has the embedding key only, the encryption key only, and both. We denote the received encrypted image containing secret data as $V$.

#### A. Data Extraction

In the first case, the receiver extracts the embedded secret data using the embedding key $K_{E}$ and the parameters $P_{k} = (L, n, r)$, where $K_{E} = (K_{SE}, K_{SS}, K_{SC})$. Divide $V$ into four sub-images $V^{(1)}$, $V^{(2)}$, $V^{(3)}$, and $V^{(4)}$ using the same algorithm of Eq. (4). Collect all bits in the MSB planes of $V^{(2)}$, $V^{(3)}$, and $V^{(4)}$, and select $L$ bits according to the selection key $K_{SE}$. Shuffle the selected bits using the shuffle key $K_{SS}$ and divide the shuffled bits into $K$ groups, each containing $n$ bits. Denote the bits in each group as $D(k,l)$ where $k=1,2,\ldots,K$ and $l=1,2,\ldots,n$. For each group, extract the last $(n-r)$ bits $[D(k,r+1), D(k,r+2), \ldots, D(k,n)]$. Thus, the secret data can be reproduced by concatenating all the extracted bits from all $K$ groups, and decrypted to the plaintext message using the key $K_{SC}$.

#### B. Image Decryption and Estimation

In the second case, since the receiver has the encryption key but not the embedding key, he can reconstruct an approximate image based on image estimation.
Denote pixels in $V$ as $V'_{i,j}$ ($1 \leq i \leq M$, $1 \leq j \leq N$). Decompose $V'_{i,j}$ into 8 bits $v'_{i,j,0}, v'_{i,j,1}, \ldots, v'_{i,j,7}$ using Eq. (1), and decrypt the bits using stream decipher
\[
b_{i,j,u} = v'_{i,j,u} \oplus k_{i,j,u}, \quad u = 0, 1, 2, \ldots, 7
\] (11)
where $k_{i,j,u}$ are the key stream bits generated by the encryption key. Values of the decrypted image can be constructed from the deciphered bits using Eq. (3). We denote the decrypted image as $A$.

Without the embedding key, the selected pixels that contain secret bits cannot be identified. To approximately recover the content, the receiver down-samples the decrypted image $A$ into four sub-images $A^{(1)}, A^{(2)}, A^{(3)}$ and $A^{(4)}$ according to Eq. (4). Pixel values in sub-image $A^{(1)}$ are the same as the pixel values in the original image, while part of the MSBs of the pixels in $A^{(2)}, A^{(3)}$ and $A^{(4)}$ may differ from that in the original image. As a result, an estimation algorithm is defined to do the approximate reconstruction.

The receiver generates a reference image $B$ sized $M \times N$ from the sub-image $A^{(3)}$ using bilinear interpolation. Then, he/she down-samples $B$ into four sub-images $B^{(1)}, B^{(2)}, B^{(3)}$ and $B^{(4)}$. With $A^{(3)}$ and $B^{(3)}$, the estimated sub-images $A^{(d)}$ ($k=2, 3, 4$) can be generated using Eq. (12). Because the last seven MSBs in the sub-images $A^{(k)}$ ($k=2, 3, 4$) are unchanged during data hiding, the interpolated values in $B^{(k)}$ is used to optimize the MSB planes in $A^{(k)}$. For each sub-image $A^{(k)}$ ($k=2, 3, 4$), on the $(i,j)$-th position ($1 \leq i \leq M/2$, $1 \leq j \leq N/2$), if the calculated value
\[
\text{mod}(A^{(k)}(i,j)+128) + 128
\]
is closer to the interpolated value
\[
\text{mod}(B^{(k)}(i,j)+128)
\]
the MSB of the $(i,j)$-th pixel is estimated to be 1; otherwise 0.

At this point, the receiver can compose sub-images $A^{(1)}, A^{(2)}, A^{(3)}$ and $A^{(4)}$ to generate an estimated image $A'$:
\[
\begin{align*}
A'(2i-1,2j-1) &= A^{(1)}(i,j) \\
A'(2i-1,2j) &= A^{(2)}(i,j) \quad i = 1,2,\ldots,M/2 \\
A'(2i,2j-1) &= A^{(3)}(i,j) \quad j = 1,2,\ldots,N/2 \\
A'(2i,2j) &= A^{(4)}(i,j)
\end{align*}
\] (13)
Here $A'(r,s)$ ($r=1,2,\ldots; M$; $s=1,2,\ldots; N$) are the pixels of the composed image. The estimated image $A'$ is the approximate image reconstructed by the receiver.

The proposed estimation algorithm can also be used to find empirical error probability $q$ of the virtual channel. With a database containing numerous natural images, we perform the estimation algorithm to generated estimated images. Calculate differences of the MSBs of the last three sub-images between the original and estimated images. The largest error probability is chosen as $q$ for generating the LDPC matrix $H$.

C. Lossless Recovery

In the third case, if the receiver has both the encryption and embedding keys, he/she can extract the secret data correctly and recover the original image perfectly. Using the embedding key, the receiver collects the $L$ selected bits in the MSBs of the sub-images, and separates the $(n-r)-K$ secret bits from the $r-K$ compressed bits. Denote the extracted $r-K$ compressed bits as $S_r$ and divide $S_r$ into $K$ groups. Groups $[S_r(k,1), S_r(k,2), \ldots, S_r(k,r)]$ ($k=1,2,\ldots,K$) are used as side information for the distributed source decoder.

With the encryption key, the receiver constructs an approximate image $A'$ using the estimation algorithm proposed in Subsection B. Encrypt $A'$ by the stream cipher defined in Eq. (2), and down-sample the encrypted image into four sub-images. Extract the $L$ selected bits in the MSBs of the sub-images and permute these bits using the shuffle key. We denote the shuffled bits as $U$. Divide $U$ into $K$ groups $[U(k,1), U(k,2), \ldots, U(k,n)]$ ($k=1,2,\ldots,K$), each containing $n$ bits.

Next, the log-likelihood ratio (LLR) [23] are calculated using the predefined $q$.
\[
\text{LLR}(k,i) = \log \frac{\Pr[R(k,i) = 0|U(k,i) \not= k]}{\Pr[R(k,i) = 1|U(k,i) \not= k]}
\] (14)

With the calculated LLR, the LDPC parity-check matrix $H$ and the side information $[S_r(k,1), S_r(k,2), \ldots, S_r(k,r)]$, the original bits can be recovered with an iterative belief propagation algorithm (BPA). Details of the iterative BPA decoding algorithm can be found in [24]. A diagram of decoding is shown in Fig. 4, where $[R(k,1), R(k,2), \ldots, R(k,n)]$ stands for the recovered source bits and $k=1,2,\ldots,K$.

VI. EXPERIMENTAL RESULTS AND DISCUSSION

The proposed method is verified in experiments using standard gray images, all sized 512×512. In the Slepian-Wolf encoding and decoding, we use the same LDPC parity-check matrix $H$ as used in [27] with $q=0.1$, $n=6336$ and $r=3840$. Error probability of the virtual channel $q$ is obtained by applying image estimation using a database containing 5000 natural images of various types. We find $q=0.1$ is large enough for all these images, which guarantees error-free LDPC decoding. Peak signal-to-noise ratio (PSNR) is used to evaluate recovered image quality.

A. Data Hiding/Extraction and Image Recovery

Fig. 5 illustrates a group of experimental results with Lena. After encrypting the original image Fig. 5(a) into Fig. 5(b) using the stream cipher, 196,608 bits of Fig. 5(b) are selected, with $L=3MN/4$ and $n=1$, and divided into 31 groups. Each group is encoded with LDPC to generate the corresponding
syndrome, and 77,376 additional bits are embedded into the encrypted image. Fig. 5(c) shows the resulting encrypted image containing secret bits. The embedding rate is 0.2952 bpp.

On the receiver side, 77,376 secret bits can be completely extracted without error when the embedding key is known. With the encryption key only, an approximate image shown in Fig. 5(d) is obtained using the estimation algorithm in Eq. (12), with PSNR= 63.1dB. Fig. 5(e) shows differences between Fig. 5(a) and Fig. 5(d), with dark dots indicating locations where the pixels in (d) differs from the corresponding ones in (a), and white area meaning no difference. With both the encryption and embedding keys, the original image is perfectly recovered as shown in Fig. 5(f), which is identical to Fig. 5(a).

B. Comparison with VRAE Methods

Because no operation is required before image encryption on the sender side, the proposed method is a kind of VRAE. We compare the embedding payload and the approximate image quality of the method with some existing VRAE methods, including [7], [8], [9], [10] and [12].

With the proposed method, large embedding payload can be achieved. Table II shows the maximal embedding payloads of the VRAE methods. For all tested images sized 512×512, at most 77376 bits can be embedded using the proposed scheme, and the embedding rate is 0.2952 bpp. With [7], [8], [9], [10] and [12], the maximal embedding rates are merely 0.0625, 0.0039, 0.0039, 0.033, and 0.1776, respectively, indicating that payload of the proposed method is much higher than the other five methods. For fair comparison, all six methods are realized under the condition that the original image be perfectly recovered and the embedded bits extracted without error on the receiver side, and we have used the parameters defined in all methods that achieve the highest embedding payload. High payload of the proposed method is attributed to the combination of DSC and estimation. In [10], three LSBs are compressed using a linear transform to accommodate the secret message. In this case, compression is not sufficient, resulting in a low payload.

With the encryption key only, an approximate image can be reconstructed with high quality. Fig. 6 shows curves of PSNR of the approximate images at different embedding rates. Because the pixels containing embedded bits cannot be identified without the embedding key, the receiver assumes that all 3MN/4 bits of the MSB planes in the last three sub-images are modified. With the proposed method, quality of the approximate image keeps unchanged for all embedding rates. It is observed that the proposed approach outperforms the state-of-the-art RDH algorithms in most cases. In the five traditional VRAE methods, the receiver directly decrypts the received image to reconstruct an approximate image in case that the receiver has the encryption key only. In this case, the recovered quality falls steeply with the increase of embedding rate. The present work uses an estimation mechanism to guarantee a high-quality reconstructed image. Compared with the other methods, which are not compatible with quality-improvement mechanism, the PSNR values are better.

C. Comparison with VRBE Methods

Comparisons of the quality of the approximate images with the VRBE methods [11] and [13] are shown in Fig. 7. The approximate image quality of the proposed method is better at relatively high embedding rates than [11]. Because [11] vacates the embedding room before image encryption, the embedding rate can reach 0.5 bpp, which is higher than the proposed method. While the approximate image quality of [13] is better in most cases, the maximal embedding rate is much less than the proposed method. It should be noted that the comparisons are based on the finally generated approximate images, because an estimation algorithm is required to be used in the proposed method, while no estimation algorithm can be used in the other methods.

D. Maximal Embedding Rate vs. Value of q

In the proposed method, we use fixed error probability q for the virtual channel between the source and the side information. The empirically generated constant q=0.1 can ensure perfect image recovery in general. However, in some cases, different q values may also be used. For example, the data-hider may have prior knowledge that the encrypted images are of the same type, e.g., X-ray images; or a feedback channel between the data-hider and the receiver for adaptive embedding may be available. In this case, a higher embedding rate can be achieved using a smaller q. In the test, we arbitrarily choose 500 different images from the UCID database [28], and turn all images into grayscale. With the proposed estimation algorithm, values of error probability q are shown in Fig. 8. The embedding rates corresponding to different q are shown in Fig. 9, in which R_E is the actual embedding rate and R_C the bound of the embedding rate. The curves show that the embedding rate increases when q gets smaller. Table III presents maximal embedding payload for some images in error-free decoding. Because we use a precise q value in this case, payloads higher than 0.2952 bpp (with q=0.1) are achieved.

If the data-hider chose an inappropriate q smaller than the estimated error probability, there would be no influences on the data extraction. In this case, content of the original image cannot be perfectly recovered since DSC decoding would fail. However, an approximate image can still be generated with good quality, if the receiver has both the embedding and encryption keys. After identifying which bits on the MSB planes of the three sub-images were selected by the data-hider using the embedding key, the receiver decrypts the image with the encryption key, and approximately recovers these bits using the proposed estimation algorithm in Eq. (12). Fig. 10 shows that quality of the recovered images keeps good albeit it declines when the embedding rate increases.

VII. DISCUSSION

The proposed method belongs to the VRAE category. Purpose of the present work is to improve the embedding payload in comparison with the other VRAE methods. To this end, the MSB plane is used as the cover to accommodate additional bits. As the error probability q of estimating MSB plane is much less than estimating the other planes, a much larger embedding payload can be achieved using DSC. During image reconstruction, although some noise would occur when directly decrypting the marked encrypted image, a provided estimation mechanism can be used to eliminate the noise. The
proposed method is different from the method of [10]. In [10], some encrypted LSB planes are compressed using a linear transform to vacate embedding room, and the original image can be reconstructed by estimating the LSB planes. However, efficiency of compression and LSB plane estimation is insufficient, resulting in a low embedding payload. To increase the payload, we have used the MSB plane and the LDPC based DSC.

As to the embedding payload, VRBE substantially outperforms the previous VRAE method because VRBE has the advantage of handling the plaintext image by the sender. When comparing with VRBE methods, embedding payload of the proposed method is not as large as that of [11], which in fact transforms the problem of RDH in encrypted image into the traditional RDH in the plaintext image. The flaws of [11] is that the sender is partly involved in the embedding computation, supposing he/she has the prior knowledge that a data embedding is to be done by the server. This might not be practical in case the sender has no idea about the possible data hiding, or has no computation capability other than image encryption. On the other hand, if the sender can do RDH himself, all tasks can be done on the sender side and the data-hider becomes redundant. Instead, embedding of the proposed method is entirely realized in the encrypted domain, and a high payload is achieved with the association of DSC decoding computation on the receiver side.

Two aspects of data security need to be considered: security of the image content and security of the additional message. The content owner does not allow the service provider to access the image content, and the data-hider does not allow adversaries to crack the embedded message. For the content owner, the original image is encrypted with a stream cipher algorithm using an encryption key $K_{ENC}$. For the data-hider, the additional bits are also encrypted with the stream cipher using another key $K_{SC}$. Many stream cipher algorithms can be used, such as RC4 or SEAL. It is infeasible for the adversary to execute an exhaustive search if appropriate keys are used, e.g., the seed keys no less than 128 bits in RC4, or no less than 160 bits in SEAL. Thus, security of the image content and the additional message can be guaranteed.

The extraction security can also be guaranteed. On the data-hider side, $L$ MSB bits are selected from $3MN/4$ pixels using a selection key $K_{SL}$, and shuffled by a shuffle key $K_{SF}$. After compressing the shuffled $L$ bits using DSC and appending the encrypted additional bits to the end of the compressed bits, a further reshuffle is done with the key $K_{SF}$ to generate the marked $L$ bits that are put into the original MSB positions. For an adversary without both the embedding and encryption keys, the success rate of correctly extracting the hidden bits is as low as $1/(C_{3MN/4}^L \cdot L!)$. Supposing the adversary has the encryption key, after directly decrypting the marked encrypted image, the amount of pixels containing noises is about $L/2$. With the only $L/2$ MSB bits, the success rate of correctly extracting the additional bits is as low as $1/(C_{3MN/4-L/2}^{L/2} \cdot L!)$. Therefore, exhaustive searching is virtually impossible for the adversaries to extract the additional message provided $M$, $N$ and $L$ are not too small.

Data extraction and image recovery is separable in the proposed method. Solutions to three different cases on the receiver side are provided: with the encryption key only, with the embedding key only, and with both the encryption and embedding keys. There is yet another special situation in which the receiver obtains the encryption key first and unexpectedly receives the embedding key much later. In the previous VRAE methods, when the receiver acquires the embedding key later, he/she still cannot extract the hidden message from the decrypted image directly, unless the marked encrypted version has been saved (this is likely the case) and is used, or extra re-encryption is carried out using the encryption key to regenerate a marked encrypted copy. In the proposed method, the receiver can save the MSB bits of the marked encrypted image before estimating the original image. After the embedding key later is acquired later, he/she can extract the additional bits from the save bits.

VIII. CONCLUSION

This paper proposes a scheme of reversible data hiding in encrypted images using distributed source coding. After encrypting the original image with a stream cipher, some bits of MSB planes are selected and compressed to make room for the additional secret data. On the receiver side, all hidden data can be extracted with the embedding key only, and the original image approximately recovered with high quality using the encryption key only. When both the embedding and encryption keys are available to the receiver, the hidden data can be extracted completely and the original image recovered perfectly.

With the idea of DSC, the proposed method substantially increases the payload as compared with the existing VRAE methods. An LDPC parity-check matrix is used to generate corresponding syndromes. Associated with the estimated image generated from the proposed estimation algorithm, the receiver can decode these syndromes back to the original bits using iterative BPA decoding.

Because embedding operations are performed to the encrypted data, the data-hider cannot access the contents of the original image. That ensures security of the contents in data hiding. As the embedding and recovery are protected by the encryption and embedding keys, an adversary is unable to break into the system without these keys.

ACKNOWLEDGEMENTS

We thank the anonymous reviewers and the editors for the thorough reviews and comments, and Professor Shuozhong Wang for the constructive discussions. The work was supported by the Natural Science Foundation of China under Grant 61472235 and Grant 61103181, Shanghai Rising-Star Program under Grant 14QA1401900, Program for Professor of Special Appointment (Eastern Scholar) at Shanghai Institutions of Higher Learning, Shanghai Pujiang Program under Grant 13PJ1403200, and Shanghai Key Laboratory of Intelligent Information Processing under Grant No. IJPL-2014-006.
REFERENCES


Zhenxing Qian received both the B.S. degree and the Ph.D. degree from University of Science and Technology of China (USTC), in 2003 and 2007, respectively. He is currently an associate professor in the School of Communication and Information Engineering, Shanghai University, China. His research interests include data hiding, image processing, and digital forensics.

Xinpeng Zhang received the B.S. degree in computational mathematics from Jilin University, China, in 1995, and the M.E. and Ph.D. degrees in communication and information system from Shanghai University, China, in 2001 and 2004, respectively. Since 2004, he has been with the faculty of the School of Communication and Information Engineering, Shanghai University, where he is currently a Professor. His research interests include information hiding, image processing and digital forensics.
Fig. 1 Sketch of the proposed system
Fig. 5 Reversible data hiding in Lena: (a) the original image, (b) the encrypted image, (c) encrypted image containing secret data, (d) the recovered image, (e) the difference between (a) and (d), white: no difference, black: difference; and (f) the perfectly recovered image.
Table II. Embedding Payload Comparison, Amount of Secret Data (Bits), Embedding Rate (bpp)  

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Lena</td>
<td>16,384, 0.0625</td>
<td>1,024; 0.0039</td>
<td>1,024; 0.0039</td>
<td>8,650; 0.0330</td>
<td>14,284, 0.0545</td>
<td>77,376; 0.2952</td>
</tr>
<tr>
<td>Man</td>
<td>16,384, 0.0625</td>
<td>625; 0.0024</td>
<td>1,024; 0.0039</td>
<td>6,554; 0.0250</td>
<td>46,557, 0.1776</td>
<td>77,376; 0.2952</td>
</tr>
<tr>
<td>Lake</td>
<td>16,384, 0.0625</td>
<td>625; 0.0024</td>
<td>1,024; 0.0039</td>
<td>3,408; 0.0130</td>
<td>14,828, 0.0566</td>
<td>77,376; 0.2952</td>
</tr>
<tr>
<td>Baboon</td>
<td>16,384, 0.0625</td>
<td>256; 0.0010</td>
<td>334; 0.0013</td>
<td>1,966; 0.0076</td>
<td>13,056, 0.0498</td>
<td>77,376; 0.2952</td>
</tr>
</tbody>
</table>

**Fig. 6** Comparisons of four different approaches using the images.
Fig. 7 Comparison of approximately recovered image quality with [11] and [13]

Fig. 8 The error probability $q$ of 500 images.

Table III. Maximal embedding payload for different images

<table>
<thead>
<tr>
<th>Image Index</th>
<th>ucid00001</th>
<th>ucid00022</th>
<th>ucid00044</th>
<th>ucid00326</th>
<th>ucid00409</th>
</tr>
</thead>
<tbody>
<tr>
<td>$q$</td>
<td>0.017</td>
<td>0.036</td>
<td>0.001</td>
<td>0.042</td>
<td>0.064</td>
</tr>
<tr>
<td>Maximal $R_E$</td>
<td>0.54</td>
<td>0.47</td>
<td>0.57</td>
<td>0.45</td>
<td>0.38</td>
</tr>
</tbody>
</table>
Fig. 9 The actual embedding rate $R_E$ and the bound $R_C$ corresponding to different $q$.

Fig. 10 Quality of approximately recovered images when Slepian-Wolf decoding fails.