Privacy-aware reversible watermarking in cloud computing environments

C C Chang, Chang-Tsun Li, Y Q Shi

01-01-2018

10536/DRO/DU:30119667

IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received September 5, 2018, accepted October 19, 2018, date of publication November 14, 2018,
date of current version December 18, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2880904

Privacy-Aware Reversible Watermarking in Cloud Computing Environments

CHING-CHUN CHANG[1], (Student Member, IEEE), CHANG-TSUN LI[2], (Senior Member, IEEE), AND YUN-QING SHI[3], (Life Fellow, IEEE)
[1]Department of Computer Science, University of Warwick, Coventry CV4 7AL, U.K.
[2]School of Computing and Mathematics, Charles Sturt University, Wagga Wagga, NSW 2678, Australia
[3]Department of Electrical and Computer Engineering, New Jersey Institute of Technology, Newark, NJ 07102, USA

Corresponding author: Ching-Chun Chang (ching-chun.chang@warwick.ac.uk)

This work was supported by the Marie Skłodowska-Curie actions of the EU Horizon 2020 programme through the project entitled Computer Vision Enabled Multimedia Forensics and People Identification (Project No. 690907, Acronym: IDENTITY).

ABSTRACT As an interdisciplinary research between watermarking and cryptography, privacy-aware reversible watermarking permits a party to entrust the task of embedding watermarks to a cloud service provider without compromising information privacy. The early development of schemes was primarily based upon traditional symmetric-key cryptosystems, which involve an extra implementation cost of key exchange. Although recent research attentions were drawn to schemes compatible with asymmetric-key cryptosystems, there were notable limitations in the practical aspects. In particular, the host signal must either be enciphered in a redundant way or be pre-processed prior to encryption, which would largely limit the storage efficiency and scheme universality. To relax the restrictions, we propose a novel research paradigm and devise different schemes compatible with different homomorphic cryptosystems. In the proposed schemes, the encoding function is recognised as an operation of adding noise, whereas the decoding function is perceived as a corresponding denoising process. Both online and offline content-adaptive predictors are developed to assist watermark decoding for various operational requirements. A three-way tradeoff between the capacity, fidelity, and reversibility is analysed mathematically and empirically. It is shown that the proposed schemes achieve the state-of-the-art performance.

INDEX TERMS Cloud computing, cyber security, homomorphic cryptosystems, information privacy, reversible watermarking, signal denoising, statistical inference, variational method.

I. INTRODUCTION

The past decades have witnessed the phenomenal prevalence of public cloud services. The seemingly unlimited storage and computational capacity of the cloud have opened up opportunities for businesses and individuals to entrust their data to cloud service providers. Meanwhile, the rapid development of cloud computing technology has also raised privacy and security concerns. A powerful solution against cyber security threats is to obfuscate the classified documents through encryption [1]. On the other hand, although encryption algorithms are widely used to protect the confidential information, some desirable functionalities might be unrealisable for the encrypted data. In response to this issue, there was a surge of research interest in secure signal processing in the encrypted domain [2]–[4]. As an emerging topic in this research field, privacy-aware watermarking has attracted a lot of attention in recent years. For the reason that many watermarking algorithms are proprietary properties, there are restrictions for commercial purposes and the availability to the general public is rather limited. An efficient solution is to request an authorised cloud service provider to encode the watermark into the given digital media, not only due to the high computational capacity of the cloud, but also the legitimate consent to carry out the algorithms. On the other hand, since the cloud server is regarded as a semi-honest entity that may collect some information from the digital contents, data privacy should be taken into account. Therefore, the research of privacy-aware watermarking addresses the problem of entrusting the task of watermarking to a cloud server without compromising data privacy. This research trend is also known as watermarking in the encrypted domain.

Watermarking is the practice of imperceptibly embedding the watermark into the host media (*e.g.* audio, image and video), where the watermark is usually a piece of information associated with the host media (*e.g.* digital identity or signature). In general, there are three parties involved in a

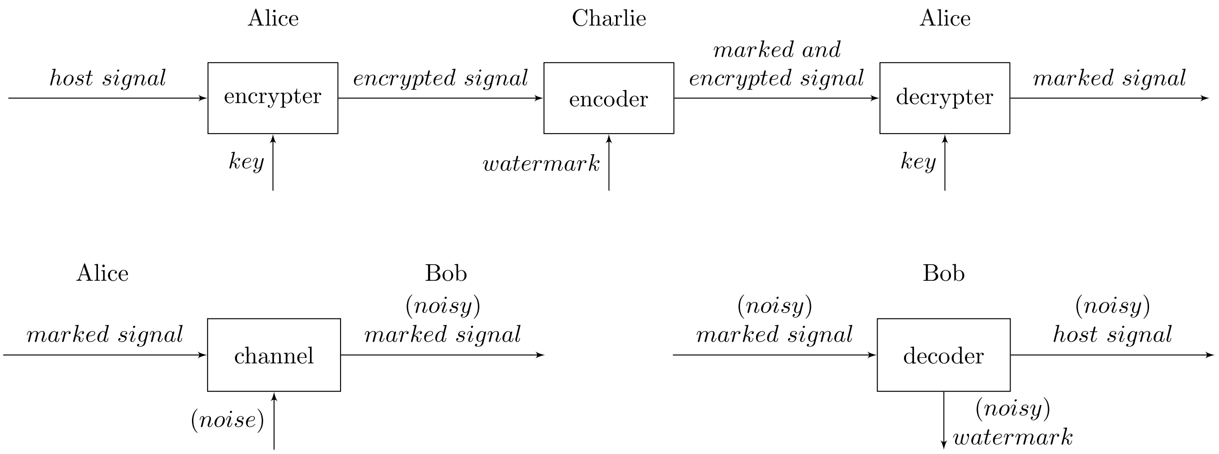


FIGURE 1. The proposed watermarking protocol. A transmitter Alice wants to protect a signal through embedding a watermark, which can be fulfilled by a cloud service provider Charlie. Then, a marked signal is transmitted over a insecure channel to a recipient Bob who is able to verify the authenticity of the received signal through analysing the extracted watermark and reproduce the original signal.

watermarking protocol: a sender who encodes the watermark into the host media, a recipient who decodes the watermark from the marked media, and an adversary who has a malicious intent against the protocol. The malicious party is often modelled as a noisy channel between the encoder and the decoder. The watermarking strategy can be either fragile or robust against the channel noise depending on the applications. Conventionally, robust watermarking is applicable to copyright protection [5]–[8], whereas fragile watermarking is advantageous to data authentication [9]–[12]. Nevertheless, fragile watermarking, as a temper detection technique, contradict itself by inevitably inflicting distortions upon the content of media. Although the distortions are generally imperceptible, the effect could be arguable in particular circumstances featuring a strict integrity requirement, especially when the host media is used for military reconnaissance or medical diagnosis. Considering these potential applications, the notion of reversible watermarking was introduced in order to permit the reproduction of the original content once the authenticity of media is verified [13]–[20].

Consider a watermarking protocol as illustrated in Fig. 1, where a sender Alice wants to communicate a digital content to a receiver Bob over an insecure channel. In order to resist malicious tampering, Alice has to embed a signature, which is sensitive against manipulations, into the content prior to dispatching it to Bob. Supposing that Alice, as a general individual, has no permission or resources to employ the watermarking algorithm, the task of watermarking has therefore to be entrusted to a licensed cloud server Charlie. As a general presumption, Charlie is an honest-but-curious party who would not deviate from the protocol specifications, though on the other hand has the interest in learning the privacy information of the content. Hence, due to privacy concerns, Alice encrypts the content prior to commissioning it to Charlie for watermarking. In summary, to utilise the complementary advantages of watermarking and cryptography, it is of great significance to develop a secure reversible watermarking system that enables the watermarking function to be operated by a licensed third party in the encrypted domain in order to preserve privacy.

The challenge of reversible watermarking in the encrypted domain is a rather difficult one. Considering that the message is concealed by encryption, we are not able to observe, analyse, and exploit data redundancy. One of the very first schemes utilises the electronic codebook (ECB) mode of block ciphers to encipher the data, which preserves the data redundancy by reusing the secret key for each block while compromising the security strength [21]. To overcome the security weakness, the researchers have developed schemes for a variety of secure cryptosystems. For instance, an attempt was made by encoding watermarks into data encrypted by the cipher block chaining (CBC) mode [22]. In a similar manner, some methods were proposed to encode watermarks into data encrypted by steam cipher [23]–[25]. In essence, these early schemes were based upon the research outcomes of lossless compression of the encrypted data [26]–[28]. Apart from this, the early development of schemes was primarily compatible with symmetric-key cryptosystems [29]–[31]. Accordingly, these schemes are confronted with a practical issue that a secret key must be pre-shared between the sender and the recipient. In order to communicate the secret key, a secure channel resistant to eavesdropping or a secure key exchange protocol must be established. Any of these solutions comes with extra costs in implementation.

In contrast to symmetric-key algorithms, asymmetric-key algorithms, also known as public-key cryptosystems, relax the restriction by using a publicly distributed key for encryption and a secret key known only to the recipient for decryption. Over the last few years, the watermarking research has extended to manage ciphers generated by public-key algorithms. The first class of strategy encrypts the data in such a manner that the input to the encryption function has a value range that is far smaller than it should be. Consider that each individual pixel is an input to be enciphered. The value range of pixels is between 0 and 255 for 8-bit greyscale images, while the message space of a given cryptosystem is, for example, between 0 and N, where N is the product of two large primes. Therefore, the input space defined by the watermarking shceme is much smaller than the actual message space defined by the cryptosystem. In implementation, a number of zeros are concatenated to the end of the input message to assure the length of input conform with message space. As a consequence, zero paddings can be viewed as additional redundancy for accommodating the payload in the encrypted domain [32]–[34]. It is nevertheless deficient for the following reasons. First, redundant bits are appended and thus the system is not space-efficient. Second, from some viewpoints it is considered problematic since the amount of padded redundancy is at least equivalent to that of the payload in most cases. One may even simply append the payload information after the host signal rather than performing complex computation to trade the pre-appended redundancy for the payload. Third, it is insecure since the watermark can be easily filtered out by decryption. To overcome these deficiencies, we suggest to encrypt a bit-stream of sufficient length each time, instead a single sample with zero-padding.

The second class of strategy, though following the space-saving principle, requires a specific type of pre-processing to be applied prior to encryption. It utilises either truncation [35] or self-embedding [36] to shrink the range of sample values and then encrypts a sufficient number of samples each time. As a result, although the input space seems to be equal to the message space defined by the cryptosystem, the actual input space is much smaller. For instance, initially the sample values range from 0 to 255 and yet subsequent to data pre-processing the possible values reduce into a subrange. The reserved space can then be exploited to accommodate the payload in the encrypted domain. This class of strategy, also known as reserving room before encryption, was first developed for watermarking schemes to work in conjunction with symmetric-key cryptosystems [37]–[39]. The practicality of this strategy is however limited because a specific type of pre-processing must be applied to the data. To remove this restriction, we suggest to manage data without any specific pre-processing.

In this paper, we address the problem of entrusting the task of watermarking to a cloud service provider without compromising information privacy. Considering the security threats of symmetric-key cryptosystems, we study reversible watermarking schemes compatible with public-key cryptosystems. The recent development of schemes requires the host data either to be enciphered in a redundant fashion or to be pre-processed prior to encryption. In response to this, we propose a novel research paradigm to overcome the aforementioned shortcomings. Two schemes are proposed to cope with ciphers generated by two types of homomorphic cryptosystems, respectively. The first scheme is constructed based upon multiplicative homomorphism and provides high content fidelity as well as high flexibility in watermarking capacity. The second scheme is devised based upon additive homomorphism and achieves an optimal reversibility. In addition to this, both online and offline content-adaptive predictors are introduced to assist watermark decoding. The former type of predictor is based upon variational method and utilises an iterative algorithm to approximate the signal, while the second type of predictor is based upon statistical inference and pre-trains a probability table for signal estimation. Experimental results shows that the proposed predictors achieve the state-of-the-art performance. In summary, several positive contributions and improvements are made, which are briefly outlined as follows:

- Modern public-key cryptosystems are adopted to avoid the security risks and implementation costs of key exchange.
- Reversible watermarking schemes compatible with different types of partially homomorphic cryptosystems are studied.
- Storage efficiency is considered by encrypting a sufficiently long sequence of bits, instead of a single sample with redundant zero padding, each time.
- Specific type of data pre-processing prior to encryption is not required, which enhances the practicality and universality.
- Both online and offline content-adaptive predictors are developed with flexibility for various operational requirements.

The remainder of this paper is organised as follows. Section II discusses various cryptosystems and how to apply them on digital images. Section III introduces the proposed watermarking scheme for multiplicative homomorphic cryptosystem. Section IV further derives a scheme for additive homomorphic cryptosystem. Section V studies content-adaptive predictors based upon variational method and statistical inference. Section VI evaluates the scheme performance and makes comparisons with state-of-the-art algorithms. Section VII concludes our work and outlines the directions for future research.

II. PRELIMINARIES

The notion of privacy homomorphisms was introduced by Rivest *et al.* [40], which was described as particular algebraic mappings between the paintext and ciphertext spaces that allow the result of operations upon the ciphertexts, when deciphered, to match the result of operations upon the plaintexts. A well-understood example of multiplicative homomorphism would be the RSA cryptosystem whose security strength is known to be equivalent to the difficulty of solving integer factorisation [41]. Suppose that p and q are two large primes and the modulus is computed as $N = p \cdot q$. Let e and d be the public and private keys of the RSA cryptosystem, respectively, such that e and d satisfy the condition that

$$e \cdot d \equiv 1 \pmod{\phi(N)}, \tag{1}$$

where ϕ is Euler's phi function, *i.e.* $\phi(N) = (p-1)(q-1)$. The RSA cryptosystem defines an encryption function

$$c \equiv m^e \pmod{N}, \tag{2}$$

and a decryption function

$$m \equiv c^d \pmod{N}, \tag{3}$$

where m denotes the message and c denotes the cipher. Consider the goal of generating the encrypted result which, when decrypted, matches the product of two messages m_1 and m_2 through the operations upon the ciphers c_1 and c_2. This goal can be achieved by multiplying two ciphers as

$$\begin{aligned} c_1 \cdot c_2 &\equiv ({m_1}^e) \cdot ({m_2}^e) \pmod{N} \\ &\equiv (m_1 \cdot m_2)^e \pmod{N}. \end{aligned} \tag{4}$$

In other words,

$$\mathcal{D}(\mathcal{E}(m_1) \cdot \mathcal{E}(m_2)) = m_1 \cdot m_2, \tag{5}$$

where $\mathcal{E}(\cdot)$ and $\mathcal{D}(\cdot)$ denote the encryption and decryption functions, respectively.

The RSA cryptosystem is limited, however, by solely permitting the multiplication of messages. As a consequence, to design homomorphic encryption algorithms that permit various mathematical operations upon ciphers has been one of the active research areas, within which there are substantial achievements such as ElGamal [42], Okamoto and Uchiyama [43], and Damgård and Jurik [44] cryptosystems. Among a variety of historical antecedents, the Paillier cryptosystem has been widely used as an example of additive homomorphism [45]. The system consists of three phases: key generation, encryption, and decryption. In the key generation phase, we choose two large primes p and q. Then, we compute $N = pq$ and $\lambda = \mathrm{lcm}(p-1, q-1)$, where 'lcm' stands for least common multiple. Next, we select a random integer $g \in \mathbb{Z}/N^2\mathbb{Z}^*$ and calculate

$$\mu \equiv (L(g^\lambda \pmod{N^2}))^{-1} \pmod{N}, \tag{6}$$

where

$$L(x) = \frac{x-1}{N}. \tag{7}$$

As a result, the public key is (n, g) and the private key is (λ, μ). In the encryption phase, let m be a message to be encrypted and r be a randomly selected integer, where $m, r \in \mathbb{Z}/N\mathbb{Z}$. The ciphertext is then computed as

$$c \equiv g^m \cdot r^N \pmod{N^2}. \tag{8}$$

In the decryption phase, the plaintext message is deciphered by

$$m \equiv L(c^\lambda \pmod{N^2}) \cdot \mu \pmod{N}. \tag{9}$$

Let m_1 and m_2 be two messages and c_1 and c_2 be two ciphers. To produce the cipher of sum of m_1 and m_2, we calculate the product of c_1 and c_2 and obtain

$$\begin{aligned} c_1 \cdot c_2 &\equiv (g^{m_1} \cdot {r_1}^N) \cdot (g^{m_2} \cdot {r_2}^N) \pmod{N^2} \\ &\equiv g^{(m_1+m_2)} \cdot (r_1 \cdot r_2)^N \pmod{N^2}. \end{aligned} \tag{10}$$

To yield the cipher of product of m_1 and m_2, we compute the exponentiation of c_1 by m_2 such that

$$\begin{aligned} c_1^{m_2} &\equiv (g^{m_1} \cdot {r_1}^N)^{m_2} \pmod{N^2} \\ &\equiv g^{(m_1 \cdot m_2)} \cdot (r_1^{m_2})^N \pmod{N^2}. \end{aligned} \tag{11}$$

As an alternative expression, we write

$$\mathcal{D}(\mathcal{E}(m_1) \cdot \mathcal{E}(m_2)) = m_1 + m_2, \tag{12}$$

and

$$\mathcal{D}(\mathcal{E}(m_1)^{m_2}) = m_1 \cdot m_2. \tag{13}$$

It can be observed that the size of encrypted data is expanded as the message space is $\mathcal{M} = \mathbb{Z}/N\mathbb{Z}$ and the ciphertext space is $\mathcal{C} = \mathbb{Z}/N^2\mathbb{Z}^*$.

As aforementioned, when encrypting a digital image, it is not advisable to consider each pixel as an individual message. Generally, the value range of pixels is far smaller than the message space, as the latter is associated with large prime numbers, which provokes dispensable redundancy. Apart from this, if the watermark is embedded into the addition redundancy created by encryption, it will be filtered out by decryption, which limits the applicability of the schemes. There are several possible approaches to image encryption without expanding image files, for example, converting the image into bit-stream and encrypting a segment of sufficient length each time. In order to preserve the visual significance, we propose to encipher each bit-plane separately. Let $\mathbf{I}$ denote an 8-bit digital image such that

$$\mathbf{I} = (\mathbf{b}_1 || \mathbf{b}_2 || \ldots || \mathbf{b}_8), \tag{14}$$

where '||' is the concatenation operator and $\mathbf{b}_1$, $\mathbf{b}_2$, …, and $\mathbf{b}_8$ are eight different bit-planes. For each bit-plane

$$\mathbf{b}_i = (b_{i,1} || b_{i,2} || \ldots), \tag{15}$$

we sample a sufficiently long bit-stream each time to form a message, which is then encrypted into a cipher. Supposing that the message space of a given cryptosystem is $\mathbb{Z}/N\mathbb{Z}$, a message is therefore a decimal number of $\log_2 N$ bits, or of $\lfloor \log_2 N \rfloor$ bits more precisely, considering that the former is not necessarily an integer.

The watermarking process is performed sequentially upon the selected messages until all the payload is embedded, while the unselected ones remain intact. Due to the fidelity constraint, we only change the messages converted from insignificant bit-planes. Intuitively, it seems that the optimal choice should be to select messages of the least significant bit-plane. However, by considering the decoding process, it is necessary that the selected messages are formed by some well-predictable bits. We shall see the reason behind this later. To pave the way for the following presentation, let us assume that the messages for watermarking have been determined. The map to record the locations of selected messages serves as the watermarking key. In the following paragraphs, we will discuss separately how we cope with ciphers generated by different types of cryptosystems.

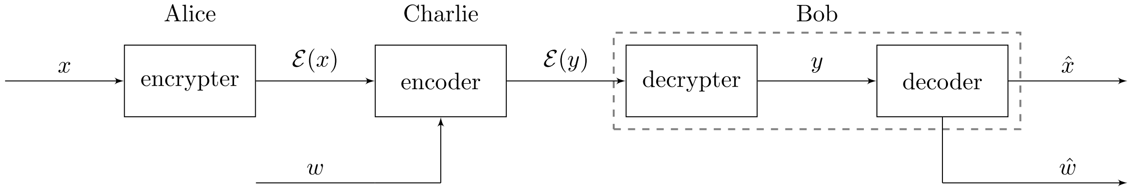


FIGURE 2. Watermarking procedures for a selected symbol. Let x be a host symbol and w be a watermark symbol. Alice encrypts x into $\mathcal{E}(x)$. Charlie encodes $\mathcal{E}(x)$ with w into $\mathcal{E}(y)$. Bob decrypts $\mathcal{E}(y)$ into y and then decodes it into $\hat{x}$ and $\hat{w}$.

III. MULTIPLICATIVE HOMOMORPHISM

Let us recapitulate the problem statement. Let Alice denote a sender, Bob a recipient, and Charlie a cloud server. Suppose that Alice wants to deliver to Bob a digital file in which a watermark is embedded for authentication purposes. Due to various constraints (*e.g.* proprietary issues), Alice has to entrust the task of watermarking to Charlie by providing the encrypted file and the watermark. The encryption key is publicly known, while the decryption key is only known to Bob. We assume that the watermark payload is a sequence of compressed and encrypted digits such that only the intended recipient Bob is able to decode it.

Consider the objective of embedding watermarks into ciphers generated by a multiplicative homomorphic cryptosystem such as RSA. The preliminary procedures are as follows. To begin with, the host data is divided into binary sequences of the length in accordance with the message space of the given cryptosystem. Then, each sequence is transformed into an integer called a symbol. After encryption, a sufficient number of enciphered symbols are selected to carry the watermark payload. We suppose that the selected symbols are all composed of bits in the l-th bit-plane. To meet the fidelity requirement, the change of bits in the l-th bit-plane must not provoke a significant degradation to the visual quality of the image. For conciseness, we describe only how a single host symbol is processed, as shown in Fig. 2.

Let p and q be two large primes and $N = p \cdot q$. The alphabet of watermark symbols is defined as the set of k distinct positive integers up to N that are coprime to N. Accordingly, we are able to embed $\log_2 k$ bits of watermark information per host symbol. The number of positive integers up to N that are relatively prime to N can be calculated by Euler's phi function $\phi(N) = (p-1)(q-1)$ and thus $k \leq \phi(N)$. This alphabet, denoted by $\mathcal{W} = \{w_i\}_{i=0}^{k-1}$, is pre-shared between Alice and Bob and serves as a codebook so that different payloads can be represented by different watermark symbols. Let a host symbol be $x \in \mathbb{Z}/N\mathbb{Z}$ and a watermark symbol be $w \in \mathcal{W}$. To begin with, Alice encrypts x into $\mathcal{E}(x)$ and uploads it along with $\mathcal{E}(w)$ to Charlie. The watermarking process is then operated by

$$\mathcal{E}(y) \equiv \mathcal{E}(x) \cdot \mathcal{E}(w) \equiv \mathcal{E}(w \cdot x) \pmod{N}. \tag{16}$$

After that, the marked and encrypted symbol $\mathcal{E}(y)$ is downloaded and sent to Bob and is then decrypted into

$$y \equiv \mathcal{D}(\mathcal{E}(y)) \equiv w \cdot x \pmod{N}. \tag{17}$$

Note that we strictly let $w_0 = 1$ in practice in order to minimise the average distortion. That is, when $w = w_0$, the marked symbol will be kept intact. The distortion only occurs when $w = w_i, \forall i \neq 0$, is embedded. We know that there exists a unique modular multiplicative inverse of an arbitrary integer a modulo N if and only if $\gcd(a, N) = 1$. Since $\mathcal{W}$ is a subset of all integers that are coprime to N, we can construct the set $\overline{\mathcal{W}} = \{w_i^{-1}\}_{i=0}^{k-1}$ that comprises of the corresponding unique inverses. We generate k possible candidates for x in such a way that each is given by

$$x_i \equiv y \cdot w_i^{-1} \pmod{N}. \tag{18}$$

The above congruence can be rewritten as

$$x_i \equiv x \cdot w \cdot w_i^{-1} \pmod{N}. \tag{19}$$

Thus, we see that $x_i = x$ if and only if $w_i^{-1} = w^{-1}$. In other words, if we are able to distinguish x from k possible candidates $\{x_i\}_{i=0}^{k-1}$, then we can determine w jointly. Recall that a selected symbol is composed of bits that can be estimated with a certain degree of accuracy from some other correlated bits. Let us denote an estimated symbol for x by $\tilde{x}$. Therefore, the original x is determined by

$$\hat{x} = \arg\min_{x_i} \Delta(x_i, \tilde{x}), \tag{20}$$

where Δ is a general distortion metric (*e.g.* Hamming distance). The reversibility, namely the ability to recover the host media, primarily depends on two factors. One is the number of candidates and the other is the estimation accuracy. The reason is straightforward. Given only a few possibilities and a highly credible clue, there is a high probability that the answer is correctly deduced. The number of candidates is governed by the capacity parameter k, whereas the estimation accuracy is related to the fidelity parameter l. Hence, we summarise that

$$\begin{aligned} \text{Capacity} &\propto k, \\ \text{Fidelity} &\propto l^{-1}, \\ \text{Reversibility} &\propto l \cdot k^{-1}. \end{aligned} \tag{21}$$

Example: Let us demonstrate the scheme with an example as follows. Assume that $p = 11$, $q = 19$, and $N = 209$. Let a payload bit be encoded by

$$w = \begin{cases} w_0, & \textit{if the bit is } 0, \\ w_1, & \textit{if the bit is } 1, \end{cases}$$

where w_0 and w_1 can be any two numbers coprime to 209. Note that w_0 is strictly set to one in practice and we let

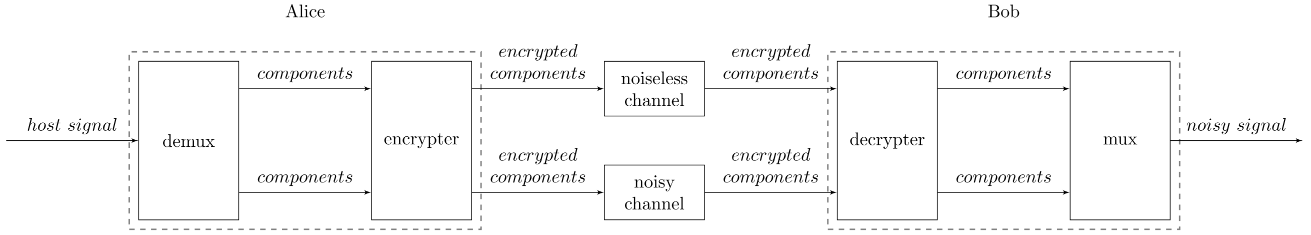


FIGURE 3. Watermarking as noise adding. Alice demultiplexes and encrypts the host signal into encrypted components and then sends each component over either a noiseless or noisy channel. Bob decrypts and multiplexes the received components into a noisy signal.

$w_1 = 3$ in this example. Suppose that Alice has a host symbol $x = 100$ and wants to embed a bit one. As a result, the watermark symbol is given by $w = w_1$. By encryption, Alice outputs $\mathcal{E}(x) = \mathcal{E}(100)$. After watermarking, Charlie produces $\mathcal{E}(y) \equiv \mathcal{E}(3) \cdot \mathcal{E}(100) \bmod 209$. By decryption, Bob obtains $y = 91$, which is computed by

$$91 \equiv 3 \cdot 100 \bmod 209.$$

The modular multiplicative inverses of w_0 and w_1 are $w_0^{-1} = 1$ and $w_1^{-1} = 70$, respectively. Hence, two candidates for x are $x_0 = 91$ and $x_1 = 100$, which are computed by

$$91 \equiv 91 \cdot 1 \bmod 209,$$
$$100 \equiv 91 \cdot 70 \bmod 209,$$

respectively. Suppose that the approximation of x is $\tilde{x} = 102$. We determine the value of x as that of the closest candidate to $\tilde{x}$ in terms of the Hamming distance and the value of w as that of the one whose inverse yields the minimum Hamming distance. The distances are calculated by

$$\Delta_0 = \Delta(91, 102) = 5,$$
$$\Delta_1 = \Delta(100, 102) = 1,$$

and therefore the decoding results are $\hat{x} = 100$ and $\hat{w} = 3$.

It can be observed that the wrong candidate is, theoretically, a random integer. Thus, if N is sufficiently large and the number of candidates to distinguish from is relatively small, then it is unlikely that randomly drawn integers would be closer to the approximation of the correct one than the correct one itself.

IV. ADDITIVE HOMOMORPHISM

We have seen that the reversibility of the previous scheme is managed by the minimum distance between possible candidates and the correct one, which is random in this very case when only the multiplicative homomorphism is permitted. To achieve a better performance, let us consider embedding watermarks into ciphers produced by an additive homomorphic cryptosystem such as the Paillier cryptosystem. For this type of cryptosystem, there exists an optimal strategy to maximise the Hamming distance between candidates in a special case that only one bit is embedded per symbol and there are only two candidates to be distinguished from in the decoding process (*i.e.* $k = 2$). Certainly, there is always a trade-off. In this case, it is the non-trivial data expansion from the message space to the cipher space due to homomorphic encryption and a little compromise on the fidelity due to distance maximisation.

Before proceeding further, let us introduce another viewpoint of this research problem, as diagramed in Fig. 3. Let Alice be a sender and Bob be a recipient. There are two communication channels between the two parties: a noiseless channel and a noisy channel. Let us assume that the noiseless channel is much more expensive than the noisy one, which complies with our intuition. Suppose that a message can be decomposed into several pieces of components and each component is transmitted over one of the channels. Alice wants the communication cost to be as low as possible. Bob, on the other hand, requires that the core idea of the message must be clear and comprehensible regardless of some errors induced by the channel noise. In other words, the noisy components must not lead to the misunderstanding upon the core idea of the message. Moreover, Bob hopes that the errors can be corrected from the context given by the other correct components. Suppose that each message component is concealed in a safe envelope during the transmission. Nevertheless, the noisy channel is still capable of adding noise to the components even without opening the envelope, imagining that raindrops pass through the envelope and wet the letter.

This research problem is analogous to that of reversible watermarking in the encrypted domain in the following senses. First, the noisy channel is analogous to the watermarking function. Second, the reciprocal of the communication cost corresponds to the watermarking capacity. Third, the requirement of preserving the core idea of the message is equivalent to the fidelity constraint of watermarking. Last, the context-based error correction is akin to the decoding process that jointly detects the payload and recovers the host media. The task of error correction will be easy if the noisy channel always modifies the components to those of the opposite meaning in the presence of noise. The reason is that given a certain context, it is much easier to sense an antonym of a word than a synonym of that word. Hence, our objective is to construct a noisy channel that always maps components to their farthest counterparts in the presence of noise, that is to say,

$$\mathcal{E}(y) = \begin{cases} \mathcal{E}(x), & \text{if no noise occurs,} \\ \mathcal{E}(\bar{x}), & \text{otherwise,} \end{cases} \tag{22}$$

where $\bar{x}$ is the farthest counterpart of x. Let x be a non-negative integer composed of bits from a certain bit-plane and $\bar{x}$ be the bitwise complement of x. Although the Paillier cryptosystem does not permit homomorphic bitwise complement operation upon the encrypted data, we are still able to obtain $\mathcal{E}(\bar{x})$ by

$$\mathcal{E}(\bar{x}) \equiv \mathcal{E}(2^t - 1) \cdot \mathcal{E}(x)^{-1} \equiv \mathcal{E}(2^t - 1 - x) \bmod N^2, \quad (23)$$

where $t = \lfloor \log_2 N \rfloor$ be the number of bits used to describe x. We will prove that $\bar{x} = 2^t - 1 - x$ later. Hence, the encoding process is carried out by

$$\mathcal{E}(y) \equiv \begin{cases} \mathcal{E}(x) \bmod N^2, & \text{if } w = 0, \\ \mathcal{E}(2^t - 1) \cdot \mathcal{E}(x)^{-1} \bmod N^2, & \text{if } w = 1, \end{cases} \quad (24)$$

and the decoding process is performed by

$$(x, w) = \begin{cases} (y, 0), & \text{if } \Delta(y, \tilde{y}) < \Delta(\bar{y}, \tilde{y}), \\ (\bar{y}, 1), & \text{otherwise}, \end{cases} \quad (25)$$

where $\bar{y}$ is the bitwise compliment of y, $\tilde{y}$ is the approximation of y, and Δ is the measure of the Hamming distance.

Lemma: Let x be a non-negative integer, $\bar{x}$ be the bitwise compliment of x, and $t = \lfloor \log_2 N \rfloor$ be the number of bits used to describe x. Then, $\bar{x} = 2^t - 1 - x$.

Proof: Let us prove that

$$x + \bar{x} = 2^t - 1$$

such that

$$x \oplus \bar{x} = \underbrace{11\ldots1}_{t \text{ bits}}, \quad \forall x, \bar{x} \in \mathbb{Z}/2^t\mathbb{Z},$$

where $\oplus$ denotes the bitwise XOR operation. This is proved by the fact that integer addition is equivalent to bitwise XOR and $2^t - 1 = 2^0 + 2^1 + \cdots + 2^{t-1} = \underbrace{11\ldots1}_{t \text{ bits}}$.

V. CONTENT-ADAPTIVE PREDICTORS

Let us demonstrate how to implement the estimation mechanisms in a practical sense. Suppose that the host data is a digital image. Recall that we have to estimate the symbols in order to activate the decoding process. A symbol is defined as an integer converted from some randomly drawn bits from the l-th bit-plane. Therefore, the objective is to estimate those selected bits. A possible strategy is to estimate the pixel itself. This strategy is advantageous in terms of the implementation cost since there are a lot of image reconstruction tools available already, such as the total variation denoising algorithm [46]. Nonetheless, most content-adaptive predictors are 'online' processing algorithms and thus it is computationally demanding for the decoder to perform those algorithms. In response to this, we further devise an 'offline' processing mechanism that efficiently infers the l-th bit of the given pixel by a pre-learned lookup table. The approach is derived from the Bayesian inference. We remark that these online and offline algorithms can be viewed as a classic example of space-time tradeoff. We begin with the implementation of the total variation denoising algorithm and then continue with the formulation of the Bayesian inference.

A. TOTAL VARIATION DENOISING

Let us consider a marked image as a noisy signal. The result of noise removal can be perceived as our expectation of the original signal. Recall that when the l-th bit-plane is taken as the watermarking channel, the l-th bit of each modifiable pixel is either flipped or kept intact. Therefore, by comparing the marked image with its denoised counterpart, we are able to infer whether an observed pixel has been modified or not. For instance, suppose that a pixel u is used to carry a watermark bit. At the receiving end, we want to know whether $u = u_0$ or $u = u_1$, where u_0 and u_1 denote the pixel values by setting the l-th bit of u to 0 and 1, respectively. By applying a denoising technique, we obtain an approximation of u denoted by $\tilde{u}$. We contend that the value of u is the one that minimises the Euclidean distance between u and $\tilde{u}$. Let $\mathbf{u}$ be a sequence of pixels to carry a watermark bit and $\tilde{\mathbf{u}}$ be an approximation of it. In practice, we tend to use an appropriately long sequence of pixels to carry one single bit of information in order to minimise the error rate, although, on the other hand, compromising the embedding rate. To decide whether $\mathbf{u} = \mathbf{u}_0$ or $\mathbf{u} = \mathbf{u}_1$, we calculate

$$\begin{aligned} \Delta_0 &= \|\mathbf{u}_0 - \tilde{\mathbf{u}}\|_1 = \sum_i |u_{0,i} - \tilde{u}_i|, \\ \Delta_1 &= \|\mathbf{u}_1 - \tilde{\mathbf{u}}\|_1 = \sum_i |u_{1,i} - \tilde{u}_i|, \end{aligned} \quad (26)$$

and choose the vector that produces a smaller L_1 norm.

Total variation denoising is based on the principle that noisy signals have high total variation. According to this principle, the denoising problem is modelled as minimising the total variation of the reconstructed signal subject to it being close to the original observed signal. Let the intensity function $u^0(x, y)$ denote the pixel intensity of an observed noisy image and $u(x, y)$ the pixel values of the desired clean image for $x, y \in \Omega$. Hence,

$$u^0(x, y) = u(x, y) + n(x, y), \quad (27)$$

where $n(x, y)$ denotes the additive noise with zero mean and standard deviation σ. The objective is to remove n and reconstruct u from u_0. The variational model is to minimise

$$\iint_\Omega |\nabla u| \, dx \, dy, \quad (28)$$

subject to

$$\frac{1}{2} \iint_\Omega (u - u^0)^2 \, dx \, dy = \sigma^2. \quad (29)$$

By introducing a Lagrange multiplier λ, this problem can be converted into an unconstrained optimisation problem, which is to minimise the functional

$$J[u] = \iint_\Omega |\nabla u| \, dx \, dy + \frac{\lambda}{2} \iint_\Omega (u - u^0)^2 \, dx \, dy. \quad (30)$$

Let us express the above equation in a compact form by

$$J = \iint_\Omega \mathcal{L}(\Omega, u, \nabla u) \, d\Omega. \quad (31)$$

To minimise J, we find the partial derivatives of $\mathcal{L}$, which are given by

$$\frac{\partial \mathcal{L}}{\partial u} = \frac{\partial}{\partial u}\frac{\lambda}{2}(u-u^0)^2 = \lambda(u-u^0), \tag{32}$$

and

$$\frac{\partial \mathcal{L}}{\partial \nabla u} = \frac{\partial \mathcal{L}}{\partial u_x} + \frac{\partial \mathcal{L}}{\partial u_y} = \frac{\nabla u}{|\nabla u|}, \tag{33}$$

where

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial u_x} &= \frac{\partial}{\partial u_x}|\nabla u| = \frac{\partial}{\partial u_x}\sqrt{u_x^2+u_y^2} \\ &= \frac{\partial}{\partial u_x}\psi^{\frac{1}{2}} = \psi^{-\frac{1}{2}}u_x = \frac{u_x}{|\nabla u|}, \end{aligned} \tag{34}$$

and similarly

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial u_y} &= \frac{\partial}{\partial u_y}|\nabla u| = \frac{\partial}{\partial u_y}\sqrt{u_x^2+u_y^2} \\ &= \frac{\partial}{\partial u_y}\psi^{\frac{1}{2}} = \psi^{-\frac{1}{2}}u_y = \frac{u_y}{|\nabla u|}. \end{aligned} \tag{35}$$

Note that we let $\psi = u_x^2 + u_y^2$. By substituting these into the Euler-Lagrange equation

$$\frac{\partial \mathcal{L}}{\partial u} - \frac{d}{d\Omega}(\frac{\partial \mathcal{L}}{\partial \nabla u}) = 0, \tag{36}$$

we obtain

$$\lambda(u-u^0) - \nabla \cdot \frac{\nabla u}{|\nabla u|} = 0. \tag{37}$$

Therefore, the steepest descent equation for J is given by

$$\frac{\partial u}{\partial t} = \lambda(u-u^0) - \nabla \cdot \frac{\nabla u}{|\nabla u|}, \tag{38}$$

and thus at the time t we update

$$\begin{aligned} u^{t+1} &= u^t + \partial u^t \\ &= u^t + \partial t(\lambda^t(u^t - u^0) - \nabla \cdot \frac{\nabla u^t}{|\nabla u^t|}), \end{aligned} \tag{39}$$

where

$$\nabla \cdot \frac{\nabla u}{|\nabla u|} = \frac{\partial}{\partial x}\frac{u_x}{\sqrt{(u_x)^2+u_y^2}} + \frac{\partial}{\partial y}\frac{u_y}{\sqrt{u_x^2+u_y^2}}. \tag{40}$$

At a given pixel u_O, let u_N, u_S, u_W and u_E denote its four neighbouring pixels at north, south, west and east directions respectively, as illustrated in Fig. 4. To simplify the notations, we denote the four neighbouring pixels altogether by u_P, where $P \in \Lambda = \{N, S, W, E\}$. In the following, we present the numerical method for updating u_O via gradient descent. The algorithm is performed iteratively until it converges to a stable state or the default maximum number of iterations is reached. For each iteration, the numerical approximation of u_O is computed by

$$u_O^{t+1} = u_O^t + \Delta t \cdot (\lambda^t(u_O^t - u_O^0) - \frac{1}{h}\frac{\nabla u_O^t}{|\nabla u_O^t|}), \tag{41}$$

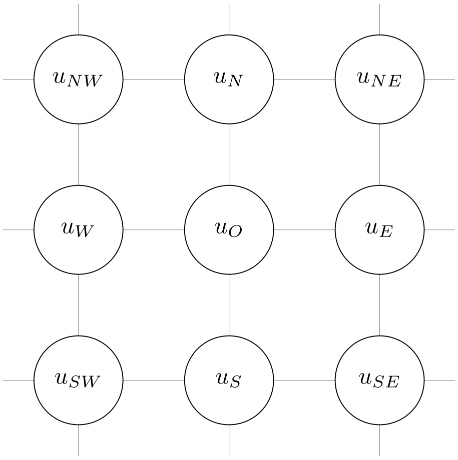


FIGURE 4. A given pixel u_O and its correlated neighbouring pixels u_N, u_S, u_W, u_E, u_{NW}, u_{NE}, u_{SW} and u_{SE}.

where Δt and h are set to 1 in our implementation. For convenience, we set λ^t to a fixed small number instead of updating it dynamically ($\lambda^t = 0.001$). We discretise

$$\frac{\nabla u_O^t}{|\nabla u_O^t|} \simeq \sum_{P \in \Lambda} \frac{u_P - u_O^t}{\sqrt{(u_P - u_O^t)^2 + \xi_P^2 + \epsilon}}, \tag{42}$$

where

$$\begin{aligned} \xi_N &= \frac{(u_W + u_{NW}) - (u_E + u_{NE})}{4}, \\ \xi_S &= \frac{(u_W + u_{SW}) - (u_E + u_{SE})}{4}, \\ \xi_W &= \frac{(u_N + u_{NW}) - (u_S + u_{SW})}{4}, \\ \xi_E &= \frac{(u_N + u_{NE}) - (u_S + u_{SE})}{4}, \end{aligned} \tag{43}$$

and ϵ is a very small number to avoid a zero divisor in practice.

B. BAYESIAN INFERENCE

A statistical approach to estimate the l-th bit of a pixel given a certain context is to collect a large number of samples under the same context and see which case, either that the bit is 0 or 1, is observed more frequently. This method will, however, encounter the so-called curse of dimensionality when the context is in a high dimensional space. That is, the amount of data required to support the sampling grows exponentially with the dimensionality of context. As aforementioned, a pixel is correlated to its eight immediate neighbours, which implies that the l-th bit of a given pixel is correlated to other 7 bits of that pixel and 8 bits of each neighbouring pixel. In total, there are 71 correlated bits, which represent 2^{71} different contexts. In this case, we require an enormous amount of data so that there are sufficient samples for each context.

In order to reduce the dimensionality of context, we model that the l-th bit of a given pixel is correlated to its immediate neighbouring bits on the current and the two

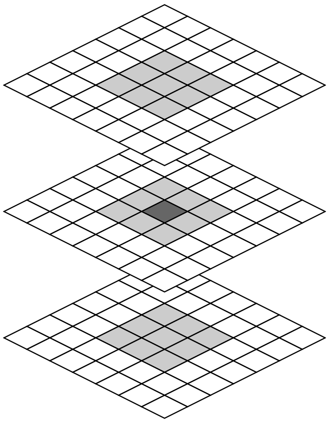


FIGURE 5. A given bit (coloured in dark grey) and its correlated neighbouring bits (coloured in light grey).

TABLE 1. Bayesian probability table, in which each column represents a possible context and each row represents a possible bit value.

	Θ_1	Θ_2	$\cdots$
$b=0$	$P(b=0\|\Theta_1)$	$P(b=0\|\Theta_2)$	$\cdots$
$b=1$	$P(b=1\|\Theta_1)$	$P(b=1\|\Theta_2)$	$\cdots$

adjacent bit-planes, as illustrated in Fig. 5. For the least significant bit-plane, we consider only the bits on the current and the second least significant bit-planes. For the most significant bit-plane, we consider only the bits on the current and the second most significant bit-planes. There are 8 correlated bits on the current bit plane, 9 on each of the two adjacent bit-planes. For a bit on the second to the seventh bit-plane, it has 26 correlated bits and thus 2^{26} possible contexts. For a bit on the first or last bit-plane, there are 17 correlated bits and accordingly 2^{17} contexts. The number of contexts is significantly reduced and yet the estimation is still remarkably accurate as demonstrated experimentally later. Let us denote the bit to be estimated by b and the collection of correlated bits, or the context, by Θ. According to the Bayes' theorem, we compute

$$\begin{aligned} P(b=0|\Theta) &\propto P(\Theta|b=0)P(b=0), \\ P(b=1|\Theta) &\propto P(\Theta|b=1)P(b=1), \end{aligned} \tag{44}$$

where $P(b)$ is the prior probability of the hypothesis of b, $P(\Theta|b)$ is the likelihood of observing the evidence Θ given the hypothesis of b, and $P(b|\Theta)$ is the posterior probability of the hypothesis of b given the observed evidence Θ. Therefore, the inference about the value of b is made by

$$\tilde{b} = \underset{b\in\{0,1\}}{\arg\max}\, P(b|\Theta). \tag{45}$$

As a result, we learn the Bayesian probability table from a large number of image samples and predict the bits by the lookup table, as illustrated in Table 1.

VI. EXPERIMENTS

In the following experiments, we test the schemes on 8-bit greyscale images of size 512 × 512, as shown in Fig. 6. We use the RSA and the Paillier cryptosystems as examples of multiplicative and additive homomorphic cryptosystems, respectively. Two prime numbers for the cryptosystems are $p = 7907$ and $q = 7919$ so that the bit-length of message is $\lfloor \log_2 pq \rfloor = 25$. Accordingly, for each bit-plane of an image, we convert every segment of 25 bits into a decimal number in order to fit the encryption function, where the marginal bits are negligible. The number of iterations for the total variation algorithm is set to 2000 since empirically it produces stable results. The Bayesian probability table is learned from a thousand image samples in the BOSSBase [47]. We begin with the analysis of three-way trade-off between capacity, fidelity, and reversibility. Then, we make comparisons between the proposed schemes and the state-of-the-art. The capacity is represented by the number of bits embedded in an image, whereas the fidelity and reversibility are measured by the peak signal-to-noise ratio (PSNR). Let $u_{i,j}$ and $\hat{u}_{i,j}$ denote a pixel and its noisy counterpart respectively, where i and j specify the pixel coordinates. The PSNR (in dB) is defined as

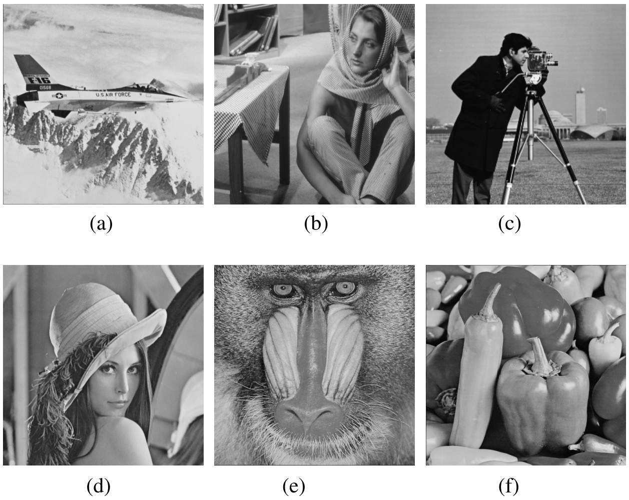


FIGURE 6. Greyscale test images of size 512 × 512 with 256 tonal options. (a) Airplane. (b) Barbara. (c) Cameraman. (d) Lena. (e) Mandrill. (f) Peppers.

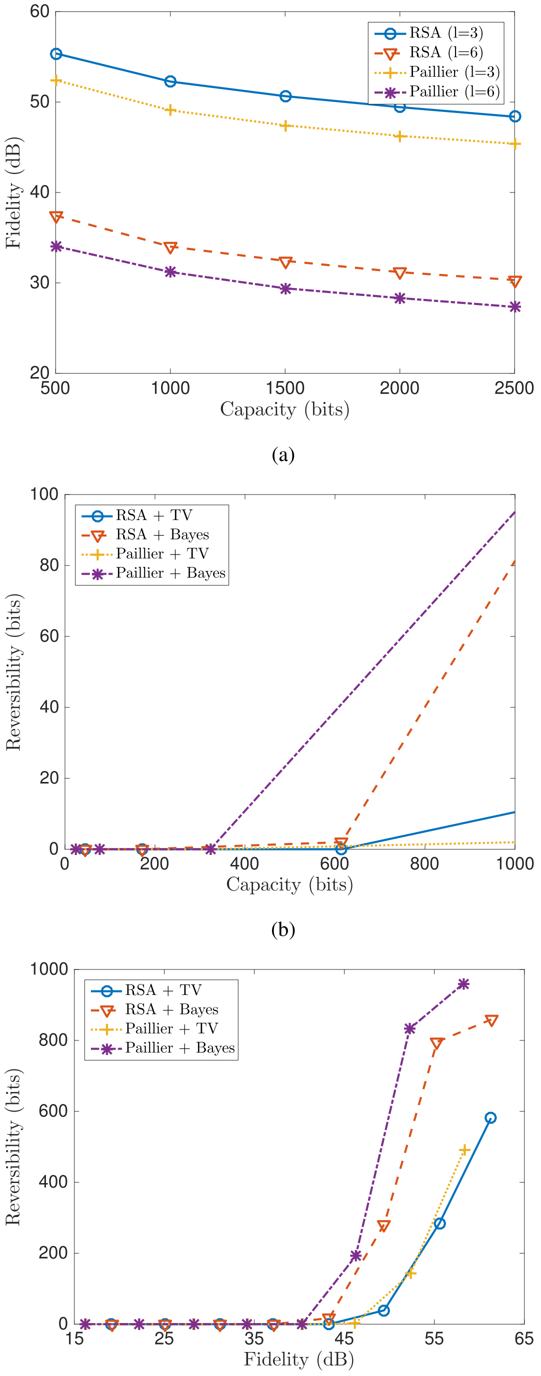
$$\text{PSNR} = 10 \times \log_{10} \frac{255^2}{\text{MSE}}. \tag{46}$$

where the mean square error (MSE) is calculated by

$$\text{MSE} = \frac{1}{512 \times 512} \sum_{i=1}^{512} \sum_{j=1}^{512} (u_{i,j} - \hat{u}_{i,j})^2. \tag{47}$$

The fidelity is quantified by the PSNR between the original and the marked images, and in a similar manner, the reversibility is evaluated by the PSNR between the original and the recovered images. For further analysis, we also measure the reversibility in terms of the number of bit errors in the following discussions.

- *Capacity-Fidelity Trade-Off:* The trade-off between capacity and fidelity is visualised in Fig. 7(a), which shows that the RSA-based scheme achieves higher fidelity than the Paillier-based scheme under the same capacity (2000 bits). The reason is straightforward: when the noise occurs, the former only introduces a random distortion while the latter



(a)

(b)

(c)

FIGURE 7. The three way trade-off between capacity, fidelity, and reversibility measured on the image Lena. (a) Tradeoff between capacity and fidelity. (b) Tradeoff between capacity and reversibility. (c) Tradeoff between fidelity and reversibility.

inflicts a maximal distortion. Note that the fidelity is not content-dependent according to our scheme design. The impact of the selection of different bit-planes as the watermarking channel is shown in Table 2. It is intuitive that the fidelity of marked images decreases drastically when a more significant bit-plane is engaged to carry watermark information.

TABLE 2. Fidelity evaluation of RSA-based and Paillier-based schemes by using different bit-planes as the watermarking channel. It shows the PSNR (in dB) of the marked image Lena. (a) Capacity: 500 bits. (b) Capacity: 1000 bits. (c) Capacity: 1500 bits. (b) Capacity: 2000 bits

	RSA	Paillier
$l = 1$	67.59	64.25
$l = 2$	61.19	58.41
$l = 3$	55.46	52.43
$l = 4$	49.46	46.28
$l = 5$	43.27	39.86
$l = 6$	37.58	34.22
$l = 7$	31.18	28.09
$l = 8$	25.31	21.96

(a)

	RSA	Paillier
$l = 1$	64.32	61.28
$l = 2$	58.32	55.10
$l = 3$	52.17	49.23
$l = 4$	46.29	43.39
$l = 5$	40.32	37.20
$l = 6$	34.11	31.25
$l = 7$	28.41	25.28
$l = 8$	22.37	19.12

(b)

	RSA	Paillier
$l = 1$	62.53	59.70
$l = 2$	56.55	53.42
$l = 3$	50.61	47.62
$l = 4$	44.40	41.47
$l = 5$	38.48	35.33
$l = 6$	32.56	29.49
$l = 7$	26.57	23.53
$l = 8$	20.29	17.53

(c)

	RSA	Paillier
$l = 1$	61.29	58.23
$l = 2$	55.57	52.26
$l = 3$	49.41	46.26
$l = 4$	43.32	40.31
$l = 5$	37.08	34.18
$l = 6$	31.18	28.32
$l = 7$	25.09	22.18
$l = 8$	19.05	16.30

(d)

- *Capacity-Reversibility Trade-Off:* The trade-off between capacity and reversibility is illustrated in Fig. 7(b). To assess the capacity-reversibility trade-off, we control the capacity by selecting different bit-planes for watermarking in such a way that the fidelity remains at 48 dB, and then observe the change of reversibility. Note that here we express the reversibility in terms of the number of bit errors in order to amplify the trend. If the reversibility is quantified by the PSNR, the proportionality may be indeterminate since a higher decoding error rate does not necessarily mean a lower visual quality. For instance, several errors occur on an insignificant bit-plane may still result in a higher visual quality than a single error on a relatively significant bit-plane. It is observed that the reversibility decreases as the capacity increases. It is due to the fact that we embed less amount of information into the more significant bit-plane and more amount into the less

TABLE 3. Reversibility evaluation of different combinations of encoding and decoding components. It shows the PSNR (in dB) and the number of bit errors of the recovered images in which 2000 random bits are embedded. (a) RSA + TV. (b) RSA + Bayes. (c) Paillier + TV. (d) Paillier + Bayes.

	Airplane		Barbara		Cameraman		Lena		Mandrill		Peppers	
	PSNR	NUM	PSNR	NUM	PSNR	NUM	PSNR	NUM	PSNR	NUM	PSNR	NUM
$l = 1$	65.38	403	63.10	667	71.61	102	63.75	581	62.36	795	62.82	715
$l = 2$	64.38	135	59.20	417	81.24	3	60.90	285	56.53	760	58.43	500
$l = 3$	69.10	13	59.18	112	∞	0	64.06	39	52.44	489	56.41	205
$l = 4$	∞	0	60.90	20	∞	0	∞	0	50.22	216	56.41	10
$l = 5$	∞	0	63.60	3	∞	0	∞	0	54.21	23	∞	0
$l = 6$	∞	0	∞	0	∞	0	∞	0	∞	0	∞	0
$l = 7$	∞	0	∞	0	∞	0	∞	0	∞	0	∞	0
$l = 8$	∞	0	∞	0	∞	0	∞	0	∞	0	∞	0

(a)

	Airplane		Barbara		Cameraman		Lena		Mandrill		Peppers	
	PSNR	NUM	PSNR	NUM	PSNR	NUM	PSNR	NUM	PSNR	NUM	PSNR	NUM
$l = 1$	62.85	706	61.82	886	67.51	249	62.02	859	61.90	882	61.68	915
$l = 2$	59.24	404	56.09	830	74.06	16	56.32	796	55.68	917	55.83	882
$l = 3$	62.04	60	52.93	434	∞	0	54.89	282	49.93	876	51.23	645
$l = 4$	∞	0	55.76	61	∞	0	61.60	17	45.54	599	54.33	87
$l = 5$	∞	0	61.60	4	∞	0	∞	0	45.42	160	∞	0
$l = 6$	∞	0	∞	0	∞	0	∞	0	49.88	15	∞	0
$l = 7$	∞	0	∞	0	∞	0	∞	0	∞	0	∞	0
$l = 8$	∞	0	∞	0	∞	0	∞	0	∞	0	∞	0

(b)

	Airplane		Barbara		Cameraman		Lena		Mandrill		Peppers	
	PSNR	NUM	PSNR	NUM	PSNR	NUM	PSNR	NUM	PSNR	NUM	PSNR	NUM
$l = 1$	64.30	253	60.78	569	74.53	24	61.41	492	59.37	787	60.09	667
$l = 2$	68.00	27	58.62	234	∞	0	60.76	143	54.06	669	56.46	385
$l = 3$	∞	0	63.28	20	∞	0	73.28	2	50.87	348	58.16	65
$l = 4$	∞	0	∞	0	∞	0	∞	0	50.68	91	∞	0
$l = 5$	∞	0	∞	0	∞	0	∞	0	∞	0	∞	0
$l = 6$	∞	0	∞	0	∞	0	∞	0	∞	0	∞	0
$l = 7$	∞	0	∞	0	∞	0	∞	0	∞	0	∞	0
$l = 8$	∞	0	∞	0	∞	0	∞	0	∞	0	∞	0

(c)

	Airplane		Barbara		Cameraman		Lena		Mandrill		Peppers	
	PSNR	NUM	PSNR	NUM	PSNR	NUM	PSNR	NUM	PSNR	NUM	PSNR	NUM
$l = 1$	59.60	746	58.42	979	65.64	186	58.51	960	58.42	981	58.31	1004
$l = 2$	57.11	331	52.69	916	∞	0	53.10	833	52.32	999	52.41	977
$l = 3$	65.15	13	51.45	305	∞	0	53.46	192	46.60	932	48.32	626
$l = 4$	∞	0	59.48	12	∞	0	∞	0	42.61	584	55.36	31
$l = 5$	∞	0	∞	0	∞	0	∞	0	45.33	78	∞	0
$l = 6$	∞	0	∞	0	∞	0	∞	0	∞	0	∞	0
$l = 7$	∞	0	∞	0	∞	0	∞	0	∞	0	∞	0
$l = 8$	∞	0	∞	0	∞	0	∞	0	∞	0	∞	0

(d)

significant bit-plane in order to keep the fidelity at a fixed level.

- *Fidelity-Reversibility Trade-Off:* The trade-off between capacity and reversibility is illustrated in Fig. 7(c). In a similar way, we evaluate the fidelity-reversibility trade-off by setting the capacity to 2000 bits, controlling the fidelity with the selection of encodable bit-plane, and observing the change of reversibility. It is evident that the Paillier-based scheme is advantageous in terms of the reversibility compared to the RSA-based scheme. This is because that the Paillier-based scheme inflicts a stronger noise to the host media, which is more distinguishable and removable than a faint noise imposed by the RSA-based scheme. On top of that, it is shown that the predictor based on total variation achieves higher reversibility than the predictor based on the Bayesian inference. Moreover, it can be observed that the reversibility is inversely proportional to the fidelity.

More comprehensive experimental results are demonstrated in Table 3, where a variety of images are tested with every possible combination of encoding and decoding mechanisms. It is witnessed that the error-free performance is achievable in practice. It can be observed that the error rate generally approaches zero when embedding payloads into a comparatively more significant bit-plane. From the previous analyses, we conclude that

- Capacity is inversely proportional to fidelity.
- Capacity is inversely proportional to reversibility.
- Fidelity is inversely proportional to reversibility.

TABLE 4. Classification of reversible watermarking schemes compatible with public-key cryptosystems.

	Proposed	Wu *et al.* [32]	Wu *et al.* [33]	Li and Li [34]	Zhang *et al.* [35]	Xiang and Luo [36]
Bit-padding	No	Yes	Yes	Yes	No	No
Pre-processing	No	No	No	No	Yes	Yes

TABLE 5. Fidelity comparison with the state-of-the-art. It shows the PSNR (in dB) of the marked images in which 2000 random bits are embedded.

	Airplane	Barbara	Cameraman	Lena	Mandrill	Peppers
RSA ($l = 3$)	49.27	49.32	49.47	49.43	49.13	49.19
RSA ($l = 6$)	31.13	31.08	31.31	31.18	31.32	31.19
Paillier ($l = 3$)	46.18	46.36	46.10	46.25	46.26	46.23
Paillier ($l = 6$)	28.14	28.37	28.17	28.32	28.26	28.16
Zhang [48]	41.11	41.04	40.98	41.06	41.04	41.0
Wu and Sun ($l = 3$) [49]	46.30	46.16	46.33	46.29	46.16	46.27
Wu and Sun ($l = 6$) [49]	28.25	28.39	28.25	28.43	28.38	28.22
Liao and Shu [50]	41.12	41.04	41.05	41.04	41.03	41.04

TABLE 6. Reversibility comparison with the state-of-the-art. It shows the PSNR (in dB) and the number of bit errors of the recovered images in which 2000 random bits are embedded.

	Airplane		Barbara		Cameraman		Lena		Mandrill		Peppers	
	PSNR	NUM	PSNR	NUM	PSNR	NUM	PSNR	NUM	PSNR	NUM	PSNR	NUM
RSA + TV ($l = 3$)	69.10	13	59.18	112	∞	0	64.06	39	52.44	489	56.41	205
RSA + TV ($l = 6$)	∞	0	∞	0	∞	0	∞	0	∞	0	∞	0
RSA + Bayes ($l = 3$)	62.04	60	52.93	434	∞	0	54.89	282	49.93	876	51.23	645
RSA + Bayes ($l = 6$)	∞	0	∞	0	∞	0	∞	0	49.88	15	∞	0
Paillier + TV ($l = 3$)	∞	0	63.28	20	∞	0	73.28	2	50.87	348	58.16	65
Paillier + TV ($l = 6$)	∞	0	∞	0	∞	0	∞	0	∞	0	∞	0
Paillier + Bayes ($l = 3$)	65.15	13	51.45	305	∞	0	53.46	192	46.60	932	48.32	626
Paillier + Bayes ($l = 6$)	∞	0	∞	0	∞	0	∞	0	∞	0	∞	0
Zhang [48]	55.09	71	48.93	169	54.38	85	57.10	28	46.16	309	54.57	47
Wu and Sun ($l = 3$) [49]	∞	0	58.58	59	∞	0	69.30	5	50.03	423	57.66	73
Wu and Sun ($l = 6$) [49]	∞	0	∞	0	∞	0	∞	0	∞	0	∞	0
Liao and Shu [50]	60.73	13	57.57	23	67.56	3	62.46	7	52.54	70	58.13	20

- The RSA-based scheme preserves higher fidelity than the Paillier-based scheme.
- The Paillier-based scheme achieves higher reversibility than the RSA-based scheme.
- The predictor based on total variation is more content-adaptive and results in higher reversibility than that based on Bayesian inference.
- The predictor based on Bayesian inference consumes less computational power in decoding than that based on total variation.

In the following, we compare the scheme performance with the state-of-the-art. We begin by classifying the existing reversible watermarking schemes compatible with public-key cryptosystems [32]–[36], as presented in Table 4. As far as we are aware, our work is one of the pioneering research on the schemes compatible with public-key homomorphic cryptosystems under the condition that neither additional bit padding nor specific pre-processing is undertaken. Although schemes addressing this strictly defined problem are hardly found, we remark that this design principle was followed in the earlier literature of schemes based on traditional symmetric-key cryptosystems. In order to make paralleled and meaningful comparisons, the proposed schemes are compared with those under the same, or at least similar, research constraint [48]–[50]. We test the fidelity and reversibility between different schemes by embedding randomly generated 2000 bits into various host images. It is observed from Table 5 that the visual qualities of various host images are similar for any listed encoding algorithm and thus it is evident that the fidelity is not content-dependent. Yet, different schemes and configurations result in different fidelity. It is shown that the RSA-based and Paillier-based schemes preserves comparatively high fidelity when embedding payloads into the third bit-plane. From Table 6, we can further observe that the proposed schemes achieve relatively high reversibility, especially when using the total variation technique for decoding. When the sixth bit-plane is used as the embedding channel, the proposed schemes show the error-free performance on most of the test images. Even when embedding payloads into the third bit-plane, the reversibility is higher than the average in most cases. The experimental results show that the proposed schemes achieve a remarkable balance between fidelity and reversibility under the given capacity constraint. On top of this, the proposed schemes share the advantages of the modern public-key cryptosystems.

VII. CONCLUSION

This paper considers the problem of entrusting the watermarking operation to a cloud service provider without undermining data privacy. Constructing reversible watermarking schemes compatible with public-key cryptosystems is of non-trivial challenge since there is virtually no data redundancy to be exploited in the encrypted domain. The recent development of various schemes required the host media either to be enciphered in a redundant fashion or to be pre-processed prior to encryption. To address these limits, we propose a novel paradigm and derive different schemes compatible with different types of public-key homomorphic cryptosystems. The host image is encrypted in such a fashion that a sufficiently long sequence of bits, instead of a pixel, is regarded as input to the encryptor. It significantly reduces the size of the encrypted data and improves the space efficiency. In addition to this, we suggest to encrypt different bit-planes separately in order to control the fidelity factor. Any specific data pre-processing is not required in order to promote the practicality and universality. For the RSA-like cryptosystems that permit multiplicative homomorphic computation, the proposed scheme is flexible in terms of embedding capacity. As for the Paillier-like cryptosystems that allow additive operation as well as partial multiplicative operation, we propose a scheme which is optimal with respect to the reversibility. Furthermore, we develop content-adaptive predictors based on variational method and statistical inference for assisting watermark decoding. Both online and offline prediction algorithms are provided in order to suit different operational requirements. We remark that the decoding errors are inevitable from a theoretical point of view and yet in practice the error rate approaches zero at a low expense of the fidelity and capacity. Experimental results show that the proposed schemes achieve remarkable balance between fidelity and reversibility under the given capacity constraint. From our perspective, it is of great significance to develop privacy-aware watermarking methodology suitable for a variety of modern cryptosystems. On top of this, more practical applications, such as privacy protection in the Internet of Things, deserves further investigation.

ACKNOWLEDGEMENT

The authors would like to thank Mr Yijun Quan for the valuable discussions on mathematical derivations, and Dr Kaimeng Chen for the helpful technical assistance. They would also like to thank anonymous reviewers for their critical comments and suggestions on improving the presentation and organisation of this paper.

REFERENCES

[1] C. E. Shannon, "Communication theory of secrecy systems," *Bell Labs Tech. J.*, vol. 28, no. 4, pp. 656–715, Oct. 1949.

[2] R. L. Lagendijk, Z. Erkin, and M. Barni, "Encrypted signal processing for privacy protection: Conveying the utility of homomorphic encryption and multiparty computation," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 82–105, Jan. 2013.

[3] C. Aguilar-Melchor, S. Fau, C. Fontaine, G. Gogniat, and R. Sirdey, "Recent advances in homomorphic encryption: A possible future for signal processing in the encrypted domain," *IEEE Signal Process. Mag.*, vol. 30, no. 2, pp. 108–117, Mar. 2013.

[4] M. Barni, G. Droandi, and R. Lazzeretti, "Privacy protection in biometric-based recognition systems: A marriage between cryptography and signal processing," *IEEE Signal Process. Mag.*, vol. 32, no. 5, pp. 66–76, Sep. 2015.

[5] I. J. Cox, J. Kilian, F. T. Leighton, and T. Shamoon, "Secure spread spectrum watermarking for multimedia," *IEEE Trans. Image Process.*, vol. 6, no. 12, pp. 1673–1687, Dec. 1997.

[6] I. J. Cox, M. L. Miller, and A. L. McKellips, "Watermarking as communications with side information," *Proc. IEEE*, vol. 87, no. 7, pp. 1127–1141, Jul. 1999.

[7] B. Chen and G. W. Wornell, "Quantization index modulation: A class of provably good methods for digital watermarking and information embedding," *IEEE Trans. Inf. Theory*, vol. 47, no. 4, pp. 1423–1443, May 2001.

[8] C.-C. Chang, P. Tsai, and C.-C. Lin, "SVD-based digital image watermarking scheme," *Pattern Recognit. Lett.*, vol. 26, no. 10, pp. 1577–1586, Jul. 2005.

[9] D. Kundur and D. Hatzinakos, "Digital watermarking for telltale tamper proofing and authentication," *Proc. IEEE*, vol. 87, no. 7, pp. 1167–1180, Jul. 1999.

[10] M. U. Celik, G. Sharma, E. Saber, and A. M. Tekalp, "Hierarchical watermarking for secure image authentication with localization," *IEEE Trans. Image Process.*, vol. 11, no. 6, pp. 585–595, Jun. 2002.

[11] C. T. Li, "Digital fragile watermarking scheme for authentication of JPEG images," *IEE Proc. Vis., Image Signal Process.*, vol. 151, no. 6, pp. 460–466, Dec. 2004.

[12] C.-C. Chang, Y.-S. Hu, and T.-C. Lu, "A watermarking-based image ownership and tampering authentication scheme," *Pattern Recognit. Lett.*, vol. 27, no. 5, pp. 439–446, Apr. 2006.

[13] C. W. Honsinger, P. W. Jones, M. Rabbani, and J. C. Stoffel, "Lossless recovery of an original image containing embedded data," U.S. Patent 6 278 791, Aug. 21, 2001.

[14] J. Fridrich, M. Goljan, and R. Du, "Invertible authentication," *Proc. SPIE*, vol. 4314, pp. 197–208 Aug. 2001.

[15] J. Tian, "Reversible data embedding using a difference expansion," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 8, pp. 890–896, Aug. 2003.

[16] A. M. Alattar, "Reversible watermark using the difference expansion of a generalized integer transform," *IEEE Trans. Image Process.*, vol. 13, no. 8, pp. 1147–1156, Aug. 2004.

[17] Z. Ni, Y.-Q. Shi, N. Ansari, and W. Su, "Reversible data hiding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 3, pp. 354–362, Mar. 2006.

[18] D. M. Thodi and J. J. Rodriguez, "Expansion embedding techniques for reversible watermarking," *IEEE Trans. Image Process.*, vol. 16, no. 3, pp. 721–730, Mar. 2007.

[19] Y. Yang, X. Sun, H. Yang, C. T. Li, and R. Xiao, "A contrast-sensitive reversible visible image watermarking technique," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 5, pp. 656–667, May 2009.

[20] W.-L. Tai, C.-M. Yeh, and C.-C. Chang, "Reversible data hiding based on histogram modification of pixel differences," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 6, pp. 906–910, Jun. 2009.

[21] W. Puech, M. Chaumont, and O. Strauss, "A reversible data hiding method for encrypted images," *Proc. SPIE*, vol. 6819, pp. 68191E-1–68191E-9, Feb. 2008.

[22] Z. Qian, X. Zhang, Y. Ren, and G. Feng, "Block cipher based separable reversible data hiding in encrypted images," *Multimedia Tools Appl.*, vol. 75, no. 21, pp. 13749–13763, Nov. 2016.

[23] X. Zhang, "Separable reversible data hiding in encrypted image," *IEEE Trans. Inf. Forensics Security*, vol. 7, no. 2, pp. 826–832, Apr. 2012.

[24] X. Zhang, Z. Qian, G. Feng, and Y. Ren, "Efficient reversible data hiding in encrypted images," *J. Vis. Commun. Image Represent.*, vol. 25, no. 2, pp. 322–328, Feb. 2014.

[25] Z. Qian and X. Zhang, "Reversible data hiding in encrypted images with distributed source encoding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 4, pp. 636–646, Apr. 2016.

[26] M. Johnson, P. Ishwar, V. Prabhakaran, D. Schonberg, and K. Ramchandran, "On compressing encrypted data," *IEEE Trans. Signal Process.*, vol. 52, no. 10, pp. 2992–3006, Oct. 2004.

[27] D. Schonberg, S. C. Draper, C. Yeo, and K. Ramchandran, "Toward compression of encrypted images and video sequences," *IEEE Trans. Inf. Forensics Security*, vol. 3, no. 4, pp. 749–762, Dec. 2008.

[28] D. Klinc, C. Hazay, A. Jagmohan, H. Krawczyk, and T. Rabin, "On compression of data encrypted with block ciphers," *IEEE Trans. Inf. Theory*, vol. 58, no. 11, pp. 6989–7001, Nov. 2012.

[29] W. Hong, T.-S. Chen, and H.-Y. Wu, "An improved reversible data hiding in encrypted images using side match," *IEEE Signal Process. Lett.*, vol. 19, no. 4, pp. 199–202, Apr. 2012.

[30] Z. Qian, X. Zhang, and G. Feng, "Reversible data hiding in encrypted images based on progressive recovery," *IEEE Signal Process. Lett.*, vol. 23, no. 11, pp. 1672–1676, Nov. 2016.

[31] F. Huang, J. Huang, and Y.-Q. Shi, "New framework for reversible data hiding in encrypted domain," *IEEE Trans. Inf. Forensics Security*, vol. 11, no. 12, pp. 2777–2789, Dec. 2016.

[32] H.-T. Wu, Y.-M. Cheung, and J. Huang, "Reversible data hiding in Paillier cryptosystem," *J. Vis. Commun. Image Represent.*, vol. 40, pp. 765–771, Oct. 2016.

[33] X. Wu, B. Chen, and J. Weng, "Reversible data hiding for encrypted signals by homomorphic encryption and signal energy transfer," *J. Vis. Commun. Image Represent.*, vol. 41, pp. 58–64, Nov. 2016.

[34] M. Li and Y. Li, "Histogram shifting in encrypted images with public key cryptosystem for reversible data hiding," *Signal Process.*, vol. 130, pp. 190–196, Jan. 2017.

[35] X. Zhang, J. Long, Z. Wang, and H. Cheng, "Lossless and reversible data hiding in encrypted images with public-key cryptography," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 9, pp. 1622–1631, Sep. 2016.

[36] S. Xiang and X. Luo, "Reversible data hiding in homomorphic encrypted domain by mirroring ciphertext group," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 11, pp. 3099–3110, Nov. 2017.

[37] 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, no. 3, pp. 553–562, Mar. 2013.

[38] W. Zhang, K. Ma, and N. Yu, "Reversibility improved data hiding in encrypted images," *Signal Process.*, vol. 94, pp. 118–127, Jan. 2014.

[39] 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. Cybern.*, vol. 46, no. 5, pp. 1132–1143, May 2016.

[40] R. L. Rivest *et al.*, "On data banks and privacy homomorphisms," in *Foundations of Secure Computation*, R. A. DeMillo, Ed. New York, NY, USA: Academic, 1978, pp. 169–180.

[41] R. L. Rivest, A. Shamir, and L. Adleman, "A method for obtaining digital signatures and public-key cryptosystems," *Commun. ACM*, vol. 21, no. 2, pp. 120–126, Feb. 1978.

[42] T. ElGamal, "A public key cryptosystem and a signature scheme based on discrete logarithms," *IEEE Trans. Inf. Theory*, vol. 31, no. 4, pp. 469–472, Jul. 1985.

[43] T. Okamoto and S. Uchiyama, "A new public-key cryptosystem as secure as factoring," in *Proc. Int. Conf. Theory Appl. Cryptograph. Techn. (EUROCRYPT)*, Espoo, Finland, May 1998, pp. 308–318.

[44] I. Damgård and M. Jurik, "A generalisation, a simplication and some applications of Paillier's probabilistic public-key system," in *Proc. Int. Workshop Pract. Theory Public Key Cryptogr. (PKC)*, Feb. 2001, pp. 119–136.

[45] P. Paillier, "Public-key cryptosystems based on composite degree residuosity classes," in *Proc. Int. Conf. Theory Appl. Cryptograph. Techn.*, Prague, Czech Republic, Mar. 1999, pp. 223–238.

[46] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys. D, Nonlinear Phenomena*, vol. 60, nos. 1–4, pp. 259–268, 1992.

[47] P. Bas, T. Filler, and T. Pevný, "'Break our steganographic system': The ins and outs of organizing BOSS," in *Proc. 13th Int. Workshop Inf. Hiding*, Prague, Czech Republic, May 2011, pp. 59–70.

[48] X. Zhang, "Reversible data hiding in encrypted image," *IEEE Signal Process. Lett.*, vol. 18, no. 4, pp. 255–258, Apr. 2011.

[49] X. Wu and W. Sun, "High-capacity reversible data hiding in encrypted images by prediction error," *Signal Process.*, vol. 104, pp. 387–400, Nov. 2014.

[50] X. Liao and C. Shu, "Reversible data hiding in encrypted images based on absolute mean difference of multiple neighboring pixels," *J. Vis. Commun. Image Represent.*, vol. 28, pp. 21–27, Apr. 2015.

CHING-CHUN CHANG received the B.B.A. degree in information management from National Central University, Taiwan, in 2015. He is currently pursuing the Ph.D. degree with the Department of Computer Science, University of Warwick, U.K. He engaged in a short-term scientific mission supported by the European Cooperation in Science and Technology actions at the Faculty of Computer Science, Otto von Guericke University Magdeburg, Germany, in 2016. He participated in a research and innovation staff exchange scheme supported by Marie Skłodowska-Curie actions at the Department of Electrical and Computer Engineering, New Jersey Institute of Technology, USA, in 2017. He has been a Visiting Scholar at the School of Computer and Mathematics, Charles Sturt University, Australia, since 2018. His research interests include digital watermarking, steganography, secret sharing, applied cryptography, and multimedia security.

CHANG-TSUN LI received the B.Eng. degree in electrical engineering from National Defence University (NDU), Taiwan, in 1987, the M.Sc. degree in computer science from the Naval Postgraduate School (NPS), USA, in 1992, and the Ph.D. degree in computer science from the University of Warwick, U.K., in 1998. He was an Associate Professor with the Department of Electrical Engineering, NDU, from 1998 to 2002, and a Visiting Professor with the Department of Computer Science, NPS, in 2001. He was a Professor with the Department of Computer Science, University of Warwick, until 2017. He is currently a Professor with the School of Computing and Mathematics, Charles Sturt University, Australia. His research interests include multimedia forensics, biometrics, bioinformatics, data mining, machine learning, image processing, computer vision, and pattern recognition. The outcomes of his multimedia forensics research have been translated into award-winning commercial products protected by a series of international patents and have been used by a number of police forces and courts of law around the world. He involved in the organization of many international conferences and workshops and also served as a member of the international program committees for several international conferences. He is also actively contributing keynote speeches and talks at various international events. He is currently an Associate Editor of the IEEE ACCESS, the *EURASIP Journal of Image and Video Processing*, and *IET Biometrics*.

YUN-QING SHI (F'06–LF'18) received the B.Sc. and M.Sc. degree from Shanghai Jiao Tong University, China, and the Ph.D. degree from the University of Pittsburgh, USA. He has been with the New Jersey Institute of Technology, USA, since 1987. He had industrial experience in a radio factory as a Principal Design and Test Engineer in numerical control manufacturing and electronic broadcasting devices. Some of his research projects have been supported by several federal and New Jersey State Funding Agencies. He has authored/co-authored over 400 papers in his research areas and is also an editor of several books, special issues, and proceedings. He has delivered over 120 invited talks and 18 tutorials around the world. He holds 30 awarded U.S. patents and was honored as a fellow of the National Academy of Inventors in 2017. His research interests include information hiding, digital forensics, signal processing, data communications, computer vision, pattern recognition, and multidimensional systems. He has been a fellow of the IEEE for his contribution to multidimensional signal processing since 2006, and has become a Life Fellow of the IEEE since 2018. He has served as an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING, the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS, and the IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY. He has also served as an editorial board member of several journals and a technical program chair of several international conferences.