Bi-level Protected Compressive Sampling

Leo Yu Zhang, Kwok-Wo Wong, *Senior Member, IEEE*, Yushu Zhang, and Jiantao Zhou, *Member, IEEE*

Abstract—Some pioneering works have investigated embedding cryptographic properties in compressive sampling (CS) in a way similar to one-time pad symmetric cipher. This paper tackles the problem of constructing a CS-based symmetric cipher under the key reuse circumstance, i.e., the cipher is resistant to common attacks even when a fixed measurement matrix is used multiple times. To this end, we suggest a bi-level protected CS (BLP-CS) model which makes use of the advantage of measurement matrix construction without restricted isometry property (RIP). Specifically, two kinds of artificial basis mismatch techniques are investigated to construct key-related sparsifying bases. It is demonstrated that the encoding process of BLP-CS is simply a random linear projection, which is the same as the basic CS model. However, decoding the linear measurements requires knowledge of both the key-dependent sensing matrix and its sparsifying basis. The proposed model is exemplified by sampling images as a joint data acquisition and protection layer for resource-limited wireless sensors. Simulation results and numerical analyses have justified that the new model can be applied in circumstances where the measurement matrix can be reused.

Index Terms—Compressive sampling (CS), encryption, known/chosen-plaintext attack, random projection, restricted isometry property (RIP).

I. INTRODUCTION

COMPRESSIVE sampling (CS) has received extensive research attention in the last decade [1]–[3]. By utilizing the fact that natural signals are either sparse or compressible, the CS theory demonstrates that such signals can be faithfully recovered from a small set of linear, nonadaptive measurements, allowing sampling at a rate lower than that required by the Nyquist-Shannon sampling theorem.

The use of CS for security purposes was first outlined in one of the foundation papers [4], in which Candes and Tao suggested that the measurement vector obtained from random subspace linear projection can be treated as ciphertext since the unauthorized user would not be able to decode it unless he knows in which random subspace the coefficients are expressed. In this way, the entire CS scheme can be considered as a variant of symmetric cipher, where the signal to be sampled, the measurement vector and the measurement matrix are treated as the plaintext, the ciphertext and the secret key, respectively.

It is a favorable characteristic that certain kind of data protection mechanism can be embedded into the data acquisition stage. Such a property of CS is of particular importance for data acquisition systems in sensor networks, where each sensor is usually resource-limited and a separate cryptographic layer is too expensive for secure data transmission. Example applications work under this circumstance include visual sensor networks [5], video surveillance networks [6] and etc. Meanwhile, CS paradigm is also found to be useful for medical systems, especially in the case that sampling speed [7] and privacy [8] are two major concerns.

There are a number of studies exploring the security that a CS-based symmetric cipher can provide from the computation point of view. For example, it was shown in [9] that the measurement matrix leads to computational secrecy under some attack scenarios, such as brute-force attack and ciphertext only attack (COA). Based on this result, there were many attempts in establishing secure measurement matrices. In [10], constructing the measurement matrix using physical layer properties and linear feedback shift register (LFSR) with the corresponding m-sequence was proposed. In [11], Tong *et al.* suggested constructing CS measurement matrix by chaotic sequence for privacy protection in video sequence. In [12], Cambareri *et al.* employed CS to provide two access levels by artificially carrying out sign flips to a subset of the measurement matrix. In this way, the first-class decoder, who can access full knowledge of the measurement matrix, can retrieve the signal faithfully while the second-class decoder, who can only access partial knowledge of the measurement matrix, subjects to a quality degradation during reconstruction. The work was later extended to multi-class low-complexity CS-based encryption [13].

Another research area of the secrecy of CS lies in the information theory frame. It was shown in [14] that CS-based cryptosystems fail to satisfy both Shannon's and Wyner's perfect Secrecy. In this context, Cambareri *et al.* [13] defined an achievable security metric, i.e., asymptotic spherical security, for CS-based ciphers. Basically, it states that the statistical properties of the random measurements only leak information about the plaintexts' energy. Based on this observation, Bianchi *et al.* [15] suggested that re-normalizing every measurement vector and treating the normalized measurements as the ciphertext will lead to a perfect "securized" CS-based cipher with the help of an auxiliary secure channel to transmit the energy of the real measurement vector.

It should be noted that all the above security features of CS-based ciphers are obtained under limited attack models, i.e., the adversary is permitted to work out the secret key or plaintext from ciphertext only or to search the entire key space. Under more threatening scenarios, such as known-plaintext attack (KPA) and chosen-plaintext attack (CPA), the adversary can easily reveal the measurement matrix (secret key in a CS-based cipher) if he is able to collect sufficient amount of independent plaintexts. As such, to maintain their respective security features, all the results mentioned above must work in a one-time-sampling (OTS) manner, i.e., the measurement matrix is never re-used.

Assume that a $K \times M$ measurement matrix is produced by using a secure deterministic random number generator (SDRNG) from a secret key shared between the encoder and decoder. We note that this is exactly the case of the traditional one-time-pad (OTP) cipher [16]. If a sparse signal belongs to $\{0, 1\}^M$, it requires exactly M bits to perfectly protect this signal when OTP cipher is applied. For the case of OTS, it requires at least $K \times M$ bits (if the Bernoulli matrix is used) to sample (encrypt) the signal. From this sense, the OTS CS-based cipher indeed reduces the service life of the SDRNG. Meanwhile, generating a different measurement matrix for every signal could be energy-consuming. Additionally, for engineering practice, using the same measurement matrix for multiple signals or signal segments flavors the subsequent source coding stage of multimedia data sensing, as discussed in [17], [18]. Based on these observations, it is concluded that investigating the behavior of CS-based ciphers under the multi-time-sampling (MTS) scenario is important from both the cryptographic and engineering points of view.

The work presented in [19] offers an intimate view for MTS CS-based cipher, where a second-class user in the two-class CS encryption [13] tries to upgrade the recovery quality by studying only one pair of known-plaintext and ciphertext. Restricting the measurement matrix to the form of Bernoulli matrix, it is shown in [19] that the number of candidate measurement matrices matching a single pair of known plaintext and ciphertext is too huge for the adversary to search for the true one. Still, the result only holds for a single plaintext-ciphertext pair while in typical KPA the adversary can access a large amount of plaintexts and the corresponding ciphertexts. Thus, the true measurement matrix may be determined uniquely. The same argument also applies to the case of CPA.

A straight forward solution to support the usage of CS in MTS scenario is to encrypt the entire or only the significant part of the quantized measurement vector using some conventional cryptographic method, such as AES or RSA. However, as we mentioned earlier, a standalone encryption layer can be too costly for a CS sensor and this approach does not take advantage of the confidentiality provided by CS itself.

Another approach to achieve this goal is to embed other efficient cryptographic primitives in the the CS encoding process. This is exactly the idea of product cipher introduced by Shannon [16], who suggested combining two or more cryptographic primitives together such that the product is more secure than individual component against cryptanalysis.

In [20], Zeng *et al.* proposed a speech encryption algorithm by scrambling the CS measurements. A similar idea was later applied for secure remote image sensing [21]. For the purpose of image acquisition and confidentiality, Zhang *et al.* [22] suggested scrambling the frequency coefficients before the CS encoding instead of scrambling the CS samples. Note that scrambling the frequency coefficients is a mature technique for multimedia confidentiality in traditional coding system [23], the main advantage of employing this technique in the CS paradigm is that a so-called "acceptable" permutation can make the column (or row) sparsity level of 2D signals uniform [24], thus relaxing the restricted isometry property (RIP) of the measurement matrix and flavoring a parallel CS (PCS) reconstruction model. The same technique is also used for privacy protection in cloud-assisted image service [25]. Another popular approach to form product cipher for MTS usage of compressive imaging is to employ an optical encryption primitive, i.e., double-random phase encoding (DRPE) technique, such as those proposed in [26]–[28]. Once the MTS usage of CS-based cipher is satisfied, it could be integrated with other data-layer cryptographic components for higher-security environments. For example, the work in [29] considered embedding secret nonlinear diffusion into the measurements quantization stage to ensure privacy for imaging.

Although the above mentioned product ciphers are efficient, generally they cannot resist CPA in MTS scenario (this issue will be discussed in detail in Sections II-B and II-C). The reason for the difficulty in applying CS-based cipher for MTS usage is due to the characteristic of CS itself: 1) the signal to be sensed must be sparse; 2) the encoding process is linear. For this reason, embedding some high-security primitives before CS encoding will probably make the signal noise-like and not sparse anymore. On the other hand, the introduction of any non-linear cryptographic primitive in CS paradigm will break the linearity of the sampling process and make the recovery infeasible.

Our work moves one step further for the usage of CS-based cipher under MTS scenario. Start with a RIPless reconstruction observation, we study how to embed security features in sparsifying bases under the sparse constraint. In more detail, we suggest a bi-level protected CS (BLP-CS) framework, which can be viewed as a product cipher of the basic CS model and transform-domain encryption technique under the sparse constraint. In particular, we propose several techniques to construct secret key-related sparsifying basis and incorporate them into our BLP-CS model. At the encoding stage, this model can be viewed as a new design of the measurement matrix, thus the encoding is the same as that of the original CS model. However, a successful decoding requires knowledge of the key-dependent sensing matrix and key-related sparsifying basis. In this way, the new product cipher can resist CPA.

This paper makes two contributions in the area of embedding secrecy in CS. On the one hand, we propose a CPA-resistant product cipher by utilizing the confidentiality provided by CS. To the best of our knowledge, this is the first reprot that the CS-based (product) cipher can resist CPA. On the other hand, we incorporate a cryptographic permutation to the CS encoding

stage, thus relaxing the RIP of the measurement matrix and flavoring a PCS reconstruction for 2D sparse signals. In this sense, our work can be considered as an extension of the work presented in [24].

The rest of this paper is organized as follows. In Section II, we first review the CS framework and present the CPA on CS-based product ciphers. In Section III, two techniques for constructing secret key-related sparsifying basis are proposed to establish the bi-level protection model. Section IV presents comparisons of the OTS CS-based cipher and our BLP-CS model from complexity and security point of view. As an application example, the new model is used to sample digital images in Section V. The superiority of the new CS-based image cipher is justified by both theoretical analyses and simulation results. Our work is concluded in Section VI.

II. SECURITY DEFECTS OF EXISTING CS-BASED CIPHERS IN MTS SCENARIO

As we mentioned earlier, there exists some effort to support CS-based cipher for MTS usage [20]–[22], [25]–[28]. In this section, we report the fact that all of them fail to resist CPA. To begin with, we briefly review the theory of CS.

A. CS Preliminaries

We denote a 1D discrete signal to be sampled as a column vector $\mathbf{x} = (x_1, x_2, \ldots, x_M)^T$. 2D signals of size $M = n \times n$, $\mathbf{X} = [\mathbf{X}_{i,j}]_{i=1,j=1}^{n,n}$, can be vectorized to 1D format as $\mathbf{x}$ by stacking the columns of $\mathbf{X}$, i.e., $\mathbf{x} = \text{vec}(\mathbf{X})$. $\mathbf{x}$ is said to be k-sparse under $\mathbf{\Psi}$ if there exists a certain sparsifying basis $\mathbf{\Psi} = \{\psi_{i,j}\}_{i=1,j=1}^{M,M}$ such that $\mathbf{x} = \mathbf{\Psi}\mathbf{s}$ and $\|\mathbf{s}\|_0 = \#\{\text{supp}\mathbf{s}\} = \#\{i : s_i \neq 0\} = k << M$.

Here, we emphasize that in almost all of the works about the secrecy of CS, such as [9], [13], [15], [19], [20], [28], the role of the basis is ignored or simply treated as an orthnormal matrix. We relax the requirement of the basis to an invertible matrix in this work. The encoding process during CS is a linear projection, i.e.

$$\mathbf{y} = \mathbf{\Phi}\mathbf{x} = \mathbf{\Phi}\mathbf{\Psi}\mathbf{s} = \mathbf{A}\mathbf{s} \tag{1}$$

if the sampling is perform in the space/time domain, or equivalently

$$\mathbf{y} = \mathbf{\Phi}\mathbf{s} = \mathbf{\Phi}\mathbf{\Psi}^{-1}\mathbf{x} = \mathbf{A}\mathbf{s} \tag{2}$$

if the sampling is performed in the frequency domain.

The revolutionary finding of CS is that the K dimensional measurement vector $\mathbf{y}$ reserves all the information required for unique and stable recovery of $\mathbf{x}$ even if $k < K \ll M$ provided that the measurement matrix $\mathbf{A}$ obeys some information-preserving guarantees [4], [30]–[32]. Since the linear systems (1) and (2) are undetermined, both of them have infinite solutions. Considering the signal is sparse, the intuitive way to restore $\mathbf{x}$ is to solve the l_0 optimization problem

$$\min \|\mathbf{s}\|_0 \text{ subject to } \mathbf{y} = \mathbf{A}\mathbf{s} \tag{3}$$

to obtain $\mathbf{s}$ and then recover $\mathbf{x}$ by $\mathbf{x} = \mathbf{\Psi}\mathbf{s}$. As stated in [33], solving this problem is NP-hard because it requires an exhaustive search over all subsets of columns of $\mathbf{A}$.

The convex relaxed form of problem (3) can be expressed as

$$\min \|\mathbf{s}\|_1 \text{ subject to } \mathbf{y} = \mathbf{A}\mathbf{s}. \tag{4}$$

As proved in [4], the solution of the l_1 problem (4) is identical to that of (3) with overwhelming probability provided that $\mathbf{A}$ satisfies RIP. Examples of widely accepted matrices satisfying RIP including Gaussian ensemble and Bernoulli ensemble with $K = O(k \log M)$ rows. Up to a logarithmic factor, the number of measurements is optimal [4]. Here we note that all the previously mentioned approaches of embedding secrecy into CS-based (product) ciphers work with RIP.

Definition 1: [30] A matrix $\mathbf{A}$ of size $K \times M$ is said to satisfy the restricted isometry property of order k if there exists a constant $\delta_k \in (0, 1)$ such that

$$(1 - \delta_k)\|\mathbf{x}^{(T)}\|_2^2 \leq \|\mathbf{A}^{(T)}\mathbf{x}^{(T)}\|_2^2 \leq (1 + \delta_k)\|\mathbf{x}^{(T)}\|_2^2$$

holds for all column indices sets T with $\#T < k$, where $\mathbf{A}^{(T)}$ is a $K \times \#T$ matrix composed of the columns indexed by T, $\mathbf{x}^{(T)}$ is a vector obtained by retaining only the entries indexed by T and $\|\cdot\|_2$ denotes the l_2 norm of a vector.

More generally, let the K rows of $\mathbf{A}$, i.e., $\mathbf{a}_1^T, \ldots, \mathbf{a}_K^T$, be i.i.d. random vectors drawn from a distribution, say F. The recently developed RIPless CS theory states that the solution of problem (4) is unique and equal to that of problem (3) if the number of measurements grows proportionally to the product of coherence parameter and the condtion number of the covariance matrix [31], [32], as given by Theorem 1.

Theorem 1: [32] Let $\mathbf{s}$ be a k-sparse vector and $\omega \geq 1$. The solution of problem (4) is unique and equal to that of problem (3) with probability at least $1 - e^{-\omega}$ if the number of measurements fulfills

$$K = O(\mu(F)\theta \cdot \omega^2 k \log M)$$

where μ, the coherence parameter, is the smallest number that

$$\max_{1 \leq i \leq M} | < \mathbf{a}^T, \mathbf{e}_i > | \leq \mu(F)$$

and θ is the condition number of the covariance matrix $\Sigma = \mathbb{E}[\mathbf{a}\mathbf{a}^T]^{1/2}$ with $\mathbf{a}^T$ being a generic row random vector draw from F and $\mathbf{e}_i$ being the canonical basis vector of dimension M.

What concerns us about the RIP CS and RIPless CS is that the quantity $\mu(F)\theta$ that governs the number of required measurements for successful l_1 reconstruction is different. For Gaussian, Bernoulli and partial Fourier matrices, it is shown that $\mu(F)\theta = O(1)$ in [31]. Moreover, it is easy to find out that $\theta = 1$ for unitary matrix and $\theta > 1$ for generic matrix.[1] Moreover, the larger the value of $\mu(F)\theta$, the more the samples we need for exact reconstruction in the RIPless setting. We make us of this fact to design the measurement matrix for security purpose.

---

[1]Recall that condition number is the absolute value of the ratio between the largest and smallest singular values.

In the subsequent sections, we will show that almost all the CS-based product ciphers mentioned above, i.e., those proposed in [20]–[22], [25]–[28], fail to resist the CPA under MTS scenario due to the fact that these product ciphers work only under the RIP framework.

B. Scrambling in the Measurements Domain or the Frequency Domain

As described in the previous sections, it is more practical if the same measurement matrix can be re-used multiple times. To this end, there are some attempts trying to incorporate other low-complexity cryptographic primitives to fix the intrinsic security defect of CS in a manner of constructing product ciphers [20]–[22], [25]. A common cryptographic technique suitable for this purpose is scrambling (also known as random permutation), which has been widely used in the field of multimedia security [6], [23]. It should be noted that the works mentioned here and Section II-C are based on the RIP theory. Here, we treat the measurement matrix as Gaussian matrix for simplicity.[2]

Roughly speaking, existing works utilizing scrambling for MTS usage of CS can be divided into two classes[3]:

1) Scrambling is performed on the measurements, such as [20], [21];
2) Scrambling is done in the frequency domain, such as [22], [25].

The scrambling process can be characterized by a permutation matrix, which is a square binary matrix that has exactly one non-zero element with value 1 in each row and each column and 0s elsewhere.

According to (1), class 1 CS-based product cipher can be expressed as

$$\hat{\mathbf{y}} = \mathbf{P}_K \mathbf{y} = \mathbf{P}_K \boldsymbol{\Phi}_K \mathbf{x} = \mathbf{P}_K \boldsymbol{\Phi}_K \boldsymbol{\Psi} \mathbf{s} \qquad (5)$$

where $\mathbf{x}$ is a k-sparse signal with dimension M to be sampled (encrypted), $\boldsymbol{\Psi}$ is an orthnormal sparsifying basis, $\mathbf{P}_K$ is a $K \times K$ permutation matrix, $\boldsymbol{\Phi}_K$ is the Gaussian ensemble and $\hat{\mathbf{y}}$ is the ciphertext to be transmitted or store. A difference between this class of product cipher and the basic CS-based ciphers is that the (equivalent) secret key for the product cipher is the permutation matrix $\mathbf{P}_K$ and the measurement matrix $\boldsymbol{\Phi}_K$ while only measurement matrix can be utilized as the key in basic CS-based ciphers. Ideally (from the designer's point of view), the decoding (decryption) is composed of a two-step reconstruction, i.e.

$$\mathbf{y} = \mathbf{P}_K \hat{\mathbf{y}}$$

$$\min \|\mathbf{s}\|_1 \text{ subject to } \mathbf{y} = \boldsymbol{\Phi}_K \boldsymbol{\Psi} \mathbf{s}.$$

However, since both $\mathbf{P}_K$ and $\boldsymbol{\Psi}$ are orthonormal, $\mathbf{P}_K \boldsymbol{\Phi}_K \boldsymbol{\Psi}$, which is a rotation of $\boldsymbol{\Phi}_K$, possess the distribution of a Gaussian ensemble. Governed by the RIP theory, we can simplify the

---

[2]This simplification will not affect the security level of the discussed product cipher.

[3]Note that embedding scrambling in the time domain actually brings no benefit to security enhancement, but it helps the construction of a structural sampling ensemble [34].

decoding as a single-step optimization

$$\min \|\mathbf{s}\|_1 \text{ subject to } \hat{\mathbf{y}} = \mathbf{P}_K \boldsymbol{\Phi}_K \boldsymbol{\Psi} \mathbf{s} = \mathbf{P}_K \boldsymbol{\Phi}_K \mathbf{x}.$$

An unauthorized decoder, who can collect ciphertext for any plaintext in CPA scenario, submits a series of artificial signals $\{\mathbf{x}_j\}_{j=1}^M = \{(0, \ldots, 0, 1_j, 0, \ldots, 0)^T\}_{j=1}^M$ to the encryption oracle and concludes $\mathbf{P}_K \boldsymbol{\Phi}_K = [\hat{\mathbf{y}}_1, \ldots, \hat{\mathbf{y}}_M]$ using (5). It is clear that any further using of the same measurement and permutation matrices for security purpose is doomed to fail.

For the class 2 CS-based product ciphers, the same treatment can be applied. According to model (2), we can rewrite the encoding (encryption) process as

$$\hat{\mathbf{y}} = \boldsymbol{\Phi}_K \mathbf{P}_K \mathbf{s} = \boldsymbol{\Phi}_K \mathbf{P}_K \boldsymbol{\Psi}^{-1} \mathbf{x}.$$

Once again, $\boldsymbol{\Phi}_K \mathbf{P}_K$ can jointly work as the measurement matrix and it can be revealed by M independent chosen plaintexts and their corresponding ciphertexts.

In the following discussion, we will explain how scrambling (known as "acceptable permutation in [24]) relaxes the RIP requirement of the measurement matrix for 2D sparse signals. Without loss of generality, let $\mathbf{X} = [\mathbf{X}_{i,j}]_{i=1, j=1}^{n,n}$ be a 2D signal sparse in the canonic sparsifying basis and $\mathbf{k} = (k_1, k_2, \ldots, k_n)$ be a row vector whose entry denotes the number of nonzero elements of the columns of $\mathbf{X}$. A column by column sampling process of $\mathbf{X}$ can be summarized as

$$\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n] = \boldsymbol{\Phi} \mathbf{X} = \boldsymbol{\Phi} [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n]$$

or equivalently

$$\text{vec}(\mathbf{Y}) = [\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n]^T = \bar{\boldsymbol{\Phi}} \text{vec}(\mathbf{X}) = \bar{\boldsymbol{\Phi}} [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n]^T$$

where

$$\bar{\boldsymbol{\Phi}} = \begin{bmatrix} \boldsymbol{\Phi} & & & \\ & \boldsymbol{\Phi} & & \\ & & \ddots & \\ & & & \boldsymbol{\Phi} \end{bmatrix}.$$

The corresponding parallel (column by column) reconstruction is given by

$$\min \|\mathbf{x}_j\|_1 \text{ subject to } \mathbf{y}_j = \boldsymbol{\Phi} \mathbf{x}_j \qquad (6)$$

where $j \in \{1, 2, \ldots, n\}$ and $\boldsymbol{\Phi}$ being a typical RIP measurement matrix with $O(\|\mathbf{k}\|_\infty \cdot \log n)$ rows. As we can see, the accurate reconstruction is proportional to $\|\mathbf{k}\|_\infty$ [24]. The smaller $\|\mathbf{k}\|_\infty$ is, the fewer rows $\boldsymbol{\Phi}$ require for correct recovery or the worse RIP constant $\boldsymbol{\Phi}$ can stand.

The remaining work is to demonstrate that $\|\mathbf{k}\|_\infty$ of $\mathbf{X}$ will decrease with large probability if $\mathbf{X}$ is randomly scrambled. Let $\text{vec}(\bar{\mathbf{X}}) = \mathbf{P} \cdot \text{vec}(\mathbf{X})$ and $\bar{\mathbf{k}} = (\bar{k}_1, \ldots, \bar{k}_n)$ be the sparsity vector of $\bar{\mathbf{X}}$, we define an acceptable permutation as follows:

Definition 2: A $n^2 \times n^2$ permutation $\mathbf{P}$ is said to be acceptable if the following two rules are satisfied:

1) the expectations of the column sparsity of $\bar{\mathbf{X}}$ are the same, i.e., each column expects the same sparsity level;
2) the probability that $\|\bar{\mathbf{k}}\|_\infty$ deviates from the expected sparsity level observe a power law decay.

The following property demonstrates the role of (secret) random scrambling for 2D signals which is sparse in space. By swapping time and frequency, reconstruction model (6) can be applied to natural 2D signals, such as images. The examples demonstrating this phenomenon will be provided in Section V.

Property 1: Uniform random permutation is an *acceptable* permutation for any $n \times n$ 2D sparse signal $\mathbf{X}$.

Proof: To prove this, we recall that uniform random permutation refers to choosing a permutation from all the $(n^2)!$ candidates with equal probability. In other words, each non-zero entry of $\mathbf{X}$ will appear at any location of $\bar{\mathbf{X}}$ with probability $1/n^2$ when $\mathbf{X}$ is processed by uniform random permutation.

Since there are $\|\mathbf{k}\|_1$ non-zero entries of $\mathbf{X}$ in total, each entry of its permutated version is nonzero with probability $\|\mathbf{k}\|_1/n^2$. Apparently, the expected sparsity level of $\bar{\mathbf{x}}_j$ is $n \times \frac{\|\mathbf{k}\|_1}{n^2} = \|\mathbf{k}\|_1/n$, which meets the requirements of rule 1).

Treat each column of $\bar{\mathbf{X}}$ as realization of n independent, identically distributed random variables, the probability that $\|\bar{\mathbf{k}}\|_\infty$ deviates from the expectation $\|\mathbf{k}\|_1/n$ by t can be characterized by

$$\mathrm{Prob}((\|\bar{\mathbf{k}}\|_\infty - \|\mathbf{k}\|_1/n \geq t) = \mathrm{Prob}((\max_j(\bar{k}_j) - \|\mathbf{k}\|_1/n) \geq t)$$
$$\leq \mathrm{Prob}((\bar{k}_j - \|\mathbf{k}\|_1/n) \geq t)$$
$$\leq e^{-2nt^2}$$

where the last inequality is obtained by applying Hoeffding inequality. Hence finishes the proof. ∎

C. Concatenation of CS and DRPE

As one of the optical information processing technique, image encryption using DRPE has received a lot of research attention since its first appearance in [35], [36]. This cipher was found insecure against various plaintext attacks [37], [38]. In a different context, CS offers a new approach for hologram compression and sensing in the optical domain [39], [40]. On the one hand, the concatenation of CS and DRPE enjoys a all-optical implementation and substantially data volume reduction. On the other hand, the secrecy provided by CS may enhance the security level of DRPE, and vice visa. These reasons making cascading CS and DRPE be a noticeable alternative to support the MTS usage of CS. In the following discussion, we will point out that the later argument is questionable in MTS scenario since the CPA complexity of this model is exactly the same as that of the basic CS model.

Considering a discrete and bounded[4] 2D data $\mathbf{I} = [\mathbf{I}_{i,j}]$, the DRPE encryption can be formulated as

$$\mathbf{C}_{i,j} = \mathcal{IF}(\mathcal{FT}(\mathbf{I}_{i,j} \cdot \exp(j2\pi p_{i,j})) \cdot \exp(j2\pi q_{u,v}))$$

where the random spatial phase mask $\mathbf{P} = [\exp(j2\pi p_{i,j})]$ and the random frequency phase mask $\mathbf{Q} = [\exp(j2\pi q_{u,v})]$ are the secret keys, and $\mathcal{FT}(\mathbf{X}) = \mathbf{FXF}^*$ with $\cdot^*$ being the conjugate transpose and $\mathcal{IF}$ being the inverse Fourier transform. The DRPE decryption is omitted here since it is similar to the encryption process. With these notations, we can also divide the

encryption schemes based on concatenation of CS and DRPE into two classes:

1) CS encryption followed by DRPE [26];
2) DRPE followed by CS encryption [27], [28].

Considering a 2D image $\mathbf{X}$ with $M = n \times n$ pixels is sensed by CS with $K = m \times m$ measurements, the algorithms of class 1 can be modeled as a separate two-step process, i.e.

$$\mathrm{vec}(\mathbf{Y}) = \boldsymbol{\Phi}\,\mathrm{vec}(\mathbf{X})$$
$$\mathbf{C} = \mathcal{IF}(\mathcal{FT}(\mathbf{Y}_{i,j} \cdot \exp(j2\pi p_{i,j})) \cdot \exp(j2\pi q_{u,v})) \quad (7)$$

where $\boldsymbol{\Phi}_{m^2 \times n^2}$, $\mathbf{P}_{m \times m} = [\exp(j2\pi p_{i,j})]$ and $\mathbf{Q}_{m \times m} = [\exp(j2\pi q_{u,v})]$ serve as the (equivalent) secret key in the whole process and $\mathbf{C}$ is the ciphertext to deliver or display. As claimed in [26], decoding $\mathbf{C}$ should observe a separate DRPE decryption and CS reconstruction, or by a reversed order in algorithms belonging to class 2 [27], [28]. As such, it is demonstrated that an unauthorized user who cannot access full knowledge of $\boldsymbol{\Phi}$, $\mathbf{P}$ and $\mathbf{Q}$ is not able to decrypt $\mathbf{X}$ [26]–[28].

We investigate the real strength against CPA for the approaches mentioned above by first rewriting (7) as a matrix form [38], i.e.

$$\mathrm{vec}(\mathbf{C}) = \mathbf{T}\,\mathrm{vec}(\mathbf{Y})$$
$$= \bar{\mathbf{F}}^* \bar{\mathbf{Q}} \bar{\mathbf{F}} \bar{\mathbf{P}} \cdot \mathrm{vec}(\mathbf{Y})$$

where $\bar{\mathbf{F}}_{m^2 \times m^2}$ is the Kronecker product of the Fourier matrices $\mathbf{F}^*$ and $\mathbf{F}$, $\bar{\mathbf{P}}_{m^2 \times m^2} = \mathrm{diag}(\mathrm{vec}(\mathbf{P}))$ and $\bar{\mathbf{Q}}_{m^2 \times m^2} = \mathrm{diag}(\mathrm{vec}(\mathbf{Q}))$ are the DRPE secret key. By construction, $\bar{\mathbf{P}}$ and $\bar{\mathbf{Q}}$ are unitary matrices. So, it is concluded $\mathbf{T}$ is also a unitary matrix. In this concern, $\mathbf{T}\boldsymbol{\Phi}$ must be a RIP matrix and thus a single-step optimization can be formulated as[5]

$$\min \|\boldsymbol{\Psi}^{-1} \cdot \mathrm{vec}(\mathbf{X})\|_1 \quad \text{subject to } \mathrm{vec}(\mathbf{C}) = \mathbf{T}\boldsymbol{\Phi}\,\mathrm{vec}(\mathbf{X}).$$

Once again, the attacker who works under CPA assumption can retrieve $\mathbf{T}\boldsymbol{\Phi}$ faithfully from M independent plaintexts and the corresponding ciphertexts. Moreover, he can use this information to decode (decrypt) any subsequent ciphertexts. Similarly, we can apply the analyses to class 2 algorithms and obtain the same conclusion.

III. THE PROPOSED SCHEME

As reviewed in the previous section, existing proposals [20]–[22], [25]–[28] targeting the MTS usage of CS as joint sampling and data protection mechanism fail to resist plaintext attacks. Similarly, it can be concluded that cascading CS, scrambling and DRPE also suffer from the same defect, such as the one suggested in [42]. The underlying reason is that all these three cryptographic primitives are linear and we can always translate the encoding components to a (equivalent) RIP-based measurement matrix. Therefore, the key question is whether it is possible to construct a more secure CS-based product cipher without introducing any computing-intensive cryp-

---

[4]This always holds true given that continuous data can be adequately sampled.

[5]We note that the multiple measurement vector CS model [41] should be adopted since $\mathbf{T}$ is a complex matrix.
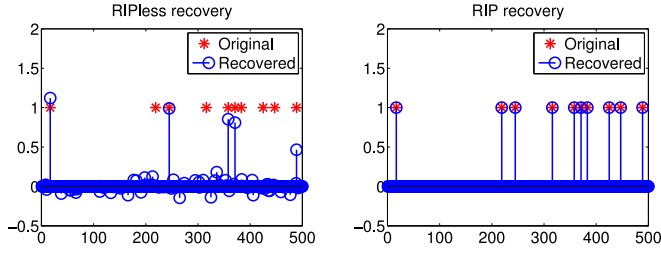
Fig. 1. Example of RIPless reconstruction and RIP reconstruction.

tographic primitives. We will give a positive solution to this problem by switching from the RIP measurement matrix construction to the RIPless matrix construction. We start with the following example.

Consider a column vector $\mathbf{x}$ of length $M = 500$ taking values from $\{0, 1\}$ has a sparsity level $k = 10$. Let F denotes an independent multivariate antipodal distribution, which is given by $F = \{\pm d_1\} \times \{\pm d_2\} \times \ldots \times \{\pm d_M\}$ with $\text{Prob}(d_j) = \text{Prob}(-d_j) = 1/2$ and $\{d_j\}_{j=1}^{M}$ be positive integers. We take 60 sensing vectors[6] from this distribution and get a measurement matrix $\mathbf{\Phi}$ which is further used to sample $\mathbf{x}$. By Definition 1, $\mathbf{\Phi}$ cannot guarantee energy-preserving property thus it is a RIPless matrix. By construction, we have $\theta = O(\max_j(d_j)/\min_j(d_j))$ and

$$\mu(F) \geq \max_{1 \leq i \leq M} | < \phi^T, \mathbf{e}_i > |$$
$$= \max_j(d_j).$$

In summary, $\mu(F)\theta = O(\max_j(d_j^2)/\min_j(d_j))$ is a non-negligible term and the following straightforward recovery dominated by RIPless theory (see Theorem 1 for detail)

$$\min \|\bar{\mathbf{x}}\|_1 \text{ subject to } \mathbf{y} = \mathbf{\Phi}\bar{\mathbf{x}}$$

returns a solution $\bar{\mathbf{x}} \neq \mathbf{x}$. Set $\mathbf{A} = \mathbf{\Phi}\mathbf{D} = \mathbf{\Phi} \cdot \text{diag}(1/d_1, \ldots, 1/d_M)$, the reconstruction can be also transformed to a two-step reconstruction, which is compliant with the RIP theory, after realizing that $\mathbf{A}$ is a Bernoulli matrix, i.e.

$$\min \|\hat{\mathbf{x}}\|_1 \text{ subject to } \mathbf{y} = (\mathbf{A}\mathbf{D}^{-1})\mathbf{x} = \mathbf{A}\hat{\mathbf{x}}$$
$$\bar{\mathbf{x}} = \mathbf{D}\hat{\mathbf{x}}.$$

We compare the recovery techniques described above. Fig. 1 depicts a typical reconstruction result with $d_j \in [1, 60]$, from which we can see that the recovery in the RIP case is exact but the RIPless case is not due to a lack of sufficient measurements.

The above example provides a preparatory understanding of how a RIPless matrix construction can be transformed to a RIP one. Still, it cannot be considered as a good CS-based cipher since an attacker can reveal $\mathbf{D}$ from $\mathbf{\Phi}$ by $d_j = |\mathbf{\Phi}_{i,j}|$. Moreover, this technique only works for vector who is sparse in the canonical basis, which is not practical for real signals. In this concern, we apply this finding to the CS model (2) and devise a so called bi-level protected CS model in a way that the measurement matrix does not satisfy RIP (we call it non-RIP matrix hereinafter) and the reconstruction works under RIP theory.

[6]Here, we take $K = 60$ because $K > 4k$ is an empirical threshold for exact CS recovery in the RIP theory [2].
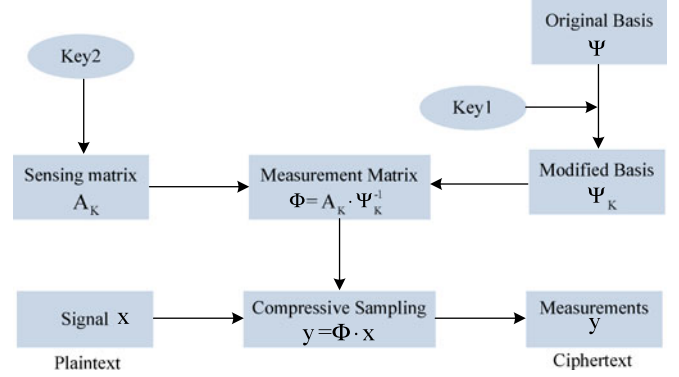

Fig. 2. Block diagram of BLP-CS.

The BLP-CS model will be described in Section III-A, which can be viewed as product of the CS-based cipher and a transform encryption. Then we propose two methods for key-related sparsifying transformation design, namely, *Type I Secret Basis* and *Type II Secret Basis*.

A. Bi-level Protection Model

The block diagram of this model is shown in Fig. 2, where we suggest using key-dependent sensing matrix, $\mathbf{A}_K$, and secret-related sparsifying basis, $\mathbf{\Psi}_K$, to determine the measurement matrix $\mathbf{\Phi} = \mathbf{A}_K \mathbf{\Psi}_K^{-1}$. Recalling the above example, we are interested in the phenomenon that the measurement matrix $\mathbf{\Phi}$ does not satisfy the RIP requirement, while the key-dependent sensing matrix $\mathbf{A}_K$ itself is a RIP matrix. Referring to (2), the sampling procedure can be expressed as

$$\mathbf{y} = \mathbf{\Phi}\mathbf{x} = \mathbf{A}_K \mathbf{\Psi}_K^{-1}(\mathbf{\Psi}_K \mathbf{s}) = \mathbf{A}_K \mathbf{s}.$$

It should be noted that the number of measurements (sampling rate) is on the order of $(k \log M)$ even though $\mathbf{\Phi}$ is a non-RIP measurement matrix. This number of measurements fails to meet the minimum requirement defined in Theorem 1, thus making the correct decoding from $\mathbf{\Phi}$ an impossible task.

To correctly decode (decrypt) $\mathbf{y}$, a legitimate user should first derive $\mathbf{A}_K$ and $\mathbf{\Psi}_K$ from the key scheduling process and then refer to the following two-step reconstruction

$$\min \|\mathbf{s}\|_1 \text{ subject to } \mathbf{y} = \mathbf{\Phi}\mathbf{x} = \mathbf{A}_K \mathbf{s}$$
$$\mathbf{x} = \mathbf{\Psi}_K \mathbf{s}$$

or equivalently

$$\min \|\mathbf{\Psi}_K^{-1}\mathbf{x}\|_1 \text{ subject to } \mathbf{y} = \mathbf{\Phi}\mathbf{x}.$$

To fulfill the security requirement, the remaining task is to design two matrices $\mathbf{A}_K$ and $\mathbf{\Psi}_K$ satisfying:

RULE a. $\mathbf{A}_K$ is a key-related matrix satisfy RIP;

RULE b. $\mathbf{\Psi}_K$ is a key-related sparsifying basis;

RULE c. $\mathbf{A}_K \mathbf{\Psi}_K^{-1}$ is a structural non-RIP matrix.

The work of designing a RIP matrix is trivial since it is already clear that Guussian/Bernoulli [4] and structurally random matrices [34] are competent for this task with overwhelming probability. Therefore, we focus our attention on the designing of $\mathbf{\Psi}_K$ in the following discussions. It is worth mentioning that

the work of designing $\mathbf{\Psi}_K$ satisfying RULE b (also known as transform encryption) is very popular in the filed of multimedia encryption, and examples can be found in [43]–[45]. However, the work of designing $\mathbf{A}_K$ and $\mathbf{\Psi}_K$ satisfying RULE c is totally new.

B. Type I Secret Basis

The first type of secret basis that drawn our attention is the parameterized construction of some familiar transform, such as parameterized discrete wavelet transform (DWT) [44], [46] and directional discrete cosine transfrom (DCT) [43], [47]. Here, we present a parameterized transform based on Fractional Fourier Transform (FrFT) as an example.

The use of FrFT for security purpose can be dated back to year 2000, when Unnikrishnan *et al.* [48] suggested to use FrFT for DRPE instead of the ordinary Fourier transform [35], in order to benefit from its extra degrees of freedom provided by the fractional orders. Generally speaking, performing an order α FrFT on a signal can be viewed as a rotation operation on the time-frequency or space-frequency distribution at an angle α. Though FrFT is very popular in optics for its easy implementation, it is not preferred in digital world since complex numbers always cause extra computational load.

To this end, Venturini and Duhamel proposed a method to construct Reality-Preserving FrFT of arbitrary order [49]. Here, we deduce the Reality-Preserving Fractional Cosine Transform (RPFrCT) by the virtue of their method. Denote the discrete cosine transform [50] of size $n \times n$ by

$$\mathbf{C} = \left(\frac{1}{\sqrt{n}} \epsilon_l \cos\left(2\pi \frac{(2i+1)l}{4n}\right) \right)$$

where $i = 0 \sim n-1$, $l = 0 \sim n-1$, $\epsilon_0 = 1$ and $\epsilon_l = \sqrt{2}$ for $l > 0$. The unitary property of $\mathbf{C}$ assures that it can be diagonalized as

$$\mathbf{C} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^* \tag{8}$$

where $\mathbf{U} = \{\mathbf{u}_i\}_{i=1}^n$ is composed of n orthonormal eigenvectors, i.e., $\mathbf{u}_m^* \mathbf{u}_i = \delta_{mi}$ and $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_i, \ldots, \lambda_n)$ with $\lambda_i = \exp(j\varphi_i)$. Replace λ_i with its α-th power λ_i^α in Eq. (8), we can express the Discrete Fractional Cosine Transform (DFrCT) matrix $\mathbf{C}_\alpha$ of order α in the compact form

$$\mathbf{C}_\alpha = \mathbf{U}\mathbf{\Lambda}^\alpha \mathbf{U}^*.$$

Having defined $\mathbf{C}_\alpha$, we can derive the RPFrCT matrix $\mathbf{R}_\alpha$ as follows:

1) For any real signal $\mathbf{x} = \{x_l\}_{l=1}^M$ of length M (M is even), construct a complex signal of length $M/2$ by

$$\widetilde{\mathbf{x}} = \{x_1 + jx_{M/2+1}, x_2 + jx_{M/2+2}, \ldots, x_{M/2} + jx_M\}.$$

2) Compute $\widetilde{\mathbf{y}} = \mathbf{B}_\alpha \widetilde{\mathbf{x}}$, where $\mathbf{B}_\alpha$ is a DFrCT matrix of size $(M/2 \times M/2)$, namely, $\mathbf{B}_\alpha = \mathbf{C}_{\alpha, M/2}$.
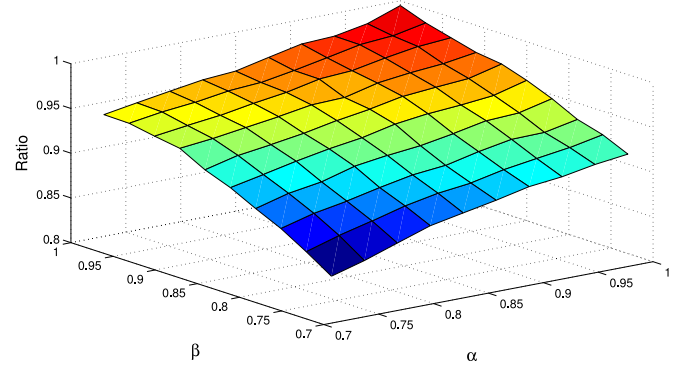

Fig. 3. Comparison between the recovery result of RPFrCT and DCT2 using the best s-term approximation at different fractional orders.

3) Determine the RPFrCT matrix $\mathbf{R}_\alpha$ by

$$
\begin{aligned}
\mathbf{y} &= (\mathrm{Re}(\widetilde{\mathbf{y}}), \mathrm{Im}(\widetilde{\mathbf{y}}))^T \\
&= \begin{pmatrix} \mathrm{Re}(\mathbf{B}_\alpha)\mathrm{Re}(\widetilde{\mathbf{x}}) - \mathrm{Im}(\mathbf{B}_\alpha)\mathrm{Im}(\widetilde{\mathbf{x}}) \\ \mathrm{Im}(\mathbf{B}_\alpha)\mathrm{Re}(\widetilde{\mathbf{x}}) + \mathrm{Re}(\mathbf{B}_\alpha)\mathrm{Im}(\widetilde{\mathbf{x}}) \end{pmatrix} \\
&= \begin{pmatrix} \mathrm{Re}(\mathbf{B}_\alpha) & -\mathrm{Im}(\mathbf{B}_\alpha) \\ \mathrm{Im}(\mathbf{B}_\alpha) & \mathrm{Re}(\mathbf{B}_\alpha) \end{pmatrix} \cdot \begin{pmatrix} \mathrm{Re}(\widetilde{\mathbf{x}}) \\ \mathrm{Im}(\widetilde{\mathbf{x}}) \end{pmatrix} \\
&= \mathbf{R}_\alpha \mathbf{x}.
\end{aligned}
$$

From the construction process listed above, we can conclude that $\mathbf{R}_\alpha$ is orthogonal, reality preserving and periodic. Then, the RPFrCT of a digital image $\mathbf{X}$ is given by

$$\mathbf{S} = \mathbf{R}_\alpha \mathbf{X} \mathbf{R}_\beta^T \tag{9}$$

where $(\cdot)^T$ represents the transpose operator, α and β are the orders of the Fractional Cosine Transform along x and y directions, respectively. Equivalently, we can express this formula as

$$\mathrm{vec}(\mathbf{S}) = \mathbf{\Psi}^{-1} \mathrm{vec}(\mathbf{X})$$

where $\mathbf{\Psi}^{-1} = \mathbf{\Psi}^T = (\mathbf{R}_\beta \otimes \mathbf{R}_\alpha)$. To study the sparsifying capability of the proposed parameterized basis, we carry out experiments on digital images at different fractional orders α and β by using the best s-term approximation, i.e., keeping the s largest coefficients and setting the remaining ones to zero. The recovered result of RPFrCT is compared with that of DCT2 using the ratio between their peak signal-to-noise ratios (PSNRs). As expected, the sparsifying capability of RPFrCT raises when α or β increases, as shown in Fig. 3. When $\alpha, \beta \in (0.9, 1]$, the sparsifying capability of RPFrCT is comparable to that of DCT2. It is worth mentioning that a similar sparsifying capability was also observed when this transform is applied to 1D signals [49].

C. Type II Secret Basis

We have demonstrated a technique for parameterized sparsifying basis construction, where the free parameter can be used as the secret key in the BLP-CS model. In this way, the resultant basis satisfies RULE b. However, it still suffers from the same CPA shown in Section II since it fails to meet RULE c. In the

subsequent discussions, we propose three kinds of operations on an existing basis to make it fulfill RULE c. We start the deviation by defining equivalent sparsifying bases.

Definition 3: Two basis matrices, $\mathbf{\Psi}$ and $\mathbf{\Psi}'$ are equivalent sparsifying bases if $\mathbf{x} = \mathbf{\Psi}\mathbf{s} = \mathbf{\Psi}'\mathbf{s}'$, $\|\mathbf{s}\|_0 = \|\mathbf{s}'\|_0 = k$ holds for any signal $\mathbf{x}$.

Property 2: $\mathbf{\Psi}'$ and $\mathbf{\Psi}$ are equivalent sparsifying bases if

$$\begin{aligned} \mathbf{\Psi}' &= \mathbb{F}_1(\mathbf{\Psi}) \\ &= (d_1\psi_1, d_2\psi_2, \ldots, d_j\psi_j, \ldots, d_M\psi_M) \end{aligned}$$

where $\{d_j\}_{j=1}^M$ are non-zero constants and ψ_j is the jth column of $\mathbf{\Psi}$.

Proof: Set $s_j' = \frac{1}{d_j}s_j$ and we have $\|\mathbf{s}\|_0 = \|\mathbf{s}'\|_0$. ∎

We demonstrate that we are able to construct a non-RIP measurement matrix satisfying RULE c. Assume $\mathbf{\Psi}$ is an orthonormal basis and set

$$\mathbf{\Psi}' = \mathbf{\Psi}\mathbf{D}$$

where $\mathbf{D} = \mathrm{diag}(1/d_1, 1/d_2, \ldots, 1/d_M)$ and $\{d_j\}_{j=1}^M$ are positive integers drawn from certain distribution independently. Let $\mathbf{A}$ denote a Gaussian matrix with i.i.d. entries and calculate $\mathbf{\Phi}$ as

$$\begin{aligned} \mathbf{\Phi} &= \mathbf{A}(\mathbf{\Psi}\mathbf{D})^{-1} \\ &= \mathbf{A}\mathbf{D}^{-1}\mathbf{\Psi}^T. \end{aligned}$$

Once again, the effect of $\mathbf{\Psi}^T$ can be viewed as a rotation of $\mathbf{A}\mathbf{D}^{-1}$ in a M dimensional space, which is energy preserving. By construction, $\mathbf{\Phi}$ is a non-RIP matrix.

Property 3: $\mathbf{\Psi}'$ and $\mathbf{\Psi}$ are equivalent sparsifying bases if

$$\mathbf{\Psi}' = \mathbb{F}_2(\mathbf{\Psi}) = \mathbf{\Psi}\mathbf{P}$$

where $\mathbf{P}$ is a random permutation matrix.

Proof: Since $\mathbf{\Psi}\mathbf{s} = \mathbf{\Psi}(\mathbf{P}\mathbf{P}^T)\mathbf{s} = \mathbf{\Psi}'(\mathbf{P}^T\mathbf{s}) = \mathbf{\Psi}'\mathbf{s}'$, $\|\mathbf{s}'\|_0 = \|\mathbf{P}^T\mathbf{s}\|_0 = \|\mathbf{s}\|_0$. ∎

In the 1D case, this property implies that random scrambling does not cause any loss of the sparsity level of any given signal. In the 2D case, as we have shown in Section II-B, it helps to uniform the column (or row) sparsity level and thus flavors a parallel CS reconstruction technique, which will be exemplified in Section V.

In addition, if we know or partially know that $\mathrm{supp}(\mathbf{s})$ is localized in a certain k-dimensional subspace rather than uniformly distributed in $\mathbb{R}^M$, we can embed more secrets into the sparsifying basis, as stated in Property 4. Here we assume that $\mathbf{\Psi}$ is an orthonormal sparsifying basis for simplicity.

Property 4: $\mathbf{\Psi}'$ and $\mathbf{\Psi}$ are equivalent sparsifying bases if

$$\begin{aligned} \mathbf{\Psi}' &= \mathbb{F}_3(\mathbf{\Psi}) \\ &= (\psi_1, \ldots, \psi_{j-1}, a\psi_j + b\psi_k, \psi_{j+1}, \ldots, \psi_M) \end{aligned}$$

where a, b are non-zero constants and $j, k \in \mathrm{supp}(\mathbf{s})$ or $j, k \notin \mathrm{supp}(\mathbf{s})$.

Proof: Since $\mathbf{\Psi}$ is orthonormal, $s_j = (\psi_j, \mathbf{x}) = \psi_j^T \mathbf{x}$ and we know $s_j = 0$ when $j \notin \mathrm{supp}(\mathbf{s})$. Then the proof for $j, k \notin \mathrm{supp}(\mathbf{s})$ is trivial. For $j, k \in \mathrm{supp}(\mathbf{s})$, set

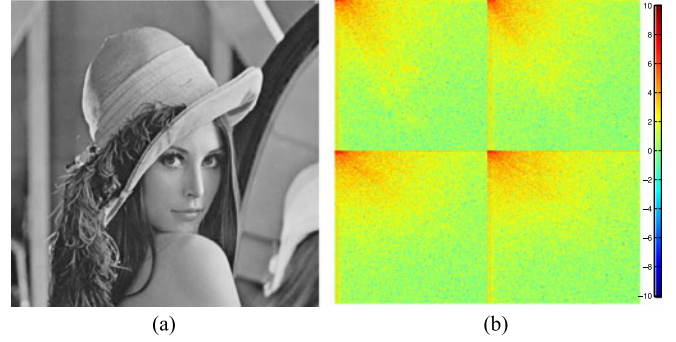$$\mathbf{s}' = (s_1', s_2', \ldots, s_j', \ldots, s_k', \ldots, s_M')^T$$

Fig. 4. (a) Original image "Lena." (b) Energy distribution of RPFrCT coefficients of "Lena" using logarithm base.

with

$$s_i' = \begin{cases} s_i/a, & \text{if } i = j \\ s_i - s_j b/a, & \text{if } i = k \\ s_i, & \text{otherwise.} \end{cases} \quad (10)$$

Then we have

$$\begin{aligned} \mathbf{x} &= \mathbf{\Psi}\mathbf{s} \\ &= \sum_{\substack{i=1 \\ i \neq j,k}}^M s_i\psi_i + s_j\psi_j + s_k\psi_k \\ &= \sum_{\substack{i=1 \\ i \neq j,k}}^M s_i\psi_i + \frac{s_j}{a}(a\psi_j + b\psi_k) + \left(s_k - \frac{bs_j}{a}\right)\psi_k \\ &= \mathbf{\Psi}'\mathbf{s}'. \end{aligned}$$

By (10), we conclude that $\|\mathbf{s}'\|_0 = \|\mathbf{s}\|_0$, hence completes the proof. ∎

Obviously, the operator $\mathbb{F}_3(\cdot)$ can be applied to three or more columns as long as all of the chosen columns are either in $\mathrm{supp}(\mathbf{s})$ or not. Finally, we provide an example to further illustrate Property 4. The grayscale image "Lena" with size 512×512, as shown in Fig. 4(a), is transformed using RPFrCT with orders $\alpha = 0.99$ and $\beta = 0.95$. Fig. 4(b) shows the absolute value of the RPFrCT coefficients under the logarithm base. It is clear that the energy of the RPFrCT coefficients matrix is localized, specifically, they are concentrated at the upper-left corner of the four sub-blocks. Thus, we can apply Property 4 to the RPFrCT basis $\mathbf{\Psi} = (\mathbf{R}_\beta \otimes \mathbf{R}_\alpha)^T$ accordingly. A similar effect can be observed in the parameterized DWT and DCT settings.

IV. DISCUSSIONS AND SECURITY ANALYSIS

We have demonstrated the possibility of using BLP-CS as a joint data acquisition and protection model for MTS purpose. This section aims to compare the basic OTS CS cipher and BLP-CS cipher from the viewpoints of complexity and security.

A. Complexity

Suppose we have constructed a RPFrCT matrix $\mathbf{R}_\alpha$ with appropriate fractional order α, a $M \times 1$ signal $\mathbf{x}$ can be sparsified

by $\mathbf{R}_\alpha \mathbf{x} = \mathbf{s}$. All the techniques on manipulating the sparsifying basis $\mathbf{R}_\alpha^T$ introduced in Section III-C can be unified to the following matrix notation,[7] i.e.

$$\mathbf{\Psi}_K = \mathbf{R}_\alpha^T \mathbf{P} \mathbf{D} \mathbf{Q}$$

where $\mathbf{D}$, $\mathbf{P}$ and $\mathbf{Q}$ are matrices determined by operators $\mathbb{F}_1$, $\mathbb{F}_2$ and $\mathbb{F}_3$, respectively. It is worth mentioning that $\mathbf{x} = \mathbf{\Psi}_K \mathbf{s}' = \mathbf{R}_\alpha^T \mathbf{s}$ with $\|\mathbf{s}'\|_0 = \|\mathbf{s}\|_0$. Recalling from Section III-A, the encoding of BLP-CS is governed by

$$\mathbf{y} = \mathbf{\Phi} \mathbf{x} = \mathbf{A}_K \mathbf{\Psi}_K^{-1} \mathbf{x} \qquad (11)$$

and the decoding should follow a two-step reconstruction, i.e.

$$\min \|\mathbf{s}'\|_1 \text{ subject to } \mathbf{y} = \mathbf{\Phi} \mathbf{x} = \mathbf{A}_K \mathbf{s}'$$
$$\mathbf{x} = \mathbf{\Psi}_K \mathbf{s}'. \qquad (12)$$

Once a well-designed key schedule is given,[8] a trusted third party can produce $\mathbf{\Phi}$, $\mathbf{A}_K$ and $\mathbf{\Psi}_K$ faithfully and transmit them to the encoder and decoder. An alternative option is that the encoder and decoder produce their own matrix key on the air using the agreed key schedule from the same root key. We assume the OTS CS model also adopts the same matrix key generation process for a fair comparison.

We first take a look at the encoder side. For the former situation, where the matrix key is produced by the trusted party and then delivered to both the CS encoder and decoder, the encoding complexity of the BLP-CS model outperforms that of the OTS CS model since it does not bring extra communication cost once the key is set. For the later situation, the encoding complexity of the OTS CS model is lower than that of the BLP-CS model at the first glimpse due to the reason that the encoding process of the second model involves a matrix multiplication, i.e., $\mathbf{A}_K \mathbf{\Psi}_K^{-1}$, in the key generation process. Nevertheless, since the OTS CS system requires updating the measurement matrix for every sampling, the BLP-CS model outperforms OTS CS after sampling $(2f' + f)/f'$ times. Here, f and f' refer to the complexity of the matrix multiplication and the matrix key generation, respectively.

At the decoder side, the Moore-Penrose pseudoinverse of the sensing matrix $\mathbf{A}_K$ needs to be calculated in every iteration of some l_1 optimization algorithms [51], for example, orthogonal matching pursuit [52]. The complexity of this operation dominates the overall complexity in CS reconstruction. As such, if some off-line techniques can be employed to calculate the pseudoinverse of $\mathbf{A}_K$, the complexity of the reconstruction can be largely reduced. For the OTS CS system, this is impossible since the measurement matrix is never re-used.

B. Security

1) Brute-Force and Ciphertext-Only Attacks: We employ the existing results presented in [9], [13] to show that the BLP-CS

preserves most secrecy features of the OTS CS-based cipher under these two attacks.

Theorem 2: [9, Theorem 1 and Corollary 1] Let $\mathbf{A}$ and $\mathbf{A}'$ be $K \times M$ Gaussian matrices. Let $\mathbf{x}$ be k-sparse with respect to the canonic basis and $\mathbf{y} = \mathbf{A}\mathbf{x}$. If $K > k$, then l_0 problem (3) and l_1 problem (4) will yield an K-sparse solution $\mathbf{x}'$ with probability one such that $\mathbf{y} = \mathbf{A}'\mathbf{x}'$.

We first examine the case of brute-force attack, i.e., the attacker tries to guess possible measurement matrices and uses them for decoding. Referring to Theorem 2, the l_0 or l_1 recovery governed by a wrong sensing matrix $\mathbf{A}_K$ will lead to an incorrect reconstruction with probability one. Thus the OTS CS-based cipher can guarantee computational secrecy if the key space is large enough to make systematic search of all the keys (sensing matrices) impossible. This result can be directly applied to our BLP-CS model. According (11) and (12), we can conclude that BLP-CS is computationally strong even if the attacker can successfully retrieved the secret sparsifying basis $\mathbf{\Psi}_K$. In this concern, the transform encryption approach enhances the security level of the basic CS paradigm.

An interesting security feature of the OTS CS cryptosystem under COA is the asymptotic spherical secrecy [13]. This type of secrecy states that any two different plaintexts (sparse signals to be sampled in this context) with equal power remain approximately indistinguishable from their measurement vectors when CS operates under the RIP framework. Alternatively, we can interpret this property as only the energy of the measurements carrying information about the signal. A bird's-eye view of why this asymptotic spherical secrecy holds for the OTS CS cipher may refer to the definition of RIP, which states that the CS encoding should obey an energy-preserving guarantee. A theoretical proof about this property can be found in [13].

As we demonstrated in (11) and (12), the proposed BLP-CS model works under the seemingly RIPless theory if one cannot determine $\mathbf{A}_K$ and $\mathbf{\Psi}_K$. Therefore, the energy-preserving constraint introduced by RIP is unapplicable to this setting. As such, we can conclude that the measurements (ciphertext) carries no information about the signal (plaintext) when a single ciphertext is observed. The BLP-CS and the OTS CS ciphers have the following major difference: when multiple ciphertexts are observed by the attacker, he is aware of the fact that two plaintexts must be similar if their corresponding ciphertexts are close to each other in the Euclidean space. This is caused by the multi-time usage of the same measurement matrix and the linear encoder. Surely the OTS CS cipher is more secure then the BLP-CS cipher from this point of view. Nevertheless, as mentioned in Section I, this is a favorable property that promotes the source coding gain from a system point-of-view [17]. This property also finds its way in privacy-preserving video surveillance systems [11]: assume the attacker happens to know some pairs of plaintext and ciphertext, such as static video scenes and their corresponding measurement vectors, and he wants to retrieve privacy-sensitive data from a new intercepted ciphertext. After studying the Euclidean distance of the new ciphertext, he comes to realize that plaintext corresponding to the new ciphertext contains privacy-sensitive data. However, the decryption of this ciphertext requires full knowledge of the matrix key $\mathbf{A}_K$ and

---

[7]We are aware of the fact that any parameterized orthonormal transform with good sparsifying capability can play the role of $\mathbf{R}_\alpha^T$.

[8]The design of an effective key scheduling process is not considered in this paper since our concern is only the secrecy of CS paradigm. We also note that this is a common treatment for all the state-of-the-art works on this topic.

$\mathbf{\Phi}_K$. This leads to our discussion of resistance of the BLP-CS cipher with respect to plaintext attacks.

2) Plaintext Attacks: As discussed in Section II, the data complexity of retrieving a general measurement matrix (the secret key) is M independent plaintexts and their corresponding ciphertexts in any basic CS-based cipher. If the used measurement matrix is Bernoulli, a single plaintext in the form $\mathbf{x} = (2^0, 2^1, \ldots, 2^M)^T$ and the corresponding ciphertext can be utilized to recover the Bernoulli measurement matrix completely.[9] Based on these knowledge, investigating the resistance of the OTS CS cryptosystem is a trivial work. We hereby focus on the BLP-CS cipher. Referring to (11), the attacker can retrieve $\mathbf{\Phi}$ from M independent plaintext-ciphertext pairs. By construction, $\mathbf{\Phi}$ is a non-RIP matrix. Thus the conclusion drawn from Theorem 1 assures that a straightforward use $\mathbf{\Phi}$ in the l_1 optimization problem (4) is not applicable. Considering that the l_0 optimization problem (3) is NP-hard [33], the attacker tries to decompose $\mathbf{\Phi}$ with the form $\mathbf{\Phi} = \mathbf{EF}$, with the constraint that entries of $\mathbf{E}$ should observe certain kind of distribution (Gaussian or Bernoulli). In particular, $\mathbf{F}$ is the product of an elementary matrix and an orthonormal matrix.

If the decomposition is unique or the possible number of decompositions is very limited, i.e., polynomial function of M, the attacker can determine the matrix key $\mathbf{A}_K$ and $\mathbf{\Psi}_K^{-1}$ and the BLP-CS cryptosystem is regarded as fail to resist plaintext attacks. To summarize, we conclude that the number of decompositions should be at least $O(M!)$, thus making the search for the true one inconclusive.[10] The conclusion is based on the simple fact $\mathbf{EF} = (\mathbf{EP})(\mathbf{P}^T\mathbf{F})$, where $\mathbf{P}$ is a $M \times M$ random permutation matrix. As we can see, distribution of all the entries of $(\mathbf{EP})$ is exactly the same as that of $\mathbf{E}$ and $\mathbf{P}^T$ represents elementary row operation on $\mathbf{F}$. As such, the attacker cannot distinguish the decomposition result $\mathbf{E}$ and $\mathbf{F}$ from $(\mathbf{EP})$ and $(\mathbf{P}^T\mathbf{F})$.

V. BLP-CS FOR DIGITAL IMAGES

In this section, the proposed BLP-CS model is applied as a joint data acquisition and protection layer for digital images. The aim is to provide an intuitive interpretation of how a cryptographic random scrambling can relax RIP of the measurement matrix and substantially reduce the decoding complexity, i.e., parallel reconstruction. Moreover, some other features owned by a basic CS paradigm, such as robust to packet loss and noise, are also observed.

We now consider a 2D image $\mathbf{X}$ with $M = n \times n$ pixels. If the chosen parameterized transform is RPFrCT, the basis for $\mathbf{X}$ is $(\mathbf{R}_\beta^T \otimes \mathbf{R}_\alpha^T)$ according to (9). Following the same approach adopted in [53], the encoding stage can be written as

$$\text{vec}(\mathbf{Y}) = [\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n]^T = \mathbf{\Phi} \text{ vec}(\mathbf{X})$$

---

[9]One can imagine the role of a $\{+1, -1\}$ matrix as that of a $\{0, 1\}$ matrix, the proof can be found in [19]. A vector composed by $\{0, 1\}$ can be recovered from the inner product of this vector and $\mathbf{x}$.

[10]This is even worse than directly solving the NP-hard l_0 problem (3), who has a complexity $\binom{M}{k}$.
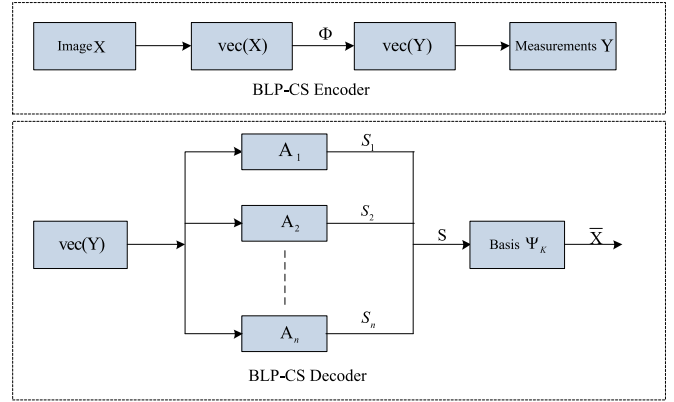

Fig. 5. Block diagram of BLP-CS for digital images.

where $\mathbf{\Phi}$ is the product of the $K \times M$ key-dependent sensing matrix $\mathbf{A}_K$ and the $M \times M$ key-dependent basis $\mathbf{\Psi}_K^{-1}$ has the form

$$\mathbf{\Psi}_K^{-1} = \mathbf{D}^{-1}\mathbf{P}^T(\mathbf{R}_\beta^T \otimes \mathbf{R}_\alpha^T)$$

and

$$\mathbf{A}_K = \begin{bmatrix} \mathbf{A}_1 & & & \\ & \mathbf{A}_2 & & \\ & & \ddots & \\ & & & \mathbf{A}_n \end{bmatrix}$$

with $\mathbf{A}_j = \mathbf{A}$ for $j \in \{1, \ldots n\}$ being Gaussian matrices. As we discussed in Section IV-A, repeatedly using the same sensing matrix for different signal segments can speed up the reconstruction if some off-line mechanism is allowed to calculate the pseudoinverse of $\mathbf{A}$ in advance.

According to Sections III-B and III-C, $\text{vec}(\mathbf{S}) = [\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_n]^T = \mathbf{\Psi}_K^{-1}\text{vec}(\mathbf{X})$ is sparse in the canonical basis. Referring to property 1 and (6), a parallel construction is applied as

$$\min \|\mathbf{s}_j\|_1 \quad \text{subject to } \mathbf{y}_j = \mathbf{A}\mathbf{s}_j \tag{13}$$

for all $j \in \{1, 2, \ldots, n\}$. Finally, the recovered image is given by $\text{vec}(\bar{\mathbf{X}}) = \mathbf{\Psi}_K \text{vec}(\mathbf{S})$. A block diagram of the whole system is depicted in Fig. 5. In summary, this system is a instance of the simplified BLP-CS model.

To further illustrate how the random scrambling $\mathbf{P}$ relaxes the RIP requirement of the sensing matrix $\mathbf{A}$, we consider another sampling configuration

$$\text{vec}(\mathbf{Y}) = \mathbf{\Phi} \text{ vec}(\mathbf{X})$$

where $\mathbf{\Phi} = \mathbf{A}_K \hat{\mathbf{\Psi}}_K^{-1}$ with $\mathbf{A}_K$ is the same as defined above and $\hat{\mathbf{\Psi}}_K^{-1} = \mathbf{D}^{-1}(\mathbf{R}_\beta^T \otimes \mathbf{R}_\alpha^T)$. Here, we note that the only difference of $\mathbf{\Psi}_K^{-1}$ and $\hat{\mathbf{\Psi}}_K^{-1}$ is the permutation matrix $\mathbf{P}$. The reconstruction is exactly the same as that of (13). By construction, this is a special form of block-based compressive sampling (BCS) [54], where each block is a column of the frequency coefficients, together with block independent recovery. We call this model BCS-In. We also note that using the smoothed projected Landweber operator can largely improve the BCS reconstruction

TABLE I
COMPARISON BETWEEN BLP-CS AND BCS-IN
IN TERMS OF APSNR AT DIFFERENT SRs

SR	10%		30%		50%		70%	
Model	BLP-CS	BCS-In	BLP-CS	BCS-In	BLP-CS	BCS-In	BLP-CS	BCS-In
"Lena"	21.6	15.5	27.5	23.3	31.4	27.3	35.7	32.1
"Peppers"	20.9	14.4	27.2	22.6	30.9	27.9	34.7	32.5
"Cameraman"	19.2	13.0	24.8	21.5	28.6	27.4	32.9	32.8
"Baboon"	17.8	9.7	20.2	17.6	22.6	21.3	25.8	25.2

TABLE II
APSNR OF THE RECONSTRUCTIONS UNDER AWGN AND VARIOUS PLRs

SR	0.1	0.3	0.5	0.7
Ideal BLP-CS	21.6	27.5	31.4	35.7
BLP-CS AWGN	21.8	27.4	31.3	34.9
BLP-CS 10% PLR	21.7	26.8	30.5	34.1
BLP-CS 20% PLR	20.9	26.2	29.5	32.7
BLP-CS 30% PLR	19.9	25.5	28.5	31.3

quality at relatively low extra computation overhead [55]. However, the study of embedding the smoothed projected Landweber operator in the BLP-CS reconstruction is out of the scope of this paper.

Four representative images, "Lena", "Peppers", "Cameraman" and "Baboon" of size 512×512 are used as our test images. The tests are carried out under different sampling rate $SR = \frac{K}{M} \times 100\%$. The reconstruction quality is evaluated in terms of average[11] peak signal-to-noise ratio, APSNR (dB) $= 10 \cdot \log_{10} \mathbb{E} \left(\frac{M \, 255^2}{\| \text{vec}(\mathbf{X}) - \text{vec}(\tilde{\mathbf{X}}) \|_2^2} \right)$. The results are listed in Table I and they support the conclusion of property 1, i.e., a cryptographic random scrambling helps make the column sparsity level of $\mathbf{S}$ uniform. The last point worth mentioning is that random scrambling is suitable for all kind of 2D sparse data (all kind of sparsifying coefficients under parameterized orthonormal transform), which extends the result that zig-zag scrambling works for DCT2 coefficients [24].

The basic CS paradigm that works under RIP theory is known to be robust with respect to transmission imperfections such as noise or packet loss [56], [57]. Since the new proposal works under the RIPless theory at only the encoder but RIP theory at the decoder, we expect the same property in our approach. To quantitatively study this, we evaluate the robustness of the proposed framework with respect to additive white Gaussian noise (AWGN) and various packet loss rates (PLRs). In the former case, we artificially add a zero-mean normal distribution random sequence with variance 1 to the measurements while in the latter we randomly discard certain number of measurements governed by PLR. Then we perform reconstruction on the corrupted measurements. In real applications, PLR can be up to 30% [58] and we measure the quality of the reconstruction in terms of APSNR at 10%, 20% and 30% PLR, respectively. These tests were carried out using the "Lena" image, but similar results were obtained using other images. As observed from Table II, our scheme is almost immune to AWGN when we compare the APSNR of the ideal case and the one with AWGN. In addition, comparing the APSNRs at different levels of PLR, we found that the reduction rate of APSNR is linear to the increasing rate of PLR, which implies that all measurements are roughly of the same importance [57].

VI. CONCLUSION

To realize the MTS usage of CS cryptosystem, some approaches have already been proposed. Typical examples include scrambling in different domains [20]–[22], [25] and cascading the DRPE technique [26]–[28]. However, we have shown that they fail to satisfy the security requirement. In this concern, we suggest a BLP-CS model by making use of the non-RIP measurement matrix construction. Our approach differs from existing ones in two aspects: 1) the RIPless CS theory is firstly applied for providing the security features of a CS-based cipher; 2) the role of the sparsifying basis for the secrecy of CS is revealed.

The security of the BLP-CS model is discussed from various aspects, such as brute-force attack, COA and plaintext attacks. Special attention has been paid to the plaintext attacks since it is widely accepted that basic CS model is immune to brute-force attack and COA [9], [13]. Under plaintext attacks, we have demonstrated that the number of candidate sensing matrices and sparsifying basis matrices that match the information inferred by the attacker is huge. Therefore, the searching of the true sensing matrix and sparsifying basis matrix is impossible.

Finally, we apply the proposed model for the purpose of secure compressive image sampling. Both theoretical analyses and experimental results support our expectation, i.e., random scrambling plays a critical role in relaxing the RIP requirement of the measurement matrix and flavoring a PCS reconstruction for 2D sparse signals. Other features of a basic CS system, such as robust to packet loss and noise, are also observed.

REFERENCES

[1] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[2] E. J. Candes and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21–30, Mar. 2008.

[3] R. Baraniuk, "Compressive sensing," *IEEE Signal Process. Mag.*, vol. 24, no. 4, pp. 118–121, Jul.2007.

[4] E. J. Candes and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?," *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5406–5425, Dec. 2006.

[5] T. Winkler and B. Rinner, "Security and privacy protection in visual sensor networks: A survey," *ACM Comput. Surv.*, vol. 47, no. 1, Jul. 2014, Art. no. 2.

[6] F. Dufaux and T. Ebrahimi, "Scrambling for privacy protection in video surveillance systems," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 8, pp. 1168–1174, Aug. 2008.

[7] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magn. Resonance Med.*, vol. 58, no. 6, pp. 1182–1195, Dec. 2007.

[11]$\mathbb{E}$ denotes calculate average over 100 tests.

[8] R. C. Barrows Jr. and P. D. Clayton, "Privacy, confidentiality, and electronic medical records," *J. Amer. Med. Inform. Assoc.*, vol. 3, no. 2, pp. 139–148, Apr. 1996.

[9] Y. Rachlin and D. Baron, "The secrecy of compressed sensing measurements," in *Proc. 46th Annu. Allerton Conf. Commun. Contr. Comput.*, 2008, pp. 813–817.

[10] R. Dautov and G. R. Tsouri, "Establishing secure measurement matrix for compressed sensing using wireless physical layer security," in *Proc. IEEE Int. Conf. Comput. Netw. Commun.*, Jan. 2013, pp. 354–358.

[11] L. Tong, F. Dai, Y. Zhang, J. Li, and D. Zhang, "Compressive sensing based video scrambling for privacy protection," in *Proc. IEEE Visual Commun. Image Process.*, Nov. 2011, pp. 1–4.

[12] V. Cambareri *et al.*, "A two-class information concealing system based on compressed sensing," in *Proc. IEEE Int. Symp. Circuits Syst.*, May 2013, pp. 1356–1359.

[13] V. Cambareri, M. Mangia, F. Pareschi, R. Rovatti, and G. Setti, "Low-complexity multiclass encryption by compressed sensing," *IEEE Trans. Signal Process.*, vol. 63, no. 9, pp. 2183–2195, May 2015.

[14] Z. Yang, W. Yan, and Y. Xiang, "On the security of compressed sensing-based signal cryptosystem," *IEEE Trans. Emerg. Topics Comput.*, vol. 3, no. 3, pp. 363–371, Sep. 2015.

[15] T. Bianchi, V. Bioglio, and E. Magli, "On the security of random linear measurements," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, May 2014, pp. 4020–4024.

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

[17] S. Mun and J. E. Fowler, "DPCM for quantized block-based compressed sensing of images," in *Proc. Eur. Signal Process. Conf.*, 2012, pp. 1424–1428.

[18] H. Liu, B. Song, F. Tian, and H. Qin, "Joint sampling rate and bit-depth optimization in compressive video sampling," *IEEE Trans. Multimedia*, vol. 16, no. 6, pp. 1549–1562, Oct. 2014.

[19] V. Cambareri, M. Mangia, F. Pareschi, R. Rovatti, and G. Setti, "On known-plaintext attacks to a compressed sensing-based encryption: A quantitative analysis," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 10, pp. 2182–2195, Oct. 2015.

[20] L. Zeng, X. Zhang, L. Chen, Z. Fan, and Y. Wang, "Scrambling-based speech encryption via compressed sensing," *EURASIP J. Adv. Signal Process.*, vol. 2012, no. 1, pp. 1–12, Dec. 2012.

[21] X. Huang, G. Ye, H. Chai, and O. Xie, "Compression and encryption for remote sensing image using chaotic system," *Security Commun. Netw.*, vol. 8, no. 18, pp. 3659–3666, Dec. 2015.

[22] Y. S. Zhang *et al.*, "Embedding cryptographic features in compressive sensing," *Neurocomput.*, vol. 205, pp. 472–480, Sep. 2016.

[23] W. Zeng and S. Lei, "Efficient frequency domain selective scrambling of digital video," *IEEE Trans. Multimedia*, vol. 5, no. 1, pp. 118–129, Mar. 2003.

[24] H. Fang, A. V. Sergiy, H. Jiang, and T. Omid, "Permutation meets parallel compressed sensing: How to relax restricted isometry property for 2D sparse signals," *IEEE Trans. Signal Process.*, vol. 62, no. 1, pp. 196–210, Jan. 2014.

[25] X. Wu, S. Tang, and P. Yang, "Low-complexity cloud image privacy protection via matrix perturbation," *CoRR*, 2014. [Online]. Available: http://arxiv.org/abs/1412.5937

[26] B. Deepan, C. Quan, Y. Wang, and C. Tay, "Multiple-image encryption by space multiplexing based on compressive sensing and the double-random phase-encoding technique," *Appl. Opt.*, vol. 53, no. 20, pp. 4539–4547, Jul.2014.

[27] N. Rawat, B. Kim, I. Muniraj, G. Situ, and B.-G. Lee, "Compressive sensing based robust multispectral double-image encryption," *Appl. Opt.*, vol. 54, no. 7, pp. 1782–1793, Mar. 2015.

[28] J. Li, J. S. Li, Y. Y. Pan, and R. Li, "Compressive optical image encryption," *Sci. Rep.*, vol. 5, May 2015, Art. no. 10374.

[29] L. Y. Zhang, K.-W. Wong, Y. Zhang, and Q. Lin, "Joint quantization and diffusion for compressed sensing measurements of natural images," *Proc. IEEE Int Symp. Circ. Syst.*, May 2015, pp. 2744–2747.

[30] R. Baraniuk, M. Davenport, R. Devore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constructive Approximation*, vol. 28, no. 3, pp. 253–263, Dec. 2008.

[31] E. J. Candes and Y. Plan, "A probabilistic and RIPless theory of compressed sensing," *IEEE Trans. Inf. Theory*, vol. 57, no. 11, pp. 7235–7254, Nov. 2011.

[32] R. Kueng and D. Gross, "RIPless compressed sensing from anisotropic measurements," *Linear Algebra Appl.*, vol. 441, pp. 110–123, Jan. 2014.

[33] E. J. Candes and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.

[34] T. T. Do, L. Gan, N. H. Nguyen, and T. Tran, "Fast and efficient compressive sensing using structurally random matrices," *IEEE Trans. Signal Process.*, vol. 60, no. 1, pp. 139–154, Jan. 2012.

[35] P. Refregier and B. Javidi, "Optical image encryption based on input plane and Fourier plane random encoding," *Opt. Lett.*, vol. 20, no. 7, pp. 767–769, Apr. 1995.

[36] B. Javidi, "Method and apparatus for encryption," U.S. Patent 5,903,648, May 11, 1999.

[37] A. Carnicer, M. Montes-Usategui, S. Arcos, and I. Juvells, "Vulnerability to chosen-cyphertext attacks of optical encryption schemes based on double random phase keys," *Opt. Lett.*, vol. 30, no. 13, pp. 1644–1646, Jul. 2005.

[38] Y. Frauel, A. Castro, T. J. Naughton, and B. Javidi, "Resistance of the double random phase encryption against various attacks," *Opt. Express*, vol. 15, no. 16, pp. 10-253–10-265, Aug. 2007.

[39] P. Clemente *et al.*, "Compressive holography with a single-pixel detector," *Opt. Lett.*, vol. 38, no. 14, pp. 2524–2527, Jul.2013.

[40] Y. Rivenson, A. Stern, and B. Javidi, "Compressive Fresnel holography," *J. Display Technol.*, vol. 6, no. 10, pp. 506–509, Oct. 2010.

[41] M. F. Duarte, S. Sarvotham, D. Baron, M. B. Wakin, and R. G. Baraniuk, "Distributed compressed sensing of jointly sparse signals," in *Proc. Asilomar Conf. Signals Syst. Comput.*, 2005, pp. 1537–1541.

[42] X. Liu, Y. Cao, P. Lu, X. Lu, and Y. Li, "Optical image encryption technique based on compressed sensing and Arnold transformation," *Optik*, vol. 124, no. 24, pp. 6590–6593, Dec. 2013.

[43] B. Zeng, S.-K. A. Yeung, S. Zhu, and M. Gabbouj, "Perceptual encryption of H. 264 videos: Embedding sign-flips into the integer-based transforms," *IEEE Trans. Inf. Forensics Security*, vol. 9, no. 2, pp. 309–320, Feb. 2014.

[44] A. Pande and J. Zambreno, "The secure wavelet transform," *J. Real-Time Image Process.*, vol. 7, no. 2, pp. 131–142, Jun. 2012.

[45] A. Pande, P. Mohapatra, and J. Zambreno, "Securing multimedia content using joint compression and encryption," *IEEE Multimedia Mag.*, vol. 20, no. 4, pp. 50–61, Oct.-Dec. 2013.

[46] D. Engel and A. Uhl, "Parameterized biorthogonal wavelet lifting for lightweight JPEG 2000 transparent encryption," in *Proc. 7th Workshop Multimedia Security*, 2005, pp. 63–70.

[47] S.-K. A. Yeung and B. Zeng, "A new design of multiple transforms for perceptual video encryption," in *Proc. 19th IEEE Int. Conf. Image Process.*, Sep.-Oct. 2012, pp. 2637–2640.

[48] G. Unnikrishnan, J. Joseph, and K. Singh, "Optical encryption by double-random phase encoding in the fractional Fourier domain," *Opt. Lett.*, vol. 25, no. 12, pp. 887–889, Jun. 2000.

[49] I. Venturini and P. Duhamel, "Reality preserving fractional transforms," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, May 2004, vol. 5, pp. V-205–V-208.

[50] G. Cariolaro, T. Ersehe, and P. Kraniaukas, "The fractional discrete cosine transform," *IEEE Trans. Signal Process.*, vol. 50, no. 4, pp. 902–911, Apr. 2002.

[51] S. Boyd and L. Vanderberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[52] J. Tropp, and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.

[53] M. F. Duarte *et al.*, "Single-pixel imaging via compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 83–91, Mar. 2008.

[54] L. Gan, "Block compressed sensing of natural images," in *Proc. 15th Int. Conf. Digit. Signal Process.*, 2007, pp. 403–406.

[55] J. E. Fowler, S. Mun, and E. W. Tramel, "Multiscale block compressed sensing with smoothed projected landweber reconstruction," in *Proc. 19th Eur. Signal Process. Conf.*, 2011, pp. 564–568.

[56] A. G. Dimakis and P. O. Vontobel, "LP decoding meets LP decoding: A connection between channel coding and compressed sensing," in *Proc. 47th Annu. Allerton Conf. Commun. Control Comput.*, 2009, pp. 8–15.

[57] J. N. Laska, P. T. Boufounos, M. A. Davenport, and R. G. Baraniuk, "Democracy in action: Quantization, saturation, and compressive sensing," *Appl. Comput. Harmon. Anal.*, vol. 31, no. 3, pp. 429–443, Nov. 2011.

[58] J. Zhao and R. Govindan, "Understanding packet delivery performance in dense wireless sensor networks," in *Proc. 1st Int. Conf. Embeded Netw. Sensor Syst.*, 2003, pp. 1–13.

Leo Yu Zhang received the B.S. and M.S. degrees in computational mathematics from Xiangtan University, Xiangtan, China, in 2009 and 2012, respectively, and is working toward the Ph.D. degree in electronic engineering at the City University of Hong Kong, Hong Kong, China.

In 2015, he spent three months at the University of Bologna, Bologna, Italy, and the University of Ferrara, Ferrara, Italy, as a Visiting Student. From 2012 to 2013, he was with the State Grid Electric Power Research Institute of China, Beijing, China, as a Development Engineer. Currently, he is a Research Assistant with the Department of Computer and Information Science, University of Macau, Macau, China. His research interests include multimedia security and compressive sensing.

Yushu Zhang received the B.S. degree in mathematics from the North University of China, Shanxi, China, in 2010, and the Ph.D. degree in computer science from Chongqing University, Chongqing, China, in December 2014.

Since January 2015, he has been an Associate Professor with the School of Electronics and Information Engineering, Southwest University, Chongqing, China. He held various research positions at the City University of Hong Kong, Hong Kong, China, and the University of Macau, Macaulay, China. His research interests include multimedia coding and security, multimedia cloud security, and compressive sensing security.

Kwok-Wo Wong (SM'93-M'95-SM'03) received the B.Sc. degree in electronic engineering from the Chinese University of Hong Kong, Hong Kong, China, and the Ph.D. degree from the City University of Hong Kong, Hong Kong, China.

He is currently an Associate Professor with the Department of Electronic Engineering, City University of Hong Kong. He has authored or coauthored more than 100 papers in 25 international mathematics, physics, and engineering journals in the fields of nonlinear dynamics, cryptography, neural networks, and optics. His current research interests include chaos, cryptography, and neural networks.

Prof. Wong is a Chartered Engineer and a Member of the Institution of Engineering and Technology.

Jiantao Zhou (M'12) received the B. Eng. degree in electronic engineering from the Dalian University of Technology, Dalian, China, in 2002, the M.Phil degree in radio engineering from Southeast University, Nanjing, China, in 2005, and the Ph.D. degree in electronic and computer engineering from the Hong Kong University of Science and Technology, Hong Kong, China, in 2009.

He is currently an Assistant Professor with the Department of Computer and Information Science, Faculty of Science and Technology, University of Macau, Macau, China. He held various research positions at the University of Illinois at Urbana-Champaign, Urbana, IL, USA, the Hong Kong University of Science and Technology, Hong Kong, China, and McMaster University, Hamilton, ON, Canada. He holds three granted U.S. patents and two granted Chinese patents. His research interests include multimedia security and forensics and high-fidelity image compression.

Prof. Zhou was a recipient of the Best Paper Award of the IEEE Pacific-Rim Conference on Multimedia in 2007.