Multi-level Reversible Data Anonymization via Compressive Sensing and Data Hiding

Mehmet Yamaç[1], Mete Ahishali[1], Nikolaos Passalis[1], Jenni Raitoharju[1], Bülent Sankur[2], and Moncef Gabbouj[1]

[1]Tampere University, Faculty of Information Technology and Communication Sciences, Tampere, Finland
[2]Boğaziçi University, Electrical and Electronics Engineering Department, Istanbul, Turkey

***Abstract*—Recent advances in intelligent surveillance systems have enabled a new era of smart monitoring in a wide range of applications from health monitoring to homeland security. However, this boom in data gathering, analyzing and sharing brings in also significant privacy concerns. We propose a Compressive Sensing (CS) based data encryption that is capable of both obfuscating selected sensitive parts of documents and compressively sampling, hence encrypting both sensitive and non-sensitive parts of the document. The scheme uses a data hiding technique on CS-encrypted signal to preserve the one-time use obfuscation matrix. The proposed privacy-preserving approach offers a low-cost multi-tier encryption system that provides different levels of reconstruction quality for different classes of users, e.g., semi-authorized, full-authorized. As a case study, we develop a secure video surveillance system and analyze its performance.**

***Index Terms*—Reversible Privacy Preservation, Multi-level Encryption, Compressive Sensing, Video Monitoring**

I. Introduction

MANY emergent smart surveillance applications (i.e., buildings, infrastructure, stores, ambient-assisted living, public areas) necessitate time-continuous data gathering and processing. Upcoming 5G and IoT technologies will enable continuous data collection and processing for persistent monitoring [1]. For example, an intelligent building system equipped with monitoring sensors such as CO2 meters, thermometers, cameras or other types of IoT devices will be instrumental in effectively automating tasks of heating, ventilation, and conditioning (HVAC) systems, or in improving the fault and hazard detection performance [2]. Another case in point is an intelligent network of cameras for continuous site surveillance or a health monitoring system [3], which gathers users' bio-signals along with video/speech data to be processed remotely. These applications and their variants that collect data via sensors or edge devices bear the concern of the privacy of people and possibly of sites. In fact, the European General Data Protection Regulation (GDPR) legislation [4] has specifically addressed these privacy concerns in data collection and processing.

Currently, there exist a plethora of privacy-preserving technologies that vary in the data type and in the application scenario. Even the definition of privacy is up to change for different application areas and use cases, depending on whether it is signal processing, a database system, secure communication, etc. [5]. Under privacy concern, documents are considered to consist of private, i.e., sensitive parts, those parts that could potentially expose compromising information to unauthorized users, and of public, i.e., non-sensitive parts. Privacy-preserving data processing then aims to encrypt the private parts of a document without deteriorating its public parts. Recent comprehensive surveys provide useful guidelines in privacy-preserving data mining [6], [7], signal processing [8], [9], and privacy metrics [5].

In principle, a naive application of strong cryptography methods such as AES [10] or RSA [11] would provide a high degree of security, in addition to privacy. However first, these encryption methods are relatively costly; but more importantly, it is neither useful nor necessary to encrypt the whole signal in real-time multimedia applications such as in video [12], image, health [13] monitoring systems or other types of IoT applications. Only the selected parts of the multimedia document deemed to carry private information need to be protected; this then gives rise to a two-tier approach. More generally in a multi-tiered approach, different parts of the document can be privacy protected at differential levels, the most strongly protected parts accessible by the highest authorization level, and so forth. We can also state the three desiderata of privacy-protection algorithms: a) The technique should be able to secure the privacy of selected sensitive portions of the data; e.g., for face hiding, it should be stronger than any automatic face recognition algorithm; b) The method should not degrade the non-sensitive parts of the documents; c) It should be able to reverse the sensitive part encryption (for authorized users) in good quality. A concomitant desideratum is that the computation cost of the data encryption should be reasonably low.

Although Compressive Sensing (CS) [14] is an alternative data acquisition strategy to conventional Nyquist/Shannon based technique, it also provides encryption with a reasonable security level via its randomized sensing mechanism. In consequence, using CS setup alone or with another lightweight encryption shell applied on top of it has recently been a popular approach for multimedia applications [15].

In this work, we pursue the approach of compressive sensing to accomplish both compression and cryptographic security on the whole data, and data hiding technology [16], [17] to hide and then recover the masked-out private parts of the document. The novel method achieves privacy protection by obfuscating the sensitive parts of the document while the CS-encryption is applied to the whole document, i.e., the combined public and private parts. We assume that the document has been preprocessed and segmented into its sensitive and non-sensitive parts. We use terms *de-identification* and *anonymization* interchangeably, in the sense of rendering unintelligible the privacy bearing segments of a document. Although our method is applicable to any document type, images, video, audio, etc., with appropriate modifications, in the sequel we will consider images as an application case.

Our scheme provides a two-tiered privacy, in which the semi-authorized user, i.e., the entity with lower authorization level can decode and view only the non-sensitive parts of the image, while the fully-authorized user decodes and sees the entire image. The semi-authorized one with only key $\mathbf{A}$ (CS-Encryption matrix) is able to recover images whose sensitive parts remain obfuscated after decoding whereas the fully authorized person with keys $\mathbf{A}$ and $\mathbf{B}$ (the latter being watermark embedding matrix) is able to recover the whole image. In both cases, the image quality is stipulated to remain close to the original quality. The significant merits of our proposed method are first to enable a low-cost, two-level encryption and second to provide reversible anonymization for the selected authorized users. Although the experiments are run only on image data, our method is general enough to be applied to any data involving privacy concerns, such as to videos as detailed in Section VI, or to bio-signals. In this work, we select face de-identification problem [18], [19] as a case study, within the context of a privacy-preserving image/video monitoring system.

The privacy protection concern in image/video has been addressed in a plethora of papers in the last decades. In summary, the technical solutions can be discussed in three groups: a) automatic blurring of faces, context-dependent blurring, e.g., bystanders only; b) blacking out of faces with random patterns, and recently; c) anonymous face substitutions or iterative regeneration schemes. Our method is in line with the noise pattern overlay methods in the literature. However, we differ from these methods in two respects: i) while we are able to fully remove the obfuscating noise pattern, we provide multi-tier differential protection; ii) we use compressive sensing for data reduction and cryptographic security, and watermark the compressed signal with the data hiding pattern).

A privacy-preserving method to which our method has some resemblance was recently described in [20]. In the method of [20], the images are first processed through a parallel group of trained auto-encoders, each generating its own sufficiently diversified sparse code. They obfuscate the sparse code by adding random noise with statistics similar to sparse code statistics to coefficients to a group of coefficients outside the sparse code support set. The support set is predefined or shared via a secret channel to the trusted user. Only the trusted user possesses the key to recover the support set of the sparse code coefficients, and thus is able to decode the sensitive image (the face). Codes from multiple auto-encoders are used to successively refine the results, i.e., incrementally improve reconstructed image quality. In contrast, our method is not face specific, does not need to find sparse codes in the encoding part, does hence not require a separate secret channel to share the obfuscation key. In addition, data reduction via CS-compression is a byproduct of our scheme.

TABLE I: From left to right: a) Symbols of the frequently used variables in the article. b) Denotations of these symbols. c) The corresponding cryptographic terminology, if applicable. d) The conditions the variables must satisfy for the encryption scheme to work properly.

Variable	Synonyms	In Crypt.	Properties
$\mathbf{s} \in \mathbb{R}^N$	Signal, document, data	Plain-text	Sparse in $\mathbf{\Phi}$
$\mathbf{\Phi} \in \mathbb{R}^{N \times N}$	Sparsifying matrix (basis)		
$\mathbf{x} \in \mathbb{R}^N$	Sparse coefficients of $\mathbf{s}$	Plain-text	$\lVert \mathbf{x} \rVert \leq k$
$\mathbf{A} \in \mathbb{R}^{m \times N}$	CS-measurement matrix	CS-encryption matrix (Key $\mathbf{A}$)	$\mathbf{H} = \mathbf{A\Phi}$ to have RIP
Λ_p	Location index set of the obfuscated sensitive part		$\Lambda_p \subset \{1, ..., N\}$
$\mathbf{s_s} \in \mathbb{R}^N$	Sensitive part of $\mathbf{s}$	$\mathbf{s_s} = \begin{cases} (\mathbf{s_s})_{\Lambda_p} &= \mathbf{s}_{\Lambda_p} \\ (\mathbf{s_s})_{\Lambda_p^c} &= \mathbf{0} \end{cases}$	
$\mathbf{s_n} \in \mathbb{R}^N$	Non-sensitive part of $\mathbf{s}$	$\mathbf{s_n} = \begin{cases} (\mathbf{s_n})_{\Lambda_p^c} &= \mathbf{s}_{\Lambda_p^c} \\ (\mathbf{s_n})_{\Lambda_p} &= \mathbf{0} \end{cases}$	
$\mathbf{s_{\Lambda_p}} \in \mathbb{R}^{\lvert \Lambda_p \rvert}$	Masked out sensitive part		$\lvert \Lambda_p \rvert$: Size of the masked region
$\mathbf{\Delta_{\Lambda_p}}$	Obfuscation matrix	Anonymization operator	$\lvert \Lambda_p \rvert \times \lvert \Lambda_p \rvert$ matrix
$\mathbf{M} \in \mathbb{R}^{m \times N}$	Masking matrix	Anonymization operator	
$\mathbf{y} \in \mathbb{R}^m$	CS-based encrypted signal	Cipher-text	$\mathbf{y} = \mathbf{As}$
$\mathbf{y_d} \in \mathbb{R}^m$	Jointly CS-based encrypted and anonymized signal	Cipher-text	$\mathbf{y_d} = (\mathbf{A} + \mathbf{M})\mathbf{s}$
$\mathbf{w}$	Ternary Watermark	Length, T, and watermark magnitude, a, are predefined: $w_i \in \{a, 0, -a\}$	
$\mathbf{B} \in \mathbb{R}^{m \times T}$	Watermark embedding matrix		$T < m$
$\mathbf{F} \in \mathbb{R}^{p \times m}$	Left annihilator matrix		$\mathbf{FB} = 0$, $p = m - T$
$\mathbf{y_w} \in \mathbb{R}^m$	Marked $\mathbf{y_d}$	Cipher-text (marked)	$\mathbf{y_w} = \mathbf{y_d} + \mathbf{Bw}$

A preliminary version of this work was presented at [21]. This early version had briefly introduced the methodology and presented some test results on a token dataset (6 faces in a controlled laboratory environment). In this article, we provide a theoretical worst-case analysis on the watermark guarantee conditions (Lemma 1, Theorem 4). We have extended the paper by incorporating a discussion on the design of alternative obfuscating matrices [21] as well as on the alternative designs of the watermark embedding matrix (see Section IV-E). Simulation experiments are run on a realistic public dataset with a much bigger size (a subset of YouTube Faces Database [22]) containing 100 classes (videos of 100 identities). We have also briefly described two extensions of the proposed method: 1. Its adaptation to video signals, beyond the simple frame-by-frame privacy processing; 2. A three-tiered privacy protection in images. In the detailed performance evaluation, we illustrate the reconstruction accuracy of masked regions as a function of watermark embedding power and the choice of obfuscating masks, both being user-defined parameters. Recognition accuracies with original faces, with de-identified faces, and with faces reverse de-identified via recovered watermark are given. The result of a test against an adversary with a strong computational capability and with access to the full labeled training set is also reported.

The rest of the paper is organized as follows. The notation is provided in Section II. We give a brief overview of compressive sensing and its usage in encryption systems in Section III. We emphasize the compressive sensing properties that we have exploited in our proposed scheme. In Section IV, the proposed two-tier privacy-preserving system is presented in detail. Section V introduces a case study of the proposed method in video monitoring and gives the results of the extensive simulation studies. Finally, conclusions are drawn in Section VII.

II. Notations

In this work, the ℓ_p norm of a vector $\mathbf{x} \in \mathbb{R}^N$ is given as $\|\mathbf{x}\|_p = \left(\sum_{i=1}^{N} |x_i|^p\right)^{1/p}$ for $p \geq 1$. We also define the ℓ_0-norm of the vector $\mathbf{x} \in \mathbb{R}^N$ as $\|\mathbf{x}\|_0 = \lim_{p \to 0} \sum_{i=1}^{N} |x_i|^p = \#\{l : x_l \neq 0\}$. The exactly (or strictly) k-sparse signal in some appropriate domain is the signal, $\mathbf{x} \in \mathbb{R}^N$ with $\|\mathbf{x}\|_0 \leq k$. On the other hand, the approximately k-sparse signal (or compressible) is a signal $\mathbf{x}$ with $\|\mathbf{x} - \check{\mathbf{x}}\|_2 \leq \kappa$, where κ is a small constant and $\check{\mathbf{x}}$ is obtained via zero-outing the elements of $\mathbf{x}$ except the ones with k-largest magnitude. For convenience, we show in Table I the list of frequently used symbols, the terminology used in paper and their synonymous definitions in the cryptography literature.

III. Preliminaries and Prior Art

Our interest in compressive sensing is twofold: to compress the signal if it is already sampled or to sample analog signals directly at rates below the Nyquist-Shannon bound and to exploit the inherent cryptographic capability of compressive sensing.

A. Compressive Sensing

Compressive sensing (CS) theory has significantly impacted the field of signal processing since its inception in 2005 [14]. According to the CS theory, a signal can be sampled using far fewer measurements than the traditional Nyquist-Shannon acquisition rate, provided it is sparse or compressible in some proper domain. CS-based MRI imaging [23], radar monitoring systems [24], [25], and ECG measurements in a health monitoring system [13] are some of its success stories. It is also seen as a potential solution for hardware/software design in the applications requiring very high sampling frequencies such as wideband spectrum sensing [26] and ultra-wideband communication schemes [27]. In fact, CS is expected to play an important role in the next-generation communications systems such as 5G [28].

Let us consider the linear mapping of a discrete signal $\mathbf{s} \in \mathbb{R}^N$ as

$$\mathbf{y} = \mathbf{A}\mathbf{s}, \tag{1}$$

where $\mathbf{A} \in \mathbb{R}^{m \times N}$ is known as the measurement matrix with $m < N$. The minimum-energy solution for the underdetermined linear system of equations (1) is given by

$$\min_{\mathbf{s}} \|\mathbf{s}\|_2^2 \text{ subject to } \mathbf{A}\mathbf{s} = \mathbf{y}. \tag{2}$$

The solution of (2) is unique and has a closed form solution, $\hat{\mathbf{s}} = \mathbf{A}^T \left(\mathbf{A}\mathbf{A}^T\right)^{-1} \mathbf{y}$ provided that $\text{rank}\,(\mathbf{A}) = m \leq N$ which makes $\mathbf{A}\mathbf{A}^T$ invertible. The minimum achievable reconstruction error is $\|\mathbf{s} - \hat{\mathbf{s}}\|_2 = \mathbf{s}^T \left(\mathbf{I} - \mathbf{A}^T \left(\mathbf{A}\mathbf{A}^T\right)^{-1} \mathbf{A}\right) \mathbf{s}$, which shows that exact recovery is not possible since $\mathbf{I} \neq \mathbf{A}^T \left(\mathbf{A}\mathbf{A}^T\right)^{-1} \mathbf{A}$ when $m < N$. The CS theory addresses signals that are sparse in a proper domain, $\mathbf{\Phi} \in \mathbb{R}^{N \times N}$, i.e., $\mathbf{s} = \mathbf{\Phi}\mathbf{x}$ with $\|\mathbf{x}\|_0 \leq k$. Therefore, (1) can be re-formulated as follows,

$$\mathbf{y} = \mathbf{A}\mathbf{s} = \mathbf{A}\mathbf{\Phi}\mathbf{x} = \mathbf{H}\mathbf{x}, \tag{3}$$

where $\mathbf{H} = \mathbf{A}\mathbf{\Phi}$, and even if (3) has infinitely many solution we can look for the sparsest one,

$$\min_{\mathbf{x}} \|\mathbf{x}\|_0 \text{ subject to } \mathbf{H}\mathbf{x} = \mathbf{y}. \tag{4}$$

Eq. (4) is also known as sparse representation of $\mathbf{y}$ in $\mathbf{H}$ and it is unique, provided that the minimum number of linearly independent columns of $\mathbf{H}$, as defined in [29], is greater than $2k$. Thus for $spark(\mathbf{H}) \geq 2k$, any two distinct k-sparse signals $\mathbf{x}', \mathbf{x}''$ can be uniquely recovered from their undersampled measurements $\mathbf{y}'$, $\mathbf{y}''$ if $m \geq 2k$. Put differently, one has the surprising result that, while it is not possible to recover $\mathbf{s}$ exactly using minimum norm decoder as in (2), exact recovery of the signal is possible in the sparsifying domain.

The nonconvex problem (4) with ℓ_0-quasi-norm can be relaxed to its closest convex form, ℓ_1 as

$$\min_{\mathbf{x}} \|\mathbf{x}\|_1 \text{ subject to } \mathbf{x} \in \Upsilon\left(\mathbf{y}\right), \tag{5}$$

where $\Upsilon\left(\mathbf{y}\right) = \{\mathbf{x} : \mathbf{H}\mathbf{x} = \mathbf{y}\}$, an optimization problem that is also known as Basis Pursuit [30]. The equivalence of ℓ_0-ℓ_1 minimization problems is well investigated in the literature in terms of the properties of $\mathbf{H}$. For instance, the Null Space Property (NSP) [31] not only satisfies the ℓ_0-ℓ_1 equivalence but also comes very handy for the recovery performance analysis when $\mathbf{x}$ is not exactly k-sparse but only compressible. In the case we deal with approximately sparse signals or/and with a case where the measurements are contaminated by additive noise, the problem can be relaxed with $\Upsilon\left(\mathbf{y}\right) = \{\mathbf{x} : \|\mathbf{H}\mathbf{x} - \mathbf{y}\|_2 \leq \epsilon\}$, where ϵ is a small positive constant. Problem (5), with this new constraint, is known as Basis Pursuit Denoising (BPDN) [32]. The stability conditions of CS signal recovery techniques are also well understood: a stable solution, $\hat{\mathbf{x}}$, is expected to obey $\|\mathbf{x} - \hat{\mathbf{x}}\|_2 \leq \kappa \|\mathbf{z}\|$ with a small constant, κ for additive noise $\mathbf{z}$ perturbation in the measurements, $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{z}$.

When approximately sparse signals are measured under noise, a property stronger than NSP gives a stable recovery guarantee. This property is called Restricted Isometry Property (RIP), which is defined as follows:

Definition 1. (Restricted Isometry Property) A matrix $\mathbf{H} \in \mathbb{R}^{m \times N}$ has RIP with order k, if there exist a smallest $\delta_k(\mathbf{H})$ that satisfies

$$\left(1 - \delta_k(\mathbf{H})\right) \|\mathbf{x}\|_2^2 \leq \|\mathbf{H}\mathbf{x}\|_2^2 \leq \left(1 + \delta_k(\mathbf{H})\right) \|\mathbf{x}\|_2^2 \tag{6}$$

for all k-sparse signal, $\mathbf{x} \in \mathbb{R}^N$. The constant, $\delta_k(\mathbf{H})$ is called the Restricted Isometry Constant (RIC) of order k for matrix $\mathbf{H}$.

The stability and ℓ_0-ℓ_1 equivalence conditions w.r.t. RIC of a measurement matrix are thoroughly studied in the literature.

The authors in [33] show that the ℓ_0-ℓ_1 equivalence is achieved when $\delta_{2k}(\mathbf{H}) \leq \sqrt{2}-1$. Likewise, the stability of the ℓ_1 minimization problem is investigated in Basis Pursuit Denoising [34] and Dantzig Selector [35]. In [34], it is shown that for $\Upsilon(\mathbf{y}) = \{\mathbf{x} : \|\mathbf{Hx}-\mathbf{y}\|_2 \leq \epsilon\}$ and $\|\mathbf{z}\|_2 \leq \epsilon$, the solution of (5) satisfies

$$\|\mathbf{x}-\hat{\mathbf{x}}\|_2 \leq C_0\epsilon, \tag{7}$$

where C_0 depends on $\delta_{2k}(\mathbf{H}) < \sqrt{2}-1$ [33]. Notice that the recovery guarantee conditions of an arbitrary k-sparse signal enforce $2k$-order RIC, $\delta_{2k}(\mathbf{H})$ instead of $\delta_k(\mathbf{H})$. The intuition behind this is simply that for noise-free measurements, the null space analysis indicates that *spark*$(\mathbf{H}) \geq 2k$ in order for $\mathbf{H}$ not to map any two arbitrary but distinct $k-$sparse signals $\mathbf{x}'$ and $\mathbf{x}''$ to the same point, so that one always has $\mathbf{Hx}' \neq \mathbf{Hx}''$. In this sense, RIP gives us a stronger guarantee that after mapping with a $\mathbf{H}$, the distance between points $\mathbf{x}', \mathbf{x}''$ should be preserved at least as follows: $(1-\delta_{2k}(\mathbf{H}))\|\mathbf{x}'-\mathbf{x}''\|_2^2 \leq \|\mathbf{Hx}'-\mathbf{Hx}''\|_2^2$.

The good measurement matrices $\mathbf{A}$ that preserve the information in the sparse domain $\mathbf{\Phi}$, or alternatively $\mathbf{A\Phi} = \mathbf{H}$ are the ones that satisfy the RIP property. Certain random measurement matrices are known to satisfy this property, one popular such case being the matrix whose elements $A_{i,j}$ are i.i.d. (independent identically distributed) and drawn from a Gaussian distribution, i.e.,

$$A_{i,j} \sim \mathcal{N}\left(0, \frac{1}{m}\right) \tag{8}$$

and for $m > k(\log(N/k))$, and $\mathbf{H}$ inherits this property as well. We recall the following lemma that gives the stability condition of BPDN for measurements under additive white Gaussian noise (AWGN) contamination, since it will be handy in the sequel for the stability analysis of our encryption scheme.

Corollary 1. *(Refined from Corrollary 1.1 of [35, p. 32].) Let $\mathbf{H} \in \mathbb{R}^{m \times N}$ satisfy the RIP of order $2k$ with $\delta_{2k}(\mathbf{H}) < \sqrt{2}-1$. Assume that measurements are corrupted by i.i.d. noise with elements z_i drawn from $\mathcal{N}\left(0, \frac{1}{\sigma^2}\right)$. Then, the error of the solution of* (5) *with $\Upsilon(\mathbf{y}) = \{\mathbf{x} : \|\mathbf{Hx}-\mathbf{y}\|_2 \leq \epsilon\}$ is upper bounded by*

$$\|\mathbf{x}-\hat{\mathbf{x}}\|_2 \leq 4\frac{\sqrt{1+\delta_{2k}(\mathbf{H})}}{1-(1+\sqrt{2})\delta_{2k}(\mathbf{H})}(1+\gamma)\sqrt{m}\sigma \tag{9}$$

with probability of at least $1-\exp(-\frac{3m}{4}\gamma^2)$ where $0 < \gamma < 1$ and $\epsilon = (1+\gamma)\sqrt{m}\sigma$.

B. Compressive Sensing Based Encryption

Since in the CS setup, a signal is linearly sampled using random or pseudo-random measurement matrices, there exists an inherent capability to provide privacy and cryptographic protection [36], [37]. One advantage of CS-based encryption is that the linearity and the dimensionality reduction of the CS scheme result in low-cost operations. This could be a crucial advantage for data encryption carried out on the edge devices before data transmission to a cloud or a fusion center. In fact, it has been reported in several works [38], [39] that CS-based encryption has a much lower cost as compared to well-established encryption standards such as AES [10] or RSA [11].

The idea of formally using CS theory in the encryption system was first introduced in [40]. These authors have considered a sparse signal $\mathbf{x}$ as a plain-text input signal and encrypted it in cipher-text $\mathbf{y}$. A Gaussian measurement matrix, as in (8), was used in the role of the CS-encryption matrix, i.e., $\mathbf{y} = \mathbf{Hx}$. They consider the Shannon perfect secrecy [41] definition as a metric of security. CS-based encryption can be viewed as a particular case of a multiplicative randomization technique, which is also a well-known privacy-preserving method. Using the definition of Shannon [41], CS-based encryption literature generally defines the perfect secrecy in the information-theoretical sense as follows:

Definition 2. (Perfect Encryption System) A perfect encryption system satisfies

$$\Pr(\mathbf{x}|\mathbf{y}) = \Pr(\mathbf{x}) \tag{10}$$

for any plain-text $\mathbf{x}$ and cipher-text $\mathbf{y}$ pair.

The authors of [40] conclude that even if the Shannon perfect secrecy is not satisfied with the CS-based encryption scheme since the CS-measurements preserve the energy of plain-text as $\mathbf{H}$ must satisfy the condition, they argue that CS-based encryption guarantees computational secrecy, i.e., an attacker with bounded time. In a later work, it is shown that the CS-based encryption with the Gaussian compression matrix used only once and re-drawn for each coding instance reveals only the energy of $\mathbf{x}$ [42]. Therefore, a Gaussian CS-encryption can be said to satisfy perfect secrecy if the cipher-text, $\mathbf{y}$ is normalized to some constant energy [36, Theorem 4]. Efforts on giving privacy guarantee conditions for both normalized and unnormalized energy cipher-texts for different measurement matrix schemes continue [43], [44] (using different security metrics). Similarly, instead of Shannon perfect secrecy, Wyner-sense perfect secrecy, or their extended version have also been used in security analysis for CS-based encryption schemes [45]. In the meantime, the robustness of the CS-based encryption against attacks is investigated in [46], [47]. In [46], the authors consider a brute force and structural attack where an adversary tries a grid search to estimate the CS-encryption matrix, $\mathbf{A}$. This attack type can be considered as a known cipher-text attack under one-time usage (or one-time secret, OTS). They conclude that the computational complexity of such an attack makes this type of brute-force attack infeasible. The known plain-text type attack (KPA) under one time usage is addressed in [47], where the adversary captures the plaintext and ciphertext pair, $(\mathbf{x}, \mathbf{y})$. Furthermore, the systems that use the same CS-encryption matrix many times are well known to be unsecure against this type of attacks [42], [40].

Due to interest in application scenarios of CS-based encryption, recently hybrid models that use both CS and conventional cipher systems have become popular. For instance, [48] applies a homomorphic cryptography function on top of the CS-encryption in a wireless sensor network system. In that sense, even in multi-usage of $\mathbf{A}$, the system can be made resilient against KPA. In another vein, authors in [49] have proposed a multi-class encryption system where the CS-encryption matrix is partially corrupted differently for each user, i.e., $\mathbf{A} = \mathbf{A} + \Delta\mathbf{A}$, $\Delta\mathbf{A}$ being the partial perturbation matrix. Their scheme suggests a framework to partially corrupt the CS-encryption matrix in order to obfuscate the sensitive region of the signal. However, it is not obvious how one transmits $\Delta\mathbf{A}$ to the receiving party for reversible de-identification. One intuitive approach would be sending $\Delta\mathbf{A}$ in a secure channel, which could be problematic, especially when the obfuscation pattern changes from usage to usage. Another solution is to use steganographic methods [16], [17], [50] to embed $\Delta\mathbf{A}$ directly on CS measurements, that is, by encoding the obfuscation matrix directly on the cipher-text $\mathbf{y}$. This is the path we follow and its details are introduced in the following section.

It is worth mentioning some recent work in the vein of compression (via sparsification) and encryption strategy. These methods extract a sparse code, $\mathbf{x}$, of the private signal and then obfuscate it. In [51], [52], [53] a ternary representation of the signal is extracted from its sparse code. Then this code is ambiguated for the privacy-protected data-sharing applications, e.g., outsourced media search or person identification applications. In [54], the authors study the reconstruction capability of sparse ternary codes given the information loss during its encoding to a ternary code. A more recent work [20] ambiguates the sparse code directly by noise addition while enabling high-quality recovery with successive refinement user.

IV. Proposed Two-tiered Encryption

The proposed method exploits techniques of compressive sampling, compressive encryption and data hiding [36], [37], [14], [55], [56], [57], [16], [17]. The advantage of the CS-based technique is, on one side, that exact recovery (in strictly sparse case) or stable recovery (in approximately sparse case) of the undersampled signal is possible, and on the other side, cryptographic security can be provided.

As shown in Fig. 1, one tier of the security consists of the generation of a random corruption mask (one-time usage) to obfuscate the sensitive parts of the image. This information is then embedded directly onto the CS-encrypted signal with a ternary watermark. This data hiding scheme provides reversibility and one-time usage of the random corruption mask, which is essential for secure de-identification. In the two-tiered protection scheme, the semi-authorized user will be able to recover only the non-sensitive part while a fully authorized user is allowed to recover the whole signal.

A. Problem Definition

In the following section, we first start by giving a formal definition of the two-tiered protection scheme in the spirit of Shannon secrecy. We will define the desiderata that the ideal triple consisting of two decoders (type A, B) and an encoder must satisfy. The problem becomes then formally the design of the three mappings that guarantee the recovery and secrecy properties. Following these definitions, we give our compressive sensing based solution to the problem with a discussion of the advantages of the proposed system.

The signals of interest, $\mathbf{s} \in \mathbb{R}^N$ is composed of a sensitive part and a non-sensitive part, denoted as an orthogonal sum

$$\mathbf{s} = \mathbf{s_n} + \mathbf{s_s}, \tag{11}$$

where $\mathbf{s_s}$ is the sensitive part of the signal that can be obtained by zero-outing the coefficients of $\mathbf{s}$ which are not indexed by the corresponding index set Λ_p, and $\mathbf{s_n}$ is the remaining non-sensitive part of the signal whose non-zero coefficients are indexed by Λ_p^c. In what follows, we state the information-theoretic desiderata of the encoder and of the two decoders.

Definition 3. Fully Secure and Stable Encoder-Decoders Triple: $\mathcal{E}^*(.)$, $\mathcal{D}_1^*(.)$, $\mathcal{D}_2^*(.)$

1) We define the data coding operator (CS-Encryption) as $\mathcal{E}^*(.)$ that encrypts both the sensitive and non-sensitive parts,

$$\mathcal{E}^*(\mathbf{s}) = \mathbf{y}, \tag{12}$$

which is perfectly secure in that the coded signal, $\mathbf{y}$ does not reveal any information about $\mathbf{s}$, i.e., $\Pr(\mathbf{s}|\mathbf{y}) = \Pr(\mathbf{s})$.

2) The first-tier decoder, $\mathcal{D}_1^*(.)$ which stably recovers the non-sensitive part while not disclosing any information about the sensitive part is characterized as follows

$$\left\| \left(\mathcal{D}_\mathbf{1}^*\left(\mathcal{E}^*(\mathbf{s}) + \mathbf{e}\right)\right)_{\mathbf{\Lambda_p^c}} - \mathbf{s}_{\mathbf{\Lambda_p^c}} \right\|_2 \leq \kappa \|\mathbf{e}\|_2 \tag{13}$$

and

$$\Pr\left(\mathbf{s}_{\mathbf{\Lambda_p}} | \left(\mathcal{D}_\mathbf{1}^*\left(\mathcal{E}^*(\mathbf{s}) + \mathbf{e}\right)\right)_{\mathbf{\Lambda_p}}\right) \approx \Pr\left(\mathbf{s}_{\mathbf{\Lambda_p}}\right), \tag{14}$$

where $\mathbf{e}$ is a possible additive perturbation on $\mathbf{y}$, i.e., $\mathbf{y} = \mathcal{E}^*(\mathbf{s}) + \mathbf{e}$.

3) Finally, the second-tier decoder that stably recovers both sensitive and non-sensitive parts is defined as

$$\left\| \left(\mathcal{D}_2^*\left(\mathcal{E}^*(\mathbf{s}) + \mathbf{e}\right)\right) - \mathbf{s} \right\|_2 \leq \kappa \|\mathbf{e}\|_2 . \tag{15}$$

The goal now is to find a practical coding operator, $\mathcal{E}(.)$ that jointly encrypts the sensitive and non-sensitive parts, which is as close as possible to the ideal operator $\mathcal{E}^*(.)$.

B. Embedding Operator, $\mathcal{E}(.)$

1) Obfuscation of the Sensitive Part within CS-Encryption: The proposed embedding operator obfuscates the sensitive part $\mathbf{s}_{\mathbf{\Lambda_p}}$ of the signal with the masking pattern $\mathbf{\Delta}_{\mathbf{\Lambda_p}}$, and then compressively samples the whole, consisting of the combination of the non-sensitive part $\mathbf{s_n}$ and the masked sensitive part. The resulting intermediate code $\mathbf{y_d}$ is given by:

$$\mathbf{y_d} = \mathbf{A}_{\mathbf{\Lambda_p^c}}\mathbf{s}_{\mathbf{\Lambda_p^c}} + \mathbf{A}_{\mathbf{\Lambda_p}}\mathbf{\Delta}_{\mathbf{\Lambda_p}}\mathbf{s}_{\mathbf{\Lambda_p}} = \mathbf{A}\mathbf{s_n} + \mathbf{A}_{\mathbf{\Lambda_p}}\mathbf{\Delta}_{\mathbf{\Lambda_p}}\mathbf{s}_{\mathbf{\Lambda_p}}, \tag{16}$$

where $\mathbf{s}_{\mathbf{\Lambda_p}}$ and $\mathbf{s}_{\mathbf{\Lambda_p^c}}$ are the extracted sensitive and non-sensitive parts of $\mathbf{s}$, respectively. Here $\mathbf{\Delta}_{\mathbf{\Lambda_p}} \in \mathbb{R}^{|\Lambda_p| \times |\Lambda_p|}$ is the multiplicative obfuscation operator, i.e., a diagonal matrix consisting of random numbers and operates only on the (vectorized) sensitive part of the signal, $\mathbf{s_s}$. In other words, $\mathbf{A}_{\mathbf{\Lambda_p}} \in \mathbb{R}^{m \times |\Lambda_p|}$ and $\mathbf{A}_{\mathbf{\Lambda_p^c}} \in \mathbb{R}^{m \times (N - |\Lambda_p|)}$ are the matrices consisting of the subsets of columns of $\mathbf{A}$ that are indexed by index sets Λ_p and Λ_p^c, respectively. The encoding in $\mathbf{y_d}$ can also be formulated as an additive mask:

$$\mathbf{y_d} = (\mathbf{A}+\mathbf{M})\,\mathbf{s}, \tag{17}$$

where $\mathbf{M}\in\mathbb{R}^{m\times N}$ is the masking matrix with all zeros except the columns, $\mathbf{M}_{\mathbf{\Lambda_p}}\in\mathbb{R}^{m\times|\Lambda_p|}$. The non-zero columns of the masking matrix form can be easily calculated from Eq. (16), i.e., $\mathbf{M}_{\mathbf{\Lambda_p}} = \mathbf{A}_{\mathbf{\Lambda_p}}\mathbf{\Delta}_{\mathbf{\Lambda_p}} - \mathbf{A}_{\mathbf{\Lambda_p}}$.

2) Data Hiding with Reversibility: The obfuscation matrix $\mathbf{\Delta}_{\mathbf{\Lambda_p}}$ and its location information (if necessary) are converted to a binary code to be secretly embedded on top of the compressively sensed (encrypted) signal $\mathbf{y_d}$. The conversion of this information to a binary code is necessary to achieve reversibility. Indeed, the exact recovery of the watermark sequence is possible [16], even in noisy case (In our scheme, noise corresponds to the masking in the sensitive part) provided the signal, $\mathbf{s}$, is sparse. In a practical application, errors in a few bits on the recovered watermark is tolerable. We can define a procedure that spits out a watermark $\mathbf{w}'$ corresponding to the binary representation of $\beta\left(\mathbf{\Delta}_{\mathbf{\Lambda_p}}\right)$,

$$\beta\left(\mathbf{\Delta}_{\mathbf{\Lambda_p}}\right) \to \mathbf{w}' \in \{-a,+a\}^{T'}, \tag{18}$$

where $\beta\left(\mathbf{\Delta}_{\mathbf{\Lambda_p}}\right)$ is sufficient information to re-produce $\mathbf{\Delta}_{\mathbf{\Lambda_p}}$. An example of such an operator is given in Eqs. (36) -(37c). We also need an inverse operator of (18) in order to reproduce $\mathbf{\Delta}_{\mathbf{\Lambda_p}}$ from watermark signal, i.e., $\mathbf{w}'$ as $\hat{\mathbf{w}} \overset{\beta^{-1}}{\to} \hat{\mathbf{\Delta}}_{\mathbf{\Lambda_p}}$. This operator is defined in Eqs. (25)- (28). Note that the length of the watermark, T', can change for each use case. To accommodate varying length watermarks one can fix a maximum watermark length, T, and extend the binary code $\mathbf{w}'$ to a ternary one by stuffing with zeros the remaining $T-T'$ bits, i.e.,

$$\beta\left(\mathbf{\Delta}_{\mathbf{\Lambda_p}}\right) \to \mathbf{w} \in \{-a,+a,0\}^{T}. \tag{19}$$

Data hiding limits [16], [17] determine the maximum steganographic capacity T one can expect to realize. Finally, a watermark embedding matrix (based on the second authorization key) $\mathbf{B}\in\mathbb{R}^{m\times T}, T<m$ is generated to linearly spread the watermark $\mathbf{w}$ directly onto the CS-encrypted signal, i.e., the cipher-text

$$\mathbf{y_w} = \mathbf{y_d} + \mathbf{B}\mathbf{w} = (\mathbf{A}+\mathbf{M})\mathbf{s} + \mathbf{B}\mathbf{w}. \tag{20}$$

An embedding power constraint $\|\mathbf{B}\mathbf{w}\| \leq P_E$ must be imposed in order to limit the degeneration of the recovered (non-sensitive) part of the image for semi-authorized users. The proposed embedding scheme, $\mathcal{E}(.)$ is given in Algorithm 1.

Algorithm 1 Proposed Embedding, $\mathcal{E}(.)$

```
Input: $\mathbf{s}$, $\mathbf{A}$, $\mathbf{B}$;
1. Determine the mask and the obfuscation matrix, $\mathbf{\Delta}_{\mathbf{\Lambda_p}}$
2. Generate the watermark: $\beta\left(\mathbf{\Delta}_{\mathbf{\Lambda_p}}\right) \to \mathbf{w} \in \{-a,+a,0\}^{T}$
3. Joint CS-encryption and sensitive part obfuscation:
$\mathbf{y_d} = \mathbf{A}_{\mathbf{\Lambda_p^c}}\mathbf{s}_{\mathbf{\Lambda_p^c}} + \mathbf{A}_{\mathbf{\Lambda_p}}\mathbf{\Delta}_{\mathbf{\Lambda_p}}\mathbf{s}_{\mathbf{\Lambda_p}}$
4. Watermark Embedding: $\mathbf{y_w} = \mathbf{y_d} + \mathbf{B}\mathbf{w}$
Return: $\mathbf{y_w}$
```

C. Design of the Two-tiered Decoders, $\mathcal{D}_1(.)$, $\mathcal{D}_2(.)$

Users (type A or B) receive the watermarked and encrypted signal, $\mathbf{y_w}$ which can be re-cast as

$$\mathbf{y_w} = (\mathbf{A}+\mathbf{M})\mathbf{s} + \mathbf{B}\mathbf{w} = \mathbf{H}\mathbf{x} + \mathbf{B}\mathbf{w} + \mathbf{n}, \tag{21}$$

where $\mathbf{H}\mathbf{x} = \mathbf{A}\mathbf{\Phi}\mathbf{x} = \mathbf{A}\mathbf{s}$, and $\mathbf{x}\in\mathbb{R}^N$ is the sparse representation of $\mathbf{s}$ in $\mathbf{\Phi}$, and the masked part can be expressed as noise term, i.e., $\mathbf{n} = \mathbf{M}\mathbf{s} = \mathbf{M}_{\mathbf{\Lambda_p}}\mathbf{s}_{\mathbf{\Lambda_p}} = \left(\mathbf{A}_{\mathbf{\Lambda_p}}\mathbf{\Delta}_{\mathbf{\Lambda_p}} - \mathbf{A}_{\mathbf{\Lambda_p}}\right)\mathbf{s}_{\mathbf{\Lambda_p}}$. For the receiver of Type-A (the semi-authorized user A) only the key $\mathbf{A}$ is available. Since this user does not have the watermark encrypting key, $\mathbf{B}$, (s)he will perceive the cyper-text as

$$\mathbf{y_w} = \mathbf{H}\mathbf{x} + \mathbf{z}, \tag{22}$$

where $\mathbf{z}$ behaves like an additive structural noise, i.e., $\mathbf{z} = \mathbf{B}\mathbf{w} + \mathbf{n}$. In the light of the discussion in Section 1, the ℓ_1-minimization scheme in (5) can be used to recover $\mathbf{x}$ with $\Upsilon(\mathbf{y}) = \{\mathbf{x} : \|\mathbf{H}\mathbf{x} - \mathbf{y}\|_2 \leq \epsilon\}$. Afterwards, using the outcome of the ℓ_1 minimization technique, $\hat{\mathbf{x}}$, one can obtain an estimate of the signal $\mathbf{s}$ with mask, $\hat{\mathbf{s}}$, straightforwardly via $\hat{\mathbf{s}} = \mathbf{\Phi}\hat{\mathbf{x}}$. The decoding algorithm for semi-authorized users, $\mathcal{D}_1(.)$ is given in Algorithm 2.

Algorithm 2 Type A Decoding Algorithm, $\mathcal{D}_1(.)$

```
Input: $\mathbf{y_w}$, $\mathbf{A}$, $\mathbf{\Phi}$;
Hyper-parameters: $\epsilon$
1. Estimate $\hat{\mathbf{x}}$ : $\tilde{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\mathbf{x}\|_1$  s.t.  $\|\mathbf{y_w} - \mathbf{H}\mathbf{x}\|_2 \leq \epsilon$
2. $\hat{\mathbf{s}} = \mathbf{\Phi}\hat{\mathbf{x}}$.
Return: $\hat{\mathbf{s}}$
```

The receiver of Type-B, (the fully-authorized user B) will possess both CS-encryption key, $\mathbf{A}$, and watermark encryption key, $\mathbf{B}$. Type-B decoder must recover the whole signal $\mathbf{s_n}+\mathbf{s_s}$ with as low a reconstruction error as possible. A three-stage recovery scheme is proposed, which is adapted from the recovery method proposed in [16]: First, a raw estimate of the sparse signal is obtained by disregarding the watermark part $\mathbf{B}\mathbf{w}$ and using the ℓ_1-minimization (5). Second, after having a preliminary estimation of $\mathbf{x}$, the watermark can be recovered from the over-determined system of linear equations by subtracting the estimated $\mathbf{x}$ component from $\mathbf{y_w}$. In the final stage, the masking matrix, $\mathbf{M}$, can be produced via the recovered watermark, and an improved estimation is obtained using the $\mathbf{A}+\mathbf{M}$ as CS-encryption matrix and ℓ_1-minimization. The details of the proposed scheme are as follows:

First, we produce a left annihilator matrix $\mathbf{F}\in\mathbb{R}^{p\times m}$ of $\mathbf{B}\in\mathbb{R}^{m\times T}$ so that $\mathbf{F}\mathbf{B} = 0$, where $p = m - T$. Left multiplying $\mathbf{y_w}$ with $\mathbf{F}$ we obtain,

$$\tilde{\mathbf{y}} = \mathbf{F}\mathbf{y_w} = \mathbf{F}\left(\mathbf{H}\mathbf{x} + \mathbf{B}\mathbf{w} + \mathbf{n}\right) = \mathbf{F}\mathbf{H}\mathbf{x} + \mathbf{n}', \tag{23}$$

where $\mathbf{n}' = \mathbf{F}\mathbf{n}$. Eq. (23) is also an underdetermined linear system of equations and can be solved via ℓ_1-minimization as discussed in Section 1:

$$\tilde{\mathbf{x}} = \arg\min \|\mathbf{x}\|_1 \;\text{ s.t. }\; \|\tilde{\mathbf{y}} - \mathbf{F}\mathbf{H}\mathbf{x}\|_2 \leq \epsilon. \tag{24}$$

After inserting the pre-estimation of $\mathbf{x}$ in $\mathbf{H}\tilde{\mathbf{x}}$ and subtracting it from the $\mathbf{y_w}$, we get an over-determined system of linear equations: $\mathbf{y_w} - \mathbf{H}\tilde{\mathbf{x}} = \mathbf{B}\mathbf{w}$. Therefore, a raw estimation of the watermark can be obtained via

$$\mathbf{w}'' = (\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T(\mathbf{y_w} - \mathbf{H}\tilde{\mathbf{x}}). \tag{25}$$

The 0's in the ternary watermark, $\mathbf{w}$ can be extracted using simple thresholding if the length of active bits T' is unknown to user B:

$$\tilde{\mathbf{w}} = \mathbf{w}'' \odot \mathbf{1}_{|w_i''|>\eta}, \tag{26}$$

where η is the threshold value

$$1_{|w_i''|>\eta, i} = \begin{cases} 1, & \text{if } |w_i''| > \eta, \quad (27a) \\ 0 & \text{else}, \quad (27b) \end{cases}$$

and $\odot$ denotes the element-wise multiplication operator between two vectors. In some practical applications such as person de-identification on video streams (details will be given in Section V), this step is simplified to $\tilde{\mathbf{w}} = \mathbf{w}'' \odot \mathbf{1}_{T'}$, where $\mathbf{1}_{T'}$ is T-length vector with the first T' elements 1's and the rest is all zeros. The locations of the non-zero elements of $\mathbf{1}_{T'}$ can be found using the information of Λ_p, inherent in the pre-estimated signal, $\mathbf{\Phi}\tilde{\mathbf{x}}$. Alternatively, a pre-allocated set from watermark, $\mathbf{w}$, can be dedicated to secretly carry information about T'. Hereafter, the finer estimation of $\mathbf{w}$ can be easily found via

$$\hat{w}_i = a \times \operatorname{sgn}(\tilde{w}_i). \tag{28}$$

Meanwhile the decoder can obtain the masking matrix, $\mathbf{M}$, i.e., $\hat{\mathbf{M}} = \left(\mathbf{A}_{\mathbf{\Lambda_p}}\hat{\mathbf{\Delta}}_{\mathbf{\Lambda_p}} - \mathbf{A}_{\mathbf{\Lambda_p}}\right)$, where $\hat{\mathbf{w}} \overset{\beta^{-1}}{\rightarrow} \hat{\mathbf{\Delta}}_{\mathbf{\Lambda_p}}$. Finally, the sensitive and non-sensitive parts can be jointly recovered as:

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \; \|\mathbf{x}\|_1 \quad \text{s.t.} \quad \left\|(\mathbf{y} - \mathbf{B}\hat{\mathbf{w}}) - (\mathbf{A} + \hat{\mathbf{M}})\mathbf{\Phi}\mathbf{x}\right\|_2 \leq \epsilon. \tag{29}$$

Algorithm 3 Type B Decoding Algorithm, $\mathcal{D}_2(.)$

```
Input: $\mathbf{y_w}$, $\mathbf{A}$, $\mathbf{B}$, $\mathbf{\Phi}$;
Hyper-parameters: $\epsilon$, $a$, $\eta$
1. Apply $\mathbf{F}$ to $\mathbf{y_w}$ : $\tilde{\mathbf{y}} = \mathbf{F}\mathbf{y_w}$
2. Estimate $\tilde{\mathbf{x}}$ : $\tilde{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\mathbf{x}\|_1$ s.t. $\|\tilde{\mathbf{y}} - \mathbf{F}\mathbf{H}\mathbf{x}\|_2 \leq \epsilon$
3. Estimate $\mathbf{w}''$ : $\mathbf{w}'' = (\mathbf{B}^{\mathrm{T}}\mathbf{B})^{-1}\mathbf{B}^{\mathrm{T}}(\mathbf{y_w} - \mathbf{H}\tilde{\mathbf{x}})$
4a. Thresholding $\mathbf{w}''$: $\tilde{\mathbf{w}} = \mathbf{w}'' \odot \mathbf{1}_{|w_i''|>\eta}$
4b. Forming $\hat{\mathbf{w}}$, where $\hat{w}_i = a * \operatorname{sgn}(\tilde{w}_i)$
5. Obtain $\hat{\mathbf{M}}$ from $\hat{\mathbf{w}}$: i) $\hat{\mathbf{w}} \overset{\beta^{-1}}{\rightarrow} \hat{\mathbf{\Delta}}_{\mathbf{\Lambda_p}}$ ii) $\hat{\mathbf{M}} = \left(\mathbf{A}_{\mathbf{\Lambda_p}}\hat{\mathbf{\Delta}}_{\mathbf{\Lambda_p}} - \mathbf{A}_{\mathbf{\Lambda_p}}\right)$
6. $\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\mathbf{x}\|_1$ s.t. $\left\|(\mathbf{y_w} - \mathbf{B}\hat{\mathbf{w}}) - (\mathbf{A} + \hat{\mathbf{M}})\mathbf{\Phi}\mathbf{x}\right\|_2 \leq \epsilon$
7. $\hat{\mathbf{s}} = \mathbf{\Phi}\hat{\mathbf{x}}$.
Return: $\hat{\mathbf{s}}$
```

D. Impact of Random Matrices on CS Encryption Performance

Generations of the CS-encryption matrix $\mathbf{A}$ and of the watermark embedding matrix $\mathbf{B}$ play an important role for the security and recovery robustness of the encryption scheme $\mathcal{E}(.), \mathcal{D}_1(.), \mathcal{D}_2(.)$. The choice of random Gaussian matrices as in (8) for $\mathbf{A}$ is convenient because they are known to be universally optimum in the sense that they satisfy both robustness and security conditions regardless of the sparsifying basis $\mathbf{\Phi}$. These matrices have been well investigated in the literature in terms of both recovery performance as in Corollary 1 and in terms of security metrics as discussed in Section III-B. In the sequel, we will consider $\mathbf{A}$ as in (8) and $\mathbf{B}$ consisting of orthonormal columns. For this scenario, we make a RIP based theoretical guarantee condition in watermark recovery for $\mathcal{D}_2(.)$. The following lemma will be useful for the stability analysis of the decoder type-B:

Lemma 1. *Consider that the embedding, $\mathcal{E}(.)$, given by Algorithm 1 produces an encrypted signal $\mathbf{y_w}$ from $\mathbf{s}$ with keys $\mathbf{A}$ and $\mathbf{B}$, i.e., $\mathcal{E}(\mathbf{s}) = \mathbf{y_w} = \mathbf{H}\mathbf{x} + \mathbf{B}\mathbf{w} + \mathbf{n}$. Let $\mathbf{s_p} \in \mathbb{R}^{|\Lambda_p| \times 1}$ denote the perturbation on the sensitive part of the signal such that $\mathbf{s_p} = \mathbf{\Delta}_{\mathbf{\Lambda_p}}\mathbf{s}_{\mathbf{\Lambda_p}} - \mathbf{s}_{\mathbf{\Lambda_p}}$. Let also $\mathbf{A}$ be an $m \times N$ CS-encryption matrix with elements $A_{i,j}$ drawn i.i.d. according to $\mathcal{N}(0, \frac{1}{m})$. Therefore the noise pattern $\mathbf{n}$ in (21) is also a Gaussian random vector which has i.i.d. elements*

$$n_i \sim \mathcal{N}(0, \frac{\|\mathbf{s_p}\|_2^2}{m}). \tag{30}$$

Proof. Let A_{i,Λ_p} be the the i^{th} row of $\mathbf{A}_{\mathbf{\Lambda_p}}$. Then the elements of the vector, $\mathbf{A}_{\mathbf{\Lambda_p}}\mathbf{s_p} \in \mathbb{R}^{m \times 1}$ will be $\mathbf{n}_i = \langle A_{i,\Lambda_p}, \mathbf{s_p} \rangle$ independent Gaussian random variables with zero means, where $\langle \mathbf{v_1}, \mathbf{v_2} \rangle$ refers to inner product of vectors $\mathbf{v_1}, \mathbf{v_2}$. Therefore, it remains to prove that $\mathbb{E}(n_i^2) = \frac{\|\mathbf{s_p}\|_2^2}{m}$, which can be straightforwardly obtained (using i.i.d. property)

$$\mathbb{E}\left(n_i^2\right) = \mathbb{E}\left(\langle A_{i,\Lambda_p}, \mathbf{s_p} \rangle^2\right) = \mathbb{E}\left(\sum_{j \in \Lambda_p} A_{i,j}^2 s_{p_j}^2\right) = \sum_{j \in \Lambda_p} s_{p_j}^2 \, \mathbb{E}\left(A_{i,j}^2\right) = \frac{1}{m}\sum_{j \in \Lambda_p} s_{p_j}^2 = \frac{\|\mathbf{s_p}\|_2^2}{m}. \tag{31}$$

□

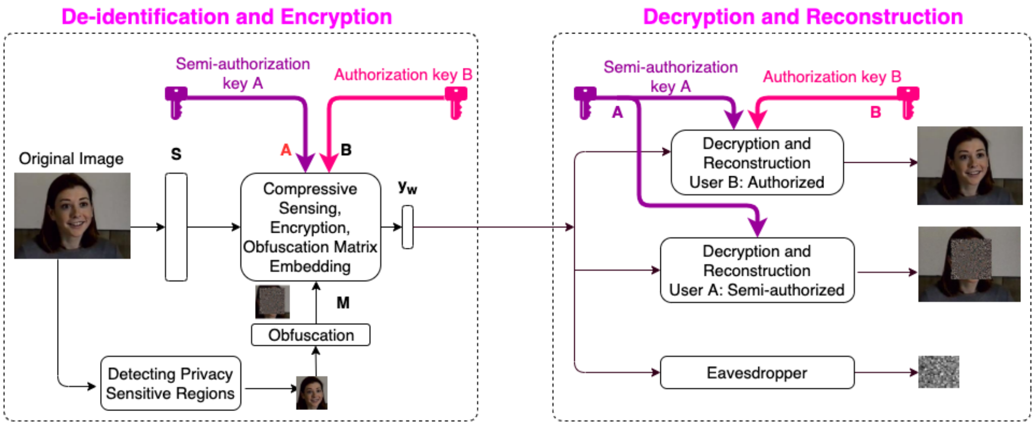
Having Lemma (1), and using Corollary 1 from the literature, we are ready to state the following theorem for watermark recovery probability of $\mathcal{D}_2(.)$:

Theorem 4. *Consider the Gaussian CS-encryption matrix defined in Eq. (8). Let the watermark-encoding matrix $\mathbf{B}$ have orthonormal columns. $\delta_{2k}(\mathbf{H}) < \sqrt{2} - 1$ and $\delta_{2k}(\mathbf{F}\mathbf{H}) < \sqrt{2} - 1$ are given. Let also the annihilator matrix $\mathbf{F}$ have orthogonal rows such that $\|F_{i,:}\|_2 = \frac{m}{p}$, where $F_{i,:}$ denotes the i^th row of $\mathbf{F}$. For a marked ciphertext, $\mathbf{y_w}$, for a particular setting of $\epsilon = (1+\gamma)\sqrt{m}\sigma_n$, Eq. (28) to be used in Algorithm 3 can recover w_i, the watermark bits, correctly $\Pr(w_i = \hat{w}_i)$ with probability at least*

$$\left(1 - 2\exp\left(\frac{-a'^2 m}{8\left\{C^2(1+\gamma)^2\right\}\|\mathbf{s_p}\|^2}\right)\right) \times \left(1 - \exp\left(-\frac{3p}{4}\gamma^2\right)\right), \tag{32}$$

where $C = 4\frac{\sqrt{1+\delta_{2k}(\mathbf{FH})}}{1-(1+\sqrt{2})\delta_{2k}(\mathbf{FH})}$ and $a' = a - \eta$, where a, ϵ and η are hyper-parameters used in Algorithm 3.

The proof of the theorem is given in Appendix A. Theorem 4 establishes a bound on the watermark recovery probability as a function of the energy of perturbation on the sensitive part, RIC of the matrix $\mathbf{FH}$ and watermark embedding strength a. This type of analysis based on RIP for the CS reconstruction


De-identification and Encryption
Decryption and Reconstruction
Semi-authorization key A
Authorization key B
Original Image
S
A
B
Compressive Sensing, Encryption, Obfuscation Matrix Embedding
y_w
M
Obfuscation
Detecting Privacy Sensitive Regions
Semi-authorization key A
Authorization key B
A
B
Decryption and Reconstruction User B: Authorized
Decryption and Reconstruction User A: Semi-authorized
Eavesdropper

User A: Semi-authorized user with only key A & User B: Authorized user with keys A and B

Fig. 1: Proposed Reversible Privacy-Preserving Video Monitoring

algorithm as in Corollary 1 is known as theoretical guarantee conditions in worst-case scenario [58]. In general, for most of the practical applications, the algorithms perform much better than the performance bounds given by this kind of RIP based analysis. Nevertheless, it gives us an indication on how to design the related matrices for the encoder (such as $\mathbf{A}$, $\mathbf{B}$, $\mathbf{H}$) and how to choose hyperparameters for the decoders. For example, choosing both $\mathbf{F}$ and $\mathbf{H}$ as Gaussian matrices may not be the right decision since the product of two random Gaussian matrices is a random matrix with coefficients drawn from a heavy-tailed distribution [59], which yields a $\delta_{2k}\left(\mathbf{FH}\right)$ bigger than the Gaussian case.

E. Choice of the encryption matrix

Although random measurement matrices are optimal in the universal sense, they become computationally unwieldy for realistic signal and measurement dimensions, N and m, respectively. Recall that the iterative signal reconstruction algorithms require transposition and multiplication of the measurement matrix several times. To ease this computational burden, one can choose the rows of the measurement (CS-encryption) matrix randomly as a subset of an orthonormal and fast implementable transform base such as Fourier, DCT, or Hadamard. In other words, one can choose m rows randomly out of the N the rows of an orthonormal transform, $\mathbf{\Theta}$. These rows are indexed by $\Omega \in \{1, 2, 3, ..., N\}$, i.e., with cardinality $|\Omega| = m$. Thanks to these types of structural CS matrices, the computational cost of $\mathbf{A}$s can be reduced significantly, i.e., down to $O(N \log N)$ flops from $O(m \times N)$ flops for general random CS matrices. For a good choice of the measurement matrix, $\mathbf{A} = \mathbf{\Theta_\Omega}$ in terms of a sparsifying basis $\mathbf{\Phi}$ the rows of $\mathbf{H}$ must be as flat (dense with nonzero elements) as possible. This can be satisfied when the rows of the measurement matrix $\mathbf{A}$ are not sparse in the sparsifying basis $\mathbf{\Phi}$. This requirement can be quantified via the ”mutual coherence” functional, i.e, $\mu(\mathbf{H}) = \max_{i,j} |H_{i,j}|$. The performance limits of the ℓ_1-decoding schemes such as (BPDN) case are given in terms of the functional $\mu(\mathbf{H})$. If one chooses randomly m rows of an orthonormal basis, $\mathbf{\Theta}$, indexed by $\Omega \in \{1, 2, 3, ..., N\}$ to build a measurement matrix $\mathbf{A}$, then a k-sparse signal can be exactly reconstructed as a solution of the ℓ_1-decoding (BP) in (5), satisfying $m \geq O(\mu^2(\mathbf{\Theta}) \times k \times \log N)$, with an overwhelming probability [60].

We have chosen the Noiselet basis and the 2-D Wavelet basis to create a CS-encryption matrix and a sparsifying matrix, respectively. First, since these two transforms are known to be maximally incoherent with each other, and second because they have fast implementations. The indices of the chosen rows are randomly drawn and then permuted to increase the security level.

F. Design of the annihilator matrix $\mathbf{F}$ and its corresponding watermark embedding matrix $\mathbf{B}$

The watermark embedding matrix $\mathbf{B}$, which must be the right null space matrix of $\mathbf{F}$, can also be chosen from a fast transform. For example, one can constitute the columns of $\mathbf{B}$ by choosing randomly a subset of the rows of DCT basis matrix, then, the rows of $\mathbf{F}$ can be made up of the remaining rows of this DCT matrix.

Theorem 4 implies that the choice of matrices $\mathbf{F}$ and $\mathbf{H}$ influences the performance of the Algorithm 3. To investigate the impact of the choice of $\mathbf{F}$ on $\mathbf{FH}$, we compare the performance of the ℓ_1 minimization on the recovery of sparse signal $\mathbf{x}$ from $\mathbf{y} = \mathbf{FHx}$, for three different settings: (i) First, with the random Gaussian measurement matrix $\mathbf{FH}$ as in Lemma 1, Theorem 4. (ii) Second, for the case where $\mathbf{F}$ is made up of a subset of the rows of DCT, $\mathbf{A}$ is similarly made of a subset of Noiselet basis, sparsifying matrix $\mathbf{\Phi}$ is chosen as Haar basis. Figure 2a shows the average mutual coherence values of $\mathbf{FH}$ under different setups. Figure 2b shows the exact recovery probabilities at different measurement rates for the three different choices of $\mathbf{F}$. These results prove that even if the random measurement matrix is universally optimum in the sense that it guarantees the exact recovery for any sparsifying basis in the worst case scenario, in practice structured matrices obtained from orthonormal transforms can perform even better. We make use of the mutual coherence functional; the formula below is slightly different from that given in the previous subsection, though related to it: $\mu\left(\mathbf{H}\right) =$

$\max_{1\leq i\leq j\leq N}\left(\frac{|<\mathbf{h_i},\mathbf{h_j}>|}{\|\mathbf{h_i}\|\|\mathbf{h_j}\|}\right)$ where $\mathbf{h_i}$ is the i^{th} column of matrix $\mathbf{H}$.

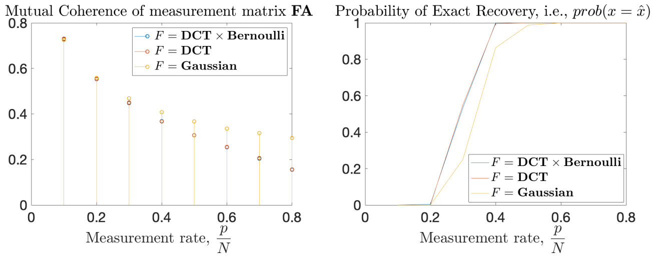


Fig. 2: Average mutual coherence of the matrix $\mathbf{FH} = \mathbf{FA\Phi}$ for different realizations of $\mathbf{A}$ and calculated probability of exact recovery over 250 trials. An exactly sparse signal is synthetically produced for $N = 256$ and $k = 30$.

(iii) Alternatively, based on the arguments in [61], a randomization matrix can be applied to $\mathbf{F}$, i.e., $\mathbf{F}' = \mathbf{FR}$, where $\mathbf{R}$ is $m \times m$ matrix of all zeros, except the diagonal terms that are drawn from the Bernoulli distribution. In [61], it is proven that the matrix $\mathbf{FRH}$ with any orthonormal basis pair, $\mathbf{F}$, $\mathbf{H}$ and randomization matrix $\mathbf{R}$ with diagonal Bernoulli elements, approaches a Gaussian matrix. This is, in fact, illustrated in Figure 3 as quantile-quantile plots. Although, this does not result in any performance increase vis-à-vis mutual coherence and recovery performance as shown in Figure 2, this scheme will enhance the security level with only negligible additional computation in the recovery part. In Figure 3, the vertical axis denotes the level at which the empirical distribution falls below a Q level (e.g., 50%), while the horizontal axis indicates the quantiles for the standard Gaussian distribution. In all cases, the similarity between the distribution of the $\mathbf{FH}$ sensing matrices and that of a Gaussian sensing matrix is obvious. Distribution of sensing matrices approaching that of a Gaussian is a desirable characteristic both for data hiding and CS-encryption purposes.

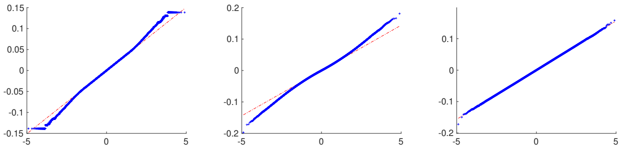


Fig. 3: Q-Q plots of the elements of the measurement matrices in reduced dimension, $\mathbf{FH}$. Vertical: sample data quantiles; horizontal: standard normal quantiles. a) $\mathbf{FH}$ = **Noiselet** × **Wavelet** b) $\mathbf{FH}$ = **DCT** × **Noiselet** × **Wavelet** c) $\mathbf{FH}$ = **DCT** × **Bernoulli** × **Noiselet** × **Wavelet**

G. Design of the obfuscation matrix

The region of interest (e.g., a face) to be obfuscated is delineated by Λ_p. Obfuscation matrix is constituted with all zero entries except for the diagonal elements that are drawn from a Bernoulli distribution with probability p_1, i.e.,

$$\Pr\left(\left(\mathbf{\Delta_{\Lambda_p}}\right)_{i,i} = \pm 1\right) = p_1 \tag{33}$$

The corresponding masking matrix $\mathbf{M}$ will be

$$M_{i,j} = \begin{cases} 0, & \text{if } j \in \Lambda_p \text{ and } \left(\mathbf{\Delta_{\Lambda_p}}\right)_{i,i} = 1 \quad \text{(34a)} \\ -2 * A_{i,j}, & \text{if } j \in \Lambda_p \text{ and } \left(\mathbf{\Delta_{\Lambda_p}}\right)_{i,i} = -1 \quad \text{(34b)} \\ 0, & j \notin \Lambda_p. \quad \text{(34c)} \end{cases}$$

Thus, the watermark generating procedure, will be

$$\beta\left(\mathbf{\Delta_{\Lambda_p}}\right) \rightarrow \left[\mathbf{w}' \ \mathbf{w}''\right], \tag{36}$$

where T'' bits $\mathbf{w}'$ are allocated for the location information of the sensitive part, i.e., the starting and ending points of rectangular region of interest including faces in the image and

$$w_i'' = \begin{cases} a, & \text{if } i \leq |\Lambda_p| \text{ and } \left(\mathbf{\Delta_{\Lambda_p}}\right)_{i,i} = 1 \quad \text{(37a)} \\ -a, & \text{if } i \leq |\Lambda_p| \text{ and } \left(\mathbf{\Delta_{\Lambda_p}}\right)_{i,i} = -1 \quad \text{(37b)} \\ 0 & i > |\Lambda_p|. \quad \text{(37c)} \end{cases}$$

Alternatively, having the intermediate estimation of image $\tilde{\mathbf{s}} = \mathbf{\Phi}\tilde{\mathbf{x}}$, the obfuscated region can be easily deduced and extracted, without the need of data hiding the location information in $\mathbf{w}'$.

H. More secure obfuscation with a key for a Gaussian vector

A semi-authorized user with only key-A may try to make a brute-force attack, by trying out all possible binary combinations of $\left(\mathbf{\Delta_{\Lambda_p}}\right)_{i,i}$'s to un-hide the obfuscated region. Even though the computational complexity of this attack is impractically high, i.e., $2^{|\Lambda_p|}$, to make the privacy protection stronger one can make use a third key, g. This can be realized using a predefined vector $\mathbf{g} \in \mathbb{R}^N$, that is known only to fully-authorized user (type B), which is used to generate another obfuscation matrix as

$$\left(\mathbf{\Delta_{\Lambda_p}}\right)_{i,i} = \begin{cases} \left(g_{\Lambda_p}\right)_i \text{ with probability } p_1, & \text{(38a)} \\ -\left(g_{\Lambda_p}\right)_i \text{ with probability } 1 - p_1, & \text{(38b)} \end{cases}$$

where $g_j \sim \mathcal{N}\left(\mu_g, \sigma_g^2\right)$.

V. A Case Study: Reversible Privacy-Preserving Video Monitoring

As a use case of the proposed two-tier image encryption algorithm, we investigate a video surveillance application where sensitive segments are to be concealed from semi-authorized users and revealed only to fully-authorized users. The sensitive parts of the image are the faces of people in the scene.

For face de-identification performance, we use two criteria: i) the Structural SIMilarity (SSIM) index [62] to measure the quality of the decoded and reconstructed image parts [63]; ii) face recognition accuracy via a machine learning algorithm as an indicator of privacy protection [64], [65]. For the semi-authorized user (with only key $\mathbf{A}$), we aim to have both minimum classification accuracy in the concealed parts and also minimum degradation in the reconstructed non-sensitive parts. For the fully-authorized user, we want to achieve the highest classification accuracy and highest reconstruction accuracy when both $\mathbf{A}$ and $\mathbf{B}$ keys are used for decoding.

We also test an attack scenario where the malicious user (e.g., a semi-authorized one or an attacker who has stolen the CS-encryption key, $\mathbf{A}$) has access to the labels of face images in the training set, so that (s)he can train a classifier to make

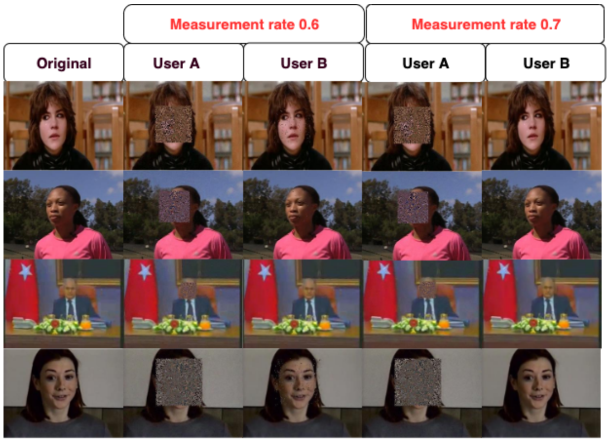


Fig. 4: Sample recovered frames for the semi-authorized (User A) and authorized (User B) (measurement rates 0.6, 0.7).

inferences from de-identified images. The experimental results (Section V-C and Table VI) show that our one-time usage of random obfuscation matrix prevents an adversarial from making an inference (identify the faces) even if the labels of the training set are captured.

A. Experimental Setup

The experimental evaluation is conducted on the YouTube Faces Database [22] to demonstrate the viability of the proposed method in such applications as video surveillance, intelligent access control, and in general, analytics for intelligent buildings. Accordingly, we have randomly chosen 5000 frames from YouTube Faces Database corresponding to 100 identities (50 frames per identity). Recovery performances are reported using 3000 frames while non-overlapping 2000 frames are collected to build the training set for privacy preservation performance evaluations. The Matlab implementation of the experiments and additional demos can be downloaded from https://github.com/mehmetyamac/CS-Privacy-Protection. We use a randomly chosen subset of the rows of noiselet basis as the measurement matrix. The implementation of the real-valued ”dragon” noiselet is borrowed from [66]. As the sparsifying matrix, we choose wavelet ”Coiflet 2” and use WaveLab850 [67] wavelet toolbox [1]. The columns of the encoding matrix $\mathbf{B}$ were chosen from the random subset of the columns of $m \times m$ DCT basis, and then were shuffled. Therefore, the rows of the annihilator matrix, $\mathbf{F}$ has been picked from the remaining columns and shuffled (i.e., $\mathbf{H} = \textbf{Noiselet} \times \textbf{Wavelet}$ and $\mathbf{F} = \textbf{DCT}$). Moreover, Gradient Projection for Sparse Reconstruction (GPSR) [68] was used for ℓ_1-minimization.

The various parameters taking place in the experiments are listed in Table II. For different watermark embedding power-to-signal ratio, $\frac{\|\mathbf{Bw}\|}{\|\mathbf{y_d}\|}$, and compression (measurement) rates, the performance of the decoders is reported in Section V-B.

TABLE II: List of the user defined parameters.

Variable	Name	Chosen Value	Mentioned Section
T	Watermark capacity	30000	Section IV-B2
T''	Number of bits allocated to encode segment location	$T'' = 0$ (Assuming that Λ_p is correctly found from $\tilde{\mathbf{s}}$)	Eq. (36)
$\frac{\|\mathbf{Bw}\|}{\|\mathbf{y_d}\|}$	Watermark embedding power	Tuned from the training set	Section V-B
μ_g	Mean of Gaussian key, g	0.9	Eq. (38a)- (38b)
σ_g^2	Variance of Gaussian key, g	0.1	Eq. (38a)- (38b)
p_1	Parameter p of Bernoulli dist. to sample obfuscation matrix.	0.5	Eq. (33)

B. Recovery Performance of $\mathcal{D}_1^*(.)$ and $\mathcal{D}_2^*(.)$

Choice of the watermark amplitude, a or alternatively the watermark embedding power is the determining factor in the watermark recovery performance (recall Theorem 4). In other words, the embedding power-to-signal ratio, $\frac{\|\mathbf{Bw}\|}{\|\mathbf{y_d}\|}$, forms the trade-off between the type A non-sensitive image recovery quality and type B sensitive image recovery quality. On the one hand, a should not be too small since the erroneous estimation of the watermark bits affects the recovery of w and $\hat{\mathbf{\Delta}}_{\mathbf{\Lambda_p}}$, hence the quality of the reconstructed sensitive part. On the other hand, increasing a could impede the decompression performance compromising the overall,$\mathbf{s_s} + \mathbf{s_{ns}}$ signal recovery, because the embedded watermark $\mathbf{Bw}$ acts as an additive noise in the decoder (Eq. (22)). This trade-off, recovery quality of sensitive regions (type B) and non-sensitive region for type A user, is observed in Figure 5. We have found empirically that good values of a are in the [0.085, 0.15] range, based on peak signal-to-noise ratios (PSNRs) and quality of recovered images.

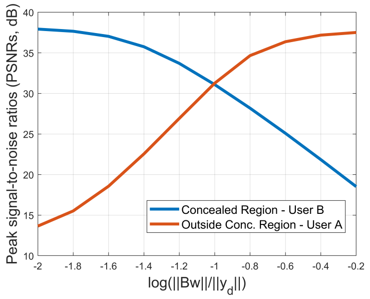


Fig. 5: Peak signal-to-noise ratios (PSNRs, dB) over recovered non-sensitive part (red curve), and sensitive part (blue curve) with the keys, respectively, of User A and User B. Measurement rate is fixed at 0.6.

In Table III, we show the recovery performance of type A and type B decoders for the concealed region, for the non-concealed region, and for the whole frame. Recovery qualities are reported for different compression rates (CS measurement rates: MR = m/N) and for two chosen values of $\frac{\|\mathbf{Bw}\|}{\|\mathbf{y_d}\|}$, namely, 0.15 and 0.085. Based on the visual assessment of the sample frames in Figure 4 and on the reported PSNR values in

[1] The original packet requires the input images to be square with dyadic sides, the Matlab modification in http://gtwavelet.bme.gatech.edu/ can be used to perform wavelet transformation with rectangular images with dyadic sides.

TABLE III: PSNR values over sensitive and non-sensitive regions of the frames for different measurement rates (MR) with a binary mask and a binary masked Gaussian for masking, and for embedding strength $\frac{\|\mathbf{Bw}\|}{\|\mathbf{y_d}\|} = 0.085$ (Tables a and b); $\frac{\|\mathbf{Bw}\|}{\|\mathbf{y_d}\|} = 0.15$ (Tables c and d), respectively.

(a)

	Sensitive Region		**Non-sensitive Region**		**Whole Frame**	
MR	**User A**	**User B**	**User A**	**User B**	**User A**	**User B**
0.3	11.90	16.61	26.37	27.31	21.67	24.24
0.4	11.65	20.23	28.61	30.91	22.20	27.36
0.5	11.48	24.63	30.51	34.59	22.50	30.92
0.6	11.36	29.79	32.06	38.17	22.66	34.92
0.7	11.28	35.0	33.29	41.30	22.75	38.83
0.8	11.22	39.37	34.25	43.69	22.80	42.02

(b)

	Sensitive Region		**Non-sensitive Region**		**Whole Frame**	
MR	**User A**	**User B**	**User A**	**User B**	**User A**	**User B**
0.3	12.06	16.82	26.56	27.34	21.81	24.37
0.4	11.84	20.31	28.78	30.93	22.35	27.49
0.5	11.68	24.49	30.64	34.62	22.67	31.02
0.6	11.57	29.34	32.18	38.22	22.84	34.96
0.7	11.49	34.18	33.40	41.33	22.94	38.79
0.8	11.43	38.17	34.36	43.71	22.99	41.90

(c)

	Sensitive Region		**Non-sensitive Region**		**Whole Frame**	
MR	**User A**	**User B**	**User A**	**User B**	**User A**	**User B**
0.3	11.93	20.08	24.94	27.57	21.11	25.58
0.4	11.70	24.54	26.46	31.46	21.57	29.38
0.5	11.53	29.46	27.64	35.37	21.83	33.46
0.6	11.40	34.33	28.54	38.83	21.98	37.31
0.7	11.30	38.14	29.22	41.65	22.06	40.47
0.8	11.22	40.72	29.73	43.85	22.11	42.79

(d)

	Sensitive Region		**Non-sensitive Region**		**Whole Frame**	
MR	**User A**	**User B**	**User A**	**User B**	**User A**	**User B**
0.3	12.09	20.21	25.11	27.68	21.24	25.74
0.4	11.88	24.45	26.61	31.57	21.71	29.52
0.5	11.72	29.10	27.77	35.48	21.98	33.56
0.6	11.60	33.61	28.66	38.90	22.13	37.32
0.7	11.50	37.07	29.32	41.69	22.22	40.39
0.8	11.43	39.42	29.83	43.88	22.27	42.67

TABLE IV: Structural Similarity Index (SSIM) over anonymized regions for different measurement rates (MR) using binary mask and binary masked Gaussian for masking for embedding strength $\frac{\|\mathbf{Bw}\|}{\|\mathbf{y_d}\|} = 0.085$ in (a) and 0.15 in (b).

(a)

	Binary		**Gaussian**	
MR	**User A**	**User B**	**User A**	**User B**
0.3	0.207	0.509	0.210	0.520
0.4	0.198	0.635	0.204	0.643
0.5	0.191	0.751	0.197	0.756
0.6	0.187	0.848	0.192	0.851
0.7	0.183	0.914	0.189	0.915
0.8	0.180	0.948	0.186	0.948

(b)

	Binary		**Gaussian**	
MR	**User A**	**User B**	**User A**	**User B**
0.3	0.205	0.646	0.211	0.656
0.4	0.196	0.768	0.202	0.775
0.5	0.190	0.864	0.195	0.867
0.6	0.185	0.921	0.190	0.921
0.7	0.181	0.951	0.186	0.951
0.8	0.177	0.966	0.183	0.965

Table III, we can say that User A's reconstructed faces are unrecognizable, whereas their outside regions have adequate quality, albeit around 5 dB lower in PSNRs as compared to those of User B, especially at low MRs. For User B, the reconstruction quality of both the concealed regions and the whole frame are satisfactory; there is only small detail losses in the privacy-sensitive parts.

In Table IV, SSIM values for the concealed region of reconstructed images are reported. It can be seen that faces in recovered images with using only Key $\mathbf{A}$ result in very low SSIM scores, making the unrecognizable, while their SSIM scores are very high for user type B, especially at MRs above 0.5.

C. Performance in Privacy Preservation

Privacy-preserving performance of the proposed method is evaluated by demonstrating its robustness against the state-of-art face recognition attacks. To this end, we employed a pre-trained Convolutional Neural Network (CNN) provided by the dlib library [69] to extract the facial features. Then, face recognition is performed as follows: We extract 128-dimensional embedded (CNN) face recognition features and build a database consisting of labeled faces for the query; then, perform a nearest-neighbor search and select the first nearest identity as the classification output. The experimental results are evaluated for two types of attacks.

1) Attack Type I: Known plain-text (original faces), known labels: In this scenario, a malicious user with the stolen Key $\mathbf{A}$ (or a malevolent type A user) may capture the training set with its labels to train a classifier to decipher the anonymized faces. The experiment designed to test the de-identification robustness against this type of attack is as follows: We construct a query database consisting of 2000 original clear frames (20 frames per identity). Then, we perform face recognition in the face regions that have been reconstructed with User A Key $\mathbf{A}$ and with the two User B keys. The recognition accuracies are reported in Table V. The performance of User A is about 1%, which is like a random guessing score while accuracies for User B are very satisfactory, i.e., around 75% for high MRs. This is comparable to the recognition rate achieved when the same face recognition software is tested on the original images.

TABLE V: Face recognition rates of User A and User B, the semi-authorized and authorized users, respectively, for different measurement rates (MR) using a binary mask and a binary masked Gaussian for masking, at $\frac{\|\mathbf{Bw}\|}{\|\mathbf{y_d}\|} = 0.085$ (Tables a and b), and at $\frac{\|\mathbf{Bw}\|}{\|\mathbf{y_d}\|} = 0.15$ (Tables c and d), respectively. The recognition accuracy on original frames is 77.37%.

(a)

MR	0.3	0.4	0.5	0.6	0.7	0.8
User A	0.0167	0.0183	0.0197	0.019	0.0187	0.0183
User B	0.0813	0.2353	0.4347	0.6107	0.7037	0.7353

(b)

MR	0.3	0.4	0.5	0.6	0.7	0.8
User A	0.017	0.0183	0.022	0.018	0.0207	0.019
User B	0.0797	0.2213	0.419	0.6073	0.707	0.7347

(c)

MR	0.3	0.4	0.5	0.6	0.7	0.8
User A	0.0147	0.021	0.0187	0.0173	0.0167	0.017
User B	0.199	0.4207	0.6323	0.705	0.7457	0.7553

(d)

MR	0.3	0.4	0.5	0.6	0.7	0.8
User A	0.0173	0.019	0.0183	0.0193	0.0173	0.0177
User B	0.1887	0.415	0.6193	0.6977	0.7417	0.7527

2) Attack Type II: Known plain-text (original faces), known anonymized and their labels: The ability of the proposed method to withstand a more challenging case, the parrot attack [70], where the user with Key $\mathbf{A}$ has captured both labeled clear images and their anonymized counterparts in the training set, is tested in the following experiment: The aforementioned query with NN-search is constructed in a way that each identity has 20 clean and 10 anonymized images with true labels. Face recognition algorithm is run over face regions in recovered images of type A. The results in Table VI reveal that the reconstructed faces for User A do not leak any useful information that can be exploited in a parrot attack since a different randomized corruption matrix was employed for each frame, i.e., the occurrence of the face with the same identity.

TABLE VI: Face recognition rates of the semi-authorized user when the corrupted images from User A are added into search space for nearest-neighbor. The accuracies are reported for different measurement rates (MR) with $\frac{\|\mathbf{Bw}\|}{\|\mathbf{y_d}\|} = 0.085$, and 0.15 using a binary mask and a binary masked Gaussian for the masking.

	Binary		Gaussian	
MR	0.085	0.15	0.085	0.15
0.3	0.017	0.0147	0.017	0.0173
0.4	0.0187	0.021	0.0183	0.019
0.5	0.0197	0.0187	0.022	0.0183
0.6	0.019	0.0173	0.018	0.0193
0.7	0.0187	0.0167	0.021	0.0177
0.8	0.0187	0.0177	0.019	0.0177

VI. Discussion

Privacy protection in video: We have so far tacitly assumed that privacy protection in video were to be realized in a frame-by-frame privacy processing mode. Thus, the sensitive part in each frame, e.g., face region was to be separately obfuscated and each such frame CS-encrypted via the $\mathbf{B}$, i.e., $\mathbf{y_w} = (\mathbf{A}+\mathbf{M})\mathbf{s}+\mathbf{Bw}$ formulation. A simple extension to a multi-frame video case would be to vectorize groups of frames, and straightforwardly adapt the above methodology, where now $\mathbf{s}_{\mathbf{\Lambda_p}}$ and $\mathbf{\Delta}_{\mathbf{\Lambda_p}}$ denote the sensitive parts and masking patterns striding over the frames in the group. A more principled way to extend the scheme to multi-frame video must leverage a tensor based CS-encryption scheme [71]. The video is considered as a 3-D signal, $\mathcal{S} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, which is a sequence of n_3 consecutive $n_1 \times n_2$ images. Then, the CS-encryption matrices, $\mathbf{A_1} \in \mathbb{R}^{m_1 \times n_1}, \mathbf{A_2} \in \mathbb{R}^{m_2 \times n_2}, \mathbf{A_3} \in \mathbb{R}^{m_3 \times n_3}$, can be applied over to $\mathcal{S}$ in order to obtain an encrypted and compressed tensor, i.e., $\mathcal{Y} = \mathcal{S} \times_1 \mathbf{A_1} \times_2 \mathbf{A_2} \times_3 \mathbf{A_3}$, where $\mathcal{S} \times_i \mathbf{A_i}$ is the i-mode product of tensor $\mathcal{S}$ and matrix $\mathbf{A_i}$. Let $\mathcal{S}_\mathbf{s}$ be the sensitive part of the video that is obtained by zero-outing the coefficients of $\mathcal{S}$ and $\mathcal{S}_\mathbf{n}$ is non-sensitive part of it. Similar to our matrix-vector notation, jointly CS-encrypted and anonymized tensor can be obtained via $\mathcal{Y}_\mathbf{d} = (\mathcal{S}_\mathbf{n} + \mathcal{P} \circ \mathcal{S}_\mathbf{s}) \times_1 \mathbf{A_1} \times_2 \mathbf{A_2} \times_3 \mathbf{A_3}$ where $\mathcal{P}$ is the degradation tensor and $\circ$ is element-wise (Hadamard) product of two tensors. Then, the marked vector, $\mathbf{y_w}$ can be easily obtained i.e., $\mathbf{y_w} = \mathrm{vec}(\mathcal{Y}_\mathbf{d}) + \mathbf{Bw}$. In the decoder part, a recovery algorithm with $\mathcal{D}_1(.)$ and $\mathcal{D}_2(.)$ similar to those in Algorithm 2 and Algorithm 3 can be used with replacing ℓ_1 based sparse vector recovery to a sparse tensor estimation method.

Multi-tier privacy protection: It is possible to extend the proposed scheme to more than two-tiers by replicating the scheme outlined in Subsection IV-C and Figure 1. Recall that the obfuscation mask encoded as $\mathbf{w}$ and embedded via an appropriate watermarking matrix $\mathbf{B}$ resulted in the expression $\mathbf{y_w} = (\mathbf{A} + \mathbf{M})\mathbf{s} + \mathbf{Bw}$. Consider, for example, a three-tier scenario, where $\mathbf{s_{s_1}}$ and $\mathbf{s_{s_2}}$ are identified as sensitive parts, the higher indexed components having, for example, a higher privacy concern. The respective obfuscation matrices, $\mathbf{M_1}$ and $\mathbf{M_2}$ are encoded by their corresponding watermarks $\mathbf{w_1}$ and $\mathbf{w_2}$. These watermark signals can be spread over $\mathbf{y_d}$, for example, as $\mathbf{y_w} = (\mathbf{A} + \mathbf{M_1} + \mathbf{M_2})\mathbf{s} + \mathbf{B_1 w_1} + \mathbf{B_2 w_2}$ or $\mathbf{y_w} = (\mathbf{A} + \mathbf{M_1} + \mathbf{M_2})\mathbf{s} + [\mathbf{B_1}\ \mathbf{B_2}]\,[\mathbf{w_1}; \mathbf{w_2}]$. If desired, the resulting signal $\mathbf{y_w}$ can be finally subjected to another layer of light-weight encryption. The decoding of the three-tier scheme follows steps similar to Section IV-C and Algorithm 3.

In this work, we have considered privacy protection in images and video as an application case. However, the proposed signal acquisition, privacy-protection and encryption scheme can be applied to any multimedia data that can be differentiated into sensitive (private) and non-sensitive (public) parts. A case in point could be the monitoring data of a wireless sensor network [72]. In such a distributed sensing mechanism, one may want to hide data in the sensor readings that would lead to traffic analysis and flow tracing. Another example would be a CS-based telehealth system [13] where health personnel with different authorization would have differential access to parts of medical data and biosignals.

Furthermore, using CS-encryption together with other lightweight encryption techniques is a common practice in the literature. For instance, in [72], the authors used Pailier cryptosystem over $\mathbf{y} = \mathbf{A}\mathbf{s}$, to strengthen the security. Similar approaches can be applied over $\mathbf{y_w}$ provided that invertibility of the applied encryption method.

VII. Conclusion

We have presented a two-tiered (potentially, multi-tiered) privacy-preserving scheme based on compressive sensing theory. The scheme accommodates two levels of users: A public user A (with only Key $\mathbf{A}$), who can recover only the non-sensitive portions of the document, and private B, i.e., the fully-authorized user who (with keys A and B) who can recover the whole document. This prioritization is enabled via a data hiding technique such that the full user in possession of (Key $\mathbf{B}$) can undo the obfuscation from within the CS-enciphered signal.

The watermark capacity of the system allows one-time usage of the obfuscation matrix, which in turn provides a higher level of security against any attacker, e.g., a curious semi-authorized user. In conclusion, the proposed approach satisfies all the criteria of privacy-protecting encoding, as itemized in the introduction section. Security can be corroborated by extra randomization as in Eq. (38a)-(38b). Extensive tests on a face anonymization use case revealed that the system is robust against cipher breaking attacks (i.e., face recognition) and that the image recovery quality is adequate for measurement rates m/N above 0.5. The experiments yielded guidelines for the selection of system parameters like compression rate and watermark embedding strength.

The proposed scheme with its experimentally proven merits of reversible anonymization provides a promising alternative of privacy-protecting encryption. An application scenario would be a video surveillance system where the collected real-time data must be transmitted and uploaded in a security monitoring center.

Appendix
Proof of Theorem 4

Using Equation 9 in Lemma 1 of [16] and the fact that $\mathbf{n'}_i \sim \mathcal{N}(0, \frac{m}{p}\sigma_n^2)$, where $\mathbf{n}' = \mathbf{F}\mathbf{n}$ we get

$$\Pr(\|\mathbf{F}\mathbf{n}\|_2 \geq (1+\gamma)\frac{\sqrt{m}}{\sqrt{p}}\sqrt{p}\sigma_n) \leq e^{-\frac{3p}{4}\gamma^2}. \tag{A.39}$$

Therefore, when we set $\epsilon = (1+\gamma)\sqrt{m}\sigma_n$ in Algorithm 3 and use the inequality that $\tilde{\mathbf{x}}$ in (24) satisfies

$$\|\mathbf{x} - \tilde{\mathbf{x}}\|_2 \leq C\epsilon \tag{A.40}$$

with probability at least $1 - \exp(-\frac{3p}{4}\gamma^2)$, where

$$C = 4\frac{\sqrt{1+\delta_{2k}(\mathbf{FH})}}{1-(1+\sqrt{2})\delta_{2k}(\mathbf{FH})} \tag{A.41}$$

Now, we define the error causing uncertainty, $\mathbf{z}$ on over-determined system, $\mathbf{y_w} - \mathbf{H}\tilde{\mathbf{x}} = \mathbf{B}\mathbf{w} + \mathbf{z}$. When we insert the $\tilde{\mathbf{x}}$ in Equation (21), we get

$$\mathbf{y_w} = \mathbf{B}\mathbf{w} + \mathbf{H}(\mathbf{x} - \tilde{\mathbf{x}}) + \mathbf{H}\tilde{\mathbf{x}} + \mathbf{n}, \tag{A.42}$$

which can be re-cast as

$$\mathbf{y_w} - \mathbf{H}\tilde{\mathbf{x}} = \mathbf{B}\mathbf{w} + \mathbf{H}(\mathbf{x} - \tilde{\mathbf{x}}) + \mathbf{n} = \mathbf{B}\mathbf{w} + \mathbf{z}', \tag{A.43}$$

where $\mathbf{z}' = \mathbf{H}(\mathbf{x} - \tilde{\mathbf{x}}) + \mathbf{n}$. Given $\|\mathbf{x} - \tilde{\mathbf{x}}\|_2 \leq C\epsilon$ and $\mathbf{H}_{i,j} \sim \mathcal{N}(0, \frac{1}{m})$. Via a similar mathematical derivation given in Lemma 1, we can claim that $\mathbf{H}(\mathbf{x} - \tilde{\mathbf{x}})$ is a Gaussian random vector where each element has a variance $\sigma^2 \leq \frac{C^2\epsilon^2}{m} = \frac{C^2(1+\gamma)^2 m\sigma_n^2}{m} = C^2(1+\gamma)^2\sigma_n^2$. Therefore, elements of $\mathbf{z}'$ is also a Gaussian vector with elements having variance $\sigma_{z'}^2 \leq 4\left\{C^2(1+\gamma)^2\right\}\sigma_n^2$. Knowing that the matrix $\mathbf{B}$ has orthonormal vectors, the pre-estimation $\mathbf{w}''$ from (25) satisfies

$$\mathbf{w} - \mathbf{w}'' = \mathbf{B}^T\mathbf{z}' = \mathbf{z}'', \tag{A.44}$$

with $z_i'' \sim \mathcal{N}\left(0, \sigma_{z'}^2\right)$. Finally, using Equation 2.17 in [73, Chapter 2], the probability of making an error in watermark bits can be easily found as

$$\begin{aligned}\Pr(w_i \neq \hat{w}_i|\{\|\mathbf{x} - \tilde{\mathbf{x}}\|_2 \leq C\epsilon\}) &= \Pr\left(|(z_i'| \geq a'\right) \\ \leq 2\exp\left(\frac{-a'^2}{2\sigma_{z'}^2}\right) &= 2\exp\left(\frac{-a'^2 m}{8\left\{C^2(1+\gamma)^2\right\}\|\mathbf{s}_p\|^2}\right),\end{aligned} \tag{A.45}$$

where $a' = a - \eta$ with a and η are user defined parameters to be used in Algorithm 3.

Acknowledgements

This work was supported by a NSF-Business Finland CVDI project (Amalia 3333/31/2018) and a Business Finland project VIRPA D (7940/31/2017) sponsored by Tieto Oyj and other companies.

References

[1] M. R. Palattella, M. Dohler, A. Grieco, G. Rizzo, J. Torsner, T. Engel, and L. Ladid, "Internet of things in the 5g era: Enablers, architecture, and business models," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 3, pp. 510–527, 2016.

[2] T. A. Nguyen and M. Aiello, "Energy Intelligent Buildings Based on User Activity: A survey," *Energy and Buildings*, vol. 56, pp. 244–257, 2013.

[3] S. R. Islam, D. Kwak, M. H. Kabir, M. Hossain, and K.-S. Kwak, "The Internet of Things for Health Care: a Comprehensive Survey," *IEEE Access*, vol. 3, pp. 678–708, 2015.

[4] P. Voigt and A. Von dem Bussche, "The EU General Data Protection Regulation (GDPR)," *A Practical Guide, 1st Ed., Cham: Springer International Publishing*, 2017.

[5] I. Wagner and D. Eckhoff, "Technical Privacy Metrics: a Systematic Survey," *ACM Computing Surveys (CSUR)*, vol. 51, no. 3, p. 57, 2018.

[6] C. C. Aggarwal and S. Y. Philip, "A General Survey of Privacy-Preserving Data Mining Models and Algorithms," in *Privacy-Preserving Data Mining*. Springer, 2008, pp. 11–52.

[7] R. Mendes and J. P. Vilela, "Privacy-Preserving Data Mining: Methods, Metrics, and Applications," *IEEE Access*, vol. 5, pp. 10 562–10 582, 2017.

[8] J. R. Padilla-López, A. A. Chaaraoui, and F. Flórez-Revuelta, "Visual Privacy Protection Methods: A Survey," *Expert Systems with Applications*, vol. 42, no. 9, pp. 4177–4195, 2015.

[9] J. R. Troncoso-Pastoriza and F. Perez-Gonzalez, "Secure Signal Processing in the Cloud: Enabling Technologies for Privacy-preserving Multimedia Cloud Processing," *IEEE Signal Processing Magazine*, vol. 30, no. 2, pp. 29–41, 2013.

[10] S. Heron, "Advanced Encryption Standard (AES)," *Network Security*, vol. 2009, no. 12, pp. 8 – 12, 2009. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1353485810700064

[11] R. L. Rivest, A. Shamir, and L. Adleman, "A method for Obtaining Digital Signatures and Public-Key Cryptosystems," *Communications of the ACM*, vol. 21, no. 2, pp. 120–126, 1978.

[12] F. Nilsson *et al.*, *Intelligent Network video: Understanding Modern Video Surveillance Systems*. CRC Press, 2016.

[13] H. Mamaghanian, N. Khaled, D. Atienza, and P. Vandergheynst, "Compressed Sensing for Real-Time Energy-Efficient ECG Compression on Wireless Body Sensor Nodes," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 9, pp. 2456–2466, 2011.

[14] E. J. Candès *et al.*, "Compressive sampling," in *Proceedings of the International Congress of Mathematicians*, vol. 3, 2006, pp. 1433–1452.

[15] Y. Zhang, Y. Xiang, and L. Y. Zhang, *Secure Compressive Sensing in Multimedia Data, Cloud Computing and IoT*. Springer, 2018.

[16] M. Yamac, Ç. Dikici, and B. Sankur, "Hiding Data in Compressive Sensed Measurements: A Conditionally Reversible Data Hiding Scheme for Compressively Sensed Measurements," *Digital Signal Processing*, vol. 48, pp. 188–200, 2016.

[17] M. Yamaç, B. Sankur, and M. Gabbouj, "Robust data hiding scheme for compressively sensed signals," in *Proceedings of the European Signal Processing Conference*, 2018, pp. 1760–1764.

[18] L. Du, M. Yi, E. Blasch, and H. Ling, "Garp-face: Balancing privacy protection and utility preservation in face de-identification," in *Proceedings of the IEEE International Joint Conference on Biometrics*, 2014, pp. 1–8.

[19] P. Agrawal and P. Narayanan, "Person De-identification in Videos," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 21, no. 3, pp. 299–310, 2011.

[20] S. Ferdowsi, B. Razeghi, T. Holotyak, F. P. Calmon, and S. Voloshynovskiy, "Privacy-preserving image sharing via sparsifying layers on convolutional groups," in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020, pp. 2797–2801.

[21] M. Yamaç, M. Ahishali, N. Passalis, J. Raitoharju, B. Sankur, and M. Gabbouj, "Reversible privacy preservation using multi-level encryption and compressive sensing," in *2019 27th European Signal Processing Conference (EUSIPCO)*, Sep. 2019, pp. 1–5.

[22] L. Wolf, T. Hassner, and I. Maoz, "Face recognition in unconstrained videos with matched background similarity," in *CVPR 2011*, 2011, pp. 529–534.

[23] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse MRI: The Application of Compressed Sensing for Rapid MR Imaging," *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, vol. 58, no. 6, pp. 1182–1195, 2007.

[24] A. C. Gurbuz, J. H. McClellan, and W. R. Scott, "A Compressive Sensing Data Acquisition and Imaging Method for Stepped Frequency GPRs," *IEEE Transactions on Signal Processing*, vol. 57, no. 7, pp. 2640–2650, 2009.

[25] M. Yamaç, M. Orhan, B. Sankur, A. S. Turk, and M. Gabbouj, "Through the wall target detection/monitoring from compressively sensed signals via structural sparsity," in *5th International Workshop on Compressed Sensing applied to Radar, Multimodal Sensing,and Imaging*, 2018.

[26] B. Hamdaoui, B. Khalfi, and M. Guizani, "Compressed Wideband Spectrum sensing: Concept, Challenges, and Enablers," *IEEE Communications Magazine*, vol. 56, no. 4, pp. 136–141, 2018.

[27] S. Gishkori, V. Lottici, and G. Leus, "Compressive Sampling-Based Multiple Symbol Differential Detection for UWB Communications," *IEEE Transactions on Wireless Communications*, vol. 13, no. 7, pp. 3778–3790, 2014.

[28] Z. Gao, L. Dai, S. Han, I. Chih-Lin, Z. Wang, and L. Hanzo, "Compressive Sensing Techniques for Next-Generation Wireless Communications," *IEEE Wireless Communications*, vol. 25, no. 3, pp. 144–153, 2018.

[29] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ_1 minimization," *Proceedings of the National Academy of Sciences*, vol. 100, no. 5, pp. 2197–2202, 2003.

[30] S. Chen and D. Donoho, "Basis pursuit," in *Proceedings of 1994 28th Asilomar Conference on Signals, Systems and Computers*, vol. 1. IEEE, 1994, pp. 41–44.

[31] A. Cohen, W. Dahmen, and R. DeVore, "Compressed Sensing and Best k-term Approximation," *Journal of the American Mathematical Society*, vol. 22, no. 1, pp. 211–231, 2009.

[32] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic Decomposition by Basis Pursuit," *SIAM review*, vol. 43, no. 1, pp. 129–159, 2001.

[33] E. J. Candès, "The Restricted Isometry Property and Its Implications for Compressed Sensing," *Comptes Rendus Mathématique*, vol. 346, no. 9, pp. 589 – 592, 2008. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1631073X08000964

[34] E. J. Candes, J. K. Romberg, and T. Tao, "Stable Signal Recovery from Incomplete and Inaccurate Measurements," *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, vol. 59, no. 8, pp. 1207–1223, 2006.

[35] Y. C. Eldar and G. Kutyniok, *Compressed Sensing: Theory and Applications*. Cambridge University Press, 2012.

[36] Y. Zhang, Y. Xiang, L. Y. Zhang, Y. Rong, and S. Guo, "Secure Wireless Communications Based on Compressive Sensing: A Survey," *IEEE Communications Surveys Tutorials*, vol. 21, no. 2, pp. 1093–1111, 2019.

[37] Y. Zhang, L. Y. Zhang, J. Zhou, L. Liu, F. Chen, and X. He, "A Review of Compressive Sensing in Information Security Field," *IEEE Access*, vol. 4, pp. 2507–2519, 2016.

[38] R. Dautov and G. R. Tsouri, "Establishing secure measurement matrix for compressed sensing using wireless physical layer security," in *2013 International Conference on Computing, Networking and Communications (ICNC)*. IEEE, 2013, pp. 354–358.

[39] R. Huang, K. Rhee, and S. Uchida, "A Parallel Image Encryption Method Based on Compressive Sensing," *Multimedia Tools and Applications*, vol. 72, no. 1, pp. 71–93, 2014.

[40] Y. Rachlin and D. Baron, "The secrecy of compressed sensing measurements," in *Proceedings of the Annual Allerton Conference on Communication, Control, and Computing*, 2008, pp. 813–817.

[41] C. E. Shannon, "Communication Theory of Secrecy Systems," *Bell System Technical Journal*, vol. 28, no. 4, pp. 656–715, 1949.

[42] T. Bianchi, V. Bioglio, and E. Magli, "Analysis of One-Time Random Projections for Privacy Preserving Compressed Sensing," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 2, pp. 313–327, 2016.

[43] M. Testa, D. Valsesia, T. Bianchi, and E. Magli, "Compressed Sensing as A Cryptosystem," in *Compressed Sensing for Privacy-Preserving Data Processing*. Springer, 2019, pp. 25–71.

[44] N. Y. Yu, "Indistinguishability and Energy Sensitivity of Gaussian and Bernoulli Compressed Encryption," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 7, pp. 1722–1735, 2018.

[45] Z. Yang, W. Yan, and Y. Xiang, "On the Security of Compressed Sensing-Based Signal Cryptosystem," *IEEE Transactions on Emerging Topics in Computing*, vol. 3, no. 3, pp. 363–371, 2014.

[46] A. Orsdemir, H. O. Altun, G. Sharma, and M. F. Bocko, "On the security and robustness of encryption via compressed sensing," in *Proceedings of the IEEE Military Communications Conference*, 2008, pp. 1–7.

[47] 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 Transactions on Information Forensics and Security*, vol. 10, no. 10, pp. 2182–2195, 2015.

[48] P. Zhang, S. Wang, K. Guo, and J. Wang, "A Secure Data Collection Scheme Based on Compressive Sensing in wireless Sensor Networks," *Ad Hoc Networks*, vol. 70, pp. 73–84, 2018.

[49] V. Cambareri, M. Mangia, F. Pareschi, R. Rovatti, and G. Setti, "Low-Complexity Multiclass Encryption by Compressed Sensing," *IEEE Transactions on Signal Processing*, vol. 63, no. 9, pp. 2183–2195, 2015.

[50] M. Yamaç, Ç. Dikici, and B. Sankur, "Robust watermarking of compressive sensed measurements under impulsive and gaussian attacks," in *Proceedings of the European Signal Processing Conference*, 2013, pp. 1–5.

[51] B. Razeghi, S. Voloshynovskiy, S. Ferdowsi, and D. Kostadinov, "Privacy-preserving identification via layered sparse code design: Distributed servers and multiple access authorization," in *2018 26th European Signal Processing Conference (EUSIPCO)*. IEEE, 2018, pp. 2578–2582.

[52] B. Razeghi, S. Voloshynovskiy, D. Kostadinov, and O. Taran, "Privacy preserving identification using sparse approximation with ambiguization," in *2017 IEEE Workshop on Information Forensics and Security (WIFS)*. IEEE, 2017, pp. 1–6.

[53] B. Razeghi and S. Voloshynovskiy, "Privacy-preserving outsourced media search using secure sparse ternary codes," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2018, pp. 1992–1996.

[54] S. Rezaeifar, B. Razeghi, O. Taran, T. Holotyak, and S. Voloshynovskiy, "Reconstruction of privacy-sensitive data from protected templates," in *2019 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2019, pp. 1163–1167.

[55] D. L. Donoho *et al.*, "Compressed Sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[56] K. Liu, H. Kargupta, and J. Ryan, "Random Projection-Based Multiplicative Data Perturbation for Privacy Preserving Distributed Data Mining," *IEEE Transactions on Knowledge and Data Engineering*, vol. 18, no. 1, pp. 92–106, 2005.

[57] H. Kargupta, S. Datta, Q. Wang, and K. Sivakumar, "Random-Data Perturbation Techniques and Privacy-Preserving Data Mining," *Knowledge and Information Systems*, vol. 7, no. 4, pp. 387–414, 2005.

[58] K. Schnass and P. Vandergheynst, "Average Performance Analysis for Thresholding," *IEEE Signal Processing Letters*, vol. 14, no. 11, pp. 828–831, 2007.

[59] B. G. Ivanoff and N. Weber, "Tail Probabilities for Weighted Sums of Products of Normal Random Variables," *Bulletin of the Australian Mathematical Society*, vol. 58, no. 2, pp. 239–244, 1998.

[60] E. Candes and J. Romberg, "Sparsity and Incoherence in Compressive Sampling," *Inverse Problems*, vol. 23, no. 3, p. 969, 2007.

[61] T. T. Do, L. Gan, N. H. Nguyen, and T. D. Tran, "Fast and Efficient Compressive Sensing Using Structurally Random Matrices," *IEEE Transactions on Signal processing*, vol. 60, no. 1, pp. 139–154, 2011.

[62] Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli *et al.*, "Image Quality Assessment: from Error Visibility to Structural Similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[63] F. Dufaux, "Video scrambling for privacy protection in video surveillance: recent results and validation framework," in *Mobile Multimedia/Image Processing, Security, and Applications 2011*, S. S. Agaian, S. A. Jassim, and Y. Du, Eds., vol. 8063, International Society for Optics and Photonics. SPIE, 2011, pp. 11 – 24. [Online]. Available: https://doi.org/10.1117/12.883948

[64] P. Korshunov, A. Melle, J.-L. Dugelay, and T. Ebrahimi, "Framework for objective evaluation of privacy filters," in *Applications of Digital Image Processing XXXVI*, A. G. Tescher, Ed., vol. 8856, International Society for Optics and Photonics. SPIE, 2013, pp. 265 – 276. [Online]. Available: https://doi.org/10.1117/12.2027040

[65] F. Dufaux and T. Ebrahimi, "A framework for the validation of privacy protection solutions in video surveillance," in *2010 IEEE International Conference on Multimedia and Expo*. IEEE, 2010, pp. 66–71.

[66] J. Romberg, "Imaging via Compressive Sampling," *Signal Processing Magazine, IEEE*, vol. 25, no. 2, pp. 14–20, 2008.

[67] D. Donoho, A. Maleki, and M. Shahram, "Wavelab 850," *Software toolkit for time-frequency analysis*, 2006.

[68] M. A. Figueiredo, R. D. Nowak, and S. J. Wright, "Gradient Projection for Sparse Reconstruction: Application to Compressed Sensing and other Inverse Problems," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 1, no. 4, pp. 586–597, 2007.

[69] D. E. King, "DLIB-ML: A Machine Learning Toolkit," *Journal of Machine Learning Research*, vol. 10, no. Jul, pp. 1755–1758, 2009.

[70] E. M. Newton, L. Sweeney, and B. Malin, "Preserving Privacy by De-identifying Face Images," *IEEE transactions on Knowledge and Data Engineering*, vol. 17, no. 2, pp. 232–243, 2005.

[71] Q. Wang, M. Wei, X. Chen, and Z. Miao, "Joint Encryption and Compression of 3D Images Based on Tensor Compressive Sensing with Non-autonomous 3D Chaotic System," *Multimedia Tools and Applications*, vol. 77, no. 2, pp. 1715–1734, 2018.

[72] K. Xie, X. Ning, X. Wang, S. He, Z. Ning, X. Liu, J. Wen, and Z. Qin, "An Efficient Privacy-Preserving Compressive Data Gathering Scheme in WSNs," *Information Sciences*, vol. 390, pp. 82–94, 2017.

[73] M. J. Wainwright, *High-Dimensional Statistics: A Non-Asymptotic Viewpoint*. Cambridge University Press, 2019, vol. 48.

Mehmet Yamaç received his B.S. degree in Electrical and Electronics Engineering from Anadolu University, Eskisehir, Turkey, in 2009 the M.S. degree in Electrical and Electronics Engineering from Bogazici University, Istanbul, Turkey, in 2014. He was research and teaching assistant at Bogazici University during 2012-2017. He is currently a Ph.D. candidate at the Department of Computing Sciences, Tampere University, Tampere, Finland. His research interests are computer and machine vision, machine learning and compressive sensing.

Mete Ahishali received the B.Sc. degree (Hons.) in Electrical and Electronics Engineering from Izmir University of Economics, Izmir, Turkey, in 2017, and the M.Sc. degree (Hons.) in Data Engineering and Machine Learning from Tampere University, Tampere, Finland, in 2019. Since 2017, he has been working as a researcher in Signal Analysis and Machine Intelligence research group under the supervision of Prof. Gabbouj, and he is currently pursuing the Ph.D. degree in Computing and Electrical Engineering at Tampere University. His research interests are pattern recognition, machine learning, and semantic segmentation with applications in computer vision, remote sensing, and biomedical images.

Nikolaos Passalis received the B.Sc. degree in informatics, the M.Sc. degree in information systems, and the Ph.D. degree in informatics from the Aristotle University of Thessaloniki, Thessaloniki, Greece, in 2013, 2015, and 2018, respectively. Since 2019, he has been a post-doctoral researcher with the Aristotle University of Thessaloniki, while from 2018 to 2019 he also conducted post-doctoral research at the Faculty of Information Sciences, Tampere University, Finland. He has (co)authored more than 30 journal articles and 45 conference papers. His research interests include deep learning, information retrieval, time-series analysis and computational intelligence.

Jenni Raitoharju received her Ph.D. degree at Tampere University of Technology, Finland in 2017. Since then, she has worked as a Postdoctoral Research Fellow at the Faculty of Information Technology and Communication Sciences, Tampere University, Finland. In 2019, she started working as a Senior Research Scientist at the Finnish Environment Institute, Jyväskylä, Finland after receiving Academy of Finland Postdoctoral Researcher funding for 2019-2022. She has co-authored 19 journal papers and 29 papers in international conferences. She is the chair of Young Academy Finland 2019-2021. Her research interests include machine learning and pattern recognition methods along with applications in biomonitoring and autonomous systems.

Bülent Sankur is presently at Bogazici University in the Department of Electrical-Electronic Engineering. His research interests are in the areas of digital signal processing, security and biometry, cognition and multimedia systems. He has served as a consultant in several industrial and government projects and has been involved in various European framework and/or bilateral projects. He has held visiting positions at the University of Ottawa, Technical University of Delft, Ecole Nationale Supérieure des Télécommunications, Paris, and Istanbul Technical University. He was the chairman in EUSIPCO'05: The European Conference on Signal Processing, in ICT'96: Inter. Conference on Telecommunications, as well as the technical chairman of ICASSP'00. Dr. Sankur is presently an associate editor in Journal of Image and Video Processing, and Journal of Image ad Vision Computing. He is a member of Academy of Sciences in Turkey.

MONCEF GABBOUJ received his BS degree in 1985 from Oklahoma State University, and his MS and PhD degrees from Purdue University, in 1986 and 1989, respectively, all in electrical engineering. Dr. Gabbouj is a Professor of Signal Processing at the Department of Computing Sciences, Tampere University, Tampere, Finland. He was Academy of Finland Professor during 2011-2015. His research interests include Big Data analytics, multimedia content-based analysis, indexing and retrieval, artificial intelligence, machine learning, pattern recognition, nonlinear signal and image processing and analysis, voice conversion, and video processing and coding. Dr. Gabbouj is a Fellow of the IEEE and member of the Academia Europaea and the Finnish Academy of Science and Letters. He is the past Chairman of the IEEE CAS TC on DSP and committee member of the IEEE Fourier Award for Signal Processing. He served as associate editor and guest editor of many IEEE, and international journals and Distinguished Lecturer for the IEEE CASS. Dr. Gabbouj is the Finland Site Director of the NSF IUCRC funded Center for Visual and Decision Informatics (CVDI) and leads the Artificial Intelligence Research Task Force of the Ministry of Economic Affairs and Employment funded Research Alliance on Autonomous Systems (RAAS).