Analysis of One-Time Random Projections for Privacy Preserving Compressed Sensing

Analysis of One-Time Random Projections for Privacy Preserving Compressed Sensing

Tiziano Bianchi, *Member, IEEE,* Valerio Bioglio, *Member, IEEE,* and Enrico Magli, *Senior Member, IEEE*

***Abstract*—In this paper, the security of the compressed sensing (CS) framework as a form of data confidentiality is analyzed. Two important properties of one-time random linear measurements acquired using a Gaussian i.i.d. matrix are outlined: i) the measurements reveal only the energy of the sensed signal; ii) only the energy of the measurements leaks information about the signal. An important consequence of the above facts is that CS provides information theoretic secrecy in a particular setting. Namely, a simple strategy based on the normalization of Gaussian measurements achieves, at least in theory, perfect secrecy, enabling the use of CS as an additional security layer in privacy preserving applications. In the generic setting in which CS does not provide information theoretic secrecy, two alternative security notions linked to the difficulty of estimating the energy of the signal and distinguishing equal-energy signals are introduced. Useful bounds on the mean square error of any possible estimator and the probability of error of any possible detector are provided and compared to simulations. The results indicate that CS is in general not secure according to cryptographic standards, but may provide a useful built-in data obfuscation layer.**

***Index Terms*—Compressed sensing, confidentiality, encryption, privacy preservation, random matrices, security.**

I. Introduction

Compressed sensing (CS) has emerged in recent years as an efficient framework for acquiring signals, able to surpass the bounds of the classical Shannon-Nyquist theory [1]–[3]. This result can be achieved observing that, if the signal satisfies certain properties of sparsity, linear measurements enable signal recovery with high probability, provided that enough measurements are available with respect to signal sparsity. The measurements have to satisfy certain incoherence properties, which is the case of linear measurements acquired using random matrices [4], [5].

Several applications can benefit from the low complexity acquisition and reduced energy consumption offered by CS, as demonstrated by recent works on spectrum sensing for cognitive radios [6], [7], wireless sensor networks [8]–[10], network anomaly detection [11]. In many recent papers [12]–[15], it has been suggested that the randomness in the acquisition process may also implicitly provide some kind of privacy preservation for similar applications. Hence, it is important to precisely characterize the confidentiality of CS measurements, since in applications where CS is already useful as a signal acquisition strategy, CS may also provide confidentiality at no cost, or a very little additional cost.

Since its early days, the structure of CS hinted the possibility that such a framework may provide some notion of confidentiality. In [16] the authors conclude that, even if CS does not provide information theoretic secrecy [17], it offers computational secrecy if viewed as a cryptosystem. Many subsequent works tried to further investigate the security of CS as a cryptosystem [18]–[21].

In this paper, we analyze the confidentiality of CS measurements in the case of one-time random projections. To begin with, we prove that, differently from [16], under specific distributions of the signal, a CS framework that exploits Gaussian i.i.d. sensing matrices can achieve secrecy in an information theoretic sense. Namely, we demonstrate that in this case the measurements reveal only the energy of the sensed signal, and conversely that only the energy of the measurements leaks information about the signal. As a consequence, a CS framework that uses Gaussian random matrices is, at least in theory, perfectly secure when sensing constant energy signals.

However, our study is not limited to Gaussian sensing matrices. When the measurements are acquired with a generic matrix, we prove that the spherical angle of the measurements reveals only the spherical angle of the signal. In the case of Gaussian sensing matrices, this allows us to propose a simple strategy based on the normalization of the measurements which achieves perfect secrecy irrespective of the distribution of the sensed signal, even though the energy of the original signal must be transmitted through an alternative secure channel if the correct signal has to be recovered.

For the generic cases in which CS does not provide information theoretic secrecy, we propose alternative security notions linked to the difficulty of solving particular signal processing tasks. The first metric is related to the minimum mean square error achievable by practical estimators when estimating the energy of the sensed signals from the measurements. The second metric links the confidentiality of CS to the performance of a detector trying to distinguish equal-energy signals with different spherical angles. For specific distributions of the sensed signal and the sensing matrix we are able to provide closed form bounds characterizing the confidentiality of CS according to both security notions. Simulation results are included to validate such bounds in simple scenarios and to provide further insight about the

Tiziano Bianchi and Enrico Magli are with the Dept. of Electronics and Telecommunications, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129, Torino, Italy. Valerio Bioglio was with the Dept. of Electronics and Telecommunications, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129, Torino, Italy. He is now with the Mathematics and Algorithmic Sciences Lab., Huawei Technologies France, Boulogne-Billancourt, France.

This work was supported by the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013) / ERC Grant agreement n. 279848.

behavior of the aforementioned metrics in the case of more general distributions.

This paper extends a previous work by Bianchi *et al.* [22] both from a theoretical and an experimental point of view. The most significant novelty is the analysis of generic sensing matrices, providing important insights on the confidentiality of non-Gaussian sensing matrices. We also discuss possible attacks to generic sensing matrices in the case of equal-energy signals and we experimentally verify their performance with respect to the security bounds. Finally, we provide detailed proofs for all the propositions in [22]. The results complement those in [22], and show that Gaussian sensing matrices are indeed a very special case, since generic sensing matrices can only provide a weak notion of secrecy.

The rest of the paper is organized as follows. In Section II, some background material about compressed sensing and standard security definitions is provided. The main results of the paper, dealing with the confidentiality of random measurements, are presented in Section III. In Section IV useful bounds on the confidentiality of the measurements are derived, while in Section V some possible attacks are introduced. Experimental results are presented in Section VI, while issues related to signal recovery, quantization, and practical implementation of the proposed system are discussed in Section VII. The main implications of the obtained results are finally discussed in Section VIII.

II. Background

A. Notations

We denote column vectors x by lowercase letters, and matrices A by uppercase letters. We denote as $[A]_{ij}$ the element at the i-th row and j-th column of matrix A. The p-norm, $p \geq 1$, of a vector x is denoted as $||x||_p$; the notation $||x||_0$ indicates the number of nonzero elements of x. The probability density function (pdf) of x is denoted as $\mathbb{P}_X(x)$, or, when the meaning is unambiguous, simply as $\mathbb{P}(x)$, while $E[x]$ and $\mathrm{Var}[x]$ denote the expectation and the variance of x, respectively. The function $\log$ is always taken to the base e.

B. Compressed Sensing

A signal $x \in \mathbb{R}^n$ is called k-sparse if there exists a $n \times n$ basis Φ and a vector $\theta \in \mathbb{R}^n$ such that

$$x = \Phi\theta \tag{1}$$

and θ has at most k nonzero entries, i.e., $||\theta||_0 \leq k$. According to the compressed sensing framework, a k-sparse signal can be exactly recovered from $m < n$ linear measurements acquired using the $m \times n$ sensing matrix A

$$y = Ax \tag{2}$$

by solving the minimization problem

$$\hat{\theta} = \arg\min_{\theta} ||\theta||_0, \quad \text{subject to } A\Phi\theta = y \tag{3}$$

as long as the smallest number of columns of the $m \times n$ matrix $A\Phi$ that are linearly dependent is strictly greater than $2k$, which implies $m \geq 2k$ [1], [2].

In practice, if the entries of A are i.i.d. variables from a sub-Gaussian distribution, then exact recovery of k-sparse signals can be achieved with very high probability by solving the convex minimization problem

$$\hat{\theta} = \arg\min_{\theta} ||\theta||_1, \quad \text{subject to } A\Phi\theta = y \tag{4}$$

as long as $m = O(k\log(n/k))$ [4]. Hence, in the following we will focus on random sensing matrices A with i.i.d. entries.

C. Security definitions

Let us call the set of possible plaintexts $\mathcal{P}$, the set of cipher texts $\mathcal{C}$ and a key K. A private key cryptosystem is a pair of functions $e_K : \mathcal{P} \to \mathcal{C}, d_K : \mathcal{C} \to \mathcal{P}$ such that, given a plain text $p \in \mathcal{P}$, and a ciphertext $c \in \mathcal{C}$, we have that $d_K(e_K(p)) = p$ and that it is unfeasible, without knowing the key K, to determine p such that $e_K(p) = c$.

A cryptosystem is said to be perfectly secure [17] if the posterior probability of the ciphertext given the plaintext p is independent of p, i.e., if

$$\mathbb{P}(c|p) = \mathbb{P}(c). \tag{5}$$

Given a perfectly secure cryptosystem, an attack can not be more successful than guessing the plaintext at random. It can be proved that a cryptosystem can be perfectly secure only if the size of the plaintext p is smaller than or equal to the size of the key K and the key is used only once; this can be achieved only by the one-time pad scheme. However, practical cryptosystems are usually only computationally secure; this means that breaking the cryptosystem is equivalent to solve an NP-hard problem, i.e., a problem whose solution can not be computed in polynomial time with respect to the size of the key.

It is possible to find the following equivalences between a CS scheme described in (2) and a private key cryptosystem: the signal x is the plain text p, the sensing matrix A is the secret key K and the measurement vector y is the cipher text c. The encryption function e_A is the matrix multiplication between the sensing matrix A and the signal x; the decryption is achieved by an algorithm able to solve the problem in (4). The notions of perfect security and computational security can be naturally extended to CS measurements.

Apart from the above security definitions, implying perfect message confidentiality, in multimedia encryption it is sometimes required that an attacker is not able to recover a copy of the plaintext with a sufficiently high quality. This leads to a different security notion with respect to the standard cryptographic definitions, which is often referred to as perceptual/transparent encryption [23]–[25]. Since this is usually an application-dependent notion, in this case there are no formal and universally agreed security definitions.

D. Security Scenarios and Attack Models

We assume a scenario in which a device acquire some signals using CS and transmit the measurements on a publicly accessible channel. The sensing matrix is communicated only to the intended receivers through a secure channel, while

possible adversaries have no knowledge of the sensing matrix. We also assume that intended receivers and adversaries do not collude.

In the case of sensing of multiple signals, the confidentiality of CS measurements is affected by the policies of generation and managing of the sensing matrix. In this paper, we will focus on the one-time sensing matrix (OTS) scenario. We will assume that each sensing matrix is used only once, and that different sensing matrices are statistically independent. Under this scenario, it is sufficient to consider the confidentiality of a single CS framework $y = Ax$, since measurements of multiple signals will be statistically independent. It is also sufficient to consider the case where the adversary has only knowledge of the measurements y, that corresponds to a ciphertext-only attack (COA) scenario.

The OTS scenario seems the most promising one for providing an effective confidentiality layer. When the sensing matrix is used multiple times, blind source separation techniques can be applied to identify the sensing matrix [26]. Moreover, other attack models are possible, *e.g.*, the known-plaintext attack (KPA), where the attacker knows both the message x and the corresponding encryption y. Obviously, CS is not secure under KPA when the same sensing matrix is used multiple times [16], since the knowledge of n linearly independent messages would be sufficient to solve the system of linear equations. An analysis of KPA under the OTS scenario, restricted to the case of Bernoulli sensing matrices, can be found in [21].

E. Related Works

The security of CS as a symmetric cryptosystem was first analyzed in [16], where the authors show that CS does not provide information theoretic secrecy but argue that it is computationally secure, as long as the sensing matrix is used only once. Similarly, in [18] the authors prove the computational security of CS against a systematic search of the sensing matrix, even in the case of known signal sparsity. More recently, a practical CS system with two confidentiality levels based on Bernoulli sensing matrices has been proposed and analyzed in [19], [20] and its security against KPAs has been investigated in [21].

Some results suggest that CS can also provide some notion of security in the wiretap channel model [27], [28]. In this model, an eavesdropper has access to a secret communication through a different channel with respect to the intended receiver. In [27], the authors show that, if the channels use different sensing matrices and the eavesdropper obtains less measurements than the intended receiver, the secrecy capacity of the intended receiver is nearly equal to the channel capacity. In [28], the authors construct the sensing matrix in such a way that the combination of the eavesdropper's channel and the sensing matrix does not guarantee signal recovery. In this way, the channel knowledge can be exploited in order to offer security against an eavesdropper through the construction of a special sensing matrix.

Finally, some authors suggest that CS can implicitly provide a privacy preserving layer [29]. In [30], random projections were proposed as a tool for enabling privacy preserving data mining, even if reconstruction of the original data from projections was not addressed. In [31], the authors show that a similar setting can be used for sparse regression and analyze its privacy properties. In [32], the authors propose a mechanism to achieve differential privacy by adding noise to CS measurements.

A number of application relying on privacy properties of random projections have been recently proposed. In [13], the authors propose to combine random projections and multi-party computation for outsourcing watermarked image data to the cloud. When a watermark, compressed with the same sensing matrix, is presented to the cloud, this enables privacy-preserving watermark detection. A similar framework is proposed in [14], where outsourced random projections can be used for privacy-preserving data mining and other signal processing tasks. Alternatevely, random projections can be used for generating a robust image hash [15].

With respect to [16], [18], our paper analyzes the statistical properties of random linear projections, so that its results hold also in the presence of computationally unbounded adversaries. In this sense, the more closely related papers are [19], [20]. However, results in our paper are based on novel confidentiality metrics based on MSE and signal distinguishability, which permit a precise characterization of the confidentiality of the measurements under different system parameters. Since we assume a perfect channel for transmitting the measurements, our results are not directly related to the wiretap channel model. Similarly, our results do not apply to most of the papers about privacy-preserving data mining applications [14], [15], [30], [32], which are not based on the OTS strategy. However, applications like that described in [13] could benefit from the proposed framework.

III. Confidentiality of the Measurements

In this section, we summarize the main results of the paper. The first two results regard Gaussian sensing matrices, which, as we will see, play a special role in the confidentiality of CS measurements. Namely, we prove that a Gaussian sensing matrix discloses the energy of the sensed signal, and that only the energy of the measurements carries information about the signal. The third result holds for generic sensing matrices and states that the spherical angle of the measurements provides information only on the spherical angle of the signal. An important consequence of the above results is that normalized measurements obtained with a Gaussian sensing matrix provide a perfectly secret channel. For the sake of conciseness, most of the proofs are given in the Appendix.

Let us consider OTS measurements defined by $y = Ax$. Let us assume that x is a random vector with an arbitary probability distribution, denote with $I(x, y)$ the mutual information between x and y [33], and define $\mathcal{E}_x = ||x||_2^2$. We have the following important result:

Proposition 1. If $[A]_{i,j}$ are i.i.d. zero-mean Gaussian variables, then OTS measurements satisfy $I(x; y) = I(\mathcal{E}_x; y)$.

Proof. See the Appendix. □

The above result says that OTS measurements obtained using an i.i.d. Gaussian sensing matrix do not reveal anything more about x than its energy and what can be inferred by knowing its energy. It is worth noting that this is true irrespective of the sparsity degree of x, that is, x does not necessarily have to be sparse. In the following, we will denote such measurements as Gaussian-OTS (G-OTS) measurements. An immediate consequence of the above proposition is that G-OTS measurements do not reveal anything about a family of signals with a constant energy.

Corollary 1. If $x \in \mathcal{S}_\beta$, where $\mathcal{S}_\beta = \{x|\mathcal{E}_x = \beta > 0\}$, then G-OTS measurements are perfectly secure.

Proof. If $\mathcal{E}_x$ is an a priori known constant, then $\mathbb{P}(y|\mathcal{E}_x) = \mathbb{P}(y)$. Hence, from Prop. 1 we have $\mathbb{P}(y|x) = \mathbb{P}(y)$. □

Since for most signals of interest the constant energy condition is usually not verified, it is interesting to evaluate the information leakage of the signal x due to the observation of y. Calling $\mathcal{E}_y = ||y||_2^2$, for the case of i.i.d. Gaussian sensing matrices, we have the following result:

Proposition 2. If $[A]_{i,j}$ are i.i.d. zero-mean Gaussian variables, then we have that $I(\mathcal{E}_x; y) = I(\mathcal{E}_x; \mathcal{E}_y)$.

Proof. See the Appendix. □

The above result says that $\mathcal{E}_y$ is a sufficient statistic for estimating $\mathcal{E}_x$, irrespective of the distribution of x. This result holds only for Gaussian sensing matrices. Nevertheless, a similar weaker property holds for generic sensing matrices. Let us define the spherical angle of x and y as $u_x = x/\sqrt{\mathcal{E}_x}$ and $u_y = y/\sqrt{\mathcal{E}_y}$, respectively. We have the following result:

Proposition 3. If $\mathcal{E}_y > 0$, then generic OTS measurements satisfies $I(x; u_y) = I(u_x; u_y)$.

Proof. See the Appendix. □

The above result says that the spherical angle of generic OTS measurements y does not reveal anything more about x than its spherical angle u_x and what can be inferred by knowing u_x.

Finally, a very interesting consequence of the above results is that they can be exploited to obtain a perfectly "secured" version of G-OTS measurements. Let us assume that normalized G-OTS measurements are transmitted according to the following strategy

$$u_y = \begin{cases} y/\sqrt{\mathcal{E}_y} & \mathcal{E}_y > 0 \\ U & \mathcal{E}_y = 0 \end{cases} \tag{6}$$

where U is a random vector uniformly distributed on a unit radius m-sphere. We denote it as SG-OTS.

Lemma 1. SG-OTS measurements are perfectly secure, i.e., $\mathbb{P}(u_y|x) = \mathbb{P}(u_y)$.

Proof. According to the proof of proposition 2, u_y is uniformly distributed on the unit radius m-sphere irrespective of x. □

IV. Security Metrics

The results of the previous section indicate that measurements acquired using linear random projections may leak some information about the sensed signal. In this Section, we will introduce some metrics linking this information leakage to the performance attainable by attacks based on simple signal processing tasks. More specifically, we will analyze the precision with which the energy of the sensed signal can be estimated and the distinguishability of equal-energy signals having different spherical angles. Practical upper bounds on the above metrics will also be derived for specific signal distributions.

For what concerns the energy of the sensed signal, we will show that the mean squared error (MSE) of the estimated energy can be bounded according to the mutual information between the measurements and the energy of the signal. Although such mutual information can be difficult to characterize in the case of generic signals and generic sensing matrices, we will provide useful closed-form bounds in the case of Gaussian sensing matrices and exactly sparse signals with Gaussian componenents. As to the distinguishability of equal-energy signals, we will first characterize the distinguishability of a fixed pair of signals in the case of sensing matrices having generic distributions. Then, we will analyze the expected behavior of the distinguishability for signals uniformly distributed on a sphere and exactly sparse signals.

A. Energy Estimation

The standard cryptographic definition of security fails to capture the fact that an attacker may try to estimate $\mathcal{E}_x$ up to a certain precision instead of recovering it exactly. Hence, we introduce an alternative notion of confidentiality linked to the MSE between $\mathcal{E}_x$ and an estimate $\hat{\mathcal{E}}_x$ obtained by observing only the measurements y. In the following, we will say that the measurements are η-MSE secure with respect to $\mathcal{E}_x$, if, for every possible estimator $\hat{\mathcal{E}}_x(y)$ of $\mathcal{E}_x$, we have that

$$\eta_{\hat{\mathcal{E}}_x} \triangleq \frac{E[||\mathcal{E}_x - \hat{\mathcal{E}}_x(y)||_2^2]}{\sigma^2_{\mathcal{E}_x}} \geq \eta \tag{7}$$

where $\sigma^2_{\mathcal{E}_x}$ is the variance of $\mathcal{E}_x$. Note that a Bayesian estimator is always at least 1-MSE secure, since, in the absence of any a posteriori information, the minimum MSE (MMSE) estimator of $\mathcal{E}_x$ is $\hat{\mathcal{E}}_x(y) = E[\mathcal{E}_x]$, yielding $E[||\mathcal{E}_x - \hat{\mathcal{E}}_x(y)||_2^2] = \sigma^2_{\mathcal{E}_x}$.

Let us consider an estimator $\hat{\mathcal{E}}_x(y)$ of $\mathcal{E}_x$ which relies on the measurement y. By using rate-distortion theory, we can link the mutual information between y and $\mathcal{E}_x$ to the MSE of the estimator through the following lower bound [33, Th. 8.6.6.]:

$$E[||\mathcal{E}_x - \hat{\mathcal{E}}_x(y)||_2^2] \geq \frac{1}{2\pi} e^{2h(\mathcal{E}_x|y)-1} = \frac{1}{2\pi} e^{2h(\mathcal{E}_x)-2I(\mathcal{E}_x;y)-1}. \tag{8}$$

This leads immediately to the following result:

Lemma 2. Generic measurements are at least η-MSE secure with respect to $\mathcal{E}_x$, where

$$\eta = \frac{e^{2h(\mathcal{E}_x|y)-1}}{2\pi\sigma^2_{\mathcal{E}_x}}. \tag{9}$$

The practical evaluation of the MSE security of CS measurements requires to specify a distribution for A and x. Though a characterization for generic distributions is difficult, in the case of Gaussian sensing matrices we have the following useful lemma:

Lemma 3. For G-OTS measurements $I(\mathcal{E}_x;\mathcal{E}_y)$ can be upper bounded as

$$I(\mathcal{E}_x;\mathcal{E}_y) \leq \xi(\kappa^*) - \xi\left(\frac{m}{2}\right) - \psi(\kappa^*) + \psi\left(\frac{m}{2}\right) \qquad (10)$$

where $\xi(\kappa) \triangleq \kappa + \log(\Gamma(\kappa)) + (1-\kappa)\psi(\kappa)$, $\Gamma(\kappa) = \int_0^\infty t^{\kappa-1}e^{-t}dt$ is the Gamma function, $\psi(\kappa) = \frac{d\log(\Gamma(\kappa))}{d\kappa}$ is the digamma function, and κ^* is the solution to the nonlinear equation $\log(\kappa^*) - \psi(\kappa^*) = \log(m/2) - \psi(m/2) + \log(E[\mathcal{E}_x]) - E[\log(\mathcal{E}_x)]$.

Proof. See the Appendix. □

It is worth noting that the upper bound in (10) does not depend on the variance of $[A]_{i,j}$.

The above lemma can be used to provide a closed form expression of η in the case of specific signal distributions. If x can be modeled as an exactly k-sparse signal, whose nonzero entries are i.i.d. Gaussian variables with zero mean and variance σ_x^2, then $\mathcal{E}_x$ is distributed as a chi-square variable with k degrees of freedom scaled by σ_x^2. The above fact, together with (8), leads immediately to the following

Corollary 2. If x is an exactly k-sparse signal with i.i.d. Gaussian nonzero entries, G-OTS measurements are at least η-MSE secure with respect to $\mathcal{E}_x$, where

$$\eta = \frac{e^{2\xi\left(\frac{k}{2}\right)+2\xi'\left(\frac{m}{2}\right)-2\xi'(\kappa^*)-1}}{k\pi} \qquad (11)$$

$\xi'(\kappa) = \xi(\kappa) - \psi(\kappa)$ and κ^* satisfies $\log(\kappa^*) - \psi(\kappa^*) = \log(m/2) - \psi(m/2) + \log(k/2) - \psi(k/2)$.

Proof. See the Appendix. □

B. Signal Distinguishability

Most of the theorems in Section III are valid only in the case of Gaussian i.i.d. sensing matrices. However, in practical CS framework, sub-Gaussian i.i.d. sensing matrices are often exploited due to their recovery guarantees when solving (4) and their simpler implementation. As an example, Bernoulli i.i.d. matrices, i.e., such that $[A]_{i,j} = \pm 1$ with equal probability, belong to this class and permit to avoid multiplications in the sensing process.

For what concerns the confidentiality of random measurements, the main drawback of using a generic non-Gaussian sensing matrix is that the measurements are not Gaussian distributed; hence, in general, it is not true that $\mathbb{P}(y|x) = \mathbb{P}(y|\mathcal{E}_x)$. Nevertheless, when n tends to infinity, according to the central limit theorem (CLT) $\mathbb{P}(y|x)$ tends to a multivariate Gaussian distribution with zero mean and covariance matrix equal to $\sigma_A^2\mathcal{E}_x I_m$. Hence, we can expect that the information leakage due to the non Gaussianity of A will decrease as n grows. The same observation was made in [19], [20], where the authors argued that a CS system using a Bernoulli matrix can achieve a sort of asymptotic secrecy.

1) Security Definition: In order to characterize this information leakage, we introduce a security notion based on the problem of distinguishing whether the measurements y comes from one of two known given signals x_1 and x_2. This security definition is inspired by indistinguishability definitions commonly used in cryptography [34]. When the two signals have the same energy, this problem is equivalent to distinguishing the two different distributions of y that are implied by either x_1 or x_2 and the fact that A is non-Gaussian.

Let us consider a signal x that belongs to a two-element set $\{x_1, x_2\}$; a detector is a function that given the measurements y outputs one of two possible signals x_1, x_2. Formally, this can be defined as $\mathcal{D} : \mathbb{R}^m \to \{x_1, x_2\}$. Given a certain detector, we define the probability of detection with respect to signal x_i as $P_{d,i} = \Pr\{\mathcal{D}(y) = x_i | x = x_i\}$ and the respective probability of false alarm as $P_{f,i} = \Pr\{\mathcal{D}(y) = x_i | x \neq x_i\}$. It is immediate to verify $P_{d,2} = 1 - P_{f,1}$ and $P_{f,2} = 1 - P_{d,1}$, so that $P_{d,1} - P_{f,1} = P_{d,2} - P_{f,2} \triangleq P_d - P_f$. In the following, we will say that CS measurements are ϑ-indistinguishable with respect to two signals x_1 and x_2 if for every possible detector $\mathcal{D}(y)$ we have

$$P_d - P_f \leq \vartheta. \qquad (12)$$

According to the above definition, lower values of ϑ correspond to higher security, with $\vartheta = 0$ being equivalent to perfect secrecy, since in the absence of any information random guessing achieves $P_d = P_f$.

Given OTS measurements acquired using a sensing matrix A with a certain distribution, we can link their ϑ-indistinguishability to $\mathbb{P}(y|x_1)$ and $\mathbb{P}(y|x_2)$. Let us denote the total variation (TV) distance between the random variables a and b as $\delta(a,b) = \delta(\mathbb{P}_A(a), \mathbb{P}_B(b)) = \frac{1}{2}\int |\mathbb{P}_A(t) - \mathbb{P}_B(t)|dt$. Let us also denote in short $\delta(\mathbb{P}(y|x_1), \mathbb{P}(y|x_2)) = \delta(Ax_1, Ax_2)$. We have the following:

Lemma 4. OTS measurements are at least $\delta(Ax_1, Ax_2)$-indistinguishable with respect to two signals x_1 and x_2.

Proof. The sum of error probabilities in a statistical hypothesis test can be lower bounded as [35]

$$\begin{aligned}\Pr\{\mathcal{D}(y) = x_2|x_1\} + \Pr\{\mathcal{D}(y) = x_1|x_2\} &= 1 - P_d + P_f \\ &\geq 1 - \delta(\mathbb{P}(y|x_1), \mathbb{P}(y|x_2))\end{aligned} \qquad (13)$$

from which it is immediate to derive $P_d - P_f \leq \delta(Ax_1, Ax_2)$. □

2) Confidentiality of Generic Sensing Matrices: The above result can be used to characterize the confidentiality of OTS measurements in the case of a generic sensing matrix. Let us assume that $[A]_{i,j}$ are independent and identically distributed with zero mean. Moreover, let us assume that the above matrix is used to sense vectors with constant energy. In the following, without loss of generality we will assume $\mathcal{E}_x = 1$. Given any two different signals x_1 and x_2, with $\mathcal{E}_{x_1} = \mathcal{E}_{x_2} = 1$, we have the following result:

Proposition 4. If the random variables $a = [A]_{i,j}$ satisfy a Poincaré inequality with constant $c > 0$, i.e. for every smooth function s with derivative s', one has $\mathrm{Var}[s(a)] \leq$

$c^{-1}E[(s'(a))^2]$, then the TV distance between Ax_1 and Ax_2 is upper bounded as

$$\begin{aligned}\delta(Ax_1, Ax_2) \leq & \vartheta_A(x_1, x_2) \\ \triangleq & \sqrt{\frac{mD(a||G)||x_1||_4^4}{c+(2-c)||x_1||_4^4}} + \sqrt{\frac{mD(a||G)||x_2||_4^4}{c+(2-c)||x_2||_4^4}}\end{aligned} \tag{14}$$

where $D(a||G) = h(G) - h(a)$ is the Kullback-Leibler (KL) divergence of a Gaussian variable G with zero mean and variance σ_A^2 from a.

Proof. See the Appendix. □

The above result is based on Theorem 1 in [36], which states that the KL divergence between the distribution of the linear combination of i.i.d. variables satisfying the Poincaré inequality and a Gaussian distribution converges to zero, where the rate of convergence depends on the sum of the fourth powers of the normalized weights. In [37], it was proved that the Poincaré inequality in the above Proposition holds for every log-concave probability density function, i.e., if $\mathbb{P}(a) \sim e^{-f(a)}$, where $f(a)$ is a convex function, which include exponential, Gaussian, and sub-Gaussian distributions.

The above result can be used to quantify the distinguishability of the measurements of any two given signals. First of all, let us examine two limit cases. The lowest value of the bound occurs when $||x_1||_4^4 = ||x_2||_4^4 = 1/n$, which corresponds to $x = [\pm 1/\sqrt{n}, \ldots, \pm 1/\sqrt{n}]$. In this case, the bound becomes

$$\delta(Ax_1, Ax_2) \leq \sqrt{\frac{4mD(a||G)}{c(n-1)+2}}. \tag{15}$$

The highest value of the bound occurs when $||x_1||_4^4 = ||x_2||_4^4 = 1$, which happens when x has only one non-zero entry equal to 1, and is given by $\delta(Ax_1, Ax_2) \leq \sqrt{2mD(a||G)}$. It is evident that in the best case the TV distance goes to zero at least as $O(n^{-1/2})$, which is consistent with Berry-Esseen type estimates [38], as recognized in [36], [39]. Moreover, some recent results on the entropic central limit theorem suggest that the convergence in the best case can be $O(n^{-1})$, as long as the sensing matrix distribution satisfies some additional constraints[1] [40]. Berry-Esseen theorem is also used in [20] to argue asymptotic spherical secrecy of CS measurements. However, Berry-Esseen theorem is about the convergence of the cumulative distribution function and can not be directly used to provide a bound on TV distance here. Conversely, in the worst case the TV distance is independent of n, since obviously each entry of y has the same distribution as the entries of A and does not converge to a Gaussian distribution.

It is worth noting that the upper bound in (14), even if asymptotically correct, is surely pessimistic. As a matter of fact, if x_1 and x_2 are equal, or, more generally, they have the same entries but in a permuted order, which implies $||x_1||_4^4 = ||x_2||_4^4$, it follows that y_1 and y_2 have exactly the same distribution, so that $\delta(y_1, y_2) = 0$. A reasonable conjecture is that tighter upper bounds may exists.

Interestingly, the above upper bounds can be used even in the case of normalized measurements, when x_1 and x_2 are not necessarily unit norm vectors. Let us define $u_{x_i} = x_i/\sqrt{\mathcal{E}_{x_i}}$ and $u_{y_i} = y_i/\sqrt{\mathcal{E}_{y_i}}$, where $y_i = Ax_i$, $i = 1, 2$. Then we have the following

Corollary 3. Under the hypotheses of Prop. 4

$$\delta(u_{y_1}, u_{y_2}) \leq \vartheta_A(u_{x_1}, u_{x_2}). \tag{16}$$

Proof. See the Appendix. □

The above result implies that if OTS measurements are ϑ-indistinguishable with respect to equal-energy signals, then the normalized version of the same measurements is at least ϑ-indistinguishable with respect to generic signals.

Finally, it is worth noting that G-OTS measurements are always 0-indistinguishable with respect to equal-energy signals (equivalently, SG-OTS measurements are always 0-indistinguishable with respect to generic signals), since for a Gaussian sensing matrix $D(a||G) = 0$, which is consistent with the fact that G-OTS measurements achieve perfect secrecy when sensing equal-energy signals.

3) Expected Confidentiality for Constant Energy Signals: A remark on the previous results is that the worst case distinguishability, for non-Gaussian sensing matrices, can be very high, meaning that in an adaptive chosen plaintext attack the adversary can always find a set of signals that are easy to distinguish. In this sense, the definition of signal distinguishability is different from the cryptographic notion of message distinguishability. However, the interesting property is that the set of bad signal is usually very small, so that the probability of finding two signals that can be easily distinguished is very low. The following results, valid when x is uniformly distributed on a unit n-sphere, tells us that with high probability the upper bound is usually close to the best case. First, we will introduce the following useful lemma:

Lemma 5. If x is uniformly distributed on a unit n-sphere, then for $\epsilon_1 > 0$ and $0 < \epsilon_2 < 1$

$$\Pr\left\{||x||_4^4 \geq \frac{3+\epsilon_1}{(1-\epsilon_2)n}\right\} \leq \frac{96}{n\epsilon_1^2} + e^{-\frac{n\epsilon_2^2}{16}} \tag{17}$$

and

$$\Pr\left\{||x||_4^4 \geq \frac{\epsilon_1}{(1-\epsilon_2)n}\right\} \leq \frac{5}{2}e^{-\frac{\sqrt{\epsilon_1}}{4}} + e^{-\frac{n\epsilon_2^2}{16}}. \tag{18}$$

Proof. See the Appendix. □

We are now ready to state the main result:

Proposition 5. If x_1, x_2 are uniformly distributed on a unit n-sphere, then with probability exceeding $1 - \frac{192}{n\epsilon_1^2} - 2e^{-\frac{n\epsilon_2^2}{16}}$ we have

$$\delta(Ax_1, Ax_2) \leq \sqrt{\frac{4m(3+\epsilon_1)D(a||G)}{c(1-\epsilon_2)n + (2-c)(3+\epsilon_1)}}. \tag{19}$$

[1] In [40], the authors show that if $E[a^3] = 0$, $E[a^6] < +\infty$, and $||x||_4^4 = 1/n$, then we have $D(Ax, G) = O(n^{-2})$. The above result, when used in the proof of Proposition 4, yields a $O(n^{-1})$ convergence rate of the TV distance.

Moreover, with probability exceeding $1-5e^{-\frac{\sqrt[4]{n}}{4}}-2e^{-\frac{n\epsilon^2}{16}}$ we have

$$\delta(Ax_1,Ax_2)\leq\sqrt{\frac{4mD(a||G)}{c(1-\epsilon)\sqrt{n}+2-c}}. \tag{20}$$

Proof. By using the union bound, from (17) we have that $\Pr\left\{\max(||x_1||_4^4,||x_2||_4^4)\leq\frac{3+\epsilon_1}{(1-\epsilon_2)n}\right\}\geq 1-\frac{192}{n\epsilon_1^2}-2e^{-\frac{n\epsilon_2^2}{16}}$, which together with (14) immediately proves (19). Moreover, by setting $\epsilon_1=\sqrt{n}$ in (18) we also have $\Pr\left\{\max(||x_1||_4^4,||x_2||_4^4)\leq\frac{1}{(1-\epsilon_2)\sqrt{n}}\right\}\geq 1-5e^{-\frac{\sqrt[4]{n}}{4}}-2e^{-\frac{n\epsilon_2^2}{16}}$, which together with (14) immediately proves (20). □

An important consequence of the above bounds on TV distance is that we can characterize the distinguishability of OTS measurements for messages uniformly distributed on the unit n-sphere:

Corollary 4. If the hypotheses of the above theorems hold, then with probability exceeding $1-O(n^{-1})$ we have that OTS measurements are at least $O\left(\sqrt{\frac{m}{n}}\right)$-indistinguishable, whereas with probability exceeding $1-O(e^{-\sqrt[4]{n}})$ we have that OTS measurements are at least $O\left(\sqrt{\frac{m}{\sqrt{n}}}\right)$-indistinguishable.

4) Expected Confidentiality for k-sparse Signals: The above result says that for most of the signals on the unit n-sphere the distinguishability per measurement goes to zero as n grows. In order to extend this result to k-sparse signals, let us consider the class of signals expressed as $x=\Phi\theta$, where θ has k nonzero components uniformly distributed on the unit k-sphere and Φ is an orthonormal basis. Obviously, such signals lies on unit n-sphere. It is quite easy to verify that for $\Phi=I_n$ the problem is equivalent to having a dense signal uniformly distributed on a unit k-sphere. Hence, in this case we have $\delta(Ax_1,Ax_2)=O(\sqrt{\frac{m}{k}})=O(1)$, i.e., the TV distance does not vanish even if $\frac{m}{n}\to 0$. On the other hand, for a different Φ it is reasonable to think that the bounds about dense signals on the unit n-sphere are still valid. If the columns ϕ_i of Φ satisfy $||\phi_i||_4^4\leq\frac{C}{n}$ where C is some constant independent of n, by using Cauchy-Schwarz inequality it is easy to prove that $||x||_4^4\leq\frac{Ck^2}{n}$. Hence, if $\frac{k^2}{n}\to 0$ we have that the TV distance vanish for this class of orthonormal basis. It is worth noting that the Hadamard basis belong to this class with $C=1$.

V. Attacks to CS Measurements

The bounds introduced in the previous section are very general, in the sense that they hold for any possible attack under the COA scenario. However, it is interesting to assess how practical attacks to CS measurements, trying to extract information about x without knowing A, will perform with respect to those upper bounds. In this section, we consider two possible kinds of attacks to CS measurements. The first attack aims at estimating the energy of the signal from the energy of the measurements. The second attack aims at distinguishing two different equal-energy signals by exploiting the fact that a generic sensing matrix may produce different probability distributions of the measurements for different signals. Since the first kind of attack is essentially an estimation problem, while the second kind of attack is a detection problem, the two classes of attacks will be referred to as *estimation attacks* and *detection attacks*, respectively.

A. Estimation Attacks

Corollary 2 gives a lower bound for the MSE of any possible estimator of $\mathcal{E}_x$ in the case of exactly k-sparse signal with i.i.d. Gaussian nonzero entries. Such a bound can be compared with the performance of practical estimators. An interesting question is whether such a bound is actually tight with respect to practical estimators. In the absence of any prior knowledge on $\mathcal{E}_x$, classical estimation theory states that the variance of any unbiased estimator is always greater than the Cramér-Rao lower bound (CRLB) [41], i.e.,

$$\sigma^2_{\hat{\mathcal{E}}_x}\geq-\frac{1}{E\left[\frac{\partial^2 L(y;\mathcal{E}_x)}{\partial\mathcal{E}_x^2}\right]} \tag{21}$$

where $L(y;\mathcal{E}_x)=\log(\mathbb{P}(y|\mathcal{E}_x))$ denotes the log-likelihood function. In the case of Gaussian sensing matrices this can be computed as

$$\sigma^2_{\hat{\mathcal{E}}_x}\geq\frac{2\mathcal{E}_x^2}{m}. \tag{22}$$

The maximum likelihood (ML) estimator of $\mathcal{E}_x$ in the case of a Gaussian sensing matrix is given by

$$\hat{\mathcal{E}}_{x,ML}=\max_{\mathcal{E}_x}\log(\mathbb{P}(y|\mathcal{E}_x))=\frac{\mathcal{E}_y}{m\sigma_A^2}. \tag{23}$$

For the considered model, it is easy to verify that the ML estimator is unbiased and achieves the CRLB, since $E\left[\hat{\mathcal{E}}_{x,ML}\right]=\mathcal{E}_x$ and $\sigma^2_{\hat{\mathcal{E}}_{x,ML}}=\frac{2\mathcal{E}_x^2}{m}$. The performance of the ML estimator depends on the value of the unknown parameter $\mathcal{E}_x$. In order to obtain the MSE of the ML estimator under a specific distribution of $\mathcal{E}_x$, we can observe that, for an unbiased estimator, $E_{\mathcal{E}_x,y}[(\mathcal{E}_x-\hat{\mathcal{E}}_x)^2]=E_{\mathcal{E}_x}[\sigma^2_{\hat{\mathcal{E}}_{x,ML}}]$. In the case of a Gaussian k-sparse source, this results in

$$E[(\mathcal{E}_x-\hat{\mathcal{E}}_{x,ML})^2]=\frac{2k(k+2)\sigma_x^4}{m}, \tag{24}$$

from which

$$\eta_{\hat{\mathcal{E}}_x,ML}=\frac{k+2}{m}. \tag{25}$$

Bayesian estimators can be obtained by assuming a prior distribution for $\mathcal{E}_x$. It is well known that in this case the MSE is minimized by the posterior mean of $\mathcal{E}_x$, *i.e.*, $\hat{\mathcal{E}}_{x,MMSE}=E_{\mathcal{E}_x|y}[\mathcal{E}_x]$ [41]. For a Gaussian k-sparse signal, a closed form of the MMSE estimator can be derived as

$$\hat{\mathcal{E}}_{x,MMSE}=\frac{\sigma_x\sqrt{\mathcal{E}_y}}{\sigma_A}\frac{K_{\frac{k}{2}-\frac{m}{2}+1}\left(\frac{\sqrt{\mathcal{E}_y}}{\sigma_A\sigma_x}\right)}{K_{\frac{k}{2}-\frac{m}{2}}\left(\frac{\sqrt{\mathcal{E}_y}}{\sigma_A\sigma_x}\right)} \tag{26}$$

where $K_\nu(x)$ denotes the modified Bessel function of the second kind of order ν (see the Appendix). Unfortunately, there is no closed form for the MSE of the above estimator. A simpler estimator can be obtained by searching for the

estimator minimizing the MSE among all estimators which can be expressed as a linear function of $\mathcal{E}_y$. The general expression of the linear MMSE (LMMSE) is $\hat{\mathcal{E}}_{x,LMMSE} = C_{\mathcal{E}_x\mathcal{E}_y} C_{\mathcal{E}_y}^{-1} (\mathcal{E}_y - E[\mathcal{E}_y]) + E[\mathcal{E}_x]$, where $C_{\mathcal{E}_y} = \sigma^2_{\mathcal{E}_y}$ and $C_{\mathcal{E}_x\mathcal{E}_y} = E[\mathcal{E}_x\mathcal{E}_y] - E[\mathcal{E}_x]E[\mathcal{E}_y]$ [41]. For a Gaussian k-sparse signal, the LMMSE estimator can be easily derived as

$$\hat{\mathcal{E}}_{x,LMMSE} = \frac{\mathcal{E}_y}{\sigma_A^2(m+k+2)} + \frac{k(k+2)\sigma_x^2}{m+k+2}. \quad (27)$$

The MSE can be evaluated as

$$E[(\mathcal{E}_x - \hat{\mathcal{E}}_{x,LMMSE})^2] = \frac{2k(k+2)\sigma_x^4}{m+k+2} \quad (28)$$

from which we obtain

$$\eta_{\hat{\mathcal{E}}_x,LMMSE} = \frac{k+2}{m+k+2}. \quad (29)$$

The performance of the above estimators will be compared to the theoretical bound (11) in Section VI.

B. Detection Attacks

We consider an hypothetical scenario in which OTS measurements are used to sense two distinct signals x_1 and x_2 having equal energy. Without loss of generality, we can assume that $\mathcal{E}_{x_1} = \mathcal{E}_{x_2} = 1$. The aim of the attacker is to guess whether the measurements conceal the signal x_1 or the signal x_2. This is a classical detection problem, where the aim is to distinguish whether the measurements y come from the probability distribution $\mathbb{P}(y|x_1)$ or from the probability distribution $\mathbb{P}(y|x_2)$.

Let us consider a detector $\mathcal{D}$. The Neyman-Pearson (NP) lemma states that for $P_f = \alpha$, the probability of detection is maximized by letting $\mathcal{D}(y) = x_1$ whenever

$$\Lambda(y) = \frac{\mathbb{P}(y|x_1)}{\mathbb{P}(y|x_2)} \geq \theta \quad (30)$$

where θ satisfies $\Pr\{\Lambda(y) \geq \theta | x_2\} = \alpha$.

When the sensing matrix is made up of i.i.d. elements, it turns out that the elements of y are i.i.d. as well. This permits to rewrite the optimal NP test as

$$\Lambda'(y) = \sum_{i=1}^{m} (\log(\mathbb{P}([y]_i|x_1)) - \log(\mathbb{P}([y]_i|x_2))) \geq \theta'. \quad (31)$$

Moreover, since each element of y is given by the sum of independent variables, this gives us a convenient way for evaluating the test function $\Lambda'(y)$. Let us consider the characteristic function of the random variable $a = [A]_{ij}$, defined as $\phi_a(t) = E[e^{jta}]$. It is well known that the pdf of a random variable a can be obtained as $\mathbb{P}(a) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \phi_a(t) e^{-jta} dt$, *i.e.*, that the characteristic function and the corresponding pdf form a Fourier transform pair. We have that the characteristic function of $[y]_i$ given a generic signal x can be computed as

$$\phi_{[y]_i|x}(t) = \prod_{j=1}^{n} \phi_a([x]_j t) \quad (32)$$

where $\phi_a(t)$ is the characteristic function of a generic element of the sensing matrix A. Hence, given x_1 and x_2, we can use (32) to evaluate the characteristic functions $\phi_{[y]_i|x_1}$ and $\phi_{[y]_i|x_2}$, find the corresponding $\mathbb{P}([y]_i|x_1)$ and $\mathbb{P}([y]_i|x_2)$ through a Fourier transform, and use them in (30) in order to compute the optimal NP test.

In general, it is difficult to find a closed form expression for the error probability of a detector implemented according to the above procedure. Under the assumption that the elements of y are i.i.d., it is possible to find an upper bound for $P_d - P_f$ by numerically evaluating the KL divergence between $\mathbb{P}([y]_i|x_1)$ and $\mathbb{P}([y]_i|x_2)$ and using Pinsker's inequality. Namely, we can estimate

$$\begin{aligned} P_d - P_f \leq & \vartheta_{A,\mathrm{KL}}(x_1, x_2) \\ & \triangleq \sqrt{\frac{m}{2} \min\left(D([Ax_1]_i || [Ax_2]_i), D([Ax_2]_i || [Ax_1]_i)\right)} \end{aligned} \quad (33)$$

where $D([Ax_1]_i||[Ax_2]_i)$ and $D([Ax_2]_i||[Ax_1]_i)$ can be computed numerically.

VI. Simulation Results

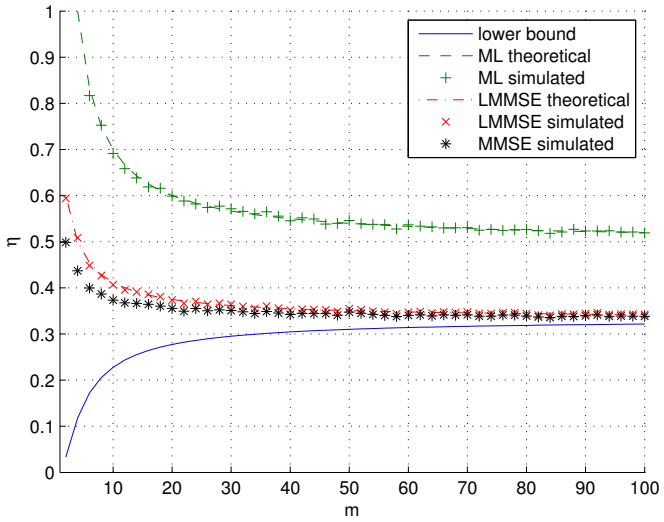
In the following sections we propose experimental results to assess the accuracy of the bounds obtained in the previous sections. Namely, we evaluate the estimation of the energy of the signal in the case of different sensing matrices, and the distinguishability of equal-energy signals for non-Gaussian sensing matrices. In both cases, all the matrices have entries that are extracted from distributions with zero mean and unitary variance.

A. Estimation of $\mathcal{E}_x$

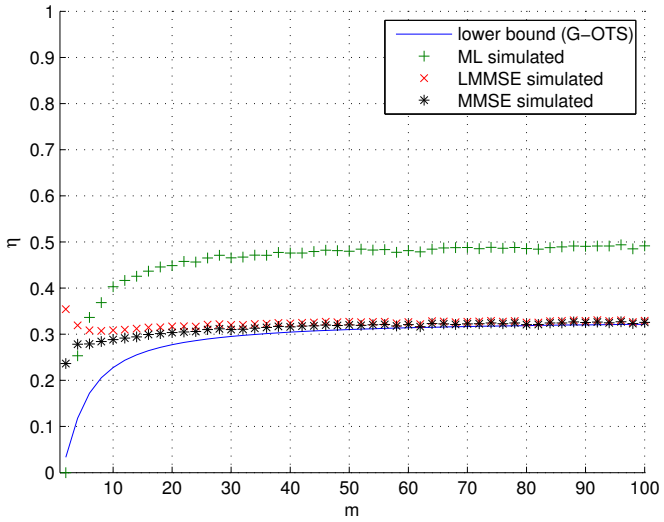
In this section, we evaluate the MSE security of OTS measurements in the case of k-sparse signals whose nonzero components are i.i.d. Gaussian. Both G-OTS measurements and OTS measurements acquires using a Bernoulli sensing matrices (B-OTS) are considered. For each experiment, empirical MSE values for the ML (25), LMMSE (29), and MMSE (26) estimators are obtained by averaging over 10^5 independent realizations of the measurements for each choice of k and m.

In a first experiment, we consider a fixed sparsity/measurement ratio $\rho = k/m = 0.5$ and we vary k in the interval $[1, 100]$. For G-OTS measurements, the obtained empirical η values versus the number of measurements m are shown in Fig. 1-(a), together with the theoretical performance of the ML and LMMSE estimator, given in (25) and (29) respectively, and the theoretical lower bound given in (11). For a fixed ρ, G-OTS measurements tend to have a constant MSE security as m grows, with the η value that does not decrease significantly for $m > 50$. It is also worth noting that the theoretical lower bound is quite loose for $m < 50$, but becomes relatively tight when m increases. For B-OTS measurements, the empirical η values are shown in Fig. 1-(b), together with the lower bound given in (11). B-OTS measurements are less secure for small m, whereas both systems have similar MSE security for $m > 50$.

In a second experiment, we consider a fixed number of measurements $m = 100$ and we vary k in the interval $[1, 50]$, obtaining different sparsity/measurement rates ρ in the interval $[0.01, 0.5]$. For G-OTS measurements, the obtained empirical

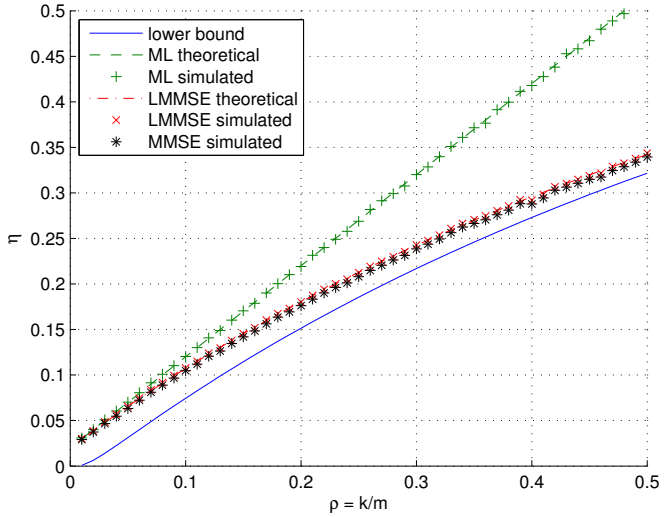


(a)

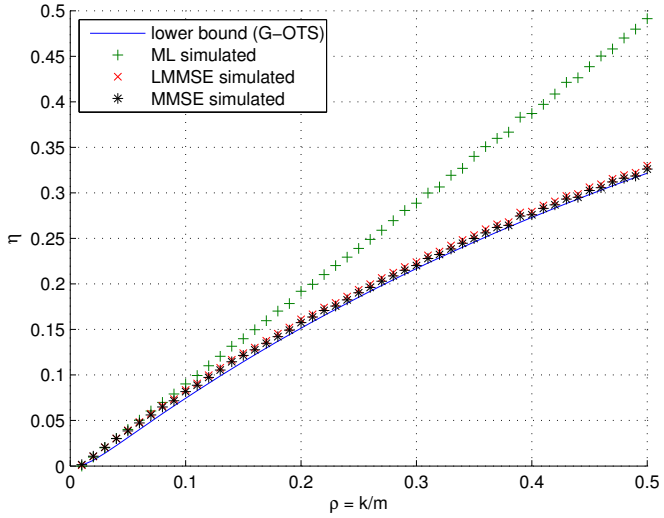


(b)

Figure 1. MSE security for different number of measurements at a fixed sparsity/measurement ratio $\rho = k/m = 0.5$. a) G-OTS; b) B-OTS.



(a)



(b)

Figure 2. MSE security for different sparsity/measurement ratios: $m = 100$, k ranges form 1 to 50. a) G-OTS; b) B-OTS.

η versus ρ are shown in Fig. 2-(a), together with the theoretical performance of the estimators and the theoretical lower bound. It is evident that the confidentiality of G-OTS measurements decreases with ρ. It is also evident that for values of ρ that are relevant to practical CS systems all the estimators can estimate the energy of x with an MSE lower than $\sigma^2_{\mathcal{E}_x}/10$, which means that the measurements permit to obtain a reasonable guess of $\mathcal{E}_x$ even if the sensing matrix is unknown. Fig. 2-(b) shows the empirical η values for B-OTS measurements, for which a similar behavior is observed. Interestingly, in both experiments the performance of B-OTS measurements is close to the lower bound obtained for G-OTS measurements, suggesting that the latter can be used to predict a sort of asymptotic behavior for large m.

As to the different estimation attacks, the above results indicate that the MMSE estimator obtains only a very slight advantage with respect to the LMMSE estimator, unless the attacker observes a very small number of measurements. Moreover, the ML estimator is comparable to the Bayesian estimators only for $\rho < 0.1$, whereas becomes quite ineffective for higher sparsity/measurement rates due to the lack of prior knowledge about the energy of x.

B. Distinguishing equal-energy signals

In this section, we evaluate the distinguishability of equal-energy signals in different scenarios. In each experiment, for the numerical evaluation of $\vartheta_{A,\mathrm{KL}}(x_1, x_2)$ and the NP test (31), the involved pdfs have been sampled on 10000 equispaced bins between -8 and 8.

The first experiment has been carried out with the aim of assessing the different upper bounds on the distinguishability of equal-energy signals. The signals have been defined as $[x_1]_i = 1/\sqrt{n}$ and $[x_2]_i = Z(\alpha)e^{-\alpha(i-1)}$, for $i = 1, \ldots, n$, where $Z(\alpha)$ is a suitable normalizing constant such that $\mathcal{E}_{x_2} = 1$. In Fig. 3 we show the theoretical upper bound $\vartheta_A(x_1, x_2)$ and the numerically evaluated upper bound $\vartheta_{A,\mathrm{KL}}(x_1, x_2)$

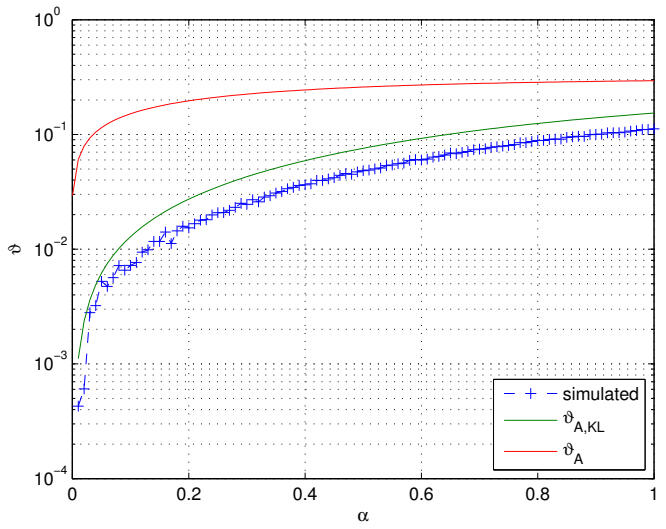


Figure 3. Distinguishability of unit energy vectors using a uniform sensing matrix, for $m = 1$, $n = 1000$.

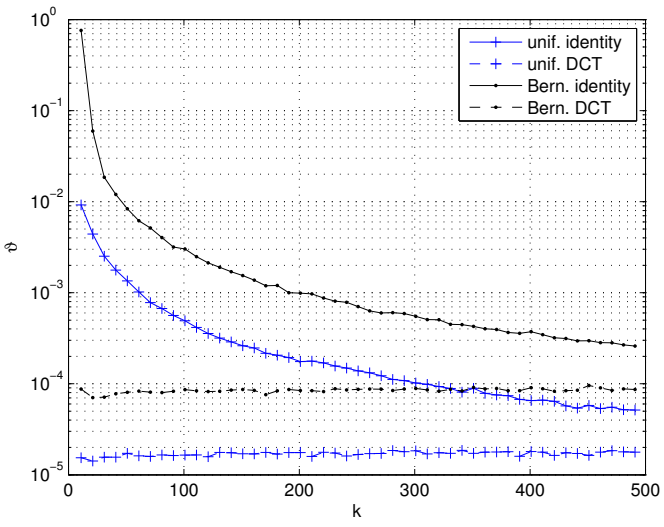


Figure 4. Distinguishability of k-sparse unit energy signals when using different sensing matrix, for $n = 1000$.

when the entries of A are i.i.d. uniform variables with unit variance (uniform sensing matrix), for $\alpha \in [0, 1]$, $m = 1$, and $n = 1000$. In the same plot, we also show the maximum value of $P_d - P_f$ achieved by the optimal NP test (31), evaluated over 10^6 independent realizations. As can be seen, the performance of the detection attack is predicted quite well by the numerical upper bound, whereas the theoretical upper bound appears overly pessimistic regarding the security of the system.

In the second experiment, we computed the numerical upper bound $\vartheta_{A,\mathrm{KL}}(x_1, x_2)$ (33) for different realizations of equal-energy signals x_1 and x_2 and different scenarios. Namely, we considered 1000 pairs θ_1, θ_2 of independent vectors with k nonzero entries uniformly distributed on a unit norm n-sphere, where the respective k-sparse signals were obtained by multiplying those vectors by a unitary matrix Φ. The first scenario considered as Φ the identity matrix, i.e., the signals were sparse in the sensing domain. The second scenario considered as Φ the discrete cosine transform (DCT) matrix. In both scenarios we computed $\vartheta_{A,\mathrm{KL}}(x_1, x_2)$ for $m = 1$, since for $m > 1$ $\vartheta_{A,\mathrm{KL}}(x_1, x_2)$ can be easily obtained by multiplying the distinguishability calculated previously by a factor $\sqrt{m}$.

In Fig. 4, we show the 0.95 percentile of $\vartheta_{A,\mathrm{KL}}(x_1, x_2)$ when $n = 1000$ and k varies in the interval $[1, 500]$. As expected, if the signals are sparse in the sensing domain the distinguishability decreases when k increases, whereas if the signal are sparse in a different domain the distinguishability is almost constant with respect to k. In Fig. 5, we show the 0.95 percentile of $\vartheta_{A,\mathrm{KL}}(x_1, x_2)$ when $k = 10$ and n varies in the interval $[20, 1000]$. As expected, the distinguishability of signals that are sparse in the DCT domain decreases when n increases, whereas if the signals are sparse in the sensing domain the distinguishability does not depend on n.

It is evident that Bernoulli matrices are less secure than uniform matrices. However, both kinds of matrix appears to have the same asymptotic behavior, both with respect to k (signal is sparse in the sensing domain) and with respect to n

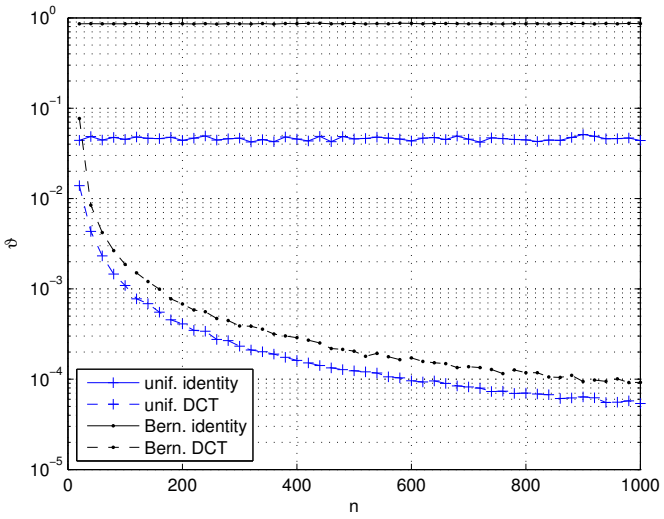


Figure 5. Distinguishability of k-sparse unit energy signals when using different sensing matrix, for $k = 10$.

(signal is sparse in the DCT domain). Namely, in the case of Fig. 4, least square fitting reports $\vartheta_{A,\mathrm{KL}}(x_1, x_2) \approx O(k^{-1.5})$ for both kinds of sensing matrix, while in the case of Fig. 5, least square fitting reports $\vartheta_{A,\mathrm{KL}}(x_1, x_2) \approx O(n^{-1.3})$. It is worth noting that this behavior is much better than the upper bound given by Corollary 4.

VII. Discussion

A. *Confidentiality vs. Signal Recovery*

The security metrics introduced in the previous sections do not consider signal recovery. However, an interesting problem is investigating the trade-off between the number of measurements required for successful signal recovery, when the sensing matrix is known at the receiver, and the confidentiality of those measurements, when the sensing matrix is kept secret. The problem of signal recovery has been extensively studied

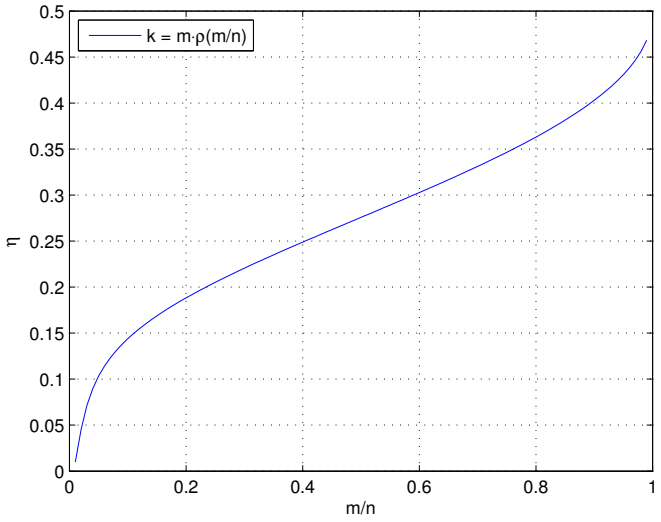


Figure 6. MSE security for different undersampling ratios m/n, assuming $n = 1000$ and $k = m \cdot \rho(m/n)$.

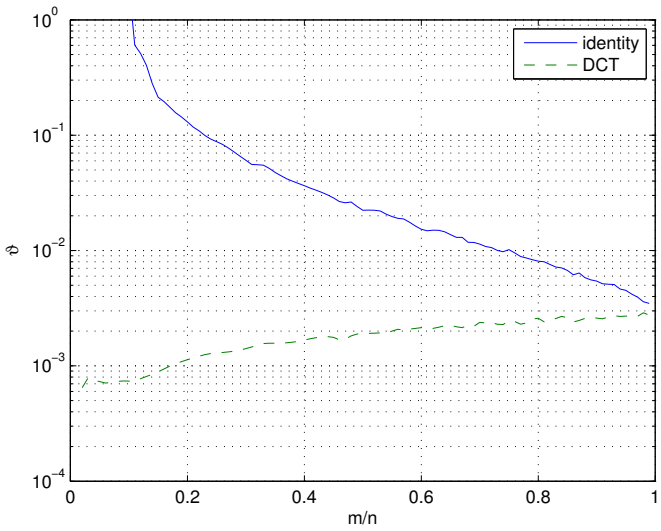


Figure 7. Distinguishability for different undersampling ratios m/n, assuming $n = 1000$ and $k = m \cdot \rho(m/n)$, when A is a Bernoulli matrix.

in CS literature [42]. In the noiseless setting we consider, an important result in this paper [43] states that, given a signal of dimension n and m measurements taken with a sensing matrix composed of i.i.d. Gaussian variables, when n tends to infinity the signal is recovered almost surely using (4) as long as its sparsity satisfies

$$k \leq m \cdot \rho(m/n). \tag{34}$$

where the curve $\rho(m/n)$ depends on polytope geometry and denotes a phase transition behavior of the recovery problem.

In Fig. 6, we report the MSE security of G-OTS measurements for different undersampling ratios m/n, assuming $n = 1000$ and $k = m \cdot \rho(m/n)$. This can be interpreted as the maximum achievable confidentiality conditional on signal recovery, i.e., when the signal can be exactly recovered from m measurements (knowing A), the MSE security of the measurements (A being secret) almost never exceeds the value given in Fig. 6. Interestingly, this value increases with m: with a higher number of measurements, less sparse signals can be recovered, for which energy estimation is less precise. It is also evident that signal recovery at very low m/n ratios implies low confidentiality of the measurements.

In Fig. 7, we report the 0.95 percentile of $\vartheta_{A,\mathrm{KL}}(x_1, x_2)$ for different undersampling ratios m/n, assuming $n = 1000$ and $k = m \cdot \rho(m/n)$, when A is a Bernoulli matrix. In this case, the curves can be interpreted as the minimum (expected) distinguishability conditional on signal recovery, i.e., when the signal can be exactly recovered from m measurements (knowing A), the (expected) distinguishability (A being secret) is almost never lower than the given curve. As can be seen, when the signal is sparse in the DCT domain we can have both signal recovery and a certain level of confidentility, whereas for signals sparse in the sensing domain signal recovery at low m/n ratios implies very low confidentiality.

B. Quantization

The analysis in the previous sections assumes that signals, sensing matrices, and measurements are represented with infinite precision. If we assume that in a practical setting an attacker observes the quantized measurements $\mathcal{Q}(y)$, the data processing inequality [33] states that $I(x; \mathcal{Q}(y)) \leq I(x; y)$, meaning that quantized measurements are at least as secure as infinite precision measurements. Hence, results based on the model in (2) can be used to lower bound the confidentility of quantized measurements.

An interesting case of quantized measurements is offered by the one-bit compressed sensing framework [44], in which only the sign of the random projections is taken as a measurement. One-bit compressed sensing has received a lot of attention because of its applications in analog to digital conversion and its connections to locality sensitive hashing [45]. In this case, it is easy to see that one-bit measurements depends only on the spherical angle of unquantized measurements. Hence, according to Lemma 1 one-bit measurements obtained with a Gaussian sensing matrix achieve perfect secrecy in the OTS setting. This is an example of a practical scenario in which CS can provide useful security properties.

As a last remark, usually the sensing hardware will be implemented relying on a digital architecture in practice, meaning that a practical sensing matrix will be composed of quantized entries. The confidentiality of a quantized matrix will in general be different that the corresponding real valued matrix. However, the security metrics in Section IV can be applied to quantized matrices as well, as shown in Section VI when dealing with Bernoulli matrices.

C. Implementation Issues

The implementation of either the G-OTS or the SG-OTS acquisition system will require to transmit a sequence of i.i.d. Gaussian sensing matrices. A solution is to use a secure random number generator (SRNG) [46] and assume that sender

and receiver synchronize their generators by sharing a secret seed, that acts like a secret key. Like a stream cipher replaces the key of the one-time pad with a keystream generated from a shared key, a practical OTS system will be based on a "keystream" of sensing matrices generated from this shared key. However, since in a OTS system the size of this keystream will be much larger than in a conventional stream cipher, a drawback is that in order to guarantee the same security level the shared key should be changed more often. The resulting acquisition system is not perfectly secure, since a brute force search of the key space will surely break it [16]. Moreover, the resulting sensing matrices will be inevitably correlated and the attacker may exploit this correlation, together with some weaknesses of the SNRG, in order to try to estimate the key. Finally, the distribution of the SRNG may slightly deviate from Gaussian, so that the hypotheses of the G-OTS/SG-OTS acquisition system are not exactly satisfied.

An alternative way for securely exchanging a private sensing matrix is provided by wireless physical layer security, which exploits the randomness of the wireless channel for extracting a common secret between sender and receiver. A solution is studied in [12], however it requires several iterations for generating a secure sensing matrix and may not be well suited for a OTS scenario.

As to the SG-OTS system, an auxiliary secure channel to transmit the value of $\mathcal{E}_y$ is required in order to exactly recover x at the intended receiver. Such a secure channel can be implemented by relying on conventional cryptographic techniques. As a consequence, the combination of SG-OTS measurements and auxiliary channel will not be perfectly secure, since a practical cryptosystem like AES offers only computational security. Alternatively, the intended receiver could exploit prior knowledge on x and avoid an auxiliary channel. However, this solution will imply the nonexact recovery of x.

VIII. Conclusions and Future Work

The results obtained in this paper give interesting insights regarding the confidentiality of CS measurements. The first important result is that a sensing matrix with zero mean i.i.d. Gaussian entries reveals only the energy of the sensed signal. As a consequence, the spherical angle of the signal is perfectly hidden by the measurements if the energy and the angle are statistically independent. Moreover, the energy of the signal is revealed only by the energy of the measurements, since Gaussian measurements have a uniformly distributed spherical angle. This result holds irrespective of the distribution of the signal x, which makes it a very general result.

The second important result is that the spherical angle of the measurements taken with a generic sensing matrix reveals only the spherical angle of the sensed signal. As a consequence, normalizing the measurements can provide an efficient way for increasing their confidentiality. As a matter of fact, normalizing Gaussian measurements actually provides a perfectly secret channel, that can be used to provide an effective confidentiality layer, offering similar security as standard cryptographic tools.

Apart from the special case of normalized measurements obtained with a Gaussian sensing matrix, CS can not provide confidentiality according to standard cryptographic definitions. Nevertheless, we provide useful criteria to assess the confidentiality of one-time CS measurements as a weak form of encryption. In the case of unnormalized Gaussian measurements, it is possible to upper bound the information leakage about the energy and predict the precision with which the energy can be estimated. In the case of normalized measurements taken with a non-Gaussian matrix, it is possible to lower bound the error of any detector trying to distinguish different signals. According to such quantities, a system designer can decide whether CS provides sufficient confidentiality for the application at hand, or standard encryption is needed.

The results obtained with the above metrics indicate that CS, even if it is not a replacement for standard encryption methods, can be used to provide a built-in data obfuscation layer complementing traditional cryptographic tools in many privacy preserving applications. As an example, let us imagine a sensor network that acquires and broadcasts CS measurements. In this setting, the use of CS is justified by the need to achieve a compact but computationally cheap signal representation for data communication. By using the proposed OTS framework, only authorized nodes knowing the actual sensing matrix can recover meaningful information. At the same time, the confidentiality of the measurements is guaranteed with respect to non-authorized nodes, without the need of an additional and power consuming encryption layer.

There are still interesting open questions regarding the confidentiality of CS measurements. Possible directions for future research include the characterization of more structured sensing matrices, the effects of quantization, and the extension to the case of sensing matrices that are used multiple times.

Appendix

1) Proof of Proposition 1: Let us consider the probability distribution function $\mathbb{P}(y|x)$ for a given x. Since $[A]_{i,j}$ are i.i.d. Gaussian, we have that $\mathbb{P}(y|x)$ is a multivariate Gaussian distribution with mean $\mu_{y|x}$ and covariance matrix $C_{y|x}$. It is immediate to find $\mu_{y|x} = E[y|x] = E[A]x = 0$, whereas by rewriting $y = (I \otimes x^T)\mathrm{vec}(A^T)$, where $\otimes$ denotes Kronecker product and $\mathrm{vec}(A)$ vectorizes matrix A by stacking its columns, we have

$$\begin{aligned} E[y \cdot y^T|x] =& (I \otimes x^T)E[\mathrm{vec}(A^T)\mathrm{vec}(A^T)^T](I \otimes x) \\ =& \sigma_A^2 (I \otimes x^T)(I \otimes x) \\ =& \sigma_A^2 x^T x I_m = \sigma_A^2 \mathcal{E}_x I_m \end{aligned} \tag{35}$$

where m is the number of measurements, I_m denotes an $m \times m$ identity matrix, and we assume that $[A]_{i,j}$ have variance σ_A^2. From the above results, it follows that $\mathbb{P}(y|x)$ depends only on $\mathcal{E}_x$, i.e. $\mathbb{P}(y|x) = \mathbb{P}(y|\mathcal{E}_x)$. The proof then follows from the following chain of mutual information equalities [33]

$$\begin{aligned} I(x;y) =& I(x, \mathcal{E}_x; y) \\ =& I(\mathcal{E}_x; y) + I(x; y|\mathcal{E}_x) \\ =& I(\mathcal{E}_x; y). \end{aligned} \tag{36}$$

since $\mathbb{P}(y|x) = \mathbb{P}(y|\mathcal{E}_x)$ implies $I(x; y|\mathcal{E}_x) = 0$.

2) Proof of Proposition 2: For a given $\mathcal{E}_x$, y is distributed as a multivariate Gaussian with a scaled identity covariance matrix, hence we have that $\mathbb{P}(y|\mathcal{E}_x)$ can be expressed as a function of $y^T y = \mathcal{E}_y$, i.e., $\mathbb{P}(y|\mathcal{E}_x) \triangleq f(\mathcal{E}_y, \mathcal{E}_x)$. Moreover, this implies that also $\mathbb{P}(y)$ can be expressed as a function of $\mathcal{E}_y$, i.e., $\mathbb{P}(y) = \int f(\mathcal{E}_y, \mathcal{E}_x)\mathbb{P}(\mathcal{E}_x)d\mathcal{E}_x \triangleq g(\mathcal{E}_y)$, meaning that y is distributed according to a spherically symmetric distribution. Let us define $u_y = y/\sqrt{\mathcal{E}_y}$. The proposition is proved by Prop. 1 and the following chain of equalities

$$\begin{aligned} I(\mathcal{E}_x; y) =& I(\mathcal{E}_x; \mathcal{E}_y, u_y) \\ =& I(\mathcal{E}_x; \mathcal{E}_y) + I(\mathcal{E}_x; u_y|\mathcal{E}_y) \\ =& I(\mathcal{E}_x; \mathcal{E}_y). \end{aligned} \tag{37}$$

since u_y is uniformly distributed on the unit m-sphere and independent of $\mathcal{E}_x$ and $\mathcal{E}_y$.

3) Proof of Proposition 3: Let us consider the equalities $y = Ax = \sqrt{\mathcal{E}_x} \cdot Au_x$ and $\mathcal{E}_y = y^T y = \mathcal{E}_x \cdot u_x^T A^T A u_x$. It follows

$$u_y = y/\sqrt{\mathcal{E}_y} = (u_x^T A^T A u_x)^{-1/2} A u_x \tag{38}$$

which implies $\mathbb{P}(u_y|x) = \mathbb{P}(u_y|u_x)$. The proof then follows on the same lines as the proof of Prop. 1.

4) Proof of Lemma 3: We have the following chain of inequalities

$$\begin{aligned} I(\mathcal{E}_x; \mathcal{E}_y) =& h(\mathcal{E}_y) - h(\mathcal{E}_y|\mathcal{E}_x) \\ =& h(\mathcal{E}_y) + E\left[\int \mathbb{P}(\mathcal{E}_y|\mathcal{E}_x) \log \mathbb{P}(\mathcal{E}_y|\mathcal{E}_x) d\mathcal{E}_y\right] \\ =& h(\mathcal{E}_y) - \xi\left(\frac{m}{2}\right) - \log(2\sigma_A^2) - E\left[\log(\mathcal{E}_x)\right] \end{aligned} \tag{39}$$

where we used the fact that $\mathbb{P}(\mathcal{E}_y|\mathcal{E}_x)$ is a chi-square distribution with m degrees of freedom scaled by $\sigma_A^2 \mathcal{E}_x$. The differential entropy of $\mathcal{E}_y$ can be upper bounded by the differential entropy of a Gamma distribution with the same expectation and log-expectation [47], i.e.,

$$h(\mathcal{E}_y) \leq \xi(\kappa^*) + \log(\vartheta^*) \tag{40}$$

where κ^* and ϑ^* satisfy $\kappa^*\vartheta^* = E[\mathcal{E}_y]$ and $\psi(\kappa^*) + \log(\vartheta^*) = E[\log(\mathcal{E}_y)]$. The moments of $\mathcal{E}_y$ can be derived as $E[\mathcal{E}_y] = m\sigma_A^2 E[\mathcal{E}_x]$ and $E[\log(\mathcal{E}_y)] = \psi\left(\frac{m}{2}\right) + \log(2\sigma_A^2) + E[\log(\mathcal{E}_x)]$. Hence, by doing some simple algebra, it is easy to derive

$$h(\mathcal{E}_y) \leq \xi(\kappa^*) - \psi(\kappa^*) + \psi\left(\frac{m}{2}\right) + \log(2\sigma_A^2) + E[\log(\mathcal{E}_x)] \tag{41}$$

with κ^* satisfying $\log(\kappa^*) - \psi(\kappa^*) = \log(m/2) - \psi(m/2) + \log(E[\mathcal{E}_x]) - E[\log(\mathcal{E}_x)]$. The solution κ^* is unique, since $\log(\kappa^*) - \psi(\kappa^*)$ is positive and strictly decreasing (see, e.g., Th. 3.1 in [48]) and $\log(E[\mathcal{E}_x]) \geq E[\log(\mathcal{E}_x)]$. The proof then easily follows by combining (39) and (41).

5) Proof of Corollary 2: Since $\mathcal{E}_x$ is a chi-square variable with k degrees of freedom scaled by σ_x^2, we have $\sigma_{\mathcal{E}_x}^2 = 2k\sigma_x^4$ and $\log(E[\mathcal{E}_x]) - E[\log(\mathcal{E}_x)] = \log(k/2) - \psi(k/2)$ [49]. Moreover, by using (10) in (8) we have

$$\begin{aligned} E[(\mathcal{E}_x - \hat{\mathcal{E}}_x)^2] \geq& \frac{1}{2\pi} e^{2h(\mathcal{E}_x) - 2\xi'(\kappa^*) + 2\xi'\left(\frac{m}{2}\right) - 1} \\ =& \frac{4\sigma_x^4}{2\pi} e^{2\xi\left(\frac{k}{2}\right) + 2\xi'\left(\frac{m}{2}\right) - 2\xi'(\kappa^*) - 1} \end{aligned} \tag{42}$$

from which the proof easily follows.

6) Proof of Proposition 4: Let us consider a generic element $[y_i]_k$ of the vector $y_i = Ax_i$, for $i = 1, 2$. Clearly, $\mathrm{Var}[[y_i]_k] = \sigma_A^2 ||x_i||_2^2 = \sigma_A^2$. The CLT states that, as n goes to infinity, the variable $[y_i]_k$ converges in distribution towards a Gaussian variable G with variance σ_A^2. Moreover, according to Theorem 1 in [36], if the entries of A are i.i.d., admit a probability density function, and satisfy the Poincaré inequality with constant $c > 0$, the following inequality holds

$$D([y_i]_k||G) \leq \frac{2||x_i||_4^4}{c + (2-c)||x_i||_4^4} D(a||G). \tag{43}$$

Let us assume that $G^{(m)}$ is a vector of m i.i.d. Gaussian variables with variance σ_A^2. The proof then follows from the following chain of inequalities

$$\begin{aligned} \delta(Ax_1, Ax_2) \leq& \delta(Ax_1, G^{(m)}) + \delta(Ax_2, G^{(m)}) \\ \leq& \sqrt{\frac{1}{2} D(Ax_1||G^{(m)})} + \sqrt{\frac{1}{2} D(Ax_2||G^{(m)})} \\ =& \sqrt{\frac{m}{2} D([y_1]_k||G)} + \sqrt{\frac{m}{2} D([y_2]_k||G)} \end{aligned} \tag{44}$$

where the first line follows from the fact TV distance satisfies the triangle inequality, the second line follows from Pinsker's inequality between TV distance and KL divergence [50], [51], and the third line follows form the fact that the entries of y_1 and y_2 are i.i.d..

7) Proof of Lemma 5: Since a vector x uniformly distributed on a unit norm n-sphere can be obtained by normalizing a vector of n i.i.d. zero mean and unit variance Gaussian variables [52], we have that its components satisfy $x_i^2 = \Gamma_i / Z$, $i = 1, \ldots, n$, where Γ_i are i.i.d. according to a chi-squared distribution with 1 degree of freedom and $Z = \sum_i^n \Gamma_i$ is a chi-squared variable with n degrees of freedom. Hence, we immediately have

$$\Pr\left\{\sum_{i=1}^n \Gamma_i^2 \geq t\right\} = \Pr\left\{Z^2 ||x||_4^4 \geq t\right\} = \Pr\left\{||x||_4^4 \geq \frac{t}{Z^2}\right\}. \tag{45}$$

Hence, we can derive the following bound

$$\begin{aligned} \Pr\left\{||x||_4^4 \geq \frac{t}{K}\right\} \leq& \Pr\left\{Z^2 > K\right\} \Pr\left\{||x||_4^4 \geq \frac{t}{Z^2}\right\} \\ &+ \Pr\left\{Z^2 \leq K\right\} \\ \leq& \Pr\left\{\sum_{i=1}^n \Gamma_i^2 \geq t\right\} + \Pr\left\{Z^2 \leq K\right\}. \end{aligned} \tag{46}$$

Using a simple Chernoff bound on chi-squared distributed variables, we obtain

$$\begin{aligned} \Pr\left\{Z^2 \leq (1-\epsilon)n^2\right\} \leq& \Pr\left\{Z^2 \leq (1-\epsilon/2)^2 n^2\right\} \\ =& \Pr\left\{Z \leq (1-\epsilon/2)n\right\} \leq e^{-\frac{n\epsilon_2^2}{16}}. \end{aligned} \tag{47}$$

As to the bound in (17), this can be derived by applying Chebyshev's inequality to $\sum_{i=1}^n \Gamma_i^2$. We have $E[\Gamma_i^2] = 3$.

Hence, since Γ_i are i.i.d,

$$\begin{aligned}\Pr\left\{\sum_{i=1}^{n}\Gamma_i^2 \geq (3+\epsilon)n\right\} \leq & \frac{\mathrm{Var}[\sum_{i=1}^{n}\Gamma_i^2]}{n^2\epsilon^2} \\ = & \frac{\sum_{i=1}^{n}\mathrm{Var}[\Gamma_i^2]}{n^2\epsilon^2} = \frac{96}{n\epsilon^2}.\end{aligned} \tag{48}$$

By using (47) and (48) in (46), we obtain the result in (17).

As to the bound in (18), let us consider the function $f(t) = \cosh\sqrt{t}$. The function $f(t)$ is nondecreasing and convex for $t > 0$, hence, for $-\frac{1}{2} < s < \frac{1}{2}$ we have

$$\begin{aligned}\Pr\left\{\sum_{i=1}^{n}\Gamma_i^2 \geq n\epsilon\right\} = & \Pr\left\{f\left(\frac{s^2}{n}\sum_{i=1}^{n}\Gamma_i^2\right) \geq f(s^2\epsilon)\right\} \\ \leq & \frac{E\left[f\left(\frac{s^2}{n}\sum_{i=1}^{n}\Gamma_i^2\right)\right]}{f(s^2\epsilon)} \\ \leq & \frac{E\left[\frac{1}{n}\sum_{i=1}^{n}f\left(s^2\Gamma_i^2\right)\right]}{f(s^2\epsilon)} \\ = & \frac{E\left[e^{s\Gamma_i}\right] + E\left[e^{-s\Gamma_i}\right]}{2\cosh(s\sqrt{\epsilon})} \\ = & \frac{(1-2s)^{-\frac{1}{2}} + (1+2s)^{-\frac{1}{2}}}{2\cosh(s\sqrt{\epsilon})}\end{aligned} \tag{49}$$

where the first inequality follows from Markov's inequality and the second inequality from Jensen's inequality. The result in (18) can then be obtained by choosing $s = \frac{1}{4}$ and using (47) and (49) in (46).

8) Proof of Corollary 3: Let us define $y_i' = Au_{x_i}$. It is easy to verify $u_{y_i'} = y_i'/\sqrt{\mathcal{E}_{y_i'}} = u_{y_i}$. We have the following inequalities involving the KL divergence

$$\begin{aligned}D(Au_{x_i}||G^{(m)}) = & D(\mathbb{P}(u_{y_i}, \mathcal{E}_{y_i'})||\mathbb{P}(u_G, \mathcal{E}_G)) \\ = & D(u_{y_i}||u_G) + D(\mathbb{P}(\mathcal{E}_{y_i'}|u_{y_i})||\mathbb{P}(\mathcal{E}_G|u_G)) \\ \geq & D(u_{y_i}||u_G)\end{aligned} \tag{50}$$

where $u_G = G^{(m)}/\sqrt{\mathcal{E}_G}$ and we exploited the chain rule for KL divergence [33] and the fact that KL divergence is always nonnegative. Hence, the proof follows from the following chain of inequalities

$$\begin{aligned}\delta(u_{y_1}, u_{y_2}) \leq & \delta(u_{y_1}, u_G) + \delta(u_{y_2}, u_G) \\ \leq & \sqrt{\frac{1}{2}D(u_{y_1}||u_G)} + \sqrt{\frac{1}{2}D(u_{y_2}||u_G)} \\ \leq & \sqrt{\frac{1}{2}D(Au_{x_1}||G^{(m)})} + \sqrt{\frac{1}{2}D(Au_{x_2}||G^{(m)})}.\end{aligned} \tag{51}$$

9) Derivation of (26)*:* The MMSE estimator can be analytically computed as

$$\begin{aligned}\hat{\mathcal{E}}_{x,MMSE} = & \frac{\int \mathcal{E}_x\mathbb{P}(y|\mathcal{E}_x)\mathbb{P}(\mathcal{E}_x)d\mathcal{E}_x}{\int \mathbb{P}(y|\mathcal{E}_x)\mathbb{P}(\mathcal{E}_x)d\mathcal{E}_x} \\ = & \frac{\int_0^\infty \mathcal{E}_x^{\frac{k}{2}-\frac{m}{2}} e^{-\frac{\mathcal{E}_y}{2\sigma_A^2\mathcal{E}_x} - \frac{\mathcal{E}_x}{2\sigma_x^2}} d\mathcal{E}_x}{\int_0^\infty \mathcal{E}_x^{\frac{k}{2}-\frac{m}{2}-1} e^{-\frac{\mathcal{E}_y}{2\sigma_A^2\mathcal{E}_x} - \frac{\mathcal{E}_x}{2\sigma_x^2}} d\mathcal{E}_x}\end{aligned} \tag{52}$$

where we exploited the fact that y, conditional on $\mathcal{E}_x$, is distributed according to a multivariate Gaussian with covariance matrix $\sigma_A^2\mathcal{E}_x I_m$ and $\mathcal{E}_x$ is distributed as a chi-square with k degrees of freedom scaled by σ_x^2. The result then follows by applying the equality $\int_0^\infty x^{\nu-1}e^{-a/x-bx}dx = 2(a/b)^{\nu/2}K_\nu(2\sqrt{ab})$ (see [53, Page 368]).

REFERENCES

[1] E. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, 2006.

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

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

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

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

[6] J. Bazerque and G. Giannakis, "Distributed spectrum sensing for cognitive radio networks by exploiting sparsity," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1847–1862, March 2010.

[7] Z. Fanzi, C. Li, and Z. Tian, "Distributed compressive spectrum sensing in cooperative multihop cognitive networks," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 1, pp. 37–48, Feb 2011.

[8] J. Haupt, W. Bajwa, M. Rabbat, and R. Nowak, "Compressed sensing for networked data," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 92–101, March 2008.

[9] L. Xiang, J. Luo, and A. Vasilakos, "Compressed data aggregation for energy efficient wireless sensor networks," in *2011 8th Annual IEEE Comm. Society Conf. on Sensor, Mesh and Ad Hoc Communications and Networks (SECON)*, June 2011, pp. 46–54.

[10] F. Fazel, M. Fazel, and M. Stojanovic, "Random access compressed sensing for energy-efficient underwater sensor networks," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 8, pp. 1660–1670, September 2011.

[11] M. Mardani, G. Mateos, and G. Giannakis, "Dynamic anomalography: Tracking network anomalies via sparsity and low rank," *IEEE J. Sel. Topics Signal Process.*, vol. 7, no. 1, pp. 50–66, Feb 2013.

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

[13] Q. Wang, W. Zeng, and J. Tian, "A compressive sensing based secure watermark detection and privacy preserving storage framework," *Image Processing, IEEE Transactions on*, vol. 23, no. 3, pp. 1317–1328, March 2014.

[14] ——, "Compressive sensing based secure multiparty privacy preserving framework for collaborative data-mining and signal processing," in *Multimedia and Expo (ICME), 2014 IEEE International Conference on*, July 2014, pp. 1–6.

[15] R. Sun and W. Zeng, "Secure and robust image hashing via compressive sensing," *Multimedia Tools and Applications*, vol. 70, no. 3, pp. 1651–1665, 2014.

[16] Y. Rachlin and D. Baron, "The secrecy of compressed sensing measurements," in *2008 46th Annual Allerton Conference on Communication, Control, and Computing*. IEEE, 2008, pp. 813–817.

[17] C. E. Shannon, "Communication theory of secrecy systems," *Bell. Syst. Tech. J.*, vol. 28, pp. 656–715, 1949.

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

[19] V. Cambareri, J. Haboba, F. Pareschi, H. Rovatti, G. Setti, and K.-W. Wong, "A two-class information concealing system based on compressed sensing," in *2013 IEEE International Symposium on Circuits and Systems (ISCAS)*, 2013, pp. 1356–1359.

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

[21] ——, "On known-plaintext attacks to a compressed sensing-based encryption: a quantitative analysis," *IEEE Trans. Inf. Forensics Security*, vol. PP, no. 99, pp. 1–1, 2015.

[22] T. Bianchi, V. Bioglio, and E. Magli, "On the security of random linear measurements," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'14)*, 2014, pp. 3992–3996.
[23] A. Said, "Measuring the strength of partial encryption schemes," in *IEEE International Conference on Image Processing, 2005. ICIP 2005.*, vol. 2, Sept 2005, pp. II–1126–9.
[24] Y. Mao and M. Wu, "A joint signal processing and cryptographic approach to multimedia encryption," *IEEE Trans. Image Process.*, vol. 15, no. 7, pp. 2061–2075, July 2006.
[25] T. Stütz and A. Uhl, "A survey of H.264 AVC/SVC encryption," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 3, pp. 325 –339, Mar. 2012.
[26] P. Georgiev, F. Theis, and A. Cichocki, "Sparse component analysis and blind source separation of underdetermined mixtures," *IEEE Trans. Neural Netw.*, vol. 16, no. 4, pp. 992–996, July 2005.
[27] G. Reeves, N. Goela, N. Milosavljevic, and M. Gastpar, "A compressed sensing wire-tap channel," in *2011 IEEE Information Theory Workshop (ITW)*, 2011, pp. 548–552.
[28] S. Agrawal and S. Vishwanath, "Secrecy using compressive sensing," in *2011 IEEE Information Theory Workshop (ITW)*, 2011, pp. 563–567.
[29] A. Abdulghani and E. Rodriguez-Villegas, "Compressive sensing: From "compressing while sampling" to "compressing and securing while sampling"," in *2010 Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Aug 2010, pp. 1127–1130.
[30] K. Liu, H. Kargupta, and J. Ryan, "Random projection-based multiplicative data perturbation for privacy preserving distributed data mining," *IEEE Trans. Knowl. Data Eng.*, vol. 18, no. 1, pp. 92–106, Jan 2006.
[31] S. Zhou, J. Lafferty, and L. Wasserman, "Compressed and privacy-sensitive sparse regression," *IEEE Trans. Inf. Theory*, vol. 55, no. 2, pp. 846–866, Feb 2009.
[32] Y. D. Li, Z. Zhang, M. Winslett, and Y. Yang, "Compressive mechanism: Utilizing sparse representation in differential privacy," in *Proceedings of the 10th Annual ACM Workshop on Privacy in the Electronic Society*, ser. WPES '11. New York, NY, USA: ACM, 2011, pp. 177–182.
[33] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley-Interscience, 2006.
[34] S. Goldwasser and S. Micali, "Probabilistic encryption," *Journal of Computer and System Sciences*, vol. 28, no. 2, pp. 270 – 299, 1984.
[35] L. LeCam, "Convergence of estimates under dimensionality restrictions," *Ann. Statist.*, vol. 1, no. 1, pp. 38–53, 01 1973.
[36] S. Artstein, K. M. Ball, F. Barthe, and A. Naor, "On the rate of convergence in the entropic central limit theorem," *Probability Theory and Related Fields*, vol. 129, no. 3, pp. 381–390, 2004.
[37] S. G. Bobkov, "Isoperimetric and analytic inequalities for log-concave probability measures," *The Annals of Probability*, vol. 27, no. 4, pp. 1903–1921, 10 1999.
[38] A. C. Berry, "The accuracy of the Gaussian Approximation to the sum of independent variates," *Transactions of the American Mathematical Society*, vol. 49, no. 1, pp. 122–136, Jan. 1941.
[39] O. Johnson and A. Barron, "Fisher information inequalities and the central limit theorem," *Probability Theory and Related Fields*, vol. 129, no. 3, pp. 391–409, 2004.
[40] S. G. Bobkov, G. P. Chistyakov, and F. Götze, "Rate of convergence and Edgeworth-type expansion in the entropic central limit theorem," *Ann. Probab.*, vol. 41, no. 4, pp. 2479–2512, 07 2013.
[41] S. M. Kay, *Fundamentals of Statistical Signal Processing, Estimation Theory*. Prentice-Hall, 1993.
[42] Y. Eldar and G. Kutyniok, *Compressed Sensing: Theory and Applications*, ser. Compressed Sensing: Theory and Applications. Cambridge University Press, 2012.
[43] D. Donoho and J. Tanner, "Precise undersampling theorems," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 913–924, June 2010.
[44] L. Jacques, J. Laska, P. Boufounos, and R. Baraniuk, "Robust 1-bit compressive sensing via binary stable embeddings of sparse vectors," *IEEE Trans. Inf. Theory*, vol. 59, no. 4, pp. 2082–2102, April 2013.
[45] A. Andoni and P. Indyk, "Near-optimal hashing algorithms for approximate nearest neighbor in high dimensions," *Commun. ACM*, vol. 51, no. 1, pp. 117–122, Jan. 2008.
[46] A. J. Menezes, P. C. van Oorschot, and S. A. Vanstone, *Handbook of Applied Cryptography*. CRC Press, 1996.
[47] D. Gokhale, "Maximum entropy characterizations of some distributions," in *A Modern Course on Statistical Distributions in Scientific Work*, ser. NATO Advanced Study Institutes Series, G. Patil, S. Kotz, and J. Ord, Eds. Springer Netherlands, 1975, vol. 17, pp. 299–304.
[48] G. D. Anderson, R. W. Barnard, K. C. Richards, M. K. Vamanamurthy, and M. Vuorinen, "Inequalities for zero-balanced hypergeometric functions," *Transactions of the American Mathematical Society*, vol. 347, no. 5, pp. 1713–1723, May 1995.
[49] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. New York: Dover Publications, 1964.
[50] S. Pinsker, *Information and information stability of random variables and processes*, ser. Holden-Day series in time series analysis. Holden-Day, 1964.
[51] S. Kullback, "A lower bound for discrimination information in terms of variation (corresp.)," *IEEE Trans. Inf. Theory*, vol. 13, no. 1, pp. 126–127, 1967.
[52] G. Marsaglia, "Choosing a point from the surface of a sphere," *Ann. Math. Statist.*, vol. 43, no. 2, pp. 645–646, 04 1972.
[53] I. S. Gradshteyn and I. M. Ryzhik, *Table of integrals, series, and products*, 7th ed. Elsevier/Academic Press, Amsterdam, 2007.

Tiziano Bianchi (S'03-M'05) received the M.Sc. degree (Laurea) in electronic engineering and the Ph.D. degree in information and telecommunication engineering from the University of Florence, Italy, in 2001 and 2005, respectively.

Since December 2012, he is with the Department of Electronics and Telecommunications, Politecnico di Torino as an Assistant Professor. From 2005 to 2012, he has been with the Department of Electronics and Telecommunications, University of Florence as a Research Assistant. His research interests have involved signal processing in communications and processing of SAR images. Current research topics include multimedia security technologies, signal processing in the encrypted domain, and security aspects of compressed sensing. He has published more than 100 papers on international journals and conference proceedings.

Valerio Bioglio has joined the Huawei French Research Centre in January 2015 as Researcher for the Team Coding for Data Networks. Previously he joined the CRISP Team as a Post-doc Researcher at the Department of Electronics and Telecommunications, Politecnico of Torino. He obtained his BSc in Mathematics from Università degli Studi di Torino (Italy) in 2006. He obtained her MSc in Applied Mathematics from Università degli Studi di Torino (Italy) in 2008. He completed his PhD in Computer Science at the Computer Science Department of the Università degli Studi di Torino (Italy) in 2012. His main research interests are information theory, communication theory and signal processing with focus on mathematical modeling. His publications include papers on rateless codes, network coding, P2P streaming, game theory for cooperative communications, codes for caching and distributed storage, compressed sensing, image processing.

Enrico Magli (S'97–M'01–SM'07) received the M.Sc. and Ph.D. degrees from Politecnico di Torino, Torino, Italy, in 1997 and 2001, respectively. He is currently an Associate Professor with Politecnico di Torino, Torino, Italy. His research interests are in the field of compressive sensing, image and video coding, and vision. He is an associate editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, the IEEE TRANSACTIONS ON MULTIMEDIA, and the EURASIP JOURNAL ON IMAGE AND VIDEO PROCESSING, and an IEEE Distinguished Lecturer for 2015-2016. He received the IEEE Geoscience and Remote Sensing Society 2011 Transactions Prize Paper Award, an IEEE ICIP 2015 Best Student Paper award (as senior author), and the 2010 and 2014 Best Associate Editor award of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY.