Compressive Sensing based Leakage Sampling and Reconstruction: A First Study

Changhai Ou, Chengju Zhou and Siew-Kei Lam

Hardware & Embedded Systems Lab, School of Computer Science and Engineering,
Nanyang Technological University, Singapore.
CHOu@ntu.edu.sg, zhou0271@e.ntu.edu.sg, ASSKLam@ntu.edu.sg

Abstract. An important prerequisite for Side-channel Attack (SCA) is leakage sampling where the side-channel measurements (e.g. power traces) of the cryptographic device are collected for further analysis. However, as the operating frequency of cryptographic devices continues to increase due to advancing technology, leakage sampling will impose higher requirements on the sampling equipment. This paper undertakes the first study to show that effective leakage sampling can be achieved without relying on sophisticated equipments through Compressive Sensing (CS). In particular, CS can obtain low-dimensional samples from high-dimensional power traces by simply projecting the useful information onto the observation matrix. The leakage information can then be reconstructed in a workstation for further analysis. With this approach, the sampling rate to obtain the side-channel measurements is no longer limited by the operating frequency of the cryptographic device and Nyquist sampling theorem. Instead it depends on the sparsity of the leakage signal. Our study reveals that there is large amount of information redundancy in power traces obtained from the leaky device. As such, CS can employ a much lower sampling rate and yet obtain equivalent leakage sampling performance, which significantly lowers the requirement of sampling equipments. The feasibility of our approach is verified theoretically and through experiments.

Keywords: compressive sensing · matching pursuit · OMP · CoSaMP · SP · GOMP · side-channel attack

1 Introduction

Traditional side-channel leakage sampling conforms to the Nyquist theorem, i.e. the sampling rate of the equipment must be more than (or at least) twice the highest operating frequency of the cryptographic device in order for the original leakage to be reconstructed completely from the samples. In order to obtain more leakage details for Side-Channel Attacks (SCAs), the sampling rate is usually several times higher than the operating frequency of the leaky device. With advancing technology, the operating frequency of cryptographic devices is increasing rapidly. Micro-controllers and Field-Programmable Gate Arrays (FPGAs) today have a clock operating frequency of a few MHz, while mobile phones, laptops and desktop computers can run in the order of several GHz.

The increase in operating speed of the cryptographic devices is expected to impose a challenge to the acquisition and storage of leakage signals for SCA as more sophisticated equipments will be required for leakage sampling. This may seem as a benefit from the security standpoint, but in this paper, we refute this presumption by demonstrating that using lower sampling rates can still enable us to obtain equivalent leakage sampling performance, which significantly lowers the requirement of sampling equipments. This is achieved through a novel use of Compressive Sensing (CS) which abandons the need to

conform to the sampling rate that is dictated by the Nyquist theorem. In particular, using CS, the leakage sampling rate depends solely on the sparsity of the leakage signal, which is a predominant characteristic in the power traces obtained from the leaky device.

While there has been several works in SCA that have attempted to eliminate the information redundancy in power traces, these techniques only process power traces after leakage sampling to reduce the efforts of side-channel analysis. Contrary to our work, they do not lower the requirements on the sampling equipments. In the following subsection, we will discuss these existing works along with CS before describing the main contributions of our work.

1.1 Related Works

Points-Of-Interest (POI) selection in [DS16] and dimensionality reductions in [CDP15, SNG+10, BHvW12] are two classical pre-processing techniques in SCA. The former finds the locations of POIs by using side-channel distinguishers such as Differential Power Analysis (DPA) [KJJ99] and Correlation Power Analysis (CPA) [BCO04], or leakage detection tools such as Welch's t-test [DCE16], ρ-test [DS16] and χ^2-test [MRSS18], while the latter such as PCA [SNG+10] and LDA [SA08], considers the global features of high-dimensional samples. Both techniques need to consider multiple power traces simultaneously. Discrete Wavelet Transform(DWT), Discrete Cosine Transform(DCT) and Fast Fourier Transform (FFT) [GHT05], transform power traces from time domain to sparse domain one at a time, but all the computations are carried out on the sampling equipments, which increases its workload and reduces the sampling rate.

To the best of our knowledge, Maximum Extraction and Integration in [MOP07] are two existing compressive sampling techniques that are used to eliminate information redundancy on power traces. The former maintains that the highest correlation occurs exactly at the position where the power consumption of each clock cycle reaches its maximum, and it is therefore reasonable to choose these points as representative points for entire clock cycles. The latter integrates points in each clock cycle or within small time intervals. They are all applicable to situations where the sampling rate of the oscilloscope is much higher than the bandwidth of the leaky device. Re-sampling is then performed on the collected power traces. As such, these techniques incur resource wastage since the high sampling rate provides observers with leakage containing a large amount of redundancy, which is discarded during compression. This led to the problem stated in [Don06]: "*why go to so much effort to acquire all the data when most of what we get will be thrown away? Can't we just directly measure the part that won't end up being thrown away?*". The rest of this paper will address this problem for SCA.

Based on the theory of functional analysis and approximation [Kas91], Romberg, Tao and Donoho established the theory of Compressive Sensing (CS) [CR06, CRT06, Don06]. Combined with information theory, CS makes it possible to sample signals at a rate far below the Nyquist sampling theorem while enabling equivalent sampling performance. CS has been widely studied and applied to many fields such as image, voice and signals in general. The sampling rate of CS no longer depends on the highest frequency of the signals, but instead it is governed by the signal's sparsity and Restricted Isometry Property (RIP) [CRT06]. As long as a signal is compressible or sparse in a certain transform domain, it can be projected from high-dimensional space into a low-dimensional space and retain the important information. Then, by solving an optimization problem, this signal can be reconstructed from a small number of projections with a high probability. In the context of SCA, sparsity provides a more intuitive and efficient representation of information in the power traces.

The research in CS mainly includes three aspects: (1) sparse signal representation, (2) observation matrices satisfying the incoherence and RIP properties, and (3) fast and robust signal reconstruction algorithms. Sparse signal representation, a precondition of CS,

signifies that the low-dimensional transformed vector is sparse or approximately sparse when a high-dimensional leakage signal is projected onto a certain orthogonal transformation basis. The aim is to find the features of signals in different functional spaces such as FFT, DWT, DCT and Bandelet [PM00], and provide a more concise representation. This is different from directly transforming the power traces into frequency domain, since the observer only needs to compute the inner product of the signal and the observation matrix, which requires very low computation. The purpose of the observation matrices such as Gauss, Bernoulli and Toeplitz [SYZ08] is to find a projection matrix that is not related to the sparse matrix, and they reflect the observation rules for leakage signals.

Fast and robust signal reconstruction algorithms are the most widely researched among the three aspects of CS. The algorithms can be divided into three categories according to the problem addressed: greedy algorithms for solving l_0-norm, convex optimization algorithms for solving l_1-norm and the combination of these two kinds of algorithms. Matching pursuit algorithms are typical examples of greedy algorithms, which aim to select one or more dimensions with the highest correlation with the current residual vector in each iteration, and approximate the original leakage signal and the new iteration error based on the current selected dimensions. Classical greedy algorithms include Matching Pursuits (MP) [MZ93], Orthogonal Matching Pursuit (OMP) [SM15], CoSaMP[NT10], GOMP [WKS12], TMP (Tree Matching Pursuit) [LD06] and ROMP [NV09], etc. Convex relaxation algorithms convert the non-convex optimization problem l_0-norm to the convex optimization l_1-norm. The representative algorithms include BP (Basis Pursuit) [CDS01], GP (Gradient Projections) [HTWM10], IHT (Iterative Hard Threshold) [Mal09], etc. The convex optimization algorithms are more accurate than the greedy algorithms, but their computational complexity is higher. The combined or hybrid algorithms mainly consist of CP (Chaining Pursuit) [GSTV06]. It is worth mentioning that even though there are many existing reconstruction algorithms, there are still a lot of problems in their convergence and robustness. Nevertheless, the existing theories and prior work are sufficient for CS to be applied in side-channel leakage sampling.

1.2 Our Contributions

The increasing operating frequency of cryptographic devices impose high requirements on the sampling equipments, bring new challenges to the storage and processing of power traces for SCA. In this paper, we focus on addressing these challenges which pertain to leakage sampling for SCA. While this problem has not been adequately discussed in the literature, we believe it is of high importance as we are able to show that the high operating frequency of advanced cryptographic devices is not a barrier to leakage sampling in the absence of powerful sampling equipments.

Specifically, we introduce a novel use of Compressive Sensing (CS), a new and highly-efficient compressive sampling technology for side-channel leakage sampling. CS performs compression and sampling simultaneously by projecting the high-dimensional signals onto a low-dimensional space to obtain the discrete leakage samples. This enables the high-dimensional signal to be reconstructed without distortion. Compared with classical sampling, CS uses a far lower sampling rate to achieve the same performance. The feasibility of our approach is verified by theory and experiments in this paper.

In addition, the practical deployment of the proposed approach is also highly feasible as the sampling process of CS is very simple. It only needs to compute the incoherent projection without any additional processing. A large amount of computations which are needed for reconstruction of the signals are transferred from the sampling equipments to the desktop computers and workstations, which reduces the workload of the sampling equipments. Many current sampling tasks in side-channel analysis are accomplished automatically by expensive oscilloscopes. The CS programs can be integrated into sampling plug-ins of these oscilloscopes. Oscilloscopes such as our *Tektronix DPO 7254* with a

Windows 7 operation system, even allow the development of independent sampling software. Researchers have also developed the sampling software on various platforms. For example, Inspector developed by Riscure, enables the leakage to be collected and stored on laptop computers and other advanced processors. CS can also be efficiently implemented on these platforms by integrating a small program to solve the inner product of observation matrix (see Section 4) and leakage sample vector.

1.3 Organization

The rest of this paper is organized as follows. Principles of classical compressive sampling and CS are introduced in Section 2. The first two main parts of CS, leakage sparse representation and observation, are described in Sections 3∼4. Section 5 uses classical greedy algorithms such as OMP, CoSaMP, SP and GOMP as examples to introduce the principle of signal reconstruction. The corresponding reconstruction performance evaluation criteria are also provided in this section. Observers can find good sparse domain and observation matrix through experiments, optimize the sparse coefficients and compression ratio, and implement CS with a much lower sampling rate than those required by existing sampling equipments or sampling software on computers. In Sections 6 and 7, we describe the experiments on an *AT89S52* micro-controller and measurements from *DPA contest v1.1* [dpa] to demonstrate the efficiency of the approach. Finally, Section 8 concludes this paper.

2 Preliminaries

2.1 Classical Compressive Sampling

Sampling and compression are carried out separately in traditional compressive sampling. The sampling rate must satisfy Nyquist theorem, which states that the bandwidth of sampling equipment must be at least twice the maximum frequency of the cryptographic device. Moreover, the sampling is uniform, and the number of samples collected in any fixed interval is the same. It is noteworthy that the sampling rate is usually much higher than the bandwidth of cryptographic devices in practice. As such, the power traces contain a large amount of information redundancy, which can be removed during compression to lower the complexity of side-channel attacks and evaluations. Mangard et al. stated that the peaks of the damped oscillation carried all the information that was necessary to perform a power analysis attack. Two classic compression techniques named Maximum Extraction and Integration were given [MOP07]. The reconstruction of the original power traces from compressed ones is usually not considered in SCA since attacks can be directly performed on the compressed low-dimensional power traces.

Dimensionality reduction methods such as PCA [SNG+10, BHvW12], LDA [SA08], KDA [CDP16] and manifold learning [OSW+17], obtain the low-dimensional samples by extracting the features from high-dimensional leakages, which consider the structure of all power traces. Compression only needs to consider a single trace. The low-dimensional samples obtained from different power trace sets may vary significantly after dimensionality reduction. However, the compression of one power trace is not affected by other power traces. This indicates that the above mentioned dimensionality reduction algorithms commonly used in SCA cannot be used in CS. We further illustrated why FFT, DWT and DCT cannot be directly used for sampling in Section 1.1. It is noteworthy that most of the sample points on the power traces do not contain sensitive information and there is still a lot of information redundancy in the two compression methods, Maximum Extraction and Integration [MOP07]. Moreover, they lose information and the original power traces cannot be accurately reconstructed after compression. It can only be re-sampled if

necessary, which requires significant effort.

2.2 Compressive Sensing

The principle of CS maintains that as long as the signal is sparse or sparse in a transform domain, and its projection vectors can be obtained from observation matrix, then it can be reconstructed nondestructively by optimization methods. Unlike the classical compressive sampling, wherein the sampling and compression are performed separately, CS compresses a power trace while sampling it and aims to use the least coefficients to represent the compressed signal. Compared with traditional compressive sampling, the sampling rate in CS theory no longer depends on the bandwidth of signal, but on its sparsity (see Table 1). The sparser the signal in a transformation domain, the fewer non-zero coefficients it possesses. These coefficients represent the smallest sampling bandwidth needed to reconstruct the original high-dimensional signals. As such, to achieve the same sampling performance, CS only requires a much lower sampling rate compared to traditional compressive sampling.

Table 1: Difference between classic compressive sampling and CS.

	classic compressive sampling	Compressive Sensing
Proposed in	1948	2006
Sampling mode	uniform sampling	non-uniform sampling
Sampling rate	highest frequency of leakage	sparsity
Observation	high-dimensional original samples	low dimensional observation vector
Reconstruction	*Sinc* interpolation	Solving optimization problems

Moreover, classical compression samples uniformly, while CS does not. The low-dimensional samples are obtained by computing the inner product between the signal and the observation matrix. Since each measured value takes the same information, this makes the measurement robust. Finally, Nyquist sampling theorem uses *Sinc* interpolation (i.e. Whittaker-Shannon interpolation) to reconstruct the original leakage, which is a linear operation with limited computation. CS uses nonlinear programming methods to recover leaky signals, and the corresponding complexity is high. Fortunately, most of the computation is eventually transferred from the sampling devices to computers, which notably reduces the workload of the sampling devices.

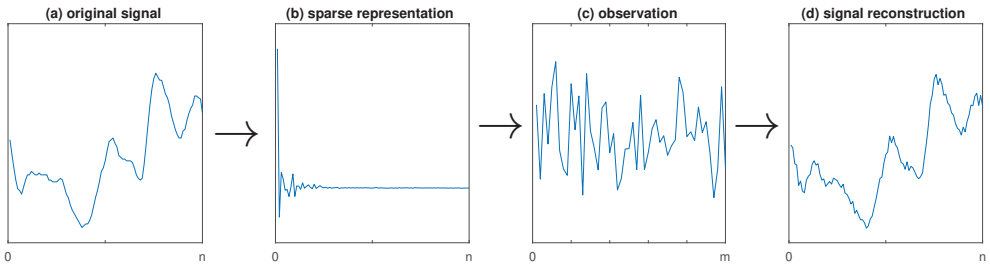

Figure 1: General procedure of CS.

The complete CS flow is shown in Fig. 1. Firstly, the m sparse coefficients of the n-dimensional original signals are obtained by sparse transformation ($m \ll n$ in Fig. 1(b)). m largest coefficients are saved and other $n - m$ coefficients are discarded. The saved coefficients are sufficient to reconstruct the original signal without distortion. Secondly, the observer encodes these saved coefficients, achieves low-dimensional observation vectors and completes the compression (Fig. 1(c)). In other words, the high-dimensional

original signal is projected onto a sparse domain to achieve low-dimensional samples. The sampling device outputs these samples as sampling results. Finally, the original signal is reconstructed by solving an optimization algorithm (Fig. 1(d)). These operations correspond to the three aspects of CS theory introduced in Section 1.1: signal sparse representation, observation matrix design, and signal reconstruction, which will be discussed in detail in Sections III∼V.

3 Leakage Sparse Representation

3.1 Sparse Decomposition

The purpose of sparse decomposition in SCA is to use as few sparse vectors as possible to represent the original leakage. If a leakage signal $\boldsymbol{x} = [x_0, x_1, \ldots, x_{n-1}] \in R^{n\times 1}$ can be represented by linear combinations of normal orthogonal bases $\Psi = [\psi_0, \psi_1, \ldots, \psi_{n-1}] \in R^{n\times n}$, then $\boldsymbol{x}$ can be represented as:

$$\boldsymbol{x} = \Psi\Theta = \sum_{i=0}^{n-1} \psi_i \theta_i. \tag{1}$$

If the number of non-zero coefficients k in Θ is much smaller than n, then $\boldsymbol{x}$ is sparse or compressible on the orthogonal basis Ψ. In other words, Θ is sparse. $\Theta = [\theta_0, \theta_1, \ldots, \theta_{n-1}]^T \in R^{n\times 1}$ is the sparse coefficients, and also the sparse representation of $\boldsymbol{x}$ on Ψ. Ψ here is the sparse base or sparse domain.

3.2 Determination Conditions

The precondition of CS is that the leakage signal is k-sparse, but most of the natural signals are not sparse. If the absolute values of the coefficients in Θ exponentially attenuate after being sorted in descending order, the observer obtains $\left|\theta_0^{'}\right| \geq \left|\theta_1^{'}\right| \geq \cdots \geq \left|\theta_{n-1}^{'}\right|$. For any $p \in (0, +\infty)$, if there exists

$$\left|\theta_i^{'}\right| \leq \frac{R}{\sqrt[p]{n}}, \tag{2}$$

then the signal is compressible or nearly sparse. Here $i \in (0, n-1)$. Parameter p controls the speed of exponential attenuation. The larger it is, the faster the attenuation, and the sparser the signal $\boldsymbol{x}$ is on Ψ. Therefore, the criterion to evaluate the performance of a set of sparse bases is to measure the attenuation tendency of the sparse coefficients. However, many signals are not strictly sparse but approximately sparse, which can still be decomposed sparsely.

Candes and Tao indicated in [CT06] that the signals with Θ exponential attenuation could be reconstructed by CS theory, and the upper bound of reconstruction error satisfies:

$$e_o = \|\boldsymbol{x} - \hat{\boldsymbol{x}}\|_2 \leq C_p \cdot R \cdot \left(\frac{k}{\log n}\right)^{-r}. \tag{3}$$

Here $r = 1/p - 1/2$, $0 < p < 1$, $R > 0$, C_p is constant which only depends on p and $\|\cdot\|_2$ is the l_2-norm.

3.3 Sparse Domains

Side-channel leakages are not sparse in time domain, but they are sparse when converted to frequency domain, wavelet domain etc. Classical transforms, such as FFT, DWT and DCT, were first applied to CS. FFT and DCT belong to global transformation and no

longer preserve features of time domain. DCT focuses on the low frequencies of power traces as critical information. DWT can better analyze the time-domain characteristics of signals and more sparsely represent the information of signals. If a suitable base is selected, the coefficients are easier to deal with than the original leakage signal.

We only consider DCT in this paper. There are 8 transform forms for one-dimensional DCT, of which the second one is the most commonly used. Let n denote the number of samples on original leakage signal, then the u-th coefficient after transformation is

$$F(u) = c(u) \sum_{i=0}^{n-1} x(i) \cos\left\{ u \frac{(2i+1)}{2n} \right\}. \tag{4}$$

$x(i)$ is the i-th time point of original leakage signal and $c(u)$ is a coefficient satisfying

$$c(u) = \begin{cases} \sqrt{\frac{1}{n}}, & u = 0 \\ \sqrt{\frac{2}{n}}, & u = 1, 2, \cdots, n-1 \end{cases} \tag{5}$$

It can be regarded as a compensation coefficient, which makes the DCT transformation matrix an orthogonal matrix. For side-channel leakage, $c(0)$ of $F(0)$ is the DC (Direct-Current) component and other coefficients are AC (Alternating Current) components. The complexity of DCT is $\mathcal{O}\left(n^2\right)$. In fact, the target of reconstruction algorithms is to reconstruct the DCT coefficients. Finally, the Inverse Discrete Cosine Transform (IDCT):

$$x(i) = \sqrt{\frac{2}{n}} \sum_{u=0}^{n-1} c(u) F(u) \cos\left\{ u \frac{(2i+1)\pi}{2n} \right\} \tag{6}$$

is performed to reconstruct the original leakage.

4 Observation

The sampling process of CS is very simple, and most of the computations lie mainly in the leakage signal reconstruction. This reduces the workload of sampling devices and facilitates the fast sampling of CS. If the leakage signal $\boldsymbol{x} \in R^{n \times 1}$ is sparse, it can be projected onto the observation matrix $\Phi = [\phi_0, \phi_1, \ldots, \phi_{m-1}] \in R^{m \times n}$:

$$\boldsymbol{y} = \Phi \boldsymbol{x}, \tag{7}$$

and obtains the low-dimensional observation vector (as shown in Fig. 2). Otherwise, the observer must project it onto orthogonal bases to make it sparse. The observation matrix Φ is independent of the sparse basis Ψ. Here $\boldsymbol{y} = [y_0, y_1, \ldots, y_{m-1}] \in R^{m \times 1}$ is the observation vector (i.e. sampling results of sampling equipment). $\boldsymbol{y} = \Lambda\Theta$ ($\Lambda = \Phi\Psi = [\gamma_0, \gamma_1, \ldots, \gamma_{n-1}]$) is defined as the sensing matrix. The principle of $\boldsymbol{y} = \Phi\boldsymbol{x}$ and $\boldsymbol{y} = \Lambda\Theta$ is similar, since we can make

$$\Theta = \Psi^T \boldsymbol{x} \tag{8}$$

and get $\Phi^{'} = \Phi\Psi^T$. The observation $\boldsymbol{y} = \Phi^{'}\boldsymbol{x}$. Matrices Ψ and Φ can be employed universally in a cryptographic implementation, and hence we only need to set them once.

Since the size m of $\boldsymbol{y}$ is much smaller than the size n of $\boldsymbol{x}$, $\boldsymbol{y} = \Lambda\Theta$ is an under-determined system of equation. This is equivalent to $\boldsymbol{x}$ being compressed, and the amount of data compressed is much lesser than the original leakage obtained by Nyquist sampling theorem. If m and n are very large, the dimensions of matrices Ψ and Φ will be very high, which need to be optimized.

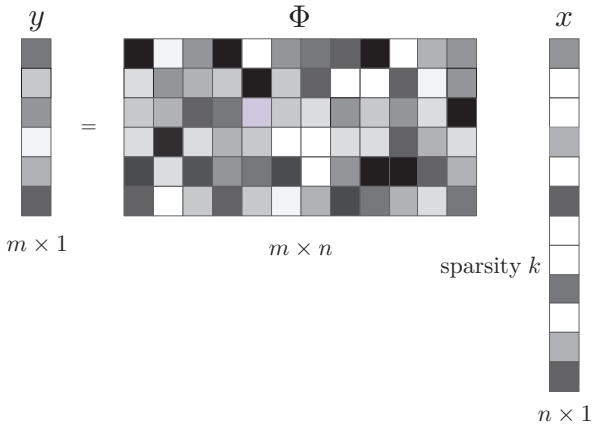

Figure 2: Low-dimensional observation from high-dimensional leakage signal x.

4.1 Constraint Conditions

Observers need to construct an observation matrix in CS theory, which plays an important role in the acquisition of observation vectors and signal reconstruction. It requires that the k measured values do not destroy the information of the original leakage signal when it is converted from $\boldsymbol{x}$ to $\boldsymbol{y}$, thus ensuring accurate signal reconstruction. This should satisfy Restricted Isometry Property (RIP) and Incoherence Property [CW08, DH01].

Restricted Isometry Property: A matrix Ψ satisfies RIP of order k (i.e. the leakage signal is k-sparse) if there exists a $\delta \in (0,1)$ that makes Λ satisfy the following inequality:

$$1-\delta \leq \frac{\|\Lambda\Theta\|_2}{\|\Theta\|_2} \leq 1+\delta. \tag{9}$$

This ensures that signals can be transformed from one domain to another without divergence. δ is the Restricted Isometry Constant (RIC) [CRT06], which is the minimum value satisfying Eq. 9.

Incoherence property: The coherence within the observation matrix Φ is the maximum absolute value of the normalized inner products, i.e.

$$\mu(\Phi) = \max_{1\leq i,j\leq n, i\neq j} \frac{|\langle \phi_i, \phi_j \rangle|}{\|\phi_i\|_2 \cdot \|\phi_j\|_2}. \tag{10}$$

Here $\langle \cdot, \cdot \rangle$ denotes dot product of two vectors. The smaller the μ, the weaker the coherence of any two atoms (i.e. columns) in Φ. The coherence coefficients of orthogonal matrices are 0.

The coherence between the observation matrix Φ and sparse transformation base Ψ can refer to the coherence between two atoms of a single matrix. Incoherence here means that the row vectors ϕ_i of Φ cannot be represented linearly by the column vectors in Ψ. The coherence of Φ and Ψ can be defined as:

$$\mu(\Phi,\Psi) = \sqrt{n} \cdot \max_{1\leq i,j\leq n} \{|\langle \phi_i, \psi_j \rangle|\} \in \left[0, \sqrt{n}\right]. \tag{11}$$

If there are coherent columns in Φ and Ψ, $\mu(\Phi,\Psi)$ is large. The smaller the $\mu(\Phi,\Psi)$ is, the more information of the original signal x is contained in the measurement samples, which leads to better reconstruction performance.

4.2 Observation Matrices

It is difficult to directly design an observation matrix satisfying the RIP condition. Most of the current observation matrices are random Gaussian matrices, wherein each element satisfies the normal distribution with mean 0 and variance $\frac{1}{m}$:

$$\phi(i,j) \sim \mathcal{N}\left(0, \frac{1}{m}\right). \tag{12}$$

Since random matrices are not related to any matrix, they can satisfy RIP conditions with high probability if $m \geq c \cdot k \cdot \log_2\left(\frac{n}{k}\right)$ [Don06]. Here c is a small constant. Another observation matrix is random Bernoulli matrix, in which elements follow Bernoulli distribution:

$$\phi(i,j) = \frac{1}{\sqrt{m}} \begin{cases} 1, & p_r = \frac{1}{2} \\ -1, & p_r = \frac{1}{2} \end{cases} \tag{13}$$

or

$$\phi(i,j) = \sqrt{\frac{3}{m}} \begin{cases} 1, & p_r = \frac{1}{6} \\ 0, & p_r = \frac{2}{3} \\ -1, & p_r = \frac{1}{6} \end{cases} \tag{14}$$

Here p_r denotes the probability of the element. Baraniuk et al. in [RMRM08] proved that Bernoulli matrix was strongly random, it could satisfy RIP condition with great probability if the number of observations $m \geq c \cdot k \cdot \log_2\left(\frac{n}{k}\right)$. Besides Gauss and Bernoulli, a large number of observation matrices have also been proposed, which will not be discussed in this paper.

5 Leakage Reconstruction

Signal reconstruction algorithms in CS aim to use the low-dimensional observation vector $\boldsymbol{y}$ to recover the high-dimensional original leakage signal $\boldsymbol{x}$. As proved in [GN03, Tro04], if the sparse coefficients in Θ satisfy

$$\|\Theta\|_{l_0} < \frac{1}{2}\left\{1 + \frac{1}{\mu\{\Phi\}}\right\}, \tag{15}$$

the low-dimensional observation is achieved by solving the optimization problem:

$$\arg\min \|\Theta\|_{l_0}, s.t. \, \boldsymbol{y} = \Lambda\Theta. \tag{16}$$

$\|\Theta\|_{l_0}$ is the number of non-zero elements in Θ. In this case, Θ has a unique solution, which is equivalent to that the minimum number of linear correlation atoms in the given matrix Φ is greater than $2k$. Thus, the sparse coefficient Θ is obtained and the side-channel leakage $\boldsymbol{x}$ is recovered by Eq. 1. The greedy algorithms aim to solve the l_0-norm. Since $m \ll n$, $\boldsymbol{y} = \Lambda\Theta$ has multiple solutions, this is NP-hard. If a reconstruction error ϵ is allowed, this model becomes

$$\arg\min \|\Theta\|_{l_0}, s.t. \, \|\boldsymbol{y} - \Lambda\Theta\| < \epsilon. \tag{17}$$

However, the new model is unstable and difficult to solve directly.

5.1 Optimization

Since Θ is sparse, the number of unknowns in $\boldsymbol{y} = \Lambda\Theta$ is greatly reduced, which makes signal reconstruction possible. The problem described in Eq. 17 can be solved by the suboptimal solution of l_0-norm:

$$\arg\min \|\Theta\|_{l_1}, s.t. \boldsymbol{y} = \Lambda\Theta \tag{18}$$

if θ satisfies certain conditions [CDS01]. The typical solutions of l_1-norm are convex optimization algorithms. The l_p-norm of vector $\boldsymbol{x} = [x_0, x_1, \ldots, x_{n-1}]^T$ is defined as

$$\|\boldsymbol{x}\|_{l_p} = \left(\sum_{i=0}^{n-1} |x_i|^p\right)^{\frac{1}{p}}. \tag{19}$$

Taking two-dimensional real space R^2 as an example, l_1-norm $\|\boldsymbol{x}\|_{l_1} = |x_1| + |x_2|$ represents the closed area surrounded by four lines shown in Fig. 3(a), and l_2-norm $\|\boldsymbol{x}\|_{l_2} = \sqrt{|x_1|^2 + |x_2|^2}$ represents a circle (as shown in Fig. 3(b)). They are hyper-prism and hyper-sphere in high-dimensional space R^n. Only if the optimal solution of $\boldsymbol{y} = \Phi\boldsymbol{x}$ falls onto the coordinate axis can the sparsity be guaranteed. In fact, if l_1-norm and l_2-norm fall onto the coordinate axis, their solutions are equivalent to the ones of l_0-norm.

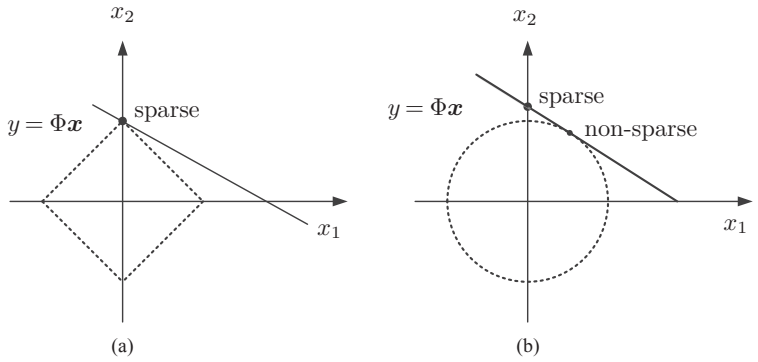

Figure 3: The geometric meaning of the optimal solutions of l_1-norm (a) and l_2-norm (b) in two-dimensional space.

From the perspective of information theory, the information cannot be infinite compressed. The number of observations (i.e. the number of measurements) m determines the compression ratio $\frac{m}{n}$. For a leakage signal with length n and sparsity k on a sparse base, the lower bound of the number of observations m required by CS is:

$$m \geq c \cdot \mu^2(\Phi, \Psi) \cdot k \cdot \log_2 n. \tag{20}$$

It is noteworthy that the above solutions all require the observer to solve constraint problems, which can be transformed into solving unconstrained problem [CT06]:

$$\arg\min \|\Theta\|_{l_1} + \eta \|\boldsymbol{y} - \Lambda\Theta\|_2. \tag{21}$$

η is a balance factor, which is used to balance reconstruction error and sparsity.

5.2 Greedy Algorithms

Signal reconstruction algorithms play a very important role in accurately reconstructing the high-dimensional original leakage signal from the low-dimensional observation vector

in CS. Greedy algorithms are the earliest and the most widely used signal reconstruction algorithms in CS. The main idea of these algorithms is to select one or more atoms (i.e. columns) having the greatest correlation with the current residual vector r in each iteration, and obtain the current optimal solution to approximate the original leakage signal and the new iteration error according to the currently selected atoms. Matching Pursuits (MP) [MZ93] and Orthogonal Matching Pursuit (OMP) (improved from MP) [SM15] are two widely used greedy algorithms in CS. MP selects the atom γ_{t-1} with highest matching degree between matrix Λ and current signal residual r_{t-1}:

$$\gamma_{t-1} = \arg\max_{\gamma_j} |\langle r_{t-1}, \gamma_j \rangle| . \tag{22}$$

Here $t-1$ is the current number of repetitions (i.e. the current number of observations). The residual is then decomposed as

$$r_{t-1} = \max |\langle r_{t-1}, \gamma_j \rangle| + r_t \tag{23}$$

after each iteration. Almost all of the matching pursuit algorithms such as CoSaMP[NT10], GOMP [WKS12], StOMP[DTDS12] and ROMP [NV09], are improved from OMP. These algorithms preserve the atom selection strategy of MP.

OMP [SM15] is the most commonly used algorithm in CS (see Algorithm 1). The atom in the sensing matrix Λ having greatest correlation with the current residual r_{t-1} is selected as a new candidate atom (Step 3). It is added to the atom matrix Λ_t and its corresponding index λ_t is added to the support set A (Step 4). For a k-sparse leakage signal $\boldsymbol{x}$, only k non-zero coefficients are involved in the operation when the sensing matrix Λ is used. These atoms are stored in matrix Λ according to the observation rules used during the iteration. OMP is then updated by subtracting its projection on the orthogonal space of the selected atom matrix from the observation vector $\boldsymbol{y}$ until the iteration $t \leq k$ is satisfied (Steps 5 and 6). Since the residual r is orthogonal to the selected atoms, an atom in Λ will not be selected twice, thus guaranteeing the convergence of the algorithm. Moreover, orthogonalization guarantees the local optimal solution in each iteration, but does not guarantees that the sum of the local optimal solutions is the global optimal solution.

Algorithm 1: Orthogonal Matching Pursuit (OMP).

Input: sensing matrix $\Lambda = \Phi\Psi$, sparse base Ψ, support matrix A, observation $\boldsymbol{y}$ and sparsity k.

Output: estimated parameters $\hat{\Theta}$ and residual r.

1 Initialization: $r_0 = \boldsymbol{y}$, $A_0 = \emptyset$, $\Lambda_0 = \emptyset$ and the number of iteration $t = 1$;
2 **while** $t \leq k$ **do**
3 $\quad \lambda_t = \arg\max_j |\langle r_{t-1}, \gamma_j \rangle|$;
4 $\quad$ find index $A_t = A_{t-1} \cup \{\lambda_t\}, \Lambda_t = \Lambda_{t-1} \cup \{\gamma_{\lambda_t}\}$;
5 $\quad \hat{\Theta}_t = \arg\min_{\hat{\Theta}_t} \left\| \boldsymbol{y} - \Lambda_t \hat{\Theta}_t \right\|_2$;
6 $\quad$ update residual $r_t = \boldsymbol{y} - \Lambda_t \hat{\Theta}_t$;
7 $\quad t = t + 1$;
8 **end**
9 recover signal $\hat{\boldsymbol{x}} = \Psi\hat{\Theta}$;

OMP uses least square method to solve the minimum $\hat{\Theta}_t$ (see Step 5 in Algorithm 1). Since $\boldsymbol{y} = \Lambda\Theta$, solving $f(\Theta) = \|\boldsymbol{y} - \Lambda_t\Theta_t\|_2$ is equivalent to solving:

$$f(\Theta) = (\boldsymbol{y} - \Lambda_t\Theta_t)^T (\boldsymbol{y} - \Lambda_t\Theta_t) . \tag{24}$$

The function $f(\Theta)$ has an extreme value at $\frac{\partial f(\Theta)}{\partial \Theta} = 0$, i.e. $\frac{\partial f(\Theta)}{\partial \Theta} = -2\Lambda_t^T(\boldsymbol{y} - \Lambda_t\Theta_t)$. We get $\Lambda_t^T\boldsymbol{y} = \Lambda_t^T\Lambda_t\Theta_t$, Step 5 can be simplified by solving

$$\Theta_t = \left(\Lambda_t^T\Lambda_t\right)^{-1}\Lambda_t^T\boldsymbol{y}. \tag{25}$$

Here Λ_t^T is the transformation matrix of Λ_t. The complexity of OMP is $\mathcal{O}(k \cdot m \cdot n)$. The advantage of OMP is its fast convergence and high accuracy under the precondition that k is known. However, if the sparsity k is unknown and the estimation is too small, then the absolute error e_o is too large, OMP falls into endless iterations or results in very low percentage recovered. Moreover, it will incur a large amount of computation and lower the quality of reconstructed signals if the sparsity k is estimated to be too large.

As we mentioned earlier, almost all of matching pursuit algorithms are improved from OMP. An obvious disadvantage of OMP algorithm is that only one atom is selected in each iteration. With the number of observations increase, the runtime increases rapidly. This can be solved by selecting multiple atoms from the observation matrix Φ or sensing matrix Λ each time. In this case, SWOMP (Stagewise Weak OMP) [BD09] which is improved from StOMP (Stagewise OMP) [DTDS12], selects all the atoms larger than a preset threshold for subsequent calculations. The difference is that the threshold of StOMP comes from residuals, while the threshold of SWOMP comes from experiential setting. In general, the threshold of SWOMP is the maximum coefficient multiplied by a parameter between 0 and 1, with a default value 0.5. The complexity of StOMP and SWOMP is $\mathcal{O}(m \cdot n)$. GOMP [WKS12] requires to set a parameter s, which denotes the number of atoms selected in each repetition (the default value is $\frac{k}{4}$). In order to reduce parameter settings, we chose GOMP to represent these improved algorithms in our experiments.

Another disadvantage of OMP algorithm is that once an atom is in the candidate set, it will never be deleted. StOMP and SWOMP also have this drawback. To improve this shortcoming, two algorithms CoSaMP (Compressive Sampling Matching Pursuit) [NT10, NV10] and SP (Subspace Pursuit) [DM09], which rely on backtracking is employed. Specifically, as the algorithm iterates, the atoms in Λ are recalculated and the non-optimal atoms are deleted. Specifically, CoSaMP [NT10, NV10] selects $2 \cdot k$ atoms most relevant to the residual in Steps 3 and 4 of Algorithm 1, and the k atoms with the largest absolute values in $\hat{\Theta}$ are selected for the next iteration in Step 5. It guarantees that there will be no more than $3 \cdot k$ atoms in Λ, $2 \cdot k$ atoms in A and at most k atoms are removed in each repetition. Compared with CoSaMP, SP [DM09] only selects k atoms most relevant to the residual in Steps 3 and 4, it guarantees that there should be no more than $2 \cdot k$ atoms in Λ and $2 \cdot k$ atoms in support vector A, and at most k atoms are removed in each repetition. The complexity of CoSaMP and SP is $\mathcal{O}(m \cdot n)$ and $\mathcal{O}(\log(k) \cdot m \cdot n)$.

5.3 Performance Criteria

There are many criteria to evaluate the performance of signal reconstruction algorithms, such as reconstruction time, reconstruction residual (also called absolute error, see Eq. 3), relative error and signal-to-noise ratio (SNR). They reflect the reconstruction performance of the algorithm from different aspects.

Relative Error: Referring to the absolute error $e_o = \|\boldsymbol{x} - \hat{\boldsymbol{x}}\|_2$ between the original leakage signal $\boldsymbol{x}$ and reconstructed signal $\hat{\boldsymbol{x}}$ given in Eq. 3, the relative error is defined as:

$$e_r = \frac{\|\boldsymbol{x} - \hat{\boldsymbol{x}}\|_2}{\|\hat{\boldsymbol{x}}\|_2}. \tag{26}$$

Signal-to-Noise Ratio: SNR is defined as the ratio of exploitable power consump-

tion component to noise component in side-channel attacks [MOP07]. It is defined as:

$$\mathrm{SNR} = 20 \times \lg \left\{ \frac{\|\boldsymbol{x}\|_2}{\|\boldsymbol{x} - \hat{\boldsymbol{x}}\|_2} \right\} \tag{27}$$

in CS, of which the molecule represents the variance of the signal. The denominator represents the absolute error of the original signal and the reconstructed signal.

Matching Degree: The matching degree α of the original signal and the recovered signal is defined as

$$\alpha = 1 - \frac{|\|\hat{\boldsymbol{x}}\|_2 - \|\boldsymbol{x}\|_2|}{|\|\hat{\boldsymbol{x}}\|_2 + \|\boldsymbol{x}\|_2|}. \tag{28}$$

It is a positive number with a value between 0 and 1. The smaller the reconstruction error $e_0 = \|\boldsymbol{x} - \hat{\boldsymbol{x}}\|_2$, the greater the matching degree, the closer to 1 the α is, and the better the reconstruction performance.

Percentage Recovered: The percentage recovered P_r in [TG07] depends on the absolute error $e_o = \|\boldsymbol{x} - \hat{\boldsymbol{x}}\|_2$ of reconstruction. If it is smaller than the pre-determined threshold, then the reconstruction is successful. Therefore, its calculation is similar to the Success Rate (SR) [SMY09] in side-channel analysis, and its value is the ratio of the number of successful reconstructions to the total number of repetitions in experiments. The setting of threshold e_o can be determined according to the required reconstruction accuracy. If the reconstructed power traces are only used for attacks, it can be set appropriately large. However, leakage evaluations require high reconstruction accuracy, so it should be set small.

6 Experiment Results On AT89S52 Micro-controller

6.1 Experimental Setups

Our first experiment is performed on an *AT89S52* micro-controller, with a clock operating frequency of 12 MHz. The shortest instructions take 12 clock cycles to execute. We use a *Tektronix DPO 7254* oscilloscope to capture leakage of the look-up table instruction "*MOVC A,@A+DPTR*", which takes 24 clock cycles. The oscilloscope has a sampling rate of up to 40 GHz, but it does not have the function to automatically collect and store waveform. We obtain the waveform acquisition plug-in from the *Tektronix* company, but the speed is very slow. We cannot even store the leakage samples of the AES encryption algorithm promptly under 500 MHz sampling rate. So, we add about 0.5 second of empty-loop instructions before look-up table operation. Finally, we acquire 20000 power traces, each of them includes 5000 samples. The rest of our experiments are performed on a *HP* desktop computer with 6 Inter(R) Xeon(R) E5-1650 v2 CPUs, 16 GB RAM and a Windows 10 operating system. It's clock frequency is 3.5 GHz. Since the power traces are affected by noise, they fluctuate significantly. We use a moving average filter with a 5-hour span to remove the noise. A random observation matrix for four algorithms is generated for each repetition.

In order to observe and compare the performance of the algorithms OMP, CoSaMP, BP and GOMP, we use the $1801^{th} \sim 2600^{th}$ samples to perform our CS using MATLAB R2016b. The time samples in this segment contain obvious leakage. The DCT coefficients of a power trace transformed from time domain to DCT domain are shown in Fig. 4. We can draw the conclusion that the leakage of *AT89S52* micro-controller is sparse in DCT domain, as most DCT coefficients are close to 0. FFT and DWT can also be used as sparse domains, although the corresponding experimental results are not given.

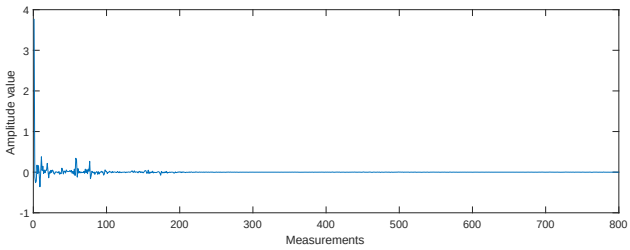

Figure 4: DCT coefficients of a power trace leaks from *AT89S52* micro-controller.

6.2 Threshold Selection

The original power trace $\boldsymbol{x}$ can be projected onto the observation matrix Φ if it satisfies RIP, and reconstructed by using at least $m \geq k \cdot \log_2\left(\frac{n}{k}\right)$ observations. The minimum number of observations m can be quickly obtained if the sparsity k is known. Otherwise, we need to test it. In order to optimize the reconstruction performance, it is necessary to adjust m appropriately. We use OMP, CoSaMP, SP and GOMP to perform our experiments and set m to 400 to test the sparsity, which ranges from 5 to 200, with step width 5. SNR, relative error, matching degree α and their runtime under different sparsity are shown in Fig. 5. The reconstruction performance is compared considering only one power trace. The SNR of OMP, CoSaMP and SP increases rapidly and reaches the highest at $k < 80$. SP and GOMP fluctuate significantly at $k > 100$. GOMP is the highest when $k < 75$, which indicates that its power trace reconstruction requires the smallest number of observations. SNR and matching degree of CoSaMP decrease rapidly when $k > 125$, the relative error of it is also much larger than other algorithms. If m is set to 320, the SNR of 4 algorithms reaches the highest at $k < 45$, and SP fluctuates significantly at $k > 60$. This also indicates that we may get very different results under different observations.

CoSaMP is the most time-consuming algorithm followed by SP (see Fig. 5(d)). It decreases rapidly to about 0 when $k > 135$, the observer should guarantee that there should be $3 \cdot k$ atoms in Λ, which indicates that very large k will affect its performance. The time consumption of OMP and GOMP only change a little under different sparsity. Through comprehensive analysis, reasonable k should be between 50 and 125, which is further set to 50 in the next experiments (if $m = 320$, k is then set to from 45 to 60). The experimental results of m from 100 to 800 are shown in Fig. 6. The SNR of the four algorithms continues to improve before the number of measurements m reach 320 (i.e. $m \geq k \cdot \log_2\left(\frac{n}{k}\right)$ is satisfied). OMP, CoSaMP and SP achieve optimal performance when $m > 300$ (SNR is about 25). This indicates that k limits the further improvement of their performance. However, the performance of GOMP is still improving and becomes the best when $m > 300$. The final SNR of GOMP is about 48. This is also reflected in Fig. 6(b) where the relative errors of OMP, CoSaMP and SP decrease to the lowest (about 0.055) when $m > 300$, while the relative error of GOMP continues to decline and finally reaches about 0.004. This also fully illustrates the superiority of GOMP.

The matching degrees of OMP, CoSaMP, SP and GOMP are greater than 0.75 in all measurements (see Fig. 6). They increase rapidly when $m < 200$ and then becomes stable and are larger than 0.9950. The runtime of GOMP increases rapidly, while other three algorithms changes little under different numbers of observations. Compared with the experimental results under different sparsity, the performance under different numbers of measurements is more stable and does not fluctuate dramatically. This is even more obvious for GOMP, which consumes more time than other 3 algorithms when $m > 350$. Choosing a reasonable number of observations can reduce the time consumption, here we set m to 320.

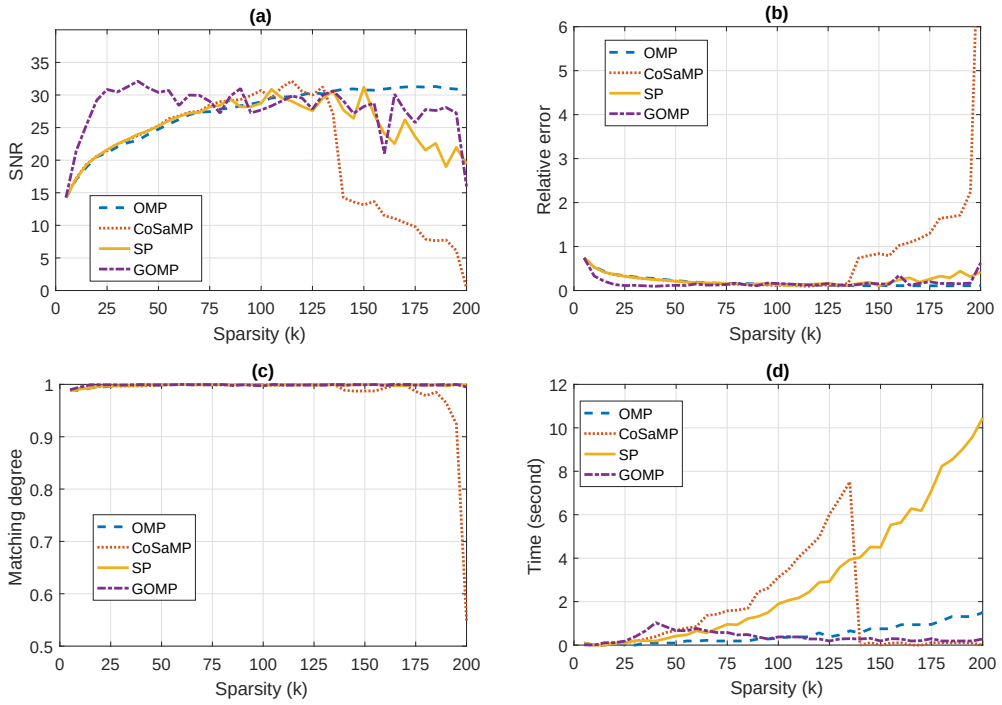

Figure 5: SNR (a), relative error (b), matching degree (c) and time consumption (d) of OMP, CoSaMP, SP and GOMP under different sparsity.

6.3 Performance Comparison

We randomly select a power trace. In fact, the reconstruction performance of power traces is almost the same under the same observation matrix. The algorithms OMP, CoSaMP, BP and GOMP can reconstruct power traces well when $k = 50$ and $m = 320$ (as shown in Table 2 and Fig. 7). The blue line and red line represent the original and the reconstructed power traces. They overlap in most regions, when the matching degree is greater than 0.9950. This shows that only 320 samples need to be collected to reconstruct the leakage of 800 samples under CS. If we collect longer traces, the compression performance is even better. GOMP performs best, since the relative error e_r is smallest, of which the corresponding SNR is also the largest. This also verifies the conclusion that SNR of GOMP is the highest when $k = 50$ in Fig. 5. Although e_r of OMP is larger than that of GOMP, the matching degree of it is better. This indicates that partial overlap has an important impact on the overall performance evaluation. Therefore, we need to integrate a number of criteria when comparing the performance of power trace reconstruction algorithms.

Table 2: Leakage reconstruction performance of OMP, CoSaMP, SP and GOMP on *AT89S52* micro-controller.

algorithms	e_r	SNR	α	time (second)
OMP	0.0647	23.7766	0.9993	0.017782
CoSaMP	0.0573	24.8361	0.9971	0.092189
SP	0.0607	24.3335	0.9958	0.044551
GOMP	0.0538	25.3873	0.9992	0.055530

The reconstruction error e_o of OMP, CoSaMP, SP and GOMP in Fig. 7 is about 0.40. Most regions of these four algorithms overlap. This indicates that 0.35 could be a very

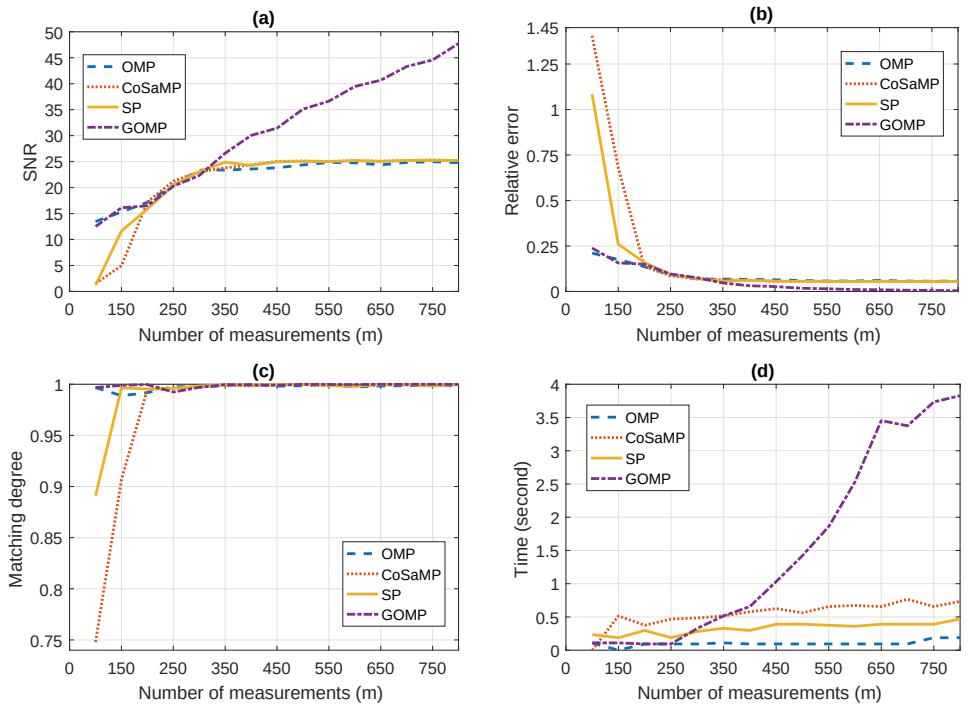

Figure 6: SNR (a), relative error (b), matching degree (c) and time consumption (d) of OMP, CoSaMP, SP and GOMP under different numbers of measurements.

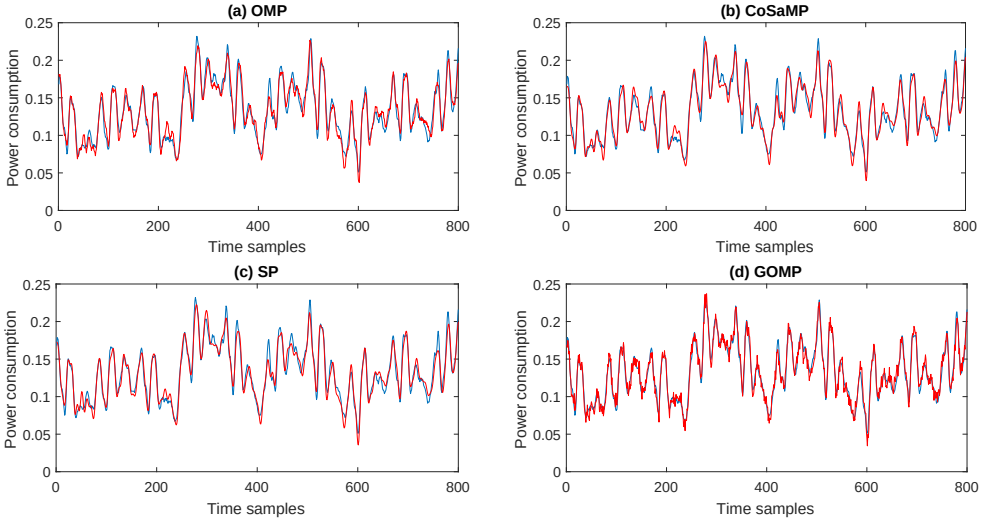

Figure 7: Leakage reconstruction using OMP, CoSaMP, SP and GOMP (the original power trace (blue) and reconstructed power trace (red)).

good threshold since it leads to high reconstruction performance. In order to compare the performance of OMP, CoSaMP, SP and GOMP more objectively, we further compare the percentage recovered (i.e. probability of successful reconstruction) under different sparsity and different numbers of observations (measurements). The corresponding experimental

results are given in Fig. 8. Let P_r denote the percentage recovered in the rest of this paper. We randomly select signals from a set of 20000 power traces for 400 repetitions in our experiments. The evaluation in Fig. 8(a) and Fig. 8(b) takes 5678.401890 and 5667.954980 seconds, respectively.

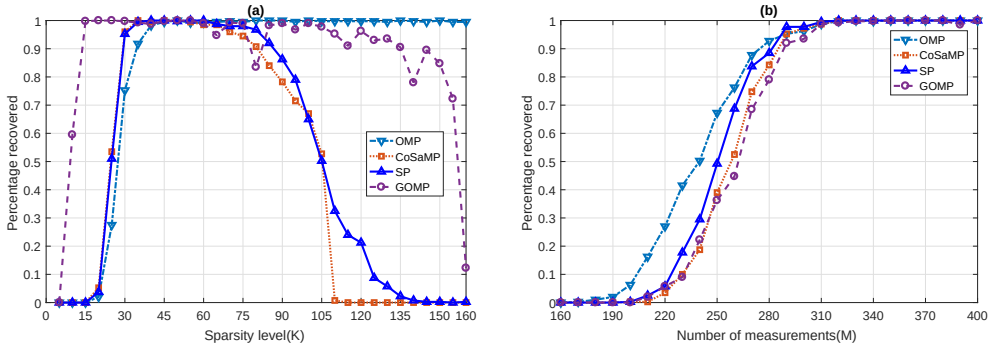

Figure 8: Percentage recovered under different sparsity levels (a) and different numbers of observations (b).

The performance of OMP, CoSaMP, SP and GOMP are very different when k varies from 5 to 160 ($m = 320$), where it first becomes better and then worse. The P_r of GOMP rapidly reaches 1.00 when k is from 5 to 15, and decreases with obvious fluctuations when k is from 105 to 145. P_r of SP and CoSaMP almost overlaps when k is from 15 to 35, decreases when k is larger than 70 and reaches 0 after $k = 110$. Compared with CoSaMP, SP and GOMP, OMP is more robust. Its P_r is constant at 1.00 when k reaches 160. The precondition of power trace reconstruction is $m >= k \cdot \log_2\left(\frac{n}{k}\right)$, which indicates that the performance of OMP will eventually decline. Compared with the performance under different sparsity k, P_r of OMP, CoSaMP, SP and GOMP increases steadily under different numbers of measurements (as shown in Fig. 8(b)). It is almost 0 when $m < 170$ and reaches 1.00 when $m = 320$. Their performance from good to poor are as follows: OMP, SP, CoSaMP and GOMP. This indicates that the observer can recover the power traces with a probability of 1.00 if $m = 320$, $k = 50$ and residual is set to 0.35. Obviously, the performance of 500MS/s sampling rate under classical compressive sampling is only equivalent to 200MS/s sampling rate under CS.

Power traces can be reconstructed accurately by fewer observations under the same sparsity when $k < 50$, which indicates superiority of the GOMP algorithm shown in Fig. 8(a). However, its performance is worse than the other three algorithms when the number of measurements is less than 310 (as shown in Fig. 8(b)). The SNR and relative error e_r of GOMP in Fig. 6 are significantly better than those of the other three algorithms when $m > 310$. However, since we set a large reconstruction error threshold 0.35, this advantage is not ultimately reflected in the percentage recovered. In fact, we can set appropriate reconstruction errors according to specific requirements. For example, in SCAs, the attacker can collect more power traces instead of accurately reconstructing each of them. The reconstruction error can be set appropriately large. However, in side-channel evaluations, for the sake of accurate security evaluation of cryptographic devices, it requires that the power traces can be accurately reconstructed, and the threshold of reconstruction error should be set sufficiently small.

7 Experiments Results On DPA Contest $v1$

7.1 Experimental Setups

In order to facilitate the re-implementation of our CS parameter adjustments, our second experiment is performed on the power trace set *SecmatV1* leaking from the unprotected DES crypto-processor implemented in ASIC provided by DPA Contest $v1.1$ [dpa]. Each power trace includes 5003 time samples. In order to observe and compare the performance of algorithms OMP, CoSaMP, BP and GOMP conveniently, we download 10000 power traces and use the times from 4001^{th} to 5000^{th} to perform our compressive sensing leakage reconstruction experiments. The sample segment includes the leakage of the last round of DES algorithm, which is commonly attacked in SCA. We also use a moving average filter with a 5-hour span to smooth all the traces simultaneously. A random observation matrix for four algorithms is generated for each repetition.

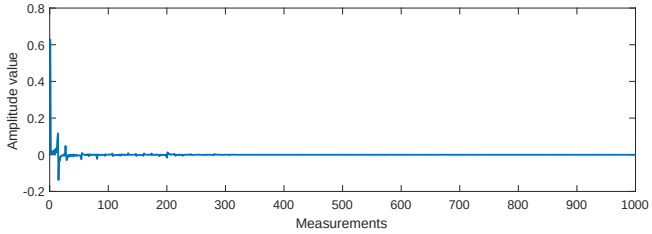

Figure 9: DCT coefficients of the second power trace in *SecmatV1*.

As we mentioned in Section 3.3, the sparser the leakage is in one domain, the better the reconstruction performance is under the same number of observations since more key information of the leakage is compressed into these dimensions. Sparse domains are also an important topic in CS. Power traces from DPA Contest $v1.1$ is well sparse in DCT domain (as shown in Fig. 9). Compared with the leakage of *AT89S52* micro-controller, its amplitude is smaller. The largest DCT coefficients occurs in the first dimension, and other DCT coefficients are very small.

7.2 Threshold Selection

SNR, relative error, matching degree and time consumption of OMP, CoSaMP, SP and GOMP under different sparsity are shown in Fig. 10. The SNR keeps increasing when $k < 25$ then reaches about 24. It is noteworthy that SNR is relatively stable when k is from 25 to 50. CoSaMP and SP fluctuate sharply when k is greater than 50, and their performance also declines sharply. Relative error and matching degree also indicate that CoSaMP can not be able to reconstruct power traces when $k > 135$. Since at least $3 \cdot k$ atoms should be guaranteed in Λ, the time consumption of CoSaMP even drops to nearly 0 at $k > 100$. The time consumption of SP and CoSaMP increases rapidly with sparsity, while OMP and GOMP changes little. Based on the above analysis, the k between 25 and 50 is a good choice if we want to test the percentage recovered.

We set k to 40, the relative error α of OMP, CoSaMP and SP decreases rapidly when m is from 50 to 300 and then becomes stable at 0.05 (as shown in Fig. 11). However, α of GOMP continues to decrease and finally reaches 0.004. The SNR of OMP, CoSaMP and SP becomes stable (about 26) when $m > 300$. The SNR of GOMP finally reaches about 48. The matching degree of all algorithms are higher than 0.92. Although we use 1000 time samples on the power traces provided by DPA contest $v1.1$ to carry out experiments, the time consumption of OMP, CoSaMP, SP and GOMP is smaller than that of 800 time

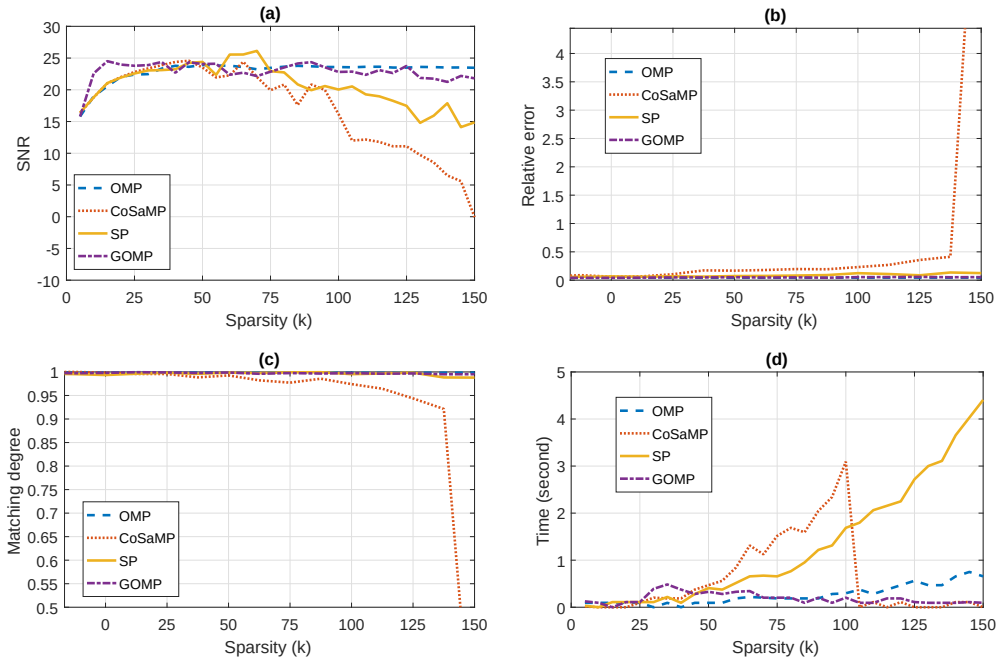

Figure 10: SNR (a), relative error (b), matching degree (c) and time consumption (d) of OMP, CoSaMP, SP and GOMP under different sparsity.

samples leaking from *AT89S52* micro-controller, which makes the reconstruction error e_o of power traces smaller. GOMP takes more time than other algorithms when $m > 300$. Based on the above performance analysis, m from 300 to 350 is a good choice.

7.3 Performance Comparison

The second power trace is selected to perform DCT transform to compare the performance of OMP, CoSaMP, BP and GOMP, and the corresponding experimental results are shown in Fig. 12. k and m are set to 40 and 300 according to experimental results given in Section 7.2. There are more clock cycles in Fig. 7 compared with the leakage in DPA Contest $v1.1$ shown in Fig. 12. Due to the preprocessing of DPA Contest $v1.1$, its power traces are also more easy to reconstruct. The reconstructed power trace of GOMP can overlap well with the original one. Compared with the other three algorithms, the reconstruction error of GOMP is smaller and the SNR is higher (as shown in Table 3). The other 3 algorithms do not overlap well in some areas where power consumption varies distinctly. α of all 4 algorithms are larger than 0.9970. This indicates that reconstruction performance of the four algorithms is very good and meets the sampling requirements very well.

Table 3: Leakage reconstruction performance of OMP, CoSaMP, SP and GOMP on DPA Contest $v1.1$.

algorithms	e_r	SNR	α	time (second)
OMP	0.0713	22.9364	0.9988	0.010863
CoSaMP	0.0612	24.2611	0.9974	0.043958
SP	0.0692	23.2020	0.9989	0.033731
GOMP	0.0601	24.4179	0.9982	0.058839

The SNR of a power trace depends on the reconstruction residual of the recovery

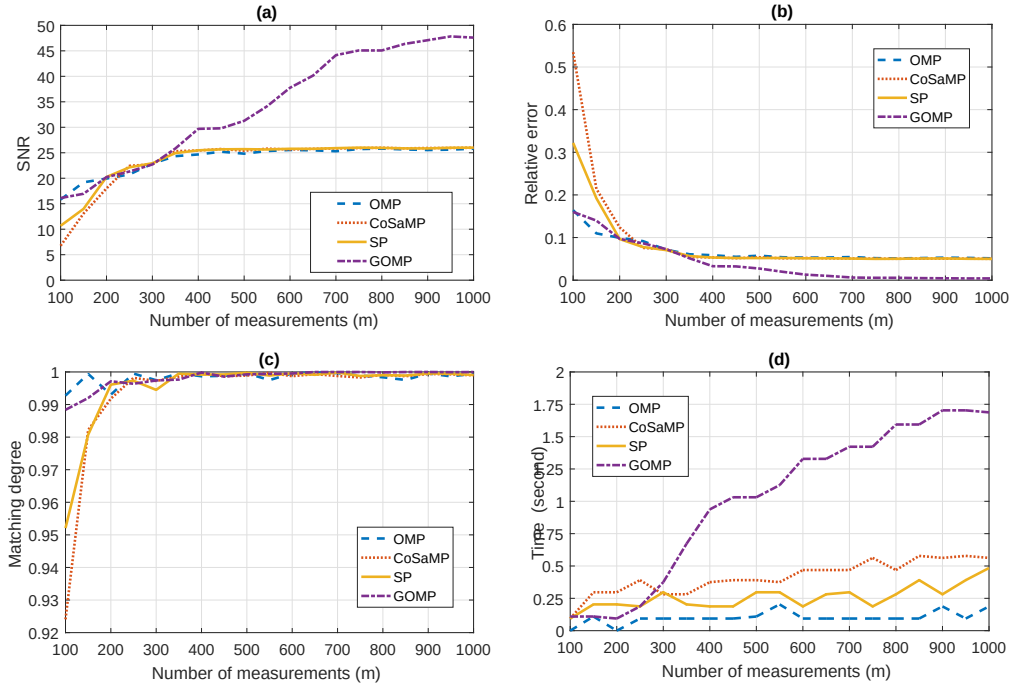

Figure 11: SNR (a), relative error (b), matching degree (c) and time consumption (d) of OMP, CoSaMP, SP and GOMP under different numbers of measurements.

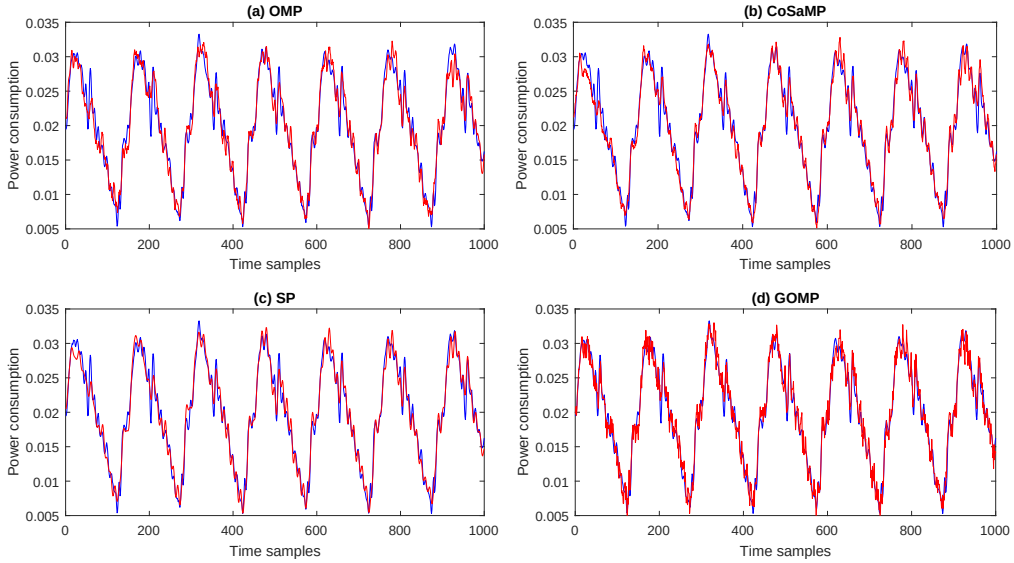

Figure 12: Leakage reconstruction using OMP, CoSaMP, SP and GOMP (the original power trace (blue) and reconstructed power trace (red)).

algorithm referring to Eq. 27. The smaller the residual, the higher the SNR of the reconstructed power traces. The reconstruction errors e_o of OMP, CoSaMP and SP in

Fig. 12 are about 0.04, compared with about 0.03 of GOMP. They are all smaller than these about 0.40 in Fig. 7, since the power consumption of ASIC used in DPA Contest $v1.1$ is smaller than *AT89S52* micro-controller. Therefore, relative error is better referential than absolute error in reconstruction performance evaluation. The relative error, SNR and matching degree in Tables 2 and 3 also illustrate that the reconstruction performance in Fig. 7 and Fig. 12 is similar. The power traces of DPA contest $v1.1$ are more compressible than these of *AT89S52* micro-controller. To achieve similar performance, we only need to collect 30% samples to reconstruct the original leakage (compared with 40% of *AT89S52* micro-controller). Moreover, the matching degree also illustrates the high performance of OMP, CoSaMP, SP and GOMP. They can quickly reconstruct the original power traces (see the time comsuption on Table 3).

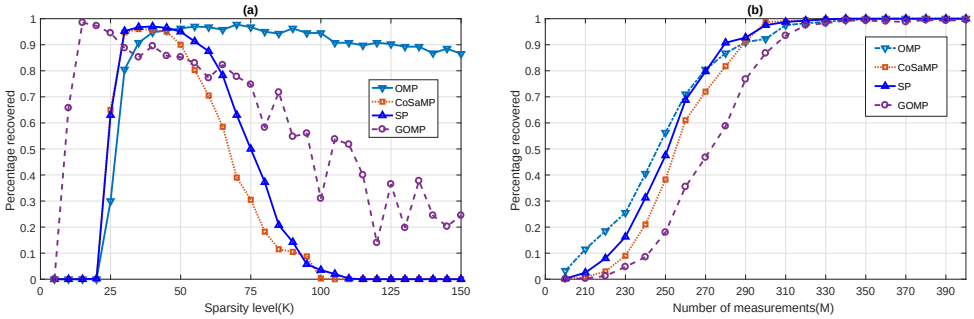

Figure 13: Percentage recovered under different sparsity levels (a) and different numbers of observations (b).

In order to further compare the performance of OMP, CoSaMP, SP and GOMP, we set $k = 40$, $m = 300$ and residual threshold to 0.04, the corresponding experimental results are shown in Fig. 13. When $k = 40$ and the step width is set to 10, it takes 5047.385466 seconds to run m from 100 to 400 (400 repetitions). When $m = 300$ and the step width is set to 5, it takes 2969.2314196 seconds to run k from 15 to 150 (400 repetitions). The P_r of OMP, CoSaMP and SP is 0 when k is from 5 to 20. k from 30 to 55 is a good choice, which also validates our conclusion in Section 7.2. P_r of GOMP increases rapidly from 0 to 1 when k is from 5 to 15, then drops and fluctuates. CoSaMP decreases faster than SP, and they drop to 0 at $k = 100$. Similar conclusion is drawn from Section 6.3. The slow performance degradation of OMP can be observed in Fig. 13, which illustrates its stability. When $k = 40$ and the number of observations is smaller than 200, all 4 algorithms fail to recover power traces. The absolute error e_o decreases with growth of m. The percentage recovered increases gradually to 1 then becomes stable when m is from 200 to 330. It is noteworthy that OMP, CoSaMP, SP and GOMP have their own advantages and disadvantages, the observer may get completely different performance from different signals. He should choose the appropriate algorithm according to the specific situation, power trace structure, observation matrix and characteristic of reconstruction algorithms.

The percentage recovered of GOMP is high when the sparsity k is small (as shown in Fig. 8 and Fig. 13). With more observations, the higher the SNR of GOMP, the better the reconstruction performance of it than other three algorithms. However, since the large reconstruction error allowed, this advantage is not ultimately reflected in Fig. 13(b). We can draw similar conclusions from Fig. 8(b). Moreover, the performance of OMP, CoSaMP and SP becomes stable after a certain number of observations, while the performance of GOMP still gradually improves with the growth of m (as shown in Fig. 6 and Fig. 11). In order to improve their performance, the sparsity can be appropriately enlarged. The SNR of OMP increases to a certain height and then becomes gradually stable, while the

performance of the other three algorithms improves first and then deteriorates sharply (as shown in Fig. 8 and Fig. 13). Enlarging the number of observations can improve their performance and make the locations of optimal performance move to the right. However, it is difficult to optimally balance k and m. There are currently many studies on the selection of them, such as the SWOMP shortly introduced in Section 5.2, of which k is the number of coefficients larger than half of the maximum.

8 Conclusions

The rapid increase in the bandwidth of cryptographic devices makes it difficult to sample, store and process leakages. In this paper, we introduce Compressive Sensing, a new and highly-efficient data sampling technology for side-channel leakage sampling and compare it with classical compressive sampling. Our experiments performed on power traces obtained from AT89S52 micro-controller and DPA contest $v1.1$ clearly demonstrate that CS can use a sampling rate much lower than the original one to obtain equivalent sampling performance. It projects the original power traces onto the observation space, and obtains the observation samples far below the original dimension. CS transfers a large amount of computation from sampling devices to advanced processors, so that the compute-intensive signal reconstruction can be carried out fast without distortion. In this paper, we only introduce the basic techniques of CS for leakage sampling and verify its superiority by experiments. There are many studies on sparse representation of signals, observation matrix design and signal reconstruction which could be applied to the leakage sampling problem. As such, we believe this work provides a new research direction in SCA which has many avenues for investigations and opportunities for further improvements.

References

[BCO04] Eric Brier, Christophe Clavier, and Francis Olivier. Correlation power analysis with a leakage model. In *Cryptographic Hardware and Embedded Systems - CHES 2004: 6th International Workshop Cambridge, MA, USA, August 11-13, 2004. Proceedings*, pages 16–29, 2004.

[BD09] Thomas Blumensath and Mike E. Davies. Stagewise weak gradient pursuits. *IEEE Trans. Signal Processing*, 57(11):4333–4346, 2009.

[BHvW12] Lejla Batina, Jip Hogenboom, and Jasper G. J. van Woudenberg. Getting more from PCA: first results of using principal component analysis for extensive power analysis. In *Topics in Cryptology - CT-RSA 2012 - The Cryptographers' Track at the RSA Conference 2012, San Francisco, CA, USA, February 27 - March 2, 2012. Proceedings*, pages 383–397, 2012.

[CDP15] Eleonora Cagli, Cécile Dumas, and Emmanuel Prouff. Enhancing dimensionality reduction methods for side-channel attacks. In *Smart Card Research and Advanced Applications - 14th International Conference, CARDIS 2015, Bochum, Germany, November 4-6, 2015. Revised Selected Papers*, pages 15–33, 2015.

[CDP16] Eleonora Cagli, Cécile Dumas, and Emmanuel Prouff. Kernel discriminant analysis for information extraction in the presence of masking. In *Smart Card Research and Advanced Applications - 15th International Conference, CARDIS 2016, Cannes, France, November 7-9, 2016, Revised Selected Papers*, pages 1–22, 2016.

[CDS01] Scott Shaobing Chen, David L. Donoho, and Michael A. Saunders. Atomic decomposition by basis pursuit. *SIAM Review*, 43(1):129–159, 2001.

[CR06] Emmanuel J. Candès and Justin K. Romberg. Quantitative robust uncertainty principles and optimally sparse decompositions. *Foundations of Computational Mathematics*, 6(2):227–254, 2006.

[CRT06] Emmanuel J. Candès, Justin K. Romberg, and Terence Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Information Theory*, 52(2):489–509, 2006.

[CT06] Emmanuel J. Candès and Terence Tao. Near-optimal signal recovery from random projections: Universal encoding strategies? *IEEE Trans. Information Theory*, 52(12):5406–5425, 2006.

[CW08] Emmanuel J. Candès and Michael B. Wakin. An introduction to compressive sampling. *Signal Processing Magazine, IEEE*, 25(2):21–30, 2008.

[DCE16] A. Adam Ding, Cong Chen, and Thomas Eisenbarth. Simpler, faster, and more robust t-test based leakage detection. In *Constructive Side-Channel Analysis and Secure Design - 7th International Workshop, COSADE 2016, Graz, Austria, April 14-15, 2016, Revised Selected Papers*, pages 163–183, 2016.

[DH01] David L. Donoho and Xiaoming Huo. Uncertainty principles and ideal atomic decomposition. *IEEE Trans. Information Theory*, 47(7):2845–2862, 2001.

[DM09] Wei Dai and Olgica Milenkovic. Subspace pursuit for compressive sensing signal reconstruction. *IEEE Trans. Information Theory*, 55(5):2230–2249, 2009.

[Don06] David L. Donoho. Compressed sensing. *IEEE Trans. Information Theory*, 52(4):1289–1306, 2006.

[dpa] Dpa contest. http://www.dpacontest.org/home/.

[DS16] François Durvaux and François-Xavier Standaert. From improved leakage detection to the detection of points of interests in leakage traces. In *Advances in Cryptology - EUROCRYPT 2016 - 35th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Vienna, Austria, May 8-12, 2016, Proceedings, Part I*, pages 240–262, 2016.

[DTDS12] David L. Donoho, Yaakov Tsaig, Iddo Drori, and Jean-Luc Starck. Sparse solution of underdetermined systems of linear equations by stagewise orthogonal matching pursuit. *IEEE Trans. Information Theory*, 58(2):1094–1121, 2012.

[GHT05] Catherine H. Gebotys, Simon Ho, and C. C. Tiu. EM analysis of rijndael and ECC on a wireless java-based PDA. In *Cryptographic Hardware and Embedded Systems - CHES 2005, 7th International Workshop, Edinburgh, UK, August 29 - September 1, 2005, Proceedings*, pages 250–264, 2005.

[GN03] Rémi Gribonval and Morten Nielsen. Sparse representations in unions of bases. *IEEE Trans. Information Theory*, 49(12):3320–3325, 2003.

[GSTV06] Anna C. Gilbert, Martin J. Strauss, Joel A. Tropp, and Roman Vershynin. Algorithmic linear dimension reduction in the l_1 norm for sparse vectors. *CoRR*, abs/cs/0608079, 2006.

[HTWM10] Zachary T. Harmany, Daniel Thompson, Rebecca Willett, and Roummel F. Marcia. Gradient projection for linearly constrained convex optimization in sparse signal recovery. In *Proceedings of the International Conference on Image Processing, ICIP 2010, September 26-29, Hong Kong, China*, pages 3361–3364, 2010.

[Kas91] B Kashin. The widths of certain finite dimensional sets and classes of smooth functions, izvestia 41 (1977), 334–351. *MR0481792 (58: 1891)*, 1891.

[KJJ99] Paul C. Kocher, Joshua Jaffe, and Benjamin Jun. Differential power analysis. In *Advances in Cryptology - CRYPTO '99, 19th Annual International Cryptology Conference, Santa Barbara, California, USA, August 15-19, 1999, Proceedings*, pages 388–397, 1999.

[LD06] Chinh La and Minh N. Do. Tree-based orthogonal matching pursuit algorithm for signal reconstruction. In *Proceedings of the International Conference on Image Processing, ICIP 2006, October 8-11, Atlanta, Georgia, USA*, pages 1277–1280, 2006.

[Mal09] Arian Maleki. Coherence analysis of iterative thresholding algorithms. *CoRR*, abs/0904.1193, 2009.

[MOP07] Stefan Mangard, Elisabeth Oswald, and Thomas Popp. *Power analysis attacks - revealing the secrets of smart cards*. Springer, 2007.

[MRSS18] Amir Moradi, Bastian Richter, Tobias Schneider, and François-Xavier Standaert. Leakage detection with the x2-test. *IACR Trans. Cryptogr. Hardw. Embed. Syst.*, 2018(1):209–237, 2018.

[MZ93] Stéphane Mallat and Zhifeng Zhang. Matching pursuits with time-frequency dictionaries. *IEEE Trans. Signal Processing*, 41(12):3397–3415, 1993.

[NT10] Deanna Needell and Joel A. Tropp. Cosamp: iterative signal recovery from incomplete and inaccurate samples. *Commun. ACM*, 53(12):93–100, 2010.

[NV09] Deanna Needell and Roman Vershynin. Uniform uncertainty principle and signal recovery via regularized orthogonal matching pursuit. *Foundations of Computational Mathematics*, 9(3):317–334, 2009.

[NV10] Deanna Needell and Roman Vershynin. Signal recovery from incomplete and inaccurate measurements via regularized orthogonal matching pursuit. *J. Sel. Topics Signal Processing*, 4(2):310–316, 2010.

[OSW+17] Changhai Ou, Degang Sun, Zhu Wang, Xinping Zhou, and Wei Cheng. Manifold learning towards masking implementations: A first study. *IACR Cryptology ePrint Archive*, 2017:1112, 2017.

[PM00] Erwan Le Pennec and Stéphane Mallat. Image compression with geometrical wavelets. In *Proceedings of the 2000 International Conference on Image Processing, ICIP 2000, Vancouver, BC, Canada, September 10-13, 2000*, pages 661–664, 2000.

[RMRM08] Baraniuk Richard, Davenport Mark, DeVore Ronald, and Wakin Michael. A simple proof of the restricted isometry property for random matrices. *Constructive Approximation*, 28(3):253–263, 2008.

[SA08] François-Xavier Standaert and Cédric Archambeau. Using subspace-based template attacks to compare and combine power and electromagnetic information leakages. In *Cryptographic Hardware and Embedded Systems - CHES 2008, 10th International Workshop, Washington, D.C., USA, August 10-13, 2008. Proceedings*, pages 411–425, 2008.

[SM15] Sujit Kumar Sahoo and Anamitra Makur. Signal recovery from random measurements via extended orthogonal matching pursuit. *IEEE Trans. Signal Processing*, 63(10):2572–2581, 2015.

[SMY09] François-Xavier Standaert, Tal Malkin, and Moti Yung. A unified framework for the analysis of side-channel key recovery attacks. In *Advances in Cryptology - EUROCRYPT 2009, 28th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Cologne, Germany, April 26-30, 2009. Proceedings*, pages 443–461, 2009.

[SNG+10] Youssef Souissi, Maxime Nassar, Sylvain Guilley, Jean-Luc Danger, and Florent Flament. First principal components analysis: A new side channel distinguisher. In *Information Security and Cryptology - ICISC 2010 - 13th International Conference, Seoul, Korea, December 1-3, 2010, Revised Selected Papers*, pages 407–419, 2010.

[SYZ08] Florian M. Sebert, Leslie Ying, and Yi Ming Zou. Toeplitz block matrices in compressed sensing. *CoRR*, abs/0803.0755, 2008.

[TG07] Joel A. Tropp and Anna C. Gilbert. Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Trans. Information Theory*, 53(12):4655–4666, 2007.

[Tro04] Joel A. Tropp. Greed is good: algorithmic results for sparse approximation. *IEEE Trans. Information Theory*, 50(10):2231–2242, 2004.

[WKS12] Jian Wang, Seokbeop Kwon, and Byonghyo Shim. Generalized orthogonal matching pursuit. *IEEE Trans. Signal Processing*, 60(12):6202–6216, 2012.