A First Study of Compressive Sensing for Side-Channel Leakage Sampling

Changhai Ou, Chengju Zhou, and Siew-Kei Lam

***Abstract*—An important prerequisite for Side-Channel Attacks (SCA) is leakage sampling where the side-channel measurements (i.e. 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 rate and storage capacity of 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). As long as the information is leaked in the low frequency component, CS can obtain low-dimensional samples by simply projecting the high-dimensional signals onto the observation matrix. The power traces can then be reconstructed in a workstation for further analysis and storage. With this approach, the sampling rate to obtain power traces 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. 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.**

***Index Terms*—compressive sensing, matching pursuit, leakage reconstruction, power trace, side-channel attack**

I. Introduction

SIDE-Channel Attack (SCA) exploits information leakage (e.g. power consumption [17], electromagnetic radiation [14], [18] and acoustic [15]) from cryptographic chips and embedded systems to infer the secret key. 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 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. This imposes higher requirements on the sampling rate and storage capacity of the sampling equipment. While this may seem as an advantage from the security standpoint, it also poses difficulty for SCA evaluations.

Several works in SCA such as Points-Of-Interest (POIs) selection [12] and time samples integration [20], have attempted to eliminate the information redundancy on power traces. Contrary to our approach, these works employ post-processing on the sampled traces to reduce the efforts of side-channel analysis instead of reducing the sampling rate. To overcome the storage limitations of sampling equipments, these existing works simply intercept and merge time samples of power traces, which will result in high information loss. In this paper, we undertake a first study on compressive sensing for side-channel leakage sampling and leakage reconstruction before describing the main contributions of our works.

A. Related Works

POI selection [12] and dimensionality reductions [3] are two classic pre-processing techniques in SCA. The former finds the locations of POIs by using side-channel distinguishers such as Differential Power Analysis (DPA) [17] and Correlation Power Analysis (CPA) [2], or leakage detection tools such as Welch's t-test [10], ρ-test [12] and χ^2-test [21]. Extracting POIs makes power traces easy to store, but it will lose most of the time samples, making it difficult to reconstruct the power traces. The latter such as Principal Component Analysis (PCA) [31] and Linear Discriminant Analysis (LDA) [32], considers the global features of high-dimensional power traces. Both techniques need to consider multiple power traces simultaneously, hence they are only suitable to be applied on compressed storage of power traces rather than directly used on sampling equipments like oscilloscopes. Discrete Wavelet Transform (DWT), Discrete Cosine Transform (DCT) and Fast Fourier Transform (FFT) [13], 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 [20] 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. These two resampling techniques can compress a large number of samples,

Manuscript received May 30, 2019; revised August 26, 2019; accepted November 20, 2019. This work was supported in part by the National Research Foundation Singapore Under Its Campus for Research Excellence and Technological Enterprise (CREATE) Programme With the Technical University of Munich at TUMCREATE. This article was recommended by Associate Editor Y. Jin. (Corresponding author: Changhai Ou.)

The authors are with the Hardware and Embedded Systems Lab, School of Computer Science and Engineering, Nanyang Technological University, Singapore (e-mail: chou@ntu.edu.sg; zhou0271@e.ntu.edu.sg; assklam@ntu.edu.sg).

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCAD.2019.2960337

significantly lower the complexity of side-channel analysis and reduce the storage requirement. However, it is difficult for both techniques to reconstruct the power traces since they lose too much information during compressive sampling. Moreover, the sampling rate of the sampling devices like oscilloscope is usually 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 [11]: "*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 [16], Romberg, Tao and Donoho established the theory of Compressive Sensing (CS) [5], [6], [11]. Combined with information theory, CS makes it possible to sample power traces 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 their sparsity and Restricted Isometry Property (RIP) [6]. As long as a power trace is compressible or sparse in a certain transform domain such as FFT, DWT and DCT, it can be projected from high-dimensional space into a low-dimensional space and retain the important information. 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. In this case, by solving an optimization problem, this power trace 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. They can be used for efficiently compressive storage of existing power traces. Moreover, they can also be integrated into sampling equipments as advanced compressive sampling techniques to increase their sampling bandwidth and reduce their sampling burdens.

Matching pursuit algorithms are typical examples of greedy algorithms used for signal reconstruction in CS, 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) [19], Orthogonal Matching Pursuit (OMP) [29], Compressive Sampling Matched Pursuit (CoSaMP) [22] and Generalized OMP (GOMP) [34], etc. Convex relaxation algorithms convert the non-convex optimization problem l_0-norm such as the pre-mentioned greedy algorithms to the convex optimization l_1-norm. It is worth mentioning that even though there are many existing reconstruction algorithms, there are still many 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 and re-sampling.

B. Our Contributions

This is the first work that aims to lower the requirements on both the sampling rate and power trace storage of sampling equipment through a novel use of Compressive Sensing (CS), which is a highly-efficient compressive sampling technology, for side-channel leakage sampling. CS performs compression and sampling simultaneously by projecting the high-dimensional leakages onto a low-dimensional space to obtain the discrete leakage samples. This enables the high-dimensional power traces to be reconstructed without distortion and significantly saves storage space for power traces.

Many current sampling tasks in side-channel analysis are accomplished by expensive oscilloscopes. However, the storage speed and capacity of oscilloscopes such as the *Tektronix DPO 7254* and $PicoScope$ -3000 cannot accommodate all the power traces when the sampling rate is high, leading to loss of power trace before the next one arrives. This problem is particularly serious when the sampled power traces are very long (e.g. the acquisition of complete AES-128 encryption). Using CS, the sampling equipment can acquire power traces quickly at a low sampling rate, thus avoiding the loss of power trace. The feasibility of our approach is verified by theory and experiments in this paper.

Our approach can be integrated into sampling plug-ins of many of these oscilloscopes. Oscilloscopes such as our *Tektronix DPO 7254* with a Windows 7 operation system allow for the development of independent sampling software. Researchers have also developed the sampling software on various platforms. For example, Inspector developed by Riscure and MathMagic [33], 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 III-C) and power traces, thus improving sampling performance.

C. Organization

The rest of this paper is organized as follows. The leakage characteristics, CPA and the principles of CS are introduced in Section II. The first two main parts of CS, leakage sparse representation (including the sparse domains) and observation, are described in Section III. Section IV uses classic greedy algorithms such as OMP, CoSaMP, SP and GOMP as examples to introduce the principle of power trace 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. Experiments are performed on an *AT89S52* micro-controller to demonstrate the efficiency of the approach in Section V. Finally, Section VI concludes this paper.

II. Preliminaries

A. Leakage Characteristics

Each time sample $\boldsymbol{x}$ of a power trace can be modeled as the sum of static power consumption component $\boldsymbol{x}_s$, exploitable power consumption component $\boldsymbol{x}_e$ and noise component $\boldsymbol{x}_n$:

$$\boldsymbol{x} = \boldsymbol{x}_s + \boldsymbol{x}_e + \boldsymbol{x}_n \tag{1}$$

according to [20]. The exploitable component $\boldsymbol{x}_e$ can be further refined into two parts: operation-dependent component $\boldsymbol{x}_o$ and data-dependent component $\boldsymbol{x}_d$:

$$\boldsymbol{x}_e = \boldsymbol{x}_o + \boldsymbol{x}_d. \tag{2}$$

Side-channel analysis can be divided into Simple Power Analysis (SPA) and Differential Power Analysis (DPA). The most widely used distinguisher in side-channel community to exploit $\boldsymbol{x}_o$ is SPA. It is often used to attack public-key systems such as RSA, which operates differently according to bits 0 and 1 of the key [24], [30]. Another technique is to use side-channel leakage to perform reverse engineering and recover the instructions executed by the chips [1], [26]. It is more convenient to collect the leakage of instruction operation as it occurs in the lower frequency component compared to the data operation. $\boldsymbol{x}_s$ and $\boldsymbol{x}_o$ can construct the basic outline of the power trace. DPA and its extensions like CPA, which exploits data-dependent component $\boldsymbol{x}_d$, are the focus of this paper.

B. Correlation Power Analysis

Side-channel leakage puts high requirements on its reconstruction, since different intermediate values do not cause drastic power changes in many chips, and there are many indicators to evaluate its performance. Besides those given in Section IV-C, we also consider the well-known model-dependent distinguisher Correlation Power Analysis (CPA) [2] (CPA), of which the principle is shown in Fig. 1. The power consumption of intermediate values satisfy:

$$\boldsymbol{x} = f\left(\mathsf{Sbox}\left(\mathbf{p} \oplus \kappa^*\right)\right) + \vartheta \tag{3}$$

when encrypting AES-128. Here Sbox and f denote the S-box operation and the leakage function, $\mathbf{p} \in \mathbb{F}_{2^8}$ and $\kappa^* \in \mathbb{F}_{2^8}$ denote the encrypted plaintext byte and its corresponding key byte, and ϑ denotes the corresponding noise. The attacker encrypts $n^{'}$ plaintexts $\mathbf{P} = \mathbf{p}_{1,\ldots,n^{'}}$ and collects $n^{'}$ power traces $\mathbf{X} = \boldsymbol{x}_{1,\ldots,n^{'}}$. Although he does not know the specific key, he can attack 16 sub-keys in the first round of AES-128 in a divide-and-conquer manner, which only requires a small amount of computation.

We usually do not profile an accurate enough leakage model like template attack [7], [27] in CPA. Hypothesis models, such as Hamming-weight model and Hamming-distance model, are not only simple but also lends themselves well towards CPA. For Hamming-weight model given in Fig. 1, CPA only needs to compute the hypothesis power consumption $\mathsf{Hw}\left(\mathsf{Sbox}\left(\mathbf{P} \oplus \kappa\right)\right)$ of the intermediate values $\mathbf{I}^{\kappa} = \boldsymbol{I}^{\kappa}_{1,\ldots,n^{'}} \in \mathbb{F}_{2^8}$ according to different guesses $\kappa \in \mathbb{F}_{2^8}$, calculates the correlation coefficient between it and the real power consumption

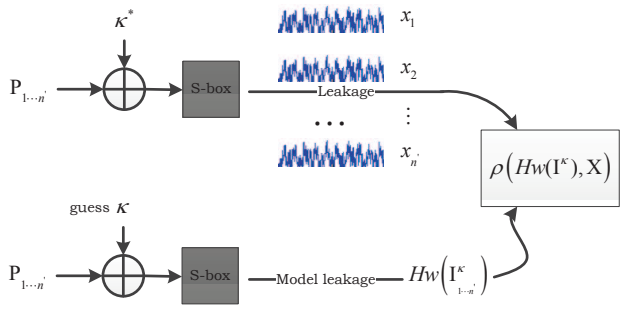

Fig. 1. The principle of CPA.

$\mathbf{X}$, and returns the candidate corresponding to the maximum correlation coefficient:

$$\kappa^{'} = \arg\max_{\kappa}\left\{\rho\left(\mathsf{Hw}\left(\mathbf{I}^{\kappa}\right), \mathbf{X}\right) | \kappa = 0, 1, \ldots, 255\right\}. \tag{4}$$

The basic CPA is performed on each time sample of power traces. We can also perform CPA on the selected POIs to improve its performance. We choose the former for the performance evaluation of leakage reconstruction.

C. Compressive Sensing

The principle of CS maintains that as long as the power trace 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. The classical compressive sampling techniques like dimensionality reduction methods such as PCA [31], LDA [32], Kernel Discriminant Analysis (KDA)[4] and manifold learning [25], obtain the low-dimensional samples by extracting the features from high-dimensional leakages. They consider the structure of all power traces, and the sampling and compression are performed separately. CS compresses a power trace while sampling it and aims to use the least coefficients to represent it. Compared with traditional compressive sampling, the sampling rate in CS theory no longer depends on the frequency of chip, but on the sparsity of power traces. CS uses nonlinear programming methods to recover power traces, 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.

The complete CS flow is shown in Fig. 2. Firstly, the m sparse coefficients of the n-dimensional original leakage signals are obtained by sparse transformation ($m \ll n$ in Fig. 2(b)). m largest coefficients are saved and other $n-m$ coefficients are discarded. The saved coefficients are sufficient to reconstruct the original leakage signal without distortion. Secondly, the high-dimensional original leakage signal is projected onto a sparse domain to achieve low-dimensional observation samples. The sampling device outputs these samples as sampling results. Finally, the original signal is reconstructed by solving

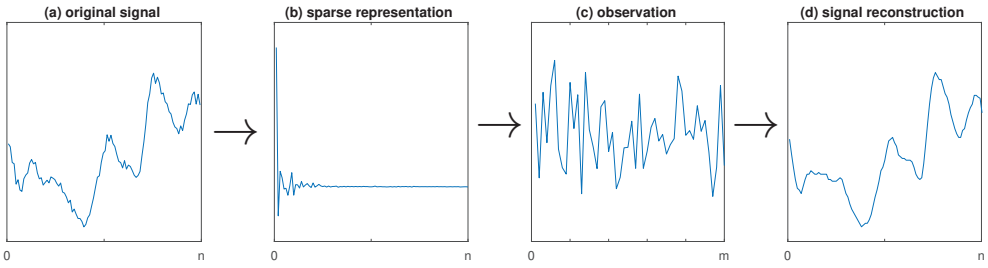

Fig. 2. General procedure of CS.

an optimization algorithm (see Fig. 2(d)). These operations will be discussed in detail in Sections III $\sim$ IV.

III. Sparse Representation of Power Traces

A. Sparse Decomposition

The purpose of sparse decomposition in SCA is to use as few sparse vectors as possible to represent the original leakage signal. If a power trace $\boldsymbol{x} = [x_0, x_1, \ldots, x_{n-1}] \in R^{n\times 1}$ including n time samples 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 the power trace $\boldsymbol{x}$ can be represented as:

$$\boldsymbol{x} = \Psi\Theta = \sum_{i=0}^{n-1} \psi_i \theta_i. \tag{5}$$

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.

The purpose of the observation matrices such as Gauss and Bernoulli, is to find a projection matrix that is not related to the sparse matrix, and they reflect the observation rules for power traces. Most of the current observation matrices are random Gaussian matrices, wherein each element satisfies the normal distribution with mean 0 and variance $\frac{1}{m}$:

$$\phi(i,j) \sim \mathcal{N}\left(0, \frac{1}{m}\right). \tag{6}$$

Since random matrices are not related to any matrix, they can satisfy RIP conditions required by CS with high probability if $m \geq c \cdot k \cdot \log_2\left(\frac{n}{k}\right)$ (see [11]). Here c is a small constant. Another observation matrix i.e. Bernoulli matrix, has also been proven to satisfy Restricted Isometry Property (RIP) condition with high probability in [28].

B. Sparse Domains

The application of frequency domain analysis is very common in the side-channel community. Power traces 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 power traces and more sparsely represent the information of them. 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:

$$F(u) = c(u) \sum_{i=0}^{n-1} x(i) \cos\left\{u\frac{(2i+1)}{2n}\right\}. \tag{7}$$

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

$$c(u) = \begin{cases} \sqrt{\frac{1}{n}}, & u = 0 \\ \sqrt{\frac{2}{n}}, & u = 1, 2, \cdots, n-1 \end{cases} \tag{8}$$

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 $\mathcal{O}\left(n^2\right)$. The Inverse Discrete Cosine Transform (IDCT):

$$x(i) = \sqrt{\frac{2}{n}} \sum_{u=0}^{n-1} c(u) F(u) \cos\left\{u\frac{(2i+1)\pi}{2n}\right\} \tag{9}$$

is performed to recover the original leakage signal $\boldsymbol{x}$.

C. Power Trace Observation

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

$$\boldsymbol{y} = \Phi\boldsymbol{x}, \tag{10}$$

and obtains the low-dimensional observation vector $\boldsymbol{y}$ (as shown in Fig.3), thus completing the sampling process. 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:

$$\Theta = \Psi^T \boldsymbol{x} \tag{11}$$

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.

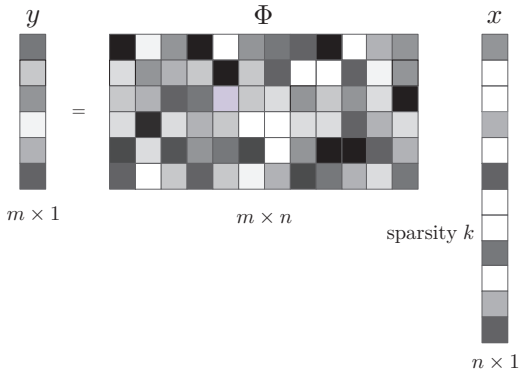

Fig. 3. Low-dimensional observation from high-dimensional power trace x in time domain.

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 number of samples compressed is much smaller 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. It is worth mentioning that matrix optimization has also been widely studied in CS, which can solve the above problem very well. For example, the long power traces can be divided into several segments, and a random matrix is established for each of them to sample leakage signals. The desired power traces can be obtained by recombining these segments. Moreover, the random matrices can remain unchanged during the sampling.

IV. Power Trace Reconstruction

A. Optimization

Different norms have different meanings. The l_p-norm of a power traces $\boldsymbol{x} = [x(0), x(1), \ldots, x(n-1)]$ is defined as

$$\|\boldsymbol{x}\|_{l_p} = \left(\sum_{i=0}^{n-1} |x(i)|^p\right)^{\frac{1}{p}}. \tag{12}$$

The l_0-norm denotes the number of non-zero values on power trace $\boldsymbol{x}$, l_1-norm denotes the sum of their absolute values, and l_2-norm denotes the square root of the sum of their squares. Taking a two-dimensional trace 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. 4(a), and l_2-norm $\|\boldsymbol{x}\|_{l_2} = \sqrt{|x_1|^2 + |x_2|^2}$ represents a circle (as shown in Fig. 4(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 other words, the two-dimensional trace can be scattered as one dimension. 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.

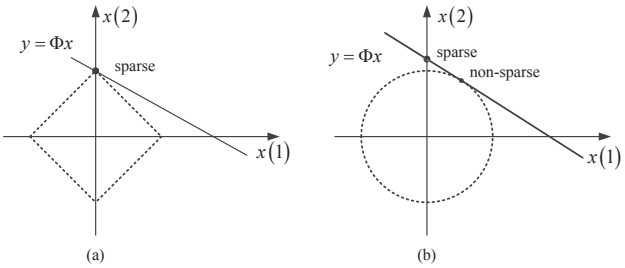

Fig. 4. The geometric meaning of the optimal solutions of l_1-norm (a) and l_2-norm (b) in two-dimensional space.

Since Θ is sparse, the number of unknown variables in $\boldsymbol{y} = \Lambda\Theta$ is greatly reduced, which makes signal reconstruction possible. The sparse coefficient Θ is obtained and a power trace $\boldsymbol{x}$ is recovered by Eq. 5. 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

$$\arg\min \|\Theta\|_{l_0}, s.t. \ \|\boldsymbol{y} - \Lambda\Theta\| < \epsilon. \tag{13}$$

However, the new model is unstable and difficult to solve directly. This can be solved by the suboptimal solution of l_0-norm: $\arg\min \|\Theta\|_{l_1}, s.t. \ \|\boldsymbol{y} - \Lambda\Theta\| < \epsilon$ if θ satisfies certain conditions [8]. The typical solutions of l_1-norm are convex optimization algorithms.

B. Greedy Algorithms

Signal reconstruction algorithms play a very important role in accurately reconstructing the high-dimensional original leakage signal $\boldsymbol{x}$ from the low-dimensional observation vector $\boldsymbol{y}$ 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 $\boldsymbol{x}$ and the new iteration error according to the currently selected atoms. These selected atoms have the greatest impact on reconstruction performance of $\boldsymbol{x}$ at present. Matching Pursuits (MP) [19] and Orthogonal Matching Pursuit (OMP) (improved from MP) [29] 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}:

$$\gamma_{t-1} = \arg\max_{\gamma_j} |\langle r_{t-1}, \gamma_j \rangle|. \tag{14}$$

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

$$r_{t-1} = \max |\langle r_{t-1}, \gamma_j \rangle| + r_t \tag{15}$$

after each iteration. Almost all of the matching pursuit algorithms such as CoSaMP [22] and GOMP [34], are improved from OMP. These algorithms preserve the atom selection strategy of MP.

OMP [29] 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 power trace $\boldsymbol{x}$ in frequency domain, 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 guarantee 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 $\quad \hat{\Theta}_t = \arg\min_{\hat{\Theta}_t} \left\| \boldsymbol{y} - \Lambda_t \hat{\Theta}_t \right\|_2^2$;
6 $\quad$ Update residual $r_t = \boldsymbol{y} - \Lambda_t \hat{\Theta}_t$;
7 $\quad t = t + 1$;
8 **end**
9 The reconstructed power trace $\hat{\boldsymbol{x}} = \Psi\hat{\Theta}$;

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:

$$f(\Theta) = (\boldsymbol{y} - \Lambda_t\Theta_t)^T (\boldsymbol{y} - \Lambda_t\Theta_t). \tag{16}$$

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

$$\Theta_t = \left(\Lambda_t^T \Lambda_t\right)^{-1} \Lambda_t^T \boldsymbol{y}. \tag{17}$$

Here Λ_t^T is the transformation matrix of Λ_t. Matching pursuit algorithm is a process of continuously selecting the atoms that are most conducive to reducing the reconstruction error. The complexity of OMP is $\mathcal{O}(k \cdot m \cdot n)$.

As we mentioned earlier, almost all of matching pursuit algorithms are improved from OMP. They are based on different mathematical principles and provide strict mathematical proofs. However, their algorithms are very similar and only several steps need to be replaced from OMP. Therefore, we just illustrate the difference between them and OMP, without elaborating on them. An obvious disadvantage of OMP algorithm is that only one atom is selected in each iteration. When 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.

Another disadvantage of OMP algorithm is that once an atom is in the candidate set, it will never be deleted. To improve this shortcoming, two algorithms CoSaMP (Compressive Sampling Matching Pursuit) [22], [23] and SP (Subspace Pursuit) [9], which rely on backtracking are employed. Specifically, as the algorithm iterates (see Steps $2 \sim 8$ in Algorithm 1), the atoms in Λ are recalculated and the non-optimal atoms are deleted. Specifically, CoSaMP [22], [23] 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 [9] 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 $\mathcal{O}(\log(k) \cdot m \cdot n)$ respectively.

C. Performance Criteria

There are many criteria to evaluate the performance of leaky signal reconstruction algorithms, such as reconstruction time, reconstruction residual $e_o = \|\boldsymbol{x} - \hat{\boldsymbol{x}}\|_2$ (also called absolute error), 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 between the original leakage signal $\boldsymbol{x}$ and reconstructed power trace $\hat{\boldsymbol{x}}$, the relative error is defined as:

$$e_r = \frac{\|\boldsymbol{x} - \hat{\boldsymbol{x}}\|_2}{\|\hat{\boldsymbol{x}}\|_2}. \tag{18}$$

Signal-to-Noise Ratio: SNR for the j^{th} time sample $\mathbf{X}^j$ is defined as the ratio of exploitable power consumption component to noise component:

$$\text{SNR} = \frac{\text{var}\left(\mathbf{X}_e^j\right)}{\text{var}\left(\mathbf{X}_n^j\right)} \tag{19}$$

in side-channel attacks [20]. Here $\mathbf{X} = \boldsymbol{x}_{1,\ldots,n'}$ denotes the collected n' power traces we introduced in Section II-B. It is defined as:

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

in CS, of which the molecule represents the variance of the original leakage signal $\boldsymbol{x}$. The denominator represents its

absolute error and the reconstructed power trace.

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

$$\alpha = 1 - \frac{|\|\hat{\boldsymbol{x}}\|_2 - \|\boldsymbol{x}\|_2|}{|\|\hat{\boldsymbol{x}}\|_2 + \|\boldsymbol{x}\|_2|}. \tag{21}$$

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.

V. Experiment Results

A. Experimental Setups

We implement the AES-128 algorithm using assembly language on an *AT89S52* micro-controller specially designed for side-channel attacks, with a clock operating frequency of 12 MHz (see Fig. 5). 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*" of AES-128, 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 storage speed is very slow. We cannot even store the power traces 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.

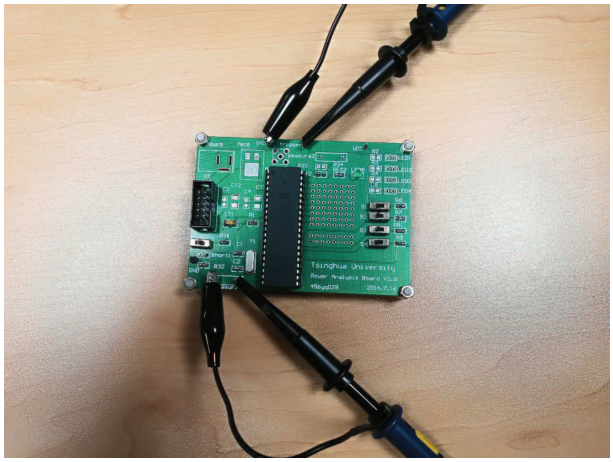

Fig. 5. Our AT89S52 micro-controller.

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-sample span to remove the noise. A random gaussian observation matrix for four algorithms is generated for each repetition.

Fig. 6 shows the power trace of a look-up table instruction (time samples from 1800 to 2500). The figure shows that different operations lead to different leakage characteristics in

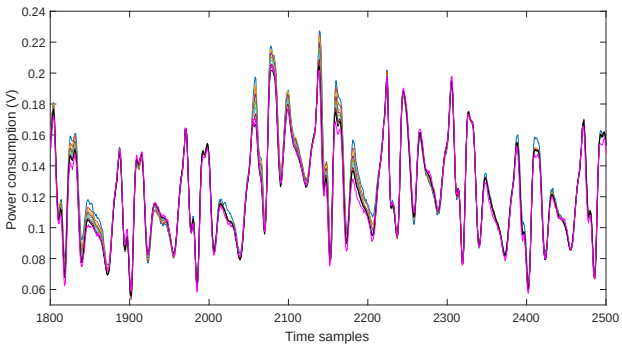

Fig. 6. Different operations lead to different leakage characteristics, and data-dependent leakage happens on peaks.

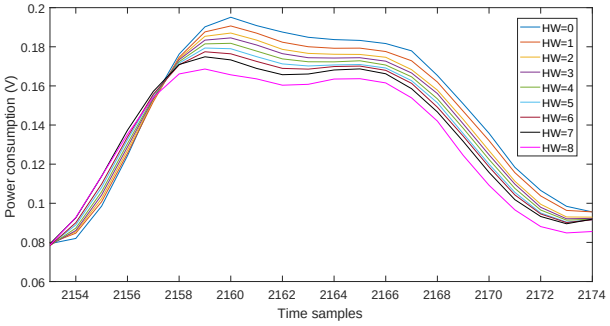

Fig. 7. The mean power consumptions of the 9 Hamming weights of S-box outputs within a clock cycle.

different clock cycles. The 9 mean power consumption traces of Hamming weights from 0 to 8 show that data-dependent information is leaked at high frequencies, which we call peaks. They leak the information of operands of $MOVC$ instructions. The time samples from 2153 to 2174 of a clock cycle in Fig. 7 clearly show the data dependent characteristics of different Hamming weights. Both classic DPA and its extensions (e.g. CPA) make use of $\boldsymbol{x}_d$.

B. Parameter Choice

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 of intermediate values of the look-up table (as shown in Fig. 9(a)). Moreover, side-channel sampling is usually targeted, such as on the first round of AES algorithm. The DCT coefficients of a power trace transformed from time domain to DCT domain are shown in Fig. 8. We can draw a conclusion that the leakage of *AT89S52* micro-controller is sparse in DCT domain, as most DCT coefficients are close to 0. We can also use CPA to filter out the frequencies that contribute more (as CPA is performed on 1400 traces on frequency domain shown in Fig. 9(b)). It can also be observed that the leaked information is mainly concentrated in the low frequency part after DCT transformation. Moreover,

the number of observations we selected should be more than 200, which will also be verified in the subsequent side-channel parameter choice and evaluations. FFT and DWT can also be used as sparse domains, although the corresponding experimental results are not given.

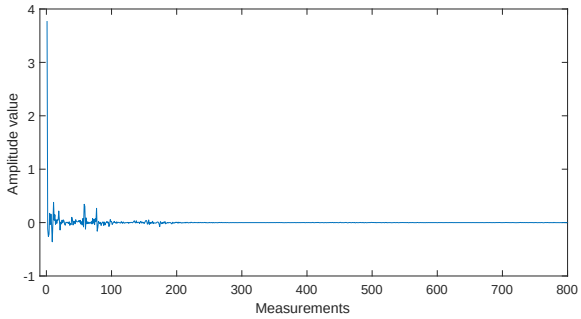

Fig. 8. DCT coefficients of a power trace leaks from *AT89S52* micro-controller.

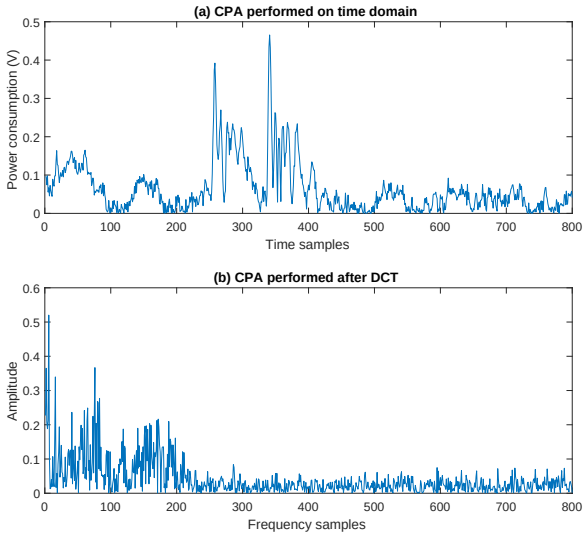

Fig. 9. CPA performed on time domain (a) and frequency domain after DCT (b).

The original leakage signal $\boldsymbol{x}$ 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. 10. We test the reconstruction performance under different power traces, and the results of most of them are similar. Therefore, we only report the reconstruction performance under a 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.

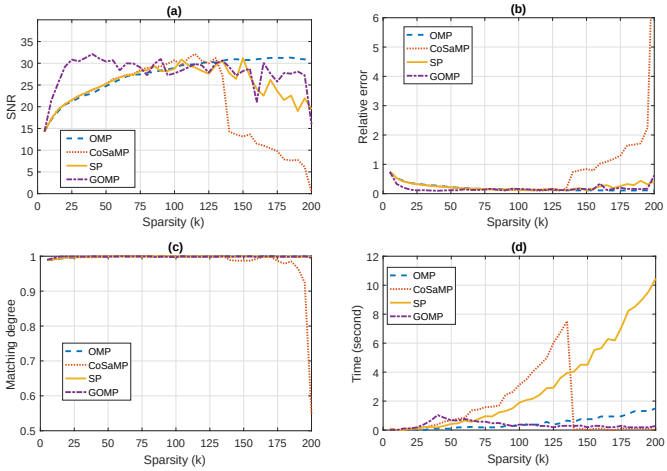

Fig. 10. SNR (a), relative error (b), matching degree (c) and time consumption (d) of OMP, CoSaMP, SP and GOMP under different sparsity.

CoSaMP is the most time-consuming algorithm followed by SP (see Fig. 10(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. 11. The SNR of the four algorithms continues to improve before the number of measurements m reaches 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. 11(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. 11). They increase rapidly when $m < 200$ and then become 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

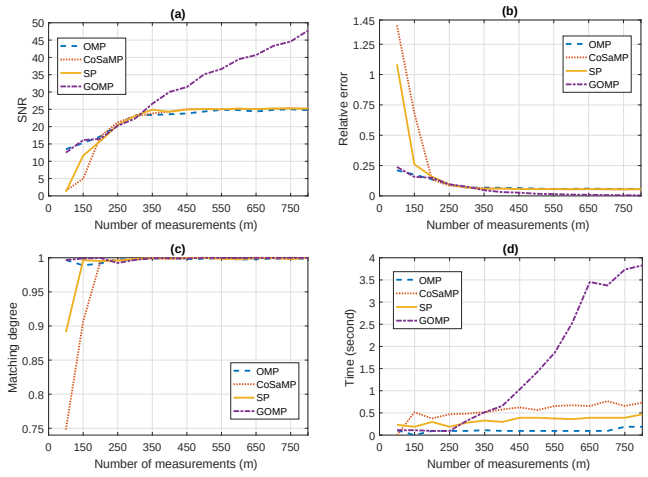

Fig. 11. SNR (a), relative error (b), matching degree (c) and time consumption (d) of OMP, CoSaMP, SP and GOMP under different numbers of measurements.

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.

C. 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 I and Fig. 12). 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. 10. 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 I
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

D. Side-Channel Evaluation

From the experiments in Section V-B, we arrived at the conclusion that $k = 50$, $m = 320$ are reasonable. Here we compare the performance of leakage reconstruction near $k = 50$, $m = 320$ under CPA. Each attack is repeated 100 times, and the success rate under different numbers of observations m are shown in Fig. 13. In order to compare the performance of the CS algorithms OMP, CoSaMP, SP and GOMP, the success rates of CPA performed on the original leakage signals are also given in Table II. Increasing the number of observations will mitigate the loss of information and significantly improve the reconstruction performance. The leakage reconstruction performance of OMP, CoSaMP and SP is similar, and much lower than the success rate of CPA performed on the original leakage signals. However, when the number of observations is about 400, GOMP achieves almost the same success rate as the one performed on the original leakage signals, which also shows that its reconstruction performance is much better than the other three schemes.

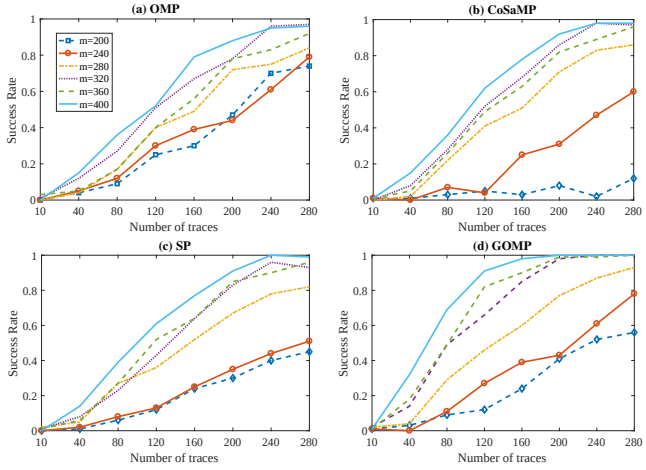

Fig. 13. Success rate of CPA under different numbers of observations m.

TABLE II
THE SUCCESS RATE OF CPA WHEN ATTACKING THE ORIGINAL LEAKAGE SIGNALS IN TIME DOMAIN.

Number of traces	10	40	80	120
Success rate	0.02	0.37	0.81	0.95
Number of traces	160	200	240	280
Success rate	0.99	1.00	1.00	1.00

The success rate under different sparsity levels k when m is set to 320 are shown in Fig. 14. Sparsity has a great impact on the performance of OMP, CoSaMP and SP, and the increase of sparsity will significantly improve their performance. However, GOMP does not change much even when the sparsity k varies from 25 to 100, and its performance is still good at very low sparsity (the performance when k is 50 can be found from Fig. 13).

As can be seen from Figs. 13 and 14, with the increase of sparsity and number of observations, the change of reconstruc-

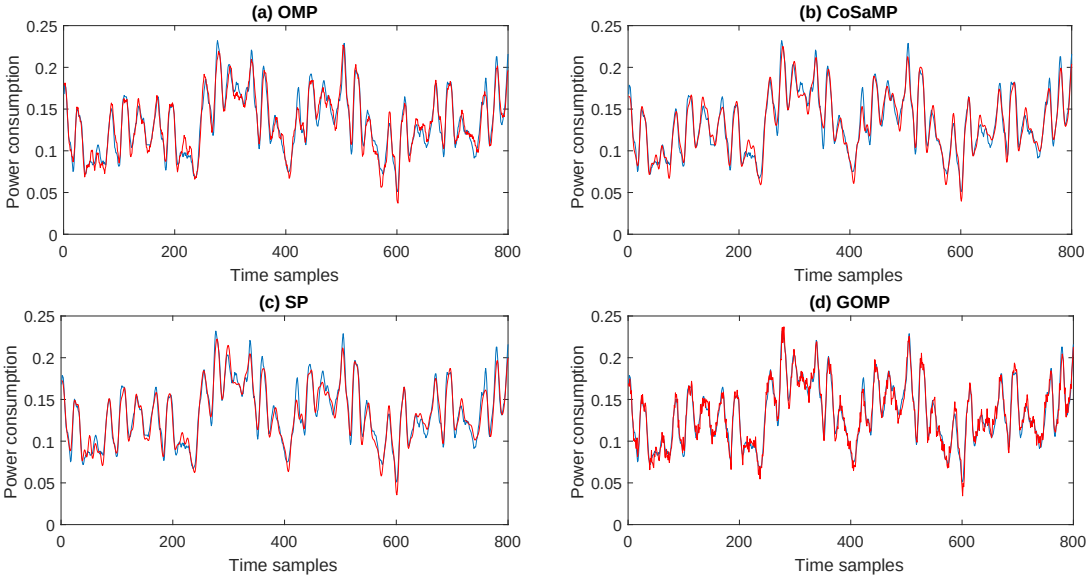

Fig. 12. Leakage reconstruction using OMP, CoSaMP, SP and GOMP (the original leakage signal (blue) and reconstructed power trace (red)).

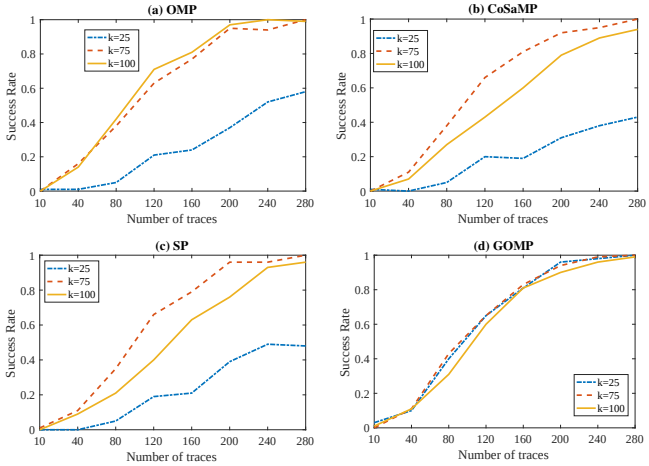

Fig. 14. Success rate of CPA under different sparsity levels k.

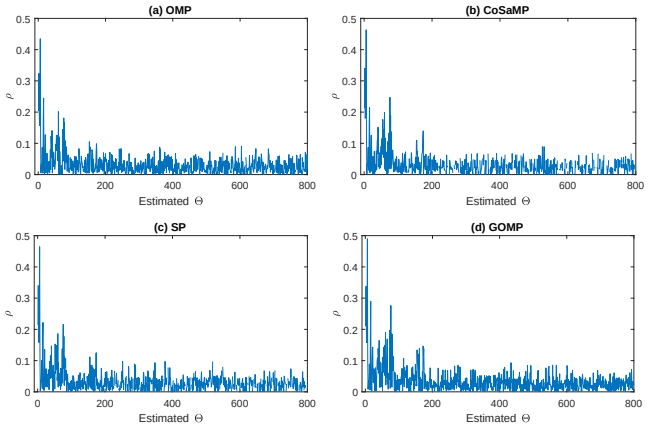

Fig. 15. CPA on the estimated Θ under OMP, CoSaMP, SP and GOMP.

tion performance gradually decreases, which also shows that the information leakage of our AT89S52 embedded system mainly occurs in the low frequency part. Compressive sensing can be used in image, voice and so on. Their requirement of signal reconstruction can be appropriately lowered, since minor changes will not significantly affect the results of reconstruction. However, the leaked information is reflected in the subtle changes in power consumption, which impose higher requirements for reconstruction. A good algorithm should be of low complexity, enables the side-channel attacks to be performed on the original leakage signals, and the reconstructed power traces should have similar performance.

Operations such as power trace reconstruction can be performed on the desktop computer. Power trace reconstruction makes it easy for us to observe information leaked in time domain, but we do not always require time domain information. After all, the purpose of side-channel attacks is to recover the key as efficiently as possible. Moreover, due to the characteristic of DCT and other algorithms, the high frequency part of the information may be lost. Since the clock frequency of our AT89S52 micro-controller is 12 MHz, and each instruction contains at least 12 clock cycles, the leakage can still be sampled by using DCT based CS. Otherwise, we have to adjust the sparse domain and compressive sampling algorithm accordingly.

Each sparse basis selected has the greatest impact on residual error in matching pursuit based power trace reconstruction algorithms. The impact of sparse coefficients gradually weaken with iterations. Therefore, we can reduce noise, improve attack success rate and save a lot of leakage reconstruction time if we select the former part of sparse coefficients to launch

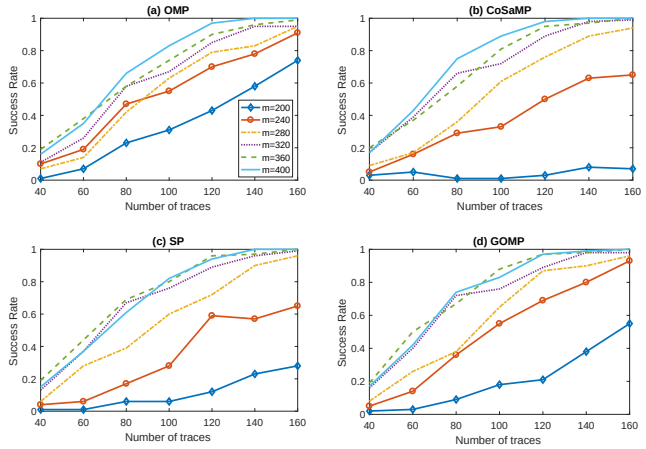

Fig. 16. Success rate of CPA under different observations m performed on sparsity coefficients.

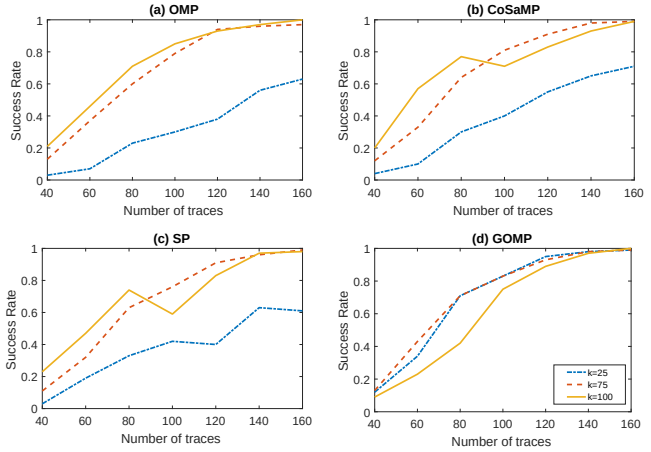

Fig. 17. Success rate of CPA under different sparsity levels k performed on sparsity coefficients.

our attacks (see Algorithm 1). The experimental results of CPA performed on the estimated sparse coefficients of 1000 power traces are shown in Fig. 15. It can be seen that the high correlation coefficients are concentrated in the first 200 dimensions. Therefore, we only attack these dimensions using CPA, and the corresponding experimental results are shown in Figs. 16 and 17. The success rate under different m and k is more efficient than that on the reconstructed power traces shown in Figs. 13 and 14, and closer to that of the original leakage signal.

VI. Conclusions

The rapid increase in the bandwidth of cryptographic devices makes it difficult to sample and process leakages, and also store power traces. In this paper, we consider compressed sampling and leakage reconstruction for the first time, and introduce Compressive Sensing, a new and highly-efficient data sampling technology for side-channel leakage sampling. Our experiments performed on an AT89S52 micro-controller clearly demonstrate that CS can use a sampling rate much lower than the original one to obtain similar or even equivalent sampling performance. It projects the original leakage signals onto the observation space, and obtains the observation samples far below the original dimension. For sampling, CS transfers a large amount of computation from sampling devices to advanced processors, so that the compute-intensive power trace reconstruction can be carried out fast without distortion. For storage, CS can compress the power traces at a higher compression ratio than classical compression, thus making the storage space more efficient. Experiments fully illustrate the advantages and practicability of compression sampling and storage of power traces.

In this paper, we only introduce the basic techniques of CS for leakage sampling and verify its effectiveness 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. Moreover, this paper only considers the compressed sampling of chips whose leakage can be well concentrated in the low-frequency part of the signal. Nowadays, many complex chips and devices induce leakages in the high-frequency part of signals. The application of CS on these devices is also an interesting research direction. As such, we believe the work in this paper provides a new research direction in SCA which has many avenues for investigations and opportunities for further improvements.

References

[1] M. Alam and D. Mukhopadhyay. How secure are deep learning algorithms from side-channel based reverse engineering? In *Proceedings of the 56th Annual Design Automation Conference 2019, DAC 2019, Las Vegas, NV, USA, June 02-06, 2019*, pages 226:1–226:2, 2019.

[2] E. Brier, C. Clavier, and F. 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.

[3] E. Cagli, C. Dumas, and E. 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.

[4] E. Cagli, C. Dumas, and E. 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.

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

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

[7] S. Chari, J. R. Rao, and P. Rohatgi. Template attacks. In *Cryptographic Hardware and Embedded Systems - CHES 2002, 4th International Workshop, Redwood Shores, CA, USA, August 13-15, 2002, Revised Papers*, pages 13–28, 2002.

[8] S. S. Chen, D. L. Donoho, and M. A. Saunders. Atomic decomposition by basis pursuit. *SIAM Review*, 43(1):129–159, 2001.

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

[10] A. A. Ding, C. Chen, and T. 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.

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

[12] F. Durvaux and F. 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.

[13] C. H. Gebotys, S. 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.

[14] D. Genkin, L. Pachmanov, I. Pipman, and E. Tromer. ECDH key-extraction via low-bandwidth electromagnetic attacks on pcs. In *Topics in Cryptology - CT-RSA 2016 - The Cryptographers' Track at the RSA Conference 2016, San Francisco, CA, USA, February 29 - March 4, 2016, Proceedings*, pages 219–235, 2016.

[15] D. Genkin, A. Shamir, and E. Tromer. RSA key extraction via low-bandwidth acoustic cryptanalysis. In *Advances in Cryptology - CRYPTO 2014 - 34th Annual Cryptology Conference, Santa Barbara, CA, USA, August 17-21, 2014, Proceedings, Part I*, pages 444–461, 2014.

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

[17] P. C. Kocher, J. Jaffe, and B. 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.

[18] J. Longo, E. D. Mulder, D. Page, and M. Tunstall. Soc it to EM: electromagnetic side-channel attacks on a complex system-on-chip. In *Cryptographic Hardware and Embedded Systems - CHES 2015 - 17th International Workshop, Saint-Malo, France, September 13-16, 2015, Proceedings*, pages 620–640, 2015.

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

[20] S. Mangard, E. Oswald, and T. Popp. *Power analysis attacks - revealing the secrets of smart cards*. Springer, 2007.

[21] A. Moradi, B. Richter, T. Schneider, and F. Standaert. Leakage detection with the x2-test. *IACR Trans. Cryptogr. Hardw. Embed. Syst.*, 2018(1):209–237, 2018.

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

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

[24] E. Oswald. Enhancing simple power-analysis attacks on elliptic curve cryptosystems. In *Cryptographic Hardware and Embedded Systems - CHES 2002, 4th International Workshop, Redwood Shores, CA, USA, August 13-15, 2002, Revised Papers*, pages 82–97, 2002.

[25] C. Ou, D. Sun, Z. Wang, X. Zhou, and W. Cheng. Manifold learning towards masking implementations: A first study. *IACR Cryptology ePrint Archive*, 2017:1112, 2017.

[26] J. Park, X. Xu, Y. Jin, D. Forte, and M. Tehranipoor. Power-based side-channel instruction-level disassembler. In *Proceedings of the 55th Annual Design Automation Conference, DAC 2018, San Francisco, CA, USA, June 24-29, 2018*, pages 119:1–119:6, 2018.

[27] C. Rechberger and E. Oswald. Practical template attacks. In *Information Security Applications, 5th International Workshop, WISA 2004, Jeju Island, Korea, August 23-25, 2004, Revised Selected Papers*, pages 440–456, 2004.

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

[29] S. K. Sahoo and A. Makur. Signal recovery from random measurements via extended orthogonal matching pursuit. *IEEE Trans. Signal Processing*, 63(10):2572–2581, 2015.

[30] Y. Sakai and K. Sakurai. Simple power analysis on fast modular reduction with NIST recommended elliptic curves. In *Information and Communications Security, 7th International Conference, ICICS 2005, Beijing, China, December 10-13, 2005, Proceedings*, pages 169–180, 2005.

[31] Y. Souissi, M. Nassar, S. Guilley, J. Danger, and F. 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.

[32] F. Standaert and C. 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.

[33] A. Wang, C. Wang, X. Zheng, W. Tian, R. Xu, and G. Zhang. Random key rotation: Side-channel countermeasure of NTRU cryptosystem for resource-limited devices. *Computers & Electrical Engineering*, 63:220–231, 2017.

[34] J. Wang, S. Kwon, and B. Shim. Generalized orthogonal matching pursuit. *IEEE Trans. Signal Processing*, 60(12):6202–6216, 2012.

Changhai Ou received his B.S. degree in Computer Science and Technology from School of Computer and Information Technology, Beijing Jiaotong University, China. He received his Ph.D. degree in Cyber Security from Institute of Information Engineering, Chinese Academy of Sciences (i.e. School of Cyber Security, University of Chinese Academy of Sciences) in July, 2018. He is now a Research Fellow in Hardware & Embedded Systems Lab (H-ESL), School of Computer Science and Engineering, Nanyang Technological University, Singapore.

His current research interests include cryptographic hardware and embedded system security, side channel attacks, machine learning and privacy preserving.

Chengju Zhou received M.S. degree in School of Computer Science and Technology from Tianjin University, China in 2015. He is currently working toward the PhD degree in the School of Computer Science and Engineering in Nanyang Technological University, Singapore.

His current research interest focuses on object detection for urban traffic scene understanding.

Siew-Kei Lam received his BASc, MEng and PhD from School of Computer Engineering, NTU. He is currently an Assistant Professor in the School of Computer Science and Engineering (SCSE), NTU and his research focuses on devising custom computing techniques to meet the increasingly challenging demands of performance, energy-efficiency, cost, and security in embedded systems. His current projects include architecture-aware algorithms for vision-enabled sensing, design methodologies for secure embedded systems, and transportation analytics.