EMD-Based Denoising for Side-Channel Attacks and Relationships between the Noises Extracted with Different Denoising Methods

Mingliang Feng[1], Yongbin Zhou[1,⋆], and Zhenmei Yu[2]

[1] State Key Laboratory of Information Security,
Institute of Information Engineering, Chinese Academy of Sciences,
89-A, Mingzhuang Rd, Beijing, 100093, P.R. China
{fengmingliang,zhouyongbin}@iie.ac.cn
[2] School of Information Technology,
Shandong Womens University,
45, Yuhan Rd, Jinan, 250002, P.R. China
yuzhenmei@gmail.com

Abstract. In essence, side-channel leakages produced during the execution of crypto implementations are noisy physical measurements. It turns out that various noises contained in leakages have, in general, negative effects on the key-recovery efficiency of side-channel attacks. Therefore, in practice, frequency-based denoising methods are presented and in wide use nowadays. However, most of them for reducing noises of high-frequency are not always effective, and they sometimes do little or even no help. On the other hand, the relationship between noises extracted with different denoising methods that target different frequencies, in time-domain, is not being discussed, which in turn will determine the potential power of combining these denoising methods. Motivated by this, we present two empirical mode decomposition (EMD) based denoising methods for side-channel attacks, and study their effectiveness in reducing noises of high frequency in real power traces. Compared with their counterparts, EMD-based denoising methods achieve both effectiveness and stability. Furthermore, we investigate the relationships between the noises extracted with denoising methods that target different frequencies, by performing attacks on real power traces denoised by multiple combinations of different denoising methods. For this purpose, we define the notion of overlapping coefficient, which measures how much that noises are overlapped with each other. Our results and observations are evidently verified by correlation power analysis attacks on multiple real power traces sets.

Keywords: Side-channel Cryptanalysis, Correlation Power Analysis, Empirical Mode Decomposition, Noise Reduction, Overlapping Coefficient.

⋆ Corresponding author.

S. Qing et al. (Eds.): ICICS 2013, LNCS 8233, pp. 259–274, 2013.

1 Introduction

Side-channel attack (SCA) aims at recovering the secret information embedded in a crypto devices from its physical leakages, including execution time[15], power consumption[1], and electromagnetic emanation[16]. Among those, power analysis attack which uses the instantaneous power consumption of crypto devices as its side-channel leakage is one of the most widely researched powerful side-channel attacks.

Generally, SCA consists of two stages: leakage acquisition and leakage exploitation. Concerning the latter, a number of power analysis attacks have been proposed so far, which are also referred to as distinguishers. Among them, differential power analysis (DPA) [1] is the most original one, which was then extended to other more powerful variants such as correlation power analysis (CPA) [2]. CPA is an effective method for finding the secret key based on the correlation between the hypothetical power consumption and the actual power consumption. Recent work [8] shows that side-channel distinguishers are not only asymptotically equivalent, but also can be rewritten one in function of the other only by modifying the power model. In particular, they have established one equivalence between most univariate side-channel distinguishers and CPA performed with different leakage models.

Even though the main focus of SCA is leakage exploitation, leakage acquisition also plays a critical role, as acquisition itself is the physical requisite for mounting power analysis attacks. The outputs of acquisition process are often referred to as power traces. Because of the electronic characteristics of the physical implementation, power traces always contain not only useful side-channel information which benefits cryptanalysts, but also a variety of noises which are found to have negative effects on side-channel attacks [14]. Therefore, to reduce noises inherent in power traces is commonly believed to be, in general, an effective approach enhancing the performance of power analysis attacks.

Up to now, a number of noise reduction methods have been proposed to reduce noises contained in power traces after sampling, i.e. to increase the signal-to-noise ratio (SNR). Generally speaking, those denoising methods can be roughly divided into two categories: frequency-based and non frequency-based. Frequency-based methods are the most popular one and in wide use in practice, which include wavelet-based methods [3] [4] and trend removal method (TR) [5]. Wavelet-based methods mainly target noise components of high frequency, while TR mainly targets that of low frequency. In [3], one applies wavelet transform to original power traces from a hardware implementation of unprotected DES on smart card, producing an approximation sub-signal. Afterwards, one performs DPA on the approximation sub-signal. In [4], one also applies wavelet transform into original power traces to obtain the approximation sub-signal and the detail sub-signal. The difference between [3] and [4] is that the latter sets a specific threshold value for the detail sub-signal, while the former sets the detail coefficients that dissatisfy the threshold to zero. Afterwards, one reconstructs the power traces and then performs power analysis attacks on the reconstructed power traces. Principal component analysis (PCA) [6] belongs to the

non frequency-based method, because it identifies trends in a whole trace set instead of a single trace. In [6], one applies PCA to original power traces, and then performs a DPA attack on a PCA-transformed power traces. Actually, the effects of PCA in practical attacks against hardware crypto implementations like that used in DPA contest v2 are very limited, and sometimes even negative [5]. However, we focus on frequency-based methods only.Frequency-based methods are most frequently used, yet there is one technical drawback: they are not always effective, and sometimes they do little or even no help, in practice. This drawback is again confirmed by one recent work of [5]. Therefore, a very natural and pertinent question arises at this point, namely, is there any effectively stable and easy-to-use denoising method dealing with high frequency noises? Another problem relates to the combination of denoising methods that target different frequencies [5]. Noise components of different frequencies will locate at distinct places with frequency domain, and they will overlap to some extent with each other in time domain. Then, how much is this overlap? This problem makes sense, because power analysis attacks examine the leakages in time domain. And this also determines how best the combination of different denoising methods would be.

Main contributions of this paper are two-fold. Firstly, we present two empirical model decomposition (EMD) [9] based denoising methods that target noises of high frequency for SCA, and address some technical issues concerning their applications. Both of these methods achieve effectiveness and stability. Secondly, we study the relationship of the noises extracted with different denoising methods. For this purpose, we define the notion of overlapping coefficient, which measures how much that noises are overlapped with each other.

The rest of this paper is organized as follows. Section 2 briefly introduces some background knowledge. Section 3 presents EMD-based denoising Methods for power analysis attacks in practice. Section 4 discusses the relationship between noises extracted with different denoising methods. Section 5 presents our experiments against real power traces from two kind of typical crypto implementations. Section 6 concludes the whole paper.

2 Preliminaries

In this section, we will present some basic knowledge, including composition of power traces, the general relationship between SNR and CPA, and EMD-based denoising methods in signal processing.

2.1 Composition of Power Trace

Power analysis attacks exploit the fact that the power consumption of cryptographic modules is correlated to the operations performed and the data processed. For each single point of a power trace, we denote the operation-dependent component by P_{op}, the data-dependent component by P_{data}. Due to the characteristics of the physical implementation, the power measurements are not always

the same even if the operation performed and data manipulated are fixed. We refer to this noise component of power consumption as $P_{el.noise}$. Besides these three components, each point in a power trace also has a constant component denoted by P_{const} (which is, for example, caused by leakage currents). Therefore, we can define each point of a power trace by (1).

$$P = P_{op} + P_{data} + P_{el.noise} + P_{const} \tag{1}$$

Given the fact that different power analysis attacks often exploit different properties of P_{op} and P_{data}, we refer to the components that exploited by a given attack as P_{exp}. And we refer to the rest part that is not exploitable of P_{op} and P_{data} combined with $P_{el.noise}$ as P_{noise}. So we can rewrite (1) to (2) in a given attack scenario.

$$P = P_{exp} + P_{noise} + P_{const} \tag{2}$$

2.2 General Relationship between SNR and CPA

SNR is the signal to noise ratio. Under our assumption and in a given attack scenario, SNR of a set of power traces at a fixed point is given by (3), in which $var(x)$ denotes the variance of x.

$$SNR = \frac{var(P_{exp})}{var(P_{noise})} \tag{3}$$

SNR quantifies the amount of information that leaks from a point of a set of power traces. The equation $\rho(H_i, P) = \rho(H_i, P_{exp})/\sqrt{1 + 1/SNR}$ [14] shows the relationship among the correlation coefficient $\rho(H_i, P)$ between the hypothetical power consumption values and the real power consumption values, the correlation coefficient $\rho(H_i, P_{exp})$ between the hypothetical power consumption values and the real side-channel leakages and SNR. It can be seen that the increase of SNR can effectively enhance the value of $\rho(H_i, P)$ with a given power traces. Besides this, in [14] the number of power traces needed to break a cryptographic implementation by CPA which is referred to as n can be estimated by (4),

$$n = 3 + 8\frac{Z_{1-\alpha}^2}{ln^2\frac{1+\rho(H_{ck},P)}{1-\rho(H_{ck},P)}} \tag{4}$$

where $Z_{1-\alpha}$ is a quintile of a normal distribution for a 2-sided confidence interval with error $1 - \alpha$. From the above formulas (3) and (4) it can be easily deduced that with the decrease of SNR, the traces number n will become bigger, and the attack will become more difficult. In order to improve the performance of power analysis attacks on given traces, attackers have to reduce the noise part P_{noise} in power traces as much as possible to enhance SNR.

2.3 Empirical Mode Decomposition and EMD-Based Denoising

In this section, we will introduce the empirical mode decomposition (EMD) method [9], and then describe two typical EMD based denoising methods: conventional EMD denoising and iterative EMD interval thresholding denoising.

Empirical Mode Decomposition in Signal Processing

The EMD method is an algorithm for the analysis of multicomponent signal [10] that breaks them down into a number of amplitude and frequency modulated (AM/FM) zero-mean signals, termed intrinsic mode functions (IMFs). In contrast to conventional decomposition methods such as wavelets, which perform the analysis by projecting the signal under consideration onto a number of predefined basis vectors, EMD expresses the signal as an expansion of basis functions that are signal-dependent and are estimated via an iterative procedure called sifting. Next we will give EMD a brief description and notation.

EMD [9] adaptively decomposes a multicomponent signal [10] $x(t)$ into a number L of the so-called IMFs $I^{(i)}(t)$ and a remainder $d(t)$ as formula (5). Here $d(t)$ is a remainder that is non-zero-mean slowly varying function with only few extrema. Each one of the IMFs, say, the ith one $I^{(i)}(t)$, is estimated with the aid of an iterative process, called sifting, applied to the residual multicomponent signal $x^{(i)}(t)$.

$$x(t) = \sum_{i=1}^{L} I^{(i)}(t) + d(t) \qquad 1 \le i \le L \tag{5}$$

$$x^{(i)}(t) = \begin{cases} x(t) & i = 1 \\ x(t) - \sum_{j=1}^{i-1} I^{(j)}(t) & i \ge 2 \end{cases} \tag{6}$$

The sifting process used in this paper is the standard one [9]. According to this, during the $(n+1)th$ sifting iteration, the temporary IMF estimate $I_n^{(i)}(t)$ is improving according to the following steps. [1]

1) Find the local maxima and mimima $I_n^{(i)}(t)$

2) Interpolate, using natural cubic splines, along the points of $I_n^{(i)}(t)$ estimated in the first step in order to form an upper and a lower envelope

3) Compute the mean of the two envelopes $m_n^{(i)}$

4) Obtain the refined estimate $I_{n+1}^{(i)}(t)$ of the IMF by subtracting the mean $m_n^{(i)}$ found in the previous step from the current IMF estimate $I_n^{(i)}(t)$.

5) Check whether a stopping criterion has been fulfilled. If not, proceed from 1) again

Supposing the procedure above runs N times before we getting the ith IMF $I^{(i)}(t)$, then the following formula must be fulfilled.

$$I^{(i)}(t) = x^{(i)}(t) - \sum_{j=1}^{N} m_j^{(i)} \tag{7}$$

What's more all IMFs have the following properties:

1) Zero mean

2) All the maxima and all the minima of $I^{(i)}(t)$ will be positive and negative respectively

[1] For the first iteration, $x^{(i)}(t)$ is used as temporary IMF estimate $I_1^{(i)}(t)$.

Often, but not always, the IMFs resemble sinusoids that are both amplitude and frequency modulated. By construction, the number of, say, $N(i)$ extrema of $I^{(i)}(t)$ positioned at time instances $r^{(i)} = [r_1^{(i)}, r_2^{(i)}, ..., r_{N(i)}^{(i)}]$ and the corresponding IMF points $I^{(i)}(r_j^{(i)}), j = 1, ..., N(i)$ will alternate between maxima and minima. As a result, in any pair of extrema, $z_j^{(i)} = [I^{(i)}(r_j^{(i)}), I^{(i)}(r_{j+1}^{(i)})]$ corresponds to a single zero-crossing interval. Whats more, each IMF occupies lower frequencies locally in the time-frequency domain than its preceding ones. Fig. 1 presents an example the EMD of a real noisy trace signal(Fig. 1(a)), and this EMD process results in seven IMFs and a final remainder(Fig. 1(b)-(i)).

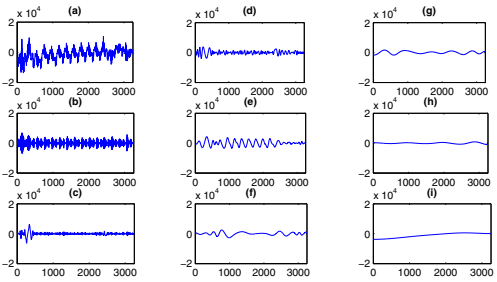


Fig. 1. A Real Noisy Trace(shown in (a)) and its EMD Components (shown in (b)-(i))

Conventional EMD Denoising

The conventional EMD denoising (EMD-Conv) method here refers to the method in [11]. The main idea is to discard the IMFs of which the main components are noises. And it is usually considered that noises exist mainly in the high frequency domain, In other words, it exists in the first few IMFs.

$$\widetilde{x}(t) = \sum_{i=M_1}^{L} I^{(i)}(t) + d(t) \tag{8}$$

In the above formula, $\widetilde{x}(t)$ is the signal after the noise reduction, and M_1 can be determined in the way that used in [11], and it can be described as below.

1) Calculate the actual IMF energies using a robust estimator based on the components median [12]

$$E_k = \left(\frac{median(I^{(k)}(t))}{0.6745} \right)^2 \qquad k = 1, 2, 3 \ldots \tag{9}$$

2) Calculate the noise-only IMF energies. And they can be approximated according to

$$\widetilde{E_k} = \frac{E_1}{\beta}\rho^{-k} \qquad k = 2, 3, 4 \ldots \tag{10}$$

where E_1 is the energy of the first IMF and β, ρ are parameters that for a specific EMD implementation, depend mainly on the number of sifting iterations used. It is suggested in [11] that setting β and ρ to be 0.719 and 2.01 respectively is a good choice. This paper also adopts this choice.

3) Compare the energies from the first IMF between the actual and the theoretical ones. If the energies significant diverge from each other at the ith IMF, indicating the presence of significant amounts of no-noise signal, then we can assign i to the parameter M_1.

Fig. 2 is an example of using conventional EMD denoising method on a noisy signal, where the blue line is the original noisy signal and the red line represents the denoised one.

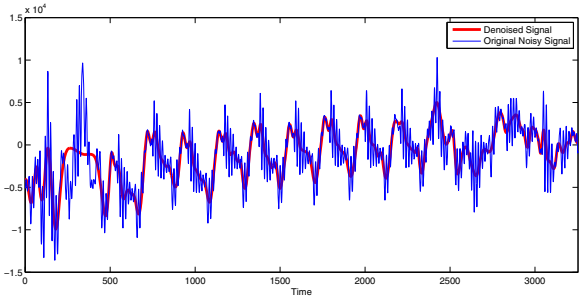


Fig. 2. Conventional EMD Denoising

Iterative EMD Interval Thresholding Denoising

The iterative EMD interval thresholding (EMD-IIT) denoising method was first introduced in [13]. The main idea of it is to enhance the tolerance against noise by averaging a number of denoised versions of the signal which are obtained iteratively. Before introducing the EMD-IIT method, we should get an understanding of the EMD interval thresholding method (EMD-IT), which is also introduced in [13]. The main idea of the EMD-IT is to reconstruct the denoised signal by

$$\widetilde{x}(t) = \sum_{i=M_1}^{M_2} \bar{I}^{(i)}(t) + \sum_{i=M_2+1}^{L} I^{(i)}(t) + d(t) \tag{11}$$

Here the $\bar{I}^{(i)}(t)$ is calculated as the formula (12)

$$\bar{I}^{(i)}(z_j^{(i)}) = \begin{cases} I^{(i)}(z_j^{(i)}) & |I^{(i)}(r_j^{(i)})| > T_i \\ 0 & |I^{(i)}(r_j^{(i)})| \leq T_i \end{cases} \tag{12}$$

where $T_i = C\sqrt{\widetilde{E}_i 2lnN}$ and C is a constant. For consant C, choosing one value between 0.6 and 0.8 is usually a good choice [13]. $\widetilde{E}_i$ is calculated using (10). N is the sample number of the signal. Now we can go to an in-depth study of EMD-IIT, it can be summarized as the following steps.

1) Perform an EMD expansion of the original noisy signal x.

2) Perform a partial reconstruction using the last $L-1$ IMFs and the remainder only, $x_p(t) = \sum_{i=2}^{L} I^{(i)}(t) + d(t)$

3) Randomly alter the sample positions of the first IMF $I_a^{(1)}(t) = Alter(I^{(1)}(t))$.

4) Construct a different noisy version of the original signal $x_a(t) = x_p(t) + I_a^{(1)}(t)$

5) Perform EMD on the new altered noisy signal $x_a(t)$.

6) Perform the EMD-IT denoising using formula (12) on the IMFs of $x_a(t)$ to obtain a denoised version $\widetilde{x}_1(t)$ of x

7) Iterate $K-1$ times between 3)-6), where K is the number of averaging iterations in order to obtain K denoised versions of x , i.e., $\widetilde{x}_1$, $\widetilde{x}_2$, , $\widetilde{x}_K$.

8) Average the resulted denoised signals $\widetilde{x}(t) = (1/K)\sum_{k=1}^{K} \widetilde{x}_k(t)$.

The altering function can take several forms, in this paper we use random permutation approach recommended by [13], in other words, the samples of the first IMF change their positions randomly.

3 EMD-Based Denoising Methods for SCA

Frequency-based denoising methods are in wide use nowadays. However, most of them for reducing noises of high-frequency are not always effective, and sometimes they do little or even no help. Therefore, for practical purpose, a more effectively stable method that target noise of high frequency is highly desirable. To address this problem, we introduce two EMD-based denoising methods into the filed of SCA. Applications of EMD-based denoising in power analysis attacks involves some technical issues and appear to be tricky, even though these methods are relatively mature in the field of signal processing.

Before talking about how to use two typical EMD-based denoising methods in power analysis attacks, we would define some key parameters concerned, which are summarized in Table 1. First of all, we present conventional EMD-Conv for SCA and show how it works. In this case, one applies EMD-Conv denoising to every single trace contained in the trace set, and then produce a new trace set. The corresponding process is shown in Algorithm 1.

Table 1. Definition of Parameters Used in EMD-Based Denoising Methods

Parameter Name	Description of Parameter
$Traceset$	a set of power traces
$Trace_i$	the i_{th} power trace in the traces set $Traceset$
$Tracenum$	the number of traces in the traces set
$Traceset'$	a set of new power traces that have been denoised
$Trace_i'$	a denoised trace that generated by $Trace_i$
M_1	the starting order of IMF that is used to reconstruct a denoised signal
IM_2	$M_2 = L - IM_2$, meaning the last $IM_2 - 1$ IMFs and the remainder do not get thresholded
C	it is a ocnstant coefficient of getting T_i, $T_i = C\sqrt{\widetilde{E_i} 2 \ln N}$
$Siftnum$	maximum number of the sifting progress to get an IMF in EMD
$Iteration$	the average number to get a denoised signal in EMD-IIT

Algorithm 1. EMD-Conv for SCA

Input: $Traceset, M_1, siftnum$
Output: $Traceset'$

```
1: function EMDConvforSCA( Traceset, M1, siftnum)
2:     i ← 1
3:     while i <= Tracenum do
4:         Trace_i' ← EMD − Conv(Trace_i, M1, siftnum)
5:     end while
6:     return Traceset'
7: end function
```

In Algorithm 1, M_1 can de determined according to the method introduced in section 2.3, and $siftnum$ is an optional parameter. If $siftnum$ is not set, the sifting progress will not end until a default stopping criterion has been fulfilled, which would be very time-consuming. Therefore, in practice, choosing one value between 10-16 of $siftnum$ is a good balance between effectiveness and time-efficiency.

Next, we will introduce more effective EMD-IIT method into SCA and show how it works. The EMD-IIT transformation from one original dataset of power traces into a new dataset, shown in Algorithm 2, is the same as that in EMD-Conv transformation. Contrary to the case of EMD-Conv, in this case,. according to [13], it has been empirically found that a very good choice of M_1 is given by $M_1 = max(1, J-2)$, where J is the order that used in EMD-Conv as a starting order to reconstruct a denoised signal. Usually, a good choice of M_2 is $L-1$. In other words, the last IMF and the remainder do not get thresholded. For parameter C, the values between 0.6 and 0.8 is often the best choice, but not always. In general, a balanced tradeoff between the number of sifting ($siftnum$) and the performance of EMD-IIT is realized with about eight sifting iterations. The final parameter $Iteration$ can be set to a value between 10 and 20. Note that, unlike using EMD-IIT in the field of signal processing, for EMD-IIT to

be correctly used in SCA, all traces must use the same permutation matrix, or it will lead to a problem of power trace misalignment that would decrease the performance of the attack or even worse make it fail.

Algorithm 2. EMD-IIT for SCA

Input: $TracesetM_1IM_2, C, Siftnum, Iteration$
Output: $Traceset^{'}$
1: **function** EMDIITFORSCA($Traceset, M_1, IM_2, C, Siftnum, Iteration$)
2: Generate a random permutation matrix $pm(Iteration * |trace|)$ according to the parameter iteration and the length of a Trace
3: $i \leftarrow 1$
4: **while** $i <= Tracenum$ **do**
5: $Trace_i^{'} \leftarrow EMD - IIT(Traceset, M_1, IM_2, C, Siftnum, Iteration, pm)$
6: **end while**
7: **return** $Traceset^{'}$
8: **end function**

Unlike the wavelet based denoising methods [3] [4] where one has to choose a wavelet basis function that affects the performance of denoising greatly, EMD based denoising methods are nonparametric. So in this respect, compared with wavelet based denoising methods, EMD based denoising methods are more easily used. Then how about the actual performance of EMD-based methods? We will study this issue through a series of experiments in section 5.

4 Relationship between Noises Extracted with Different Denoising Methods

Intuitively, the combination of denoising methods that target different frequencies will be more effective than any one of them [5]. In this section, we will examine the overlap between these noise components. And the overlap could reflect how best the combination will be in practice. Actually, this problem could also be naturally extended into the case of noise components extracted with different methods that target similar frequencies, as those of Wavelet-based and EMD-based methods.

From the perspective of set theory, there are three kinds of relationships between two sets A and B. In order to measure the overlap between two sets, we can use the formula (13), where $|set|$ is the number of elements contained in the set. If $d = 0$, then A and B do not intersect; if $0 < d < 1$, then A and B intersect, but they do not have a containment relationship; if $d = 1$, then they have a containment relationship, namely, A contains B or B contains A.

$$d = \frac{|A \cap B|}{min(|A|, |B|)} \tag{13}$$

Inspired by formula (13), we will use a similar idea in analyzing the different parts of a noisy leakages as shown in Fig. 3, where P_{exp} is the exploitable

component by a given attack, P_{noise1} is the noise extracted with a denoising method m_1, P_{noise2} is the noise extracted with another denoising method m_2. Then what is the relationship between P_{noise1} and P_{noise2}? In other words, how much the noises are overlapped with each other? Currently, there is no direct metric available to measure this overlap. Therefore, we try another indirect yet useful way. Specifically, we define the notation of overlapping coefficient, which could serve as a quantitative metric to measure the overlap rate between two noise components. The definition of overlapping coefficient is based on success rate (SR)[7], and is shown in formula (14), where ΔSR_1 is the improvement of SR achieved by m_1 on a given number of power traces compared with that acquired on the raw power traces; similarly, ΔSR_2 is achieved by m_2 and ΔSR_3 is achieved by the combination use of m_1 and m_2. In practice, this indirect quantitative metric could well reflect the relationship of the noises, which is verified by the experiments in Section 5.

$$oc = \frac{\Delta SR_1 + \Delta SR_2 - \Delta SR_3}{min(\Delta SR_1, \Delta SR_2)} \tag{14}$$

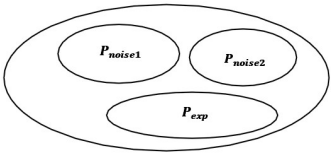


Fig. 3. Components of Noisy Side-Channel Leakages

As per its definition, it always holds that $0 \leq oc \leq 1$.Apprently, the smaller the value of oc is, the better the effect of the combination. If $oc \approx 1$, it indicates that the two denoising are almost of the same capability to remove the noise extracted with the method that generates a smaller ΔSR. In this case, using the method which makes a higher improvement of SR alone is enough, and using the combination of the two does not help.

5 Experiments

In this section, we will firstly examine the stability and effectiveness of the two EMD-based denoising methods for SCA(EMD-Conv and EMD-IIT respectively) in eliminating noise of high frequency, by performing a series of CPA attacks on real power traces from the second stage of DPA Contest and PowerSuite 4.0 (one software bechnark evalution board we designed and developed ourselves, and its CPU is an 8-bit microcontroller STC89C58RD+). And then we will explore the potential of combination of different denoising methods and study the overlapping relationship of the noises extracted with different denoising methods by performing CPA attacks on the DPA Contest v2 traces.

5.1 Settings

Hardware Implementation
The traces from the DPA Contest v2 are acquired with a sampling rate of 5G sample/s from a SASEBO-GII board, which implements an unprotected hardware AES implementation over a Xilinx Virtex-5 FPGA. Ideally, it is better to perform the denoising methods on all 32 traces sets from DPA Contest v2 public database to evaluate their performance. However, in actual cases, it is too time consuming to perform this. Therefore, we turn to another way of randomly choosing eight datasets of power traces from DPA Contest v2 public database. And then, we target the last round of the AES on these raw sets of power measurements to calculate SR on a given number of traces for the first S-box by mounting a CPA attack 500 times. The evaluation results are shown in Fig. 4. After these, we choose dataset1, which matches the average and median of the eight different SRs best, as the representative dataset to analyze.

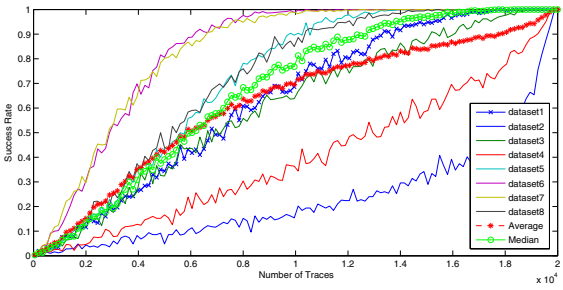


Fig. 4. Results of CPA Attacks on Eight Sets of Traces from DPA Contest v2

Software Implementation
The other three traces sets are acquired with a sampling rate of 50 M sample/s from PowerSuite 4.0 board, which contains an unprotected software AES implementation. For simulating different levels of SNR, the traces sets differ only in the average number during the sampling process. The average number is one time, four times and eight times, respectively.

Next, we will use SR to evaluate the stability and effectiveness of each denoising method or their combinations, by mounting CPA key recovery attacks 500 times for the traces set from DPA Contest v2 and 1,000 times for the traces sets from PowerSuite 4.0. For clarity, we name the denoising method in [3] as Wavelet and another denoising method in [4] as Wavelet1. And all the description of the experiments' labels are shown in Table 2. Note that our experiments' results shown that the combination order of different denoising methods has little influence on the final results.

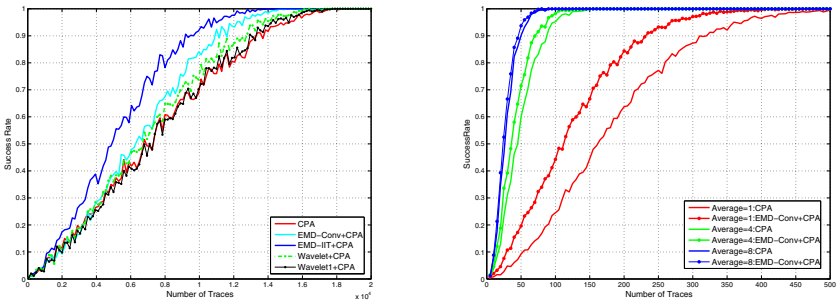
Table 2. Description of Our Experiments

Experiment Label	Description of the Experiment
CPA	perform CPA attacks on the original power traces
EMD-Conv+CPA	perform EMD-Conv method to the original power traces, and then perform CPA attacks on the resultant power traces
EMD-IIT+CPA	perform EMD-IIT method to the original power traces, and then perform CPA attacks on the resultant power traces
Wavelet+CPA	perform wavelet transform in [3] to the original power traces, and then perform CPA attacks on the resultant power traces
Wavelet1+CPA	perform wavelet transform in [4] to the original power traces, and then per-form CPA attacks on the resultant power traces
TR+CPA	perform detrending method in [5] to the original power traces , and then perform CPA attacks on the resultant power traces
EMD-IIT+TR+CPA	remove noise in the original power traces using EMD-IIT, perform detrending method in [5] to the resultant power traces, and then perform CPA attacks on the final power traces.
EMD-IIT+Wavelet+CPA	remove noise in the original power traces using EMD-IIT, perform wavelet transform in [3] to the resultant power traces, and then perform CPA attacks on the final power traces.
EMD-IIT+Wavelet1+CPA	remove noise in the original power traces using EMD-IIT, perform wavelet transform in [4] to the resultant power traces, and then perform CPA attacks on the final power traces.

5.2 Results and Analysis

Firstly, we evaluate the stability and effectiveness of the two EMD-based denoising methods on the trace set from DPA Contest v2. The results are shown in Fig. 5(a). From Fig. 5(a), it is shown that both EMD-Conv and EMD-IIT denoising methods are capabale of improving the SRs of CPAs effectively. With repsect to achieving a partial stable SR of 80%, compared with CPA which needs 12,050 traces, EMD-Conv+CPA needs 9,350, which reduces the traces needed by 22.4%. EMD-IIT+CPA works even better than EMD+CPA, and it needs only 8,150 traces, gaining an improvement of 32.3%. Meanwhile, the Wavelet-Based methods used to remove noise of high frequency do little or even no help. Specifically, Wavelet reduces the trace number less than 10%, and Wavelet1 less than 1%. That is to say, in our case, the EMD-based methods are more effective than the Wavelet-based ones.

After the effectiveness of EMD-based methods have been proved, we would like to further study their stability and performance under different SNRs. Since under this scenario, the SNRs are relatively high compared with that of DPA Contest v2, the performances of EMD-Conv and EMD-IIT are almost the same. Therefore, in this part, we only focus on EMD-Conv which is more time efficient. As is shown in Fig. 5(b), with the increase of SNR (or average times), the percentage of the decrease of trace number to achieve a partial success rate of 80% becomes smaller and smaller, from 26.9% to 14.3% to less than 2%. This phenomena can be explained like this: with the increase of SNR, noises of high frequency also decrease. In this case, the performance decrease of the EMD-Based methods is reasonable. From another perspective, though the performance of EMD-Based methods is not always significant, they can always remove noises of high frequency, which is the evidence of their stability.



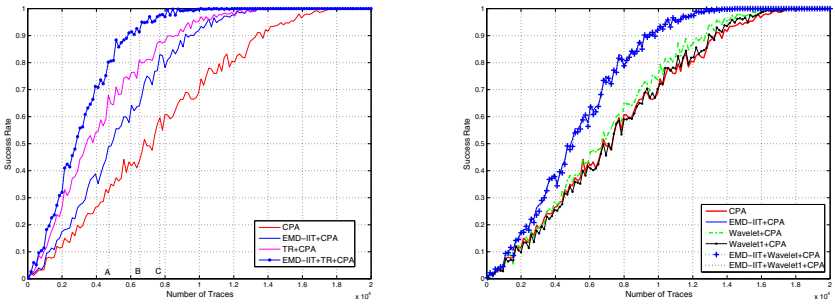
(a) (b)

Fig. 5. (a)SRs of EMD-Based and Wavelet-based Denoising on Traces from DPA Contest v2 (b)SRs of EMD-Conv Denoising on Traces with Different Noise Levels

Next, we examine the potential power of combining denoising methods and study the relationship of the noises reduced with different denoising methods on the traces set from DPA Contest v2. Firstly, we make a combination of two denoising methods that target different frequencies: one is EMD-IIT which reduces mainly noise of high frequency, and another is TR [5] which reduces the trend noise of low frequency. The results are shown in Fig. 6(a). The SR is improved greatly on a given traces number when we make a combination of EMD-IIT and TR. In terms of achieving a partial success rate of 80%, this combination reduces as many as 58.5% of the traces that needed before denoising. To study the relationship between the noises reduced by EMD-IIT and TR, we choose three numbers, A, B and C, on the abscissa axis first. Given the number of each trace set, the SR of one denoising strategy reaches 80% or little more, i.e., the SR of EMD-IIT+TR+CPA reaches 80% when given A traces. Then we calculate the overlapping coefficient for using three different number of traces respectively, The results are shown in Table 3, where ΔSR_1 is for EMD-IIT, ΔSR_2 is for TR, and ΔSR_3 is for their combination. For more accuracy, we calculate the mean value of the three overlapping cofficients, and the result is 0.6136, meaning that about 61% noise that extracted by EMD-IIT can also be extracted by TR. Secondly we make a combination between denoising methods that target noises of high frequency and the results are shown in Fig. 6(b). Clearly, these combinations do little help in reducing noise, and the overlapping coefficient values calculated by (14) are both very close to zero, which means that EMD-IIT can remove almost all the noises extracted by the Wavelet-based methods. So in these cases, choosing a more effective one, i.e. EMD-IIT, will be more reasonable. Based on the analysis of the above experiments, combination of denoising methods that target noise of different frequencies, may improve the denoising performance a lot. As to the combination of denoising methods that target the same frequency domain, it usually makes little or no improvement in removing the noise. Therefore, in this scenario, choosing a better one alone is enough.

Table 3. Overlapping Coefficients for Different Number of Traces

ΔSR and oc / Traces number	ΔSR_1	ΔSR_2	ΔSR_3	oc
A	0.142	0.330	0.406	0.4648
B	0.206	0.364	0.408	0.7864
C	0.234	0.284	0.380	0.5897



(a) (b)

Fig. 6. (a)SRs of the Combination of EMD-IIT and TR on the Traces from DPA Contest v2 (b)SRs of the Combination of EMD-IIT and Wavelet-bsed Methods on Traces from DPA Contest v2

6 Conclusions and Future Work

Reducing noise serves an important way to enhance the performance of the side-channel attacks. Considering the fact that frequency-based denoising methods are dominant and in wide use in practice and that most of existing of these methods suffers instability in performance enhancement, stable and effective frequency-based denoising methods make a lot of sense. In this paper, we proposed EMD-based denoising methods for use in side-channel attacks. Results of practical attacks against real power traces from two kind of typical crypto implementations (i.e. hardware and software implementations of AES) proves that these methods are superior to their counterparts (say, for example, Wavelet-based approaches). On the other hand, through a series of experiments of combination, it proves that the combination of methods that dealing with noises of different frequencies may improve the denoising performance a lot. At the same time we define the notion of overlapping coefficient, which is an indirect yet helpful quantitative metric to measure to what extent that noises extracted with different methods are overlapped with each other.

Additionally, EMD-based denoising methods seem not so good at dealing with side-channel leakages with high SNR. Therefore, the study of improvements of EMD-based methods in this case would be one of the relevant future works.

Acknowledgments. This work was supported in part by National Natural Science Foundation of China (No. 61272478, 61073178 and 61170282), Beijing Natural Science Foundation (No. 4112064), Strategic Priority Research Program of the Chinese Academy of Sciences (No.XDA06010701).

References

1. Kocher, P., Jaffe, J., Jun, B.: Differential Power Analysis. In: Wiener, M. (ed.) CRYPTO 1999. LNCS, vol. 1666, pp. 388–397. Springer, Heidelberg (1999)
2. Brier, E., Clavier, C., Olivier, F.: Correlation Power Analysis with a Leakage Model. In: Joye, M., Quisquater, J.-J. (eds.) CHES 2004. LNCS, vol. 3156, pp. 16–29. Springer, Heidelberg (2004)
3. Charvet, X., Pelletier, H.: Improving the DPA attack using Wavelet transform. In: Non-Invasive Attack Testing Workshop 2005 (2005)
4. Souissi, Y., Aabid, M., Debande, N., Guilley, S., Danger, J.: Novel Applications of Wavelet Transforms based Side-Channel Analysis. In: Non-Invasive Attack Testing Workshop 2011 (2011)
5. Cao, Y., Zhou, Y., Yu, Z.: On the Negative Effects of Trend Noise and Its Applications in Side-Channel Cryptanalysis, `http://eprint.iacr.org/2013/102.pdf`
6. Batina, L., Hogenboom, J., van Woudenberg, J.G.J.: Getting More from PCA: First Results of Using Principal Component Analysis for Extensive Power Analysis. In: Dunkelman, O. (ed.) CT-RSA 2012. LNCS, vol. 7178, pp. 383–397. Springer, Heidelberg (2012)
7. Standaert, F., Malkin, T., Yung, M.: A Unified Framework for the Analysis of Side-Channel Key Recovery Attacks. In: Joux, A. (ed.) EUROCRYPT 2009. LNCS, vol. 5479, pp. 443–461. Springer, Heidelberg (2009)
8. Doget, J., Prouff, E., Rivain, M., Standaert, F.X.: Univariate side channel attacks and leakage modeling. Journal of Cryptographic Engineering 1, 123–144 (2011)
9. Huang, N.E., et al.: The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. Proc. Roy. Soc. London A 454, 903–995 (1998)
10. Cohen, L.: Time-Frequency Analysis. Prentice- Hall, Englewood Cliffs (1995)
11. Flandrin, P., Rilling, G., Goncalves, P.: EMD equivalent filter banks, from interpetation to applications. In: Huang, N.E., Shen, S. (eds.) Hilbert-Huang Transform and Its Applications, 1st edn. World Scientific, Singapore (2005)
12. Mallat, S.: A Wavelet Tour of Signal Processing, 2nd edn. Academic, New York (1999)
13. Kopsinis, Y., McLaughlin, S.: Development of EMD-Based Denoising Methods Inspired by Wavelet Thresholding. IEEE Transactions on Signal Processing 57(4) (April 2009)
14. Mangard, S., Oswald, E., Popp, T.: Power Analysis Attacks: Revealing the Secrets of Smart Cards. Springer (2007)
15. Kocher, P.C.: Timing Attacks on Implementations of Diffie-Hellman, RSA, DSS, and Other Systems. In: Koblitz, N. (ed.) CRYPTO 1996. LNCS, vol. 1109, pp. 104–113. Springer, Heidelberg (1996)
16. Agrawal, D., Archambeault, B., Rao, J., Rohatgi, P.: The EM side-channel(s). In: Kaliski Jr., B.S., Koç, Ç.K., Paar, C. (eds.) CHES 2002. LNCS, vol. 2523, pp. 29–45. Springer, Heidelberg (2003)