How to Compare Selections of Points of Interest for Side-Channel Distinguishers in Practice?

Yingxian Zheng[1], Yongbin Zhou[1](✉), Zhenmei Yu[2], Chengyu Hu[3], and Hailong Zhang[1]

[1] State Key Laboratory of Information Security,
Institute of Information Engineering, Chinese Academy of Sciences,
89-A, Mingzhuang Rd, Beijing 100093, People's Republic of China
{zhengyingxian,zhouyongbin,zhanghailong}@iie.ac.cn

[2] Shandong Womens University, 45, Yuhan Rd,
Jinan 250002, People's Republic of China
yuzhenmei@gmail.com

[3] Shandong University, 27, Shanda South Rd,
Jinan 250100, People's Republic of China
hcy@sdu.edu.cn

Abstract. Side-channel distinguishers aim to reveal the secrets used in crypto devices by utilizing the subtle dependence between some sensitive intermediate values and physical leakages produced during its executions. For this purpose, one or more points of interest (POIs) corresponding to manipulations of one sensitive intermediate value are usually selected and then fed into distinguishers. However, it turns out in practice that POIs selected, even they are from the same leakage traces, will have significant impacts on the key recovery efficacy of distinguishers. Therefore, it makes a very practical sense to investigate the concrete impacts of POIs selections on side-channel distinguishers, and then pick out from those POIs selections available the most appropriate one for a certain distinguisher. In order to address these problems, we propose an evaluation framework for the analysis of POIs selections for side-channel distinguishers. Basically, our framework consists of two stages: the first stage captures the validity of points selected, while the second one reflects their quality with respect to a certain distinguisher. Specifically, on the one hand, in order to measure the goodness of one POIs selection, we introduce a quantitative metric of accuracy rate, from a perspective of statistics; on the other hand, we adopt the widely accepted security metric of success rate proposed by Standaert et al. at EUROCRYPT 2009 to reflect the quality of the points selected. Eventually, taking five typical POIs selections and three popular side-channel distinguishers as concrete study cases, we perform simulated attacks and practical attacks as well, the results of which not only fully justify our proposed methods but also reveal some interesting observations.

Keywords: Accuracy rate · Evaluation framework · Distinguisher · Selection of points of interest · Side-channel analysis

L.C.K. Hui et al. (Eds.): ICICS 2014, LNCS 8958, pp. 200–214, 2015.
DOI: 10.1007/978-3-319-21966-0_15

1 Introduction

Side-channel attacks aim at revealing the secret information embedded in a cryptographic device from its physical leakages, including execution time [1], power consumption [2], and electromagnetic emanation [3]. Among them, power analysis attack which makes use of instantaneous power consumptions of a cryptographic device is one of the most widely researched side-channel attacks. Therefore, for ease and simplicity of presentation, we concentrate on power analysis attack ONLY for illustrative purposes in this paper.

Side-channel distinguisher plays a crucial role in recovering reveal the secrets in side-channel attacks. It refers to the process during which the adversary uses some statistical tools to exploit the subtle dependence between one sensitive intermediate value and its corresponding power consumptions of cryptographic device. For real-world crypto implementations, one side-channel leakage trace usually contains multiple samples corresponding to manipulations of one sensitive intermediate value. This is quite natural because the manipulations of the sensitive intermediate value targeted usually takes more than one instruction cycle. In addition, according to Nyquist−Shannon sampling theorem, the acquisition rate of the signal acquisition device is always set to be several times faster than the working frequency of the targeted cryptography device. Those samples that exactly correspond to the manipulations of one sensitive intermediate value targeted in one leakage trace are referred to points of interest (POIs).

Based on analysis of values and of distributions, side-channel distinguishers can be divided into two categories. Distinguishers based on values include differential power analysis (DPA) [2], correlation power analysis (CPA) [4], differential cluster analysis (DCA) [5], template attack [6], stochastic method [7], and etc. Mangard et al. showed in [8] that denoted as standard univariate DPA, a number of these type of distinguishers are in fact asymptotically equivalent, given that they are provided with the same a priori information about the leakages. Therefore, in this paper, we choose CPA to be the representative of those distinguishers based on values. Distinguishers based on distributions consist of mutual information analysis (MIA) [9], KS-test based analysis (KSA) [17], MPC-KSA [10], and etc. Considering their popularity, we choose MIA and KSA to be the representatives of those distinguishers based on distributions.

Currently, there are several POIs selections available. In principle, side-channel attacks themselves could serve as the tools for POIs selection, as is already done in the field of side-channel attacks. For example, CPA, MIA and KSA all can be used to select the POIs. In addition, there are also non-attack based POIs selections. Two of them are the Sum Of Squared Pairwise Differences (sosd) [9] and the Sum Of Squared Pairwise T-Differences (sost) [11]. An important observation is that applying different POIs selections onto the same leakage traces could lead to distinct points selected, even if it is explicitly required that all POIs selected must correspond to one sensitive intermediate value targeted, which will have significant impacts on the key recovery efficacy of distinguishers. Therefore, it makes a very practical sense to investigate the concrete impacts of POIs selections on side-channel distinguishers, and then pick out from those POIs selections available the most appropriate one for a certain distinguisher.

For comparison of distinguishers, some well-known frameworks were already proposed. The first one in [13] by Standaert et al. suggests to use a leakage metric to qualify the maximal chance that an optimal attacker would have to extract the secrets. For the comparison of different distinguishers, [13] suggests metrics like o^{th}-order success rate or guessing entropy. In another framework of [14], the distance to the nearest rival is suggested. In [15], Maghrebi et al. proposed a methodology to compare two side-channel distinguishers based on simulations. In [16], some analyses showed pitfalls in the evaluation methodologies for distinguisher, including estimation bias, estimation algorithm, success rate error, and sample errors. To the best of our knowledge, all frameworks known so far concern distinguishers alone; and none of them take POIs selections themselves into serious consideration, let alone any comprehensive evaluation work about concrete impacts of POIs selections over distinguishers in practice.

1.1 Contributions

The contributions of this paper are threefold. First, we propose an two-stage evaluation framework for the analysis of POIs selections for side-channel distinguishers. Second, in order to measure the goodness of the POIs selection, we introduce the notion of accuracy rate. Third, taking five POIs selections commonly used and three typical distinguishers as concrete study cases, we perform simulated attacks and practical attacks. The experimental results not only fully justify our proposed methods, but also reveal some interesting observations.

The rest of this paper is organized as follows. Sect. 2 briefly recalls three typical distinguishers and five POIs selections commonly used; Sect. 3 introduces our proposed two-stage framework; Sect. 4 presents details and results of simulated and practical attacks, together with some useful discussions and interesting observations; Sect. 5 concludes the whole paper.

2 Preliminaries

This section will briefly recall CPA, MIA, and KSA distinguishers. These three distinguishers can also be used for selecting POIs. Besides, we will also briefly introduce sosd and sost POIs selections.

2.1 CPA

CPA identifies the correct key by calculating the Pearson correlation coefficient between real power traces and hypothetical power consumptions. The adversary chooses a sensitive intermediate value $v_i^* = g\left(x_i, k^*\right)$, where x_i is the ith plaintext (totally NT traces), k^* is a key guess. For every key guess k^*, the adversary predicates the hypothetical power consumption by $h_i^{k^*} = f\left(v_i^*\right)$, where f is a hypothetical leakage function. H^{k^*} denotes a vector of hypothetical power consumptions. L denotes a vector of real power traces. The adversary computes the Pearson correlation coefficient between H^{k^*} and L as

$$\rho(H^{k^*}, L) = \frac{\sum\limits_{i=1}^{NT} (H_i^{k^*} - \overline{H^{k^*}})(L_i - \overline{L})}{\sqrt{\sum\limits_{i=1}^{NT} \left(H_i^{k^*} - \overline{H^{k^*}}\right)^2 \cdot \sum\limits_{i=1}^{NT} \left(L_i - \overline{L}\right)^2}} \tag{1}$$

where $\overline{H^{k^*}}$ and $\overline{L}$ are the mean of H^{k^*} and that of L. Maximal correlation coefficient indicates the most likely candidate key guess as $k = \arg\max\limits_{k^*} \rho(H^{k^*}, L)$.

In practice, CPA can also be used for selecting POIs. The maximal correlation coefficient indicates the location of POIs as $[k,t] = \arg\max\limits_{k^*,t'} \rho(H^{k^*}, T(t'))$. Where $T(t')$ is point t' column of traces matrix T. In order to avoid confusion, hereafter throughout the whole paper, we use CPA-P to stand for CPA for the purpose of selecting POIs. MIA-P and KSA-P have the same meaning.

2.2 MIA

In MIA, one can compute the mutual information (MI) between the real power traces L and a hypothetical power consumption H^{k^*} as

$$I(L; H^{k^*}) = H(L) - H(L|H^{k^*}) = H(L) - \underset{h \in H^{k^*}}{E}[H(L|H^{k^*} = h)] \tag{2}$$

In this paper, for the estimation of the probability density function, we will use histogram method [9]. The largest MI indicated the most likely key guess as $k = \arg\max\limits_{k^*} I(L; H^{k^*})$.

Similarly, as is shown in [9], MIA distinguisher can also be used for selecting POIs (MIA-P) as $[k,t] = \arg\max\limits_{k^*,t'} I(T(t'); H^{k^*})$.

2.3 KSA

The KS test quantifies a distance between the empirical cumulative distribution function of two samples to determine the similarity of them. The central idea of KSA distinguisher proposed in [17] is to measure the maximum distance between the global trace distribution L and the conditional trace distribution $L|H^{k^*}$ as

$$D_{KS}(k^*) = E[KS(L||(L|H^{k^*}))] = \underset{h \in H^{k^*}}{E}[KS(L||(L|H^{k^*} = h))] \tag{3}$$

The largest distance indicates the most likely key guess as $k = \arg\max\limits_{k^*} D_{KS}(k^*)$.

Similarly, KSA distinguisher can also be used for selecting POIs (KSA-P) as $[k,t] = \arg\max\limits_{k^*,t'} D_{KS}(k^*) = E[KS(T(t')||(T(t')|H^{k^*}))]$

2.4 Sosd and Sost

Denote hypothesis power consumption by $h_i^{k*} = f(v_i^*)$. In Hamming weight model, $h_i^{k*} \in [0, 8]$. We partition all traces T according to h_i^{k*}, $G_j = \{T_i | h_i^{k*} = j\}, (j = 0, 1, 2, ..., 8)$. We calculate the mean m_j and the standard deviation σ_j of every partition G_j. For the sosd, we sum up their squared pairwise differences, $sosd = \sum_{i,j=0}^{8} (m_i - m_j)^2$. The sost is based on the T-Test. n_j is the number of traces in partition G_j. The location of POI is where the sosd or sost is biggest.

$$sost = \sum_{i,j=0}^{8} \left(\frac{m_i - m_j}{\sqrt{\frac{\sigma_i^2}{n_i} + \frac{\sigma_j^2}{n_j}}} \right)^2 \tag{4}$$

3 Evaluation Framework

In this section, we will present our framework for analysis of POIs selections for univariate side-channel distinguishers.

We argue that a real-word side-channel attack consists of two essential procedures, namely point extraction and key recovery. Unlike in those works of theoretical analysis, POIs selection really matters in real-world practices and distinguishers are sensitive to the POIs selected. In univariate case, this means once a "bad" point is fed into a certain distinguisher, key recovery efficiency of the attack will lower down, or sometimes even a wrong key guess will be made. Therefore, it is of great help if there is a method for picking out from those POIs selections available the most appropriate one for a certain distinguisher. In order to do this task, we need to measure the quality of the POIs selected. Fortunately, we could use security metric like success rate proposed in [13] to reflect the quality of POIs selected, with respect to a certain distinguisher. Yet, we have to notice that in some cases a successful attack using a certain POIs selected might also be falsity. For example, take for example a CPA attack against an unprotected AES software implementation. In this case, we choose the output of Sbox of the first round of AES encryption to the sensitive intermediate value. The outcome of performing a CPA attack will be the same as that against 4 bytes (i.e. 1st, 5th, 9th, and 13th byte) of outputs of ShiftRow operation. If only partial success rate is considered, this could lead to misleading results. As what we really need in real-world practices are those POIs selected that exactly correspond to the sensitive intermediate value targeted, which could be viewed to be a necessary requirement for a sound POIs selection. This means that success rate really makes sense only when this requirement holds.

For this requirement, we define the validity of a point with respect to a certain sensitive intermediate value. One point is said to be "*valid*" if it fall into the set of all points corresponding to the manipulations of the sensitive intermediate value; otherwise, it is said to be "*invalid*". Under the condition that two points are both valid, we say that one point is to have a "*better*" quality than another point, if the success rate of a key recovery attack using this point is higher than that of using another point, with respect to a certain distinguisher. Now, we can think of how to measure the goodness of one POIs selection. For this purpose, we introduce the notion of accuracy rate, from the perspective of statistics. Intuitively, the accuracy rate is to capture how well one POIs selection method is capable of extracting from side-channel leakage traces those points that exactly correspond to manipulations of one sensitive intermediate value targeted. The formal definition of accuracy rate will presented in Sect. 3.1.

Put above-mentioned ideas together, we put forward our evaluation framework. Basically, our framework contains of two stages. In the first stage, we measure the goodness of POIs selections through capturing the validity of points selected. The second stage reflects the quality of points selected, with respect to a certain distinguisher.

One feature of our framework is that it provides both designers and evaluators a more fine-grained way of examining two essential procedures (i.e. point extraction and key recovery) of real-world side-channel attacks. Another feature of our framework is that it could be jointly used in a very natural way with other well-known frameworks in the field for comparison of distinguishers themselves, including those of Standaert et al. [13] and Whitnall et al. [14,16]. With the help of this powerful framework, we can objectively and fairly compare different POIs selections, and then find the most suitable one for a certain distinguisher afterwards.

3.1 Metrics

We will provide the formal definition of accuracy rate of POIs selection, and then briefly discuss success rate.

Accuracy Rate (AR) of POIs Selection. The accuracy rate of one POIs selection is a expected probability of event S, if the points selected are in the POIs set corresponding to the manipulations of the sensitive intermediate value, we say event S occurs. It is straightforward that if the points are not in the POIs set, they are not pertinent to the chosen sensitive intermediate value, and they are not points we need even though distinguishers can recover the key using them in some cases. We can use this metric to measure the goodness of POIs selections. This is independent of key recovery of distinguishers.

Obviously, there is an important prerequisite, i.e. we need to know the POIs set. The example scenarios include that one performs POIs selection in simulated scenarios where he can control the generation of traces; or one knows all details about the cryptographic algorithm and cryptographic device, then he can calculate the positions or range of POIs by clock frequency of device and sampling frequency of oscilloscope. This metric is designed to be used in evaluation scenarios, because in adversarial scenarios, once we know the POIs set, we need not perform POIs selection any more.

We define a POIs selection adversary as an algorithm $A_{E_K,L}$ with time complexity τ, memory complexity m and q queries to the target physical computer.

The real POIs set tc is determined by the definition of a function β, i.e. $tc = \beta(k)$. In order to select the POIs, we assume that the output of the adversary $A_{E_K,L}$ is a sorted points vector $tg = [tg_1, tg_2, ..., tg_{|W|}]$, where W is the number of points in one whole trace. According to the selection result: the most likely POI being tg_1. Finally, we define a POIs selection of order o with the experiment:

$$\begin{aligned}&ExpA^{ps-ar-o}_{A_{E_K,L}}[tg \leftarrow A_{E_K,L}; tc = \beta(k); k \overset{R}{\leftarrow} \kappa;]\\&if\ (tg_1, tg_2, ..., tg_o) \in tc\ then\ return\ 1 \quad else\ return\ 0\end{aligned} \tag{5}$$

The o^{th}-order accuracy rate of $A_{E_K,L}$ against known POIs set is defined as:

$$AR^{ps-ar-o}_{A_{E_K,L}}(\tau, m, q) = \Pr[ExpA^{ps-ar-o}_{A_{E_K,L}} = 1] \tag{6}$$

Success Rate (SR) of Key Recovery [13]**.** Let $E_K = \{E_k(.)\}_{k \in \kappa}$ be a family of cryptographic abstract computers indexed by a variable key K. Let (E_K, L) be the physical computers corresponding to the association of E_K with a leakage function L. In general, the attack defines a function $\gamma : \kappa \rightarrow S$ which maps each key k onto an equivalent key class $s = \gamma(k)$, such that $|S| << |\kappa|$. We define a side-channel key recovery adversary as an algorithm $A_{E_K,L}$. Its goal is to guess a key class $s = \gamma(k)$ with non negligible probability. For this purpose, we assume that the output of the adversary $A_{E_K,L}$ is a guess vector $g = [g_1, g_2, ..., g_{|S|}]$ with the different key candidates sorted according to the attack result: the most likely candidate being g_1. Finally, we define a side-channel key recovery of order o with the experiment:

$$\begin{aligned}&ExpB^{sc-kr-o}_{A_{E_K,L}}[g \leftarrow A_{E_K,L}; s = \gamma(k); k \overset{R}{\leftarrow} \kappa;]\\&if\ s \in [g_1, ..., g_o]\ then\ return\ 1 \quad else\ return\ 0\end{aligned} \tag{7}$$

The o^{th}-order success rate of $A_{E_K,L}$ against a key class is defined as:

$$SR^{sc-kr-o}_{A_{E_K,L}}(\tau, m, q) = \Pr[ExpB^{sc-kr-o}_{A_{E_K,L}} = 1] \tag{8}$$

In this paper, we only consider the 1^{st}-order AR and 1^{st}-order SR.

3.2 Factors

In practice, there are some other factors to affect the key recovery of distinguishers.

Noise Level. Generally, the higher noise level increases, the worse POIs selections perform. POIs selections have different ability to adapt various noise level. We assume that the noise follow the Gaussian distribution with mean 0, and noise level is measured by standard deviation.

Leakage Type and Hypothetical Model. Leakage type refers to leakage model of crypto device. In this paper, we considered four types, i.e. Hamming Weight (HW) leakage, Hamming Distance (HD) leakage, Unevenly Weighted Sum of the Bits (UWSB) leakage and Highly Non-Linear (HNL) leakage [12].

In terms of hypothetical model, in this paper, we consider two kind of adversaries with different characterization ability to the leakage model of crypto device. An adversary with strong ability uses hypothetical leakage model (denoted by HL) as same as real leakage type (denoted by RL) to calculate the hypothetical power consumption, while an adversary with limited ability uses Hamming weight as hypothetical leakage model. We define a tuple <RL,HL> to represent a specific analysis or evaluation scenarios.

4 Experimental Evaluation

In this section, we will conduct comprehensive empirical evaluation. Our experiments will carry out in simulated scenarios and practical scenarios.

4.1 Simulated Experiments

In simulated scenarios, we choose the output of the first Sbox of the first round unprotected AES operation as the target intermediate value. In these scenarios, we know all details about the cryptographic algorithm and cryptographic device, and we control the generation of traces. Therefore, we can use the accuracy rate to evaluate the goodness of five popular POIs selections, i.e. CPA-P, sosd, sost, MIA-P, KSA-P. Three typical leakage types are adopted, i.e. HW leakage[1], UWSB leakage and HNL leakage.

The simulated traces are composed of the signal part and noise part. Firstly, we generate the signal part of 10,000 traces which contain five points corresponding to every intermediate value and five independent points. The intermediate values contains the plaintext, plaintext xor key, the output of Sbox and the result of Shift-Row. The plaintexts are random, and the key is fixed. Secondly, we add Gaussian noise varying in standard deviation to the signal part.

Our experiments are carefully divided into two stages in order to justify our proposed framework. Specifically, stage one, in each of the noise level, five POIs selections run 500 times using 1,000 and 5,000 random selected traces, and count the ARs according to definition of it respectively. Stage two, three distinguishers run 500 times using the points selected by five POIs selections, and count the SRs according to definition of it respectively.

Hamming Weight Leakage. In this scenario, the tuple is <HW,HW>. The ARs of five POIs selections are shown in Fig. 1. The quality of points selected with respect to three distinguishers are shown in Fig. 2. Specifically, we divide

[1] HD is another usual leakage types, but it is a linear leakage like Hamming weight. In simulated scenarios, we took Hamming weight as a typical leakage example.

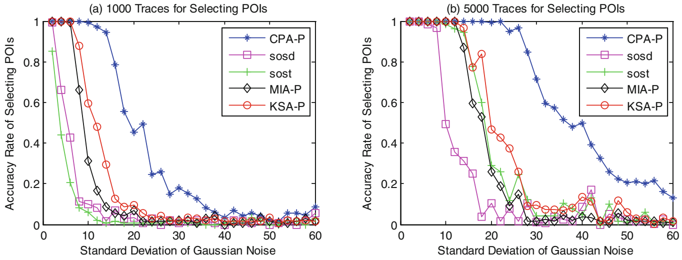

Fig. 1. ARs of five POIs selections in HW leakage simulated scenario

the results of three distinguishers into three groups according to the standard deviation of Guassian noise, and denote these groups by A, B and C, respectively. The standard deviation in Group A, B, C is 4, 8, 16, respectively.

Figure 1 shows the ARs of five POIs selections decrease rapidly with the increase of noise level. When noise level increases highly, all selections fail. However, the ARs of CPA-P is obviously higher than those of other four selections. This implies that CPA-P is the relatively strongest capacity to tolerate noise, while sosd and sost are the poorest. Compared with (a) in Fig. 1, (b) shows that the ability to tolerate noise of all selections improves.

Figure 2 shows the results of stage two in our framework. In Group A, points selected by five POIs selections are all "*valid*" to recover the key, but the SRs using points selected by five POIs selections for a certain distinguisher have subtle differences. When standard deviation of Gaussian noise is 8, points selected are still "*valid*", but the points selected by CPA-P is much better than those selected by other four selections. When standard deviation is 16, the point selected by sosd and sost are "*invalid*", CPA-P is still the best one, followed by KSA-P and MIA-P. The observations above suggest that different points in POIs set could make different key recovery efficiency.

An Unevenly Weighted Sum of the Bits Leakage. In this scenario, the least significant bit dominates in the leakage function with a relative weight of 10 and other bits with a relative weight of 1. An adversary with limited ability to describe the leakage type of device (i.e. the scenario tuple is <UWSB,HW>) and another adversary with strong ability (i.e. the scenario tuple is <UWSB,UWSB>) can get the ARs of five POIs selections. Our experiments show that the curves of ARs have exactly the same trend as those in Fig. 1.

Highly Non-Linear Leakage. In this scenario, the leakage function of cryptographic device is a highly non-linear function. Without loss of generality, S-box is used in this leakage scenario [12,20]. An adversary with limited ability to describe the leakage type of device (i.e. the scenario tuple is <HNL,HW>) can get the goodness evaluation results of POIs selections. Our experiments show that five POIs selections all fail.

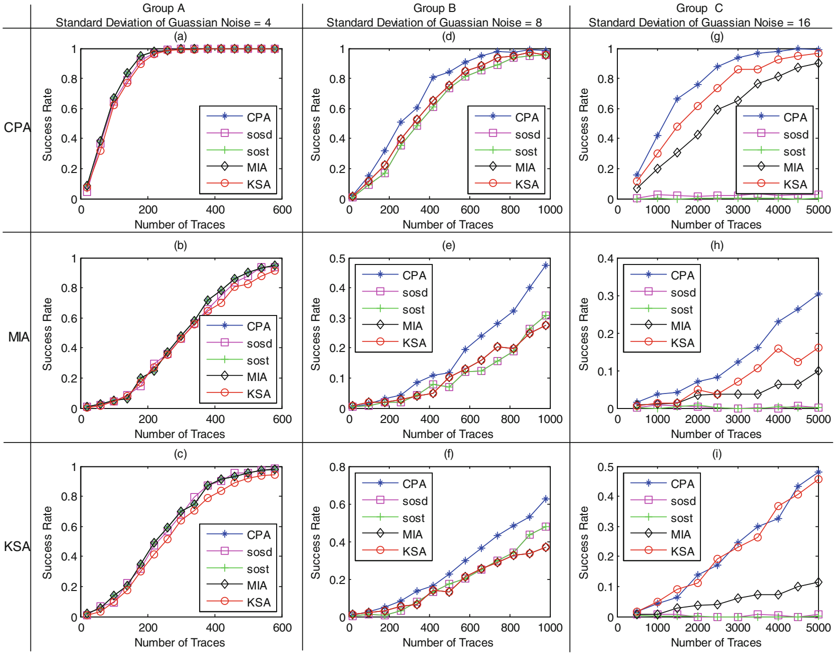

Fig. 2. SRs of CPA, MIA, and KSA using points selected by CPA-P, sosd, sost, MIA-P, KSA-P in HW leakage simulated scenarios

4.2 Practical Experiments.

In practical scenarios, we perform attacks against AES-256 RSM [18] implemented in software on an Atmel ATMega-163 smart card (Case 1) and unprotected AES implemented in hardware on Xilinx Vertex-5 FPGA (Case 2), and we use traces from DPA Contest v4 and DPA Contest v2, respectively. Especially, we ONLY focus on the POIs selection and key recovery against unprotected implementation in this paper. In Case 1, we converted the traces of protected implementation into traces of unprotected implementation using the known masks.

In the view of an adversary, we will choose hypothetical model according to priori knowledge. Specifically, we will use HW model in Case 1, and HD model in Case 2. In these practical scenarios, we cannot obtain the locations or range of POIs, the ARs cannot be computed. However, we can follow the second stage of framework, and utilize the SRs to evaluate the quality of points selected by five methods. For three distinguishers, we respectively perform key recovery attacks 300 times using every points selected by five POIs selections and count the SRs.

Case 1: Attacks Against an Unprotected AES Software Implementation. In this scenario, the output of the first S-box of the first round of AES operation is chosen as the target. The noise level of the traces from software implementation on the Atmel ATMega-163 smart card is very low. In order to

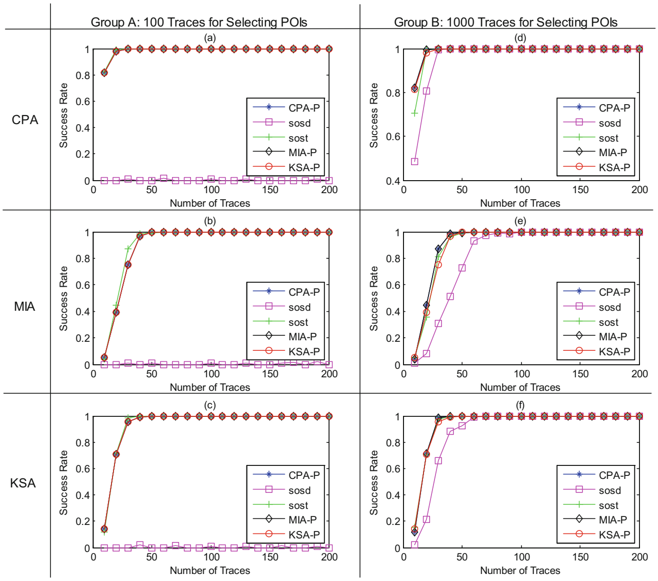

Fig. 3. SRs of CPA, MIA, and KSA using POIs selected by CPA-P, sosd, sost, MIA-P, KSA-P using original traces in Case 1

study the influence of noise level on POIs selections and distinguishers, we use additional Gaussian noise. Particularly, we employed five standard deviations of additional Gaussian noise, i.e. 0, 4, 8, 16, 32, where 0 denotes the original traces.

According to the results on DPA Contest website [19], the number of traces needed to recover the key in non-profiling attacks is at most 130. In order to study the influence of the number of traces on POIs selections, we set up two scenarios i.e. limited (100 traces) scenarios and sufficient (1,000 traces) scenarios.

Using original traces with limited number (100), five POIs selections get three points. Group A of Fig. 3 shows that, CPA, MIA, KSA can achieve 100 % SRs using the points selected by CPA-P, sost, MIA-P, KSA-P. Three distinguishers all fail using point selected by sosd. When the number of traces increases to 1,000, Group B shows that, the quality of points selected by all POIs selections are "good" enough to help three distinguishers achieve 100 % SRs.

When standard deviation of additional Gaussian noise is 4, using 100 traces, our experiments show that the most obvious change compared with Group A of Fig. 3 is that sosd and sost both fail to select a "good" point. When standard deviation 8 (Fig. 4) and 16, MIA-P fails, too. However, when 1,000 traces are used, MIA-P will be "good". We argue that this is because 100 traces are not sufficient to get satisfying probability density function, while 1000 traces do.

When standard deviation of additional Gaussian noise is 32, using limited traces, all POIs selections fail. It is because the noise level is too high and traces is too little. Using 1,000 traces, MIA-P fails. Possibly, it is because the noise

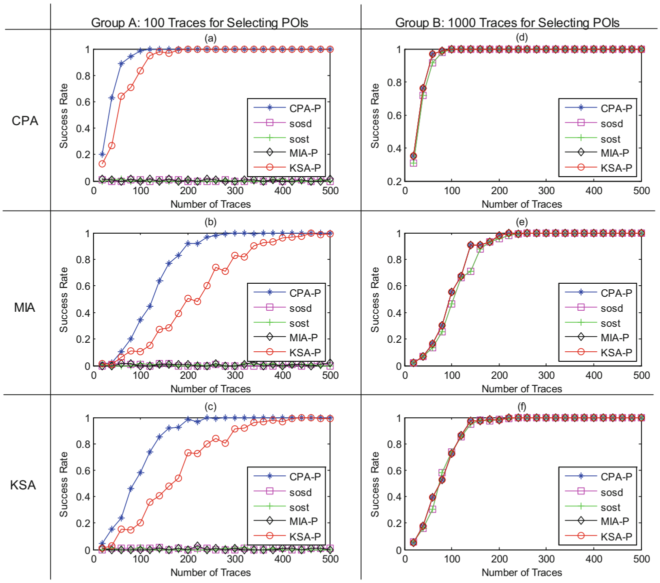

Fig. 4. SRs of CPA, MIA, and KSA using POIs selected by CPA-P, sosd, sost, MIA-P, KSA-P using traces with additional Gaussian noise of standard deviation 8 in Case 1

level is too high to get correct probability density function. This implies that MIA-P has weaker ability to tolerate noise than CPA-P and KSA-P.

Comprehensive analysing the experimental observations above, the quality of points selected by five POIs selections becomes worse with the increase of noise level. In the overall trend, sosd never has selected an "good" POI, sost affected by the noise level mostly, followed by MIA-P. The points selected by CPA-P and KSA-P are relatively more excellent. Moreover, comparing Group A with Group B, an important observation is that the negative impact of noise level could be decreased through increasing the number of traces. In addition, CPA distinguisher needs the least traces to achieve 100 % SR.

Case 2: Attacks Against an Unprotected AES FPGA Implementation. In this scenario, the input of the first S-box of the last round of AES operation is chosen as the target. As we known, the noise level of the traces from hardware implementation on Xilinx Vertex-5 FPGA is relatively high. That can factually represent a kind of common scenarios, so we will not use additional noise.

According to the results on DPA Contest website [19], the numbers of traces needed to achieve 80 % SR in non-profiling attacks range from 5,000 to 16,000. We set up two scenarios, i.e. limited (5,000) traces and sufficient (20,000) traces for selecting POIs. In this case, the SRs of three distinguishers using points selected by five POIs selections are presented in Fig. 5.

Group A of Fig. 5 shows that, using 5,000 traces, CPA can achieve 80 % SR by feeding point selected by only CPA-P. Other four POIs selections fail. MIA and KSA cannot recover the key using any points selected. Group B shows that,

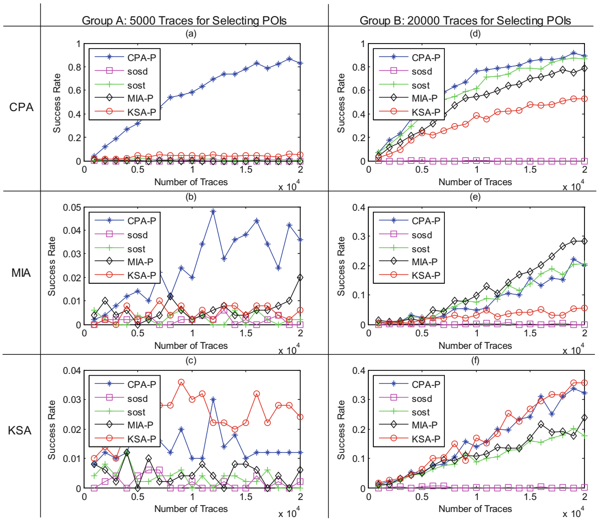

Fig. 5. SRs of CPA, MIA, and KSA using POIs selected by CPA-P, sosd, sost, MIA-P, KSA-P in Case 2

when traces for selecting POIs are sufficient, CPA can achieve 90 % SR using point selected by CPA-P; MIA can achieve 30 % SR using point selected by MIA-P; KSA can achieve 35 % SR using point selected by KSA-P. The SRs in Group B are limited with the trace number provided by DPA Contest v2 official.

4.3 Experimental Observations

According to the results of evaluation experiments in simulated scenarios and practical scenarios, we have the following observations.

1. When evaluations perform in the scenarios RL=HL={HW,UWSB,NL} or <UWSB,HW>,
   - Observation 1: The points selected by five POIs selections are not the same, sometimes differ greatly. The goodness of POIs selections significantly depends on noise level: with the increase of noise level, the goodness of five POIs selections decrease. Specifically, when the number of traces is limited, CPA-P> KSA-P > MIA-P> sosd> sost; when traces are sufficient, CPA-P> KSA-P> sost> MIA-P> sosd ("$A > B$" denotes POIs selection A is better than POIs selection B in a certain scenario).
   - Observation 2: In same noise level, the quality of point selected by CPA-P is the best. Specially, in the scenario <HD,HD>, this conclusion is incorrect.
2. When evaluations perform in the scenarios <NL,HW>,
   - Observation 3: All five POIs selections fail.
   - Observation 4: When in the scenarios <NL,UWSB or HD>, we guess that the conclusions are the same to those of Observation 3.
3. When evaluations perform in the scenario <HD,HD> and noise level is high,
   - Observation 5: We guess that the goodness of five POIs selections are the same to those of Observation 1.

- Observation 6: An interesting pattern is that the quality of points selected depend on certain distinguisher. Specifically, if choosing CPA, the optimal POIs selection is CPA-P; if choosing MIA, the optimal POIs selection is MIA-P; if choosing KSA, the optimal POIs selection is KSA-P.

Some Hints. Generally speaking, the adversary usually does not have powerful enough ability to identify a non-linear leakage. According to the observations above, we suggest that a crypto device with a highly non-linear leakage might be more secure. Some possible ways to implement it contain increasing noise level, making a device with non-linear leakage itself, or some other special methods.

5 Conclusions

In the field of side-channel attacks, POIs selection really matters much more in real-world practices than it is in those of theoretical analysis. In order to investigate the concrete impacts of POIs selections on distinguishers, and then pick out from those selection methods available the most appropriate one for a certain distinguisher, we proposed a two-stage evaluation framework which aims to separate the validity of POIs selected and their quality with respect to a certain distinguisher. This framework equips both designers and evaluators with a powerful tool to examine, in a more fine-grained way, two essential procedures (i.e. point extraction and key recovery). For the goodness of the POIs selection being used, we introduced the accuracy rate. It captures how well one POIs selection is capable of extracting from leakage traces those points that exactly correspond to the manipulations of one sensitive intermediate value targeted. In order to justify our proposed methods, we performed simulated attacks and practical attacks, taking five typical POIs selections and three distinguishers as concrete study cases. The results of these experiments also revealed some interesting observations.

Acknowledgments. This work was supported in part by National Natural Science Foundation of China (Nos. 61472416, 61272478, 61100225 and 61170282), National Key Scientific and Technological Project(No.2014ZX01032401-001), Strategic Priority Research Program of the Chinese Academy of Sciences (No. XDA06010701), One Hundred Talents Project of the Chinese Academy of Sciences, and SKLOIS Research Funding (No.2014-ZD-03).

References

1. 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)
2. Kocher, P.C., Jaffe, J., Jun, B.: Differential power analysis. In: Wiener, M. (ed.) CRYPTO 1999. LNCS, vol. 1666, p. 388. Springer, Heidelberg (1999)
3. Quisquater, J.-J., Samyde, D.: Electro magnetic analysis (EMA): measures and counter-measures for smart cards. In: Attali, S., Jensen, T. (eds.) E-smart 2001. LNCS, vol. 2140, p. 200. Springer, Heidelberg (2001)

4. 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)
5. Batina, L., Gierlichs, B., Lemke-Rust, K.: Differential cluster analysis. In: Clavier, C., Gaj, K. (eds.) CHES 2009. LNCS, vol. 5747, pp. 112–127. Springer, Heidelberg (2009)
6. Chari, S., Rao, J.R., Rohatgi, P.: Template attacks. In: Kaliski, B.S., Koç, C.K., Paar, C. (eds.) CHES 2002. LNCS, vol. 2523, pp. 13–28. Springer, Heidelberg (2003)
7. Schindler, W., Lemke, K., Paar, C.: A stochastic model for differential side channel cryptanalysis. In: Rao, J.R., Sunar, B. (eds.) CHES 2005. LNCS, vol. 3659, pp. 30–46. Springer, Heidelberg (2005)
8. Mangard, S., Oswald, E., Standaert, F.-X., One for All - All for One: Unifying Standard DPA Attacks. IET, pp. 100–111 (2010). ISSN: 1751–8709. doi:10.1049/iet-ifs.2010.0096
9. Gierlichs, B., Batina, L., Tuyls, P., Preneel, B.: Mutual information analysis. In: Oswald, E., Rohatgi, P. (eds.) CHES 2008. LNCS, vol. 5154, pp. 426–442. Springer, Heidelberg (2008)
10. Zhao, H., Zhou, Y., Standaert, F.-X., Zhang, H.: Systematic construction and comprehensive evaluation of kolmogorov-smirnov test based side-channel distinguishers. In: Deng, R.H., Feng, T. (eds.) ISPEC 2013. LNCS, vol. 7863, pp. 336–352. Springer, Heidelberg (2013)
11. Gierlichs, B., Lemke-Rust, K., Paar, C.: Templates vs. stochastic methods. In: Goubin, L., Matsui, M. (eds.) CHES 2006. LNCS, vol. 4249, pp. 15–29. Springer, Heidelberg (2006)
12. Whitnall, C., Oswald, E.: A fair evaluation framework for comparing side-channel distinguishers. J. Cryptographic Eng. **1**(2), 145–160 (2011)
13. Standaert, F.-X., Malkin, T.G., 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)
14. Whitnall, C., Oswald, E.: A comprehensive evaluation of mutual information analysis using a fair evaluation framework. In: Rogaway, P. (ed.) CRYPTO 2011. LNCS, vol. 6841, pp. 316–334. Springer, Heidelberg (2011)
15. Maghrebi, H., Rioul, O., Guilley, S., Danger, J.-L.: Comparison between side-channel analysis distinguishers. In: Chim, T.W., Yuen, T.H. (eds.) ICICS 2012. LNCS, vol. 7618, pp. 331–340. Springer, Heidelberg (2012)
16. Whitnall, C., Oswald, E., Mather, L.: An exploration of the kolmogorov-smirnov test as a competitor to mutual information analysis. In: Prouff, E. (ed.) CARDIS 2011. LNCS, vol. 7079, pp. 234–251. Springer, Heidelberg (2011)
17. Veyrat-Charvillon, N., Standaert, F.-X.: Mutual information analysis: how, when and why? In: Clavier, C., Gaj, K. (eds.) CHES 2009. LNCS, vol. 5747, pp. 429–443. Springer, Heidelberg (2009)
18. Nassar, M., Souissi, Y., Guilley, S., Danger, J.-L.: RSM: A small and fast countermeasure for AES, secure against 1st and 2nd-order zero-offset SCAs. In: IEEE DATE 2012, pp. 1173–1178 (2012)
19. DPA Contest Available at http://www.dpacontest.org
20. Reparaz, O., Gierlichs, B., Verbauwhede, I.: Generic DPA attacks: curse or blessing? In: Prouff, E. (ed.) COSADE 2014. LNCS, vol. 8622, pp. 98–111. Springer, Heidelberg (2014)