Generic Side-channel Distinguisher Based on Kolmogorov-Smirnov Test: Explicit Construction and Practical Evaluation*

LIU Ji-ye[1,2], ZHOU Yong-bin[1,♀], YANG Shu-guo[1,2], FENG Deng-guo[1]

(1. State Key Laboratory of Information Security, Institute of Software of Chinese Academy of Sciences, Beijing 100190, China;

2. Graduate University of Chinese Academy of Sciences, Beijing, 100049, China)

Abstract: Construction and evaluation of efficient distinguishers with broad generality is one fundamental problem in the area of side-channel cryptanalysis. Due to their capabilities to deal with general correlations, MIA-like distinguishers have received wide attention from academia. In this paper, we conduct a comprehensive comparison investigation of existing MIA-like distinguishers, and then propose a new generic side-channel distinguisher based on partial Kolmogorov-Smirnov test, namely PKS distinguisher. Theoretical analysis and experimental attacks unanimously justify that PKS distinguisher works remarkably well with both linear and non-linear leakage models. Specifically, PKS distinguisher has obvious advantages over existing MIA-like distinguishers in terms of both success rate and guessing entropy. Additionally, lower computational complexity of PKS distinguisher further shows its better applicability than MIA-like distinguishers.

Keywords: Side-Channel Cryptanalysis; Power Analysis Attack; Distinguisher; Distribution Similarity; Kolmogorov-Smirnov Test

1. Introduction

Cryptographic algorithm is traditionally viewed as a mathematical object (or an oracle) replying to message encryption (and/or decryption) queries according to a secret value: the key. On the contrary, side-channel cryptanalysis takes into account the physical implementation of a cryptosystem, and then tries to deduce the intermediate values from various observable side-channel leakages[1][2][3]. Once some intermediate results are deduced, the secret key will be recovered at an acceptable cost. Typically, in this context, the adversary divides the whole secret key into some smaller sub-keys. For each smaller sub-key, one distinguisher is used to tell the correct sub-key from those incorrect ones. Apparently, distinguisher with better effectiveness allows performing an in-depth study of the sources of side-channel leakages and their characteristics as well in a more comprehensive way.

Power analysis attack (PAA for short) is commonly accepted to be one of the most powerful side-channel attacks. It takes the instantaneous power consumption of cryptographic implementation to be its side-channel leakage. The most pioneering example of PAA is Differential Power Analysis(DPA for short)[2]. DPA was later extended to Correlation Power Analysis(CPA for short)[4], which is essentially an extension of DPA on multi-bits. However, CPA is based on rigorous leakage models, namely Hamming Weight model and Hamming Distance model, which do not always hold for some mainstream devices, e.g. nano-scale devices[5]. Recently, a generic form of power analysis was introduced under the name of Mutual Information Analysis(MIA for short)[8]. Due to its capability to deal with both linear and non-linear leakage models, MIA soon became a research focus. Since then, many extensions and enhancements to MIA distinguishers are proposed [7][8][9][10], and they are collectively called MIA-like distinguishers in this paper.

However, how to precisely estimate the mutual information in practice, specifically, the estimation of probability density function (PDF for short), still remains to be a very difficult problem [11]. To address this problem, some non-parametric MIA-like distinguishers are built in certain way to avoid the estimation of PDF, e.g. distribution similarity based MIA[8][10], KL-divergence based MIA[9], etc.

Therefore, a natural yet basic question arises: Is it possible to directly construct more effective distinguishers based on distribution similarity test, without the need of evaluating mutual information? To answer this question, we construct a naive KS distinguisher which simply adopts Kolmogorov-Smirnov Test(KS Test for short) as one side-channel distinguisher. After comprehensive analysis, we point out some technical obstacles that this naïve KS distinguisher has to overcome. Then, we refine KS distinguisher and propose PKS distinguisher which applies KS test on a small portion of somehow sorted power samples. Experimental results reveal that PKS distinguisher has obvious advantages over existing MIA-like distinguishers in terms of both success rate (SR for short) and guessing entropy (GE for short)[14]. More interestingly, this advantage tends to be more striking as noise increases up. Additionally, lower computational complexity of PKS distinguisher shows its better applicability than MIA-like distinguishers.

The rest of this paper is organized as follows. Section 2 briefly introduces preliminaries. The construction of PKS distinguisher as well as its brief analysis is given in Section 3. In Section 4, a comprehensive comparison among PKS and several typical MIA-like distinguishers is performed. Section 5 concludes the whole paper.

2. Preliminaries

Roughly speaking, side-channel attacks are usually classified into two categories: profiled and non-profiled. Profiled attacks are best known as Template Attacks[13][14], which need the help of reference devices for preprocessing to characterize the power consumption. By contrast, non-profiled attacks, which we will concern in this paper, rather aim to capture the behavior of more realistic adversaries who do not have a precise prior characterization of the target devices.

* This work is supported by National Natural Science Foundation of China (No. 61073178) and Beijing Natural Science Foundation (No. 4112064). ***The abridged version of this paper was accepted by Chinese Journal of Electronics on June 20, 2011 after anonymous peer reviews, and will appear in 2012.***

♀ Corresponding Author(e-mail: yongbin@icas.ac.cn)

2.1. Non-profiled Power Analysis Distinguisher

Typical procedures of non-profiled power analysis attack are as depicted in Figure 1. In order to verify the correctness of one key hypothesis, first of all, an adversary chooses the message set used to query the target device. Afterwards, he records the power consumptions while the target device is performing cryptographic operations, and evaluates the hypothetical intermediate values with key hypothesis and the chosen set of messages. Then, he maps the hypothetical intermediate values to hypothetical power consumptions using particular leakage model. Eventually, he uses distinguisher to determine the correctness of the key hypothesis. When the key hypothesis is correct, there will be a strong correlation between the sampled and the key-dependent hypothetical power values; otherwise, there will be no dependence.

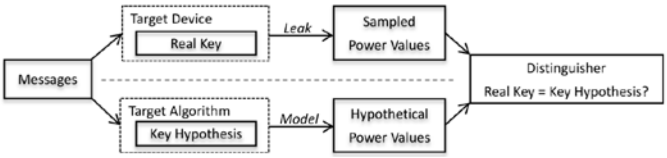

Figure 1. Schematic overview of non-profiled power analysis attack

2.2. MIA-like Distinguishers

The evaluation of mutual information strongly depends on accurate estimation of the PDF[11]. Initially, histogram approach is applied as an estimator of PDF[6]. In general, however, this approach involves determining a parameter in advance, e.g. bin-width, which straightly impacts the efficiency of the attacks.

In view of this, some representative refinements are proposed, as listed in Table 1. Broadly speaking, two main approaches are available. The first is to calculate mutual information based on PDF as usual. On one hand, several refinements are proposed for better estimation of PDF, including KDE[7], B-spline[10], etc. On the other hand, KL divergence[9] is proposed to evaluate the mutual information without using any information-theoretic metrics. Other than first PDF-based approaches, some recent works try to evaluate mutual information using distribution similarity test[8][10], in order to avoid estimation of PDF. Most of these works increase the effectiveness of original MIA more or less.

Table 1. Summary of existing refinements on MIA

	MIA	KL	KS	CvM
Histogram	[6] [7] [8]	[8]	[8]	[10]
KDE	[6] [7] [8]	[8]		
B-Spline	[10]			
Edgeworth		[9]		

3. Side-channel Distinguisher Based on Distribution Similarity Test

In constructions of some MIA-like distinguishers, distribution similarity tests are used to evaluate the similarity between the cumulative distribution function (CDF) of sampled power values $F(O)$ and the conditional CDF $F(O|S)$. The derived similarity can be used to quantify mutual information[8][10]. Inspired by this, we argue that, under ideal condition (zero noise and perfect leakage model), distribution similarity tests can be used to explicitly construct distinguisher. This section will provide two side-channel distinguishers based on distribution similarity test, namely KS and PKS respectively. Theoretical analysis of these two distinguishers will be given as well.

3.1. KS Test

In statistics, KS test is a non-parametric test to compare a sample with a reference CDF(as known as one-sample KS test), or to compare two samples(as known as two-sample KS test). The null statistic of this statistic is calculated under the null hypothesis that the samples are drawn from the same distribution (in the two-sample case) or that the sample is drawn from the reference distribution (in the one-sample case).

Let $F_n(x)$ be the empirical CDF of sample X with n elements, $F(x)$ be the reference distribution. Then, KS statistic for $F(x)$ is shown as follows:

$$D_n = \sup_x |F_n(x) - F(x)|$$

where $\sup$ is the supremum of the set of distances.

Let K_α be the percentile of Kolmogorov distribution. If $\sqrt{n}D_n > K_\alpha$, the null hypothesis is rejected; otherwise, the null hypothesis is accepted.

One variant of one-sample KS test is two-sample KS test, shown as follows:

$$D_{n,n'} = \sup_x |F_{1,n}(x) - F_{2,n'}(x)|$$

where $F_{1,n}(x)$ and $F_{2,n'}(x)$ are the empirical CDFs for the 1st and 2nd sample, respectively.

For a two-sample KS test, if $\sqrt{nn'/(n+n')}D_{n,n'} > K_\alpha$, the null hypothesis is rejected; otherwise, the null hypothesis is accepted.

p-value is the probability of obtaining a test statistic at least as extreme as the one that was actually observed, assuming the null hypothesis is true. Hence, *p*-value can be utilized to quantify the probability that a key hypothesis is the correct key.

3.2. KS Distinguisher

Ideally, in the case of noise-free and precise leakage modeling, the sampled power values are linearly correlated with the hypothetical power values if the key hypothesis is correct, which results in same shapes of CDF curves of sampled and hypothetical power values. The similarity can be quantified by two-sample KS test.

Let the real key and the key hypothesis be k and $k*$ respectively. When considering the correctness of $k*$, two inputs to distinguisher are denoted by T_k and T_{k*}, which are the vectors of sampled and hypothetical power values respectively. These two vectors are of size n, which is the number of messages used to query the target device. If $k* = k$, T_k and T_{k*} will follow the same distribution after normalization(Z-score transformation); otherwise, they will follow different distributions. In view of this, KS distinguisher is formally constructed as follows:

$$k_{ks} = \arg\max_{k*\in K} KS2(Z(T_k), Z(T_{k*})) \qquad (1)$$

where k_{ks} is the correct key identified by KS test, Z is the Z-score transformation, and $KS2$ is two-sample KS test with *p*-value as its return value.

In fact, CDFs used in KS distinguisher can be more precisely estimated than PDFs[8]. In addition, as a rule of thumb, smaller bin width usually leads to higher accuracy.

3.3. Analysis of KS Distinguisher

This subsection will analyze some practical obstacles that KS distinguisher will confront.

Our attack scenario is as follows. The target algorithm is a software implementation of AES, with the target intermediate value to be the output of the 1st S-box in the 1st AES round. The "message" M is one message byte input to the S-box, while the "key" K means one subkey byte XORed with "message byte".

Factor 1: Distribution Uniformity of S-box

When messages M follow uniform distribution, the distribution uniformity property of S-box makes its own outputs V follow uniform distribution too, which implies that hypothetical power values follow binomial distribution (under HW model). Binomial distribution can be approximated by normal distribution, when sample size is large enough. So, weather the key hypothesis is correct or not, both sampled and hypothetical power values follow normal distribution. In this case, to precisely identify the subtle difference between these two sets of power values by KS test is not feasible.

Factor 2: Meaningful Information Loss

Basically, KS distinguisher only considers the shape of CDF curves. Therefore, the order of elements in a sample do NOT affect the shape of its CDF, which directly implies that, the correlation between the pair-wise hypothesis and sampled values is substantially neglected. This greatly decreases the utilization of side-channel leakages. Experimental result shows that each key is indistinguishable from 16.4 other keys in average, using KS distinguisher when the adversary runs over all 2^8 messages.

We obtain Table 2 by simulated experiments to show the effects of these two factors. The GEs of KS distinguisher are listed with different number of messages and different levels of noise.

Table 2. GEs of KS distinguisher

Noise Level	**Message number**					
	50	100	150	200	250	300
0.1	74	33	23	19	23	22
0.3	109	96	107	94	102	117
0.5	109	114	110	116	100	91
0.7	131	93	102	112	115	113
0.9	118	102	120	131	121	132

Table 2 manifests the following observations. First of all, GE bears no obvious correlation with the number of samples, which is due to factor 1. Secondly, GE will not reach an acceptable level even under an extremely low level of noise (0.1), which is due to factor 2. Finally, GE will on the whole increase as the level of noise goes up.

3.4. PKS Distinguisher

As has been turned out in previous subsection, it is not feasible to directly apply KS test as distinguisher. Fortunately, many existing approaches are available to alleviate the impact of noise[15][16], however, the two aforementioned factors have to be dealt with. This subsection will construct another distinguisher, namely PKS distinguisher. As we will see, PKS distinguisher amazingly increases the practicability of KS test for side-channel distinguishers.

In order to make KS test practical to be a side-channel distinguisher, we consider the pair-wise correlation between hypothetical and sampled power values. In this way, we construct a new distinguisher as follows. Donate the inputs to distinguisher by a matrix $T_{2\times n} = [T_{k*}, T_k]^T$. Sort $T_{2\times n}$ in ascending (or descending) order according to the values of T_{k*}, and the resulting matrix is denoted by $T'_{2\times n} = [T'_{k*}, T_k^{k*}]^T$. In this way, one serialized sampled power value vector T_k^{k*} is produced. Then, perform one-sample KS test, with standard normal distribution as reference distribution, on the first half of T_k^{k*}, denoted by $T_k^{k*}[1, 0.5\cdot n]$ [1]. The key hypothesis with the smallest *p*-value is taken to be the correct one, as

[1] In this paper, $T[a,b]$ stands for a sub-sector consisting of a^{th} element, $(a+1)^{th}$ element, and up to b^{th} element of vector T of size n, where $1 \le a < b \le n$.

shown in Equation (2). Essentially, this distinguisher makes use of only parts of sampled power vector, hence the name partial KS distinguisher (PKS).

$$k_{pks} = \arg\min_{k^*\in K} KS(Z(T_k^{k^*}[1, 0.5\cdot n])) \qquad (2)$$

where KS is one-sample KS test with standard normal distribution as its reference distribution, k_{pks} is the correct key recovered.

Next, we will investigate the effectiveness of PKS under both linear and non-linear leakages. In these two cases, computations taken uniformly distributed messages as inputs result in normally distributed leakages, as are the hypothetical power values and sampled power values.

-Linear leakage

Consider the fact that sorting of elements in a vector is a permutation over the vector itself. Therefore, sorting does NOT change the distribution of the elements in one vector. Thus, when $k^* = k$, the smaller half of T_{k^*} is almost contained in $T_k^{k^*}[1, 0.5\cdot n]$, which theoretically follows half-normal distribution. On the other hand, when $k^* \neq k$, $T_k^{k^*}[1, 0.5\cdot n]$ contains random half elements of $T_k^{k^*}$, and $T_k^{k^*}[1, 0.5\cdot n]$ still follows the same normal distribution as that of $T_k^{k^*}$. Consequently, one can use PKS distinguisher to tell the correct key guess from those wrong ones, by observing the difference between the distribution of a partial vector and normal distribution.

Figure 2 depicts the CDF curves, when $k^* = k$, for $T_k^{k^*}[1, 0.25\cdot n]$, $T_k^{k^*}[1, 0.5\cdot n]$ and standard normal distribution. In Figure 2, the distributions of these two partial vectors are clearly distinguishable from standard normal distribution.

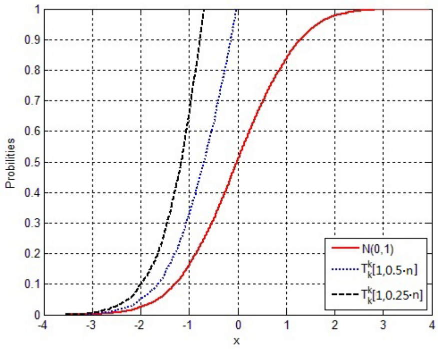

Figure 2. CDFs for different parts in T_k^k - HW model

Actually, the effectiveness of PKS can be further enhanced by selecting a smaller portion of $T_k^{k^*}$. As shown in Figure 3, the maximum distance between the CDF of $T_k^k[1, 0.25\cdot n]$ and that of normal distribution is larger than that of $T_k^k[1, 0.5\cdot n]$. Note that the selection of portion size is strongly related to number of messages. On one hand, when the message number is small, portion with too few elements may affect the accuracy of CDF, which subsequently decreases the effectiveness of PKS distinguisher. On the other hand, when the message number is large, a smaller portion is recommended to show clearer distinctiveness.

Additionally, considering the symmetry between the distribution of elements in $T_k^{k^*}[1, pn]$ and that in $T_k^{k^*}[(1-p)n, n]$, another PKS can be applied to $T_k^{k^*}[(1-p)n, n]$. In this case, the product of two *p*-values for these two PKSs (as shown in Equation (3)) will surely enhance the effectiveness of the attack.

$$k_{pks} = \arg\min_{k^*\in K}(KS(Z(T_k^{k^*}[1, pn]))\cdot KS(Z(T_k^{k^*}[(1-p)n, n]))) \qquad (3)$$

-Non-linear leakage

In the case of non-linear leakage, we take bit-weighted model[16] as a representative example to demonstrate the effectiveness of PKS distinguisher.

We will prove that the CDF of T_k under bit-weighted model is centrosymmetric, with point (0,0.5) being its center. This property is the same as that of normal distribution, which means that we can also directly use PKS distinguisher (using $N(0,1)$ as reference distribution).

Under bit-weighted model, we have

$$P_v = \sum_{i-1}^{8}(\omega_i v[i] + \overline{\omega}_i(1 - v[i])) \qquad (4)$$

where P_v is the total power consumption, $v[i]$ is the i$^{\text{th}}$ bit of v, ω_i and $\overline{\omega}_i$ are the power consumption by the i$^{\text{th}}$ bit when $v[i] = 1$ and $v[i] = 0$ respectively. The noise and constant components are neglected in order to simplify the problem.

Without loss of generality, assume that $\omega_i > \overline{\omega}_i$, then we have $\sum_{i=1}^{8}\overline{\omega}_i \le p_v \le \sum_{i=1}^{8}\omega_i$. Denote all intermediate values that satisfy $p_v = x$ by a set $V_{p_v=x} = \{v_1, v_2, ..., v_n\}$, and we have

$$\Pr(p_v = x) = \frac{n}{\#V} \qquad (5)$$

Bitwisely complement all values in $V_{p_v=x}$, and denote the new set by $\overline{V_{p_v=x}} = \{\overline{v_1}, \overline{v_2}, ..., \overline{v_n}\}$. Obviously, all elements in $\overline{V_{p_v=x}}$ correspond to power consumption value $\sum_{i=1}^{8}(\omega_i + \overline{\omega}_i) - x$. In addition, all elements outside $\overline{V_{p_v=x}}$ have different power consumption value from those in $\overline{V_{p_v=x}}$, so we have

$$\Pr(p_v = \sum_{i=1}^{8}(\omega_i + \overline{\omega_i}) - x) = \frac{n}{\#V} \qquad (6)$$

Combine Equation (5) and Equation (6), and we have

$$\Pr(p_v = x) = \Pr(p_v = \sum_{i=1}^{8}(\omega_i + \overline{\omega_i}) - x) \qquad (7)$$

Assume that $\Pr(v[i]=0) = \Pr(v[i]=1) = 0.5$ holds for intermediate value v, then we have

$$E(p_v) = \sum_{i=1}^{8}(\omega_i + \overline{\omega_i}) \Big/ 2 \qquad (8)$$

Equation (7) together with equation (8) shows that the PDF of p_v is centrosymmetric, with point (0,0.5) being its center, as depicted in Figure 3.

For most devices that follow bit weighted model, either $\omega_i > \overline{\omega_i}$ or $\omega_i < \overline{\omega_i}$ holds. This means that intermediate value with higher HW is expected to have higher (resp. lower) power consumption than that with lower HW. For this reason, we still use HW model to predict hypothetical power consumption in this scenario.

Similar to the case of linear leakage, one can use PKS distinguisher to tell the correct key hypothesis from those wrong ones by observing the difference between the distribution of a partial vector and normal distribution.

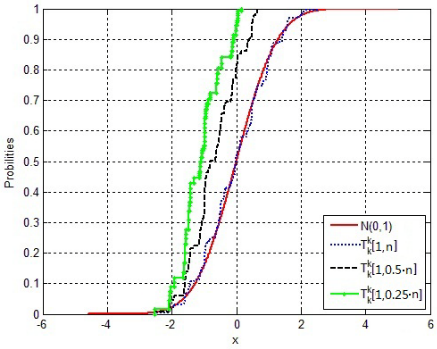

Figure 3. CDFs for different parts in T_k^k - bit weighted model

We use a set of bit weights in [16] for one STC89C52 micro-controller to investigate the effectiveness of PKS distinguisher. In this setting, we use HW model to predict the hypothetical power consumptions, and set noise level $\delta = 1$ to simulate the sampled power values. Figure 3 depicts the CDF curves, when $k* = k$, for $T_k^{k*}[1,n]$, $T_k^{k*}[1,0.25\cdot n]$, $T_k^{k*}[1,0.5\cdot n]$ and standard normal distribution. As depicted in Figure 3, the distributions of these two partial vectors are clearly distinguishable from that of the overall vector, which is similar to standard normal distribution. The smaller the selected portion is, the more significant distance between the portion's distributions and standard normal distribution is. Obviously, the technique combining two symmetric partial vectors (as shown in Equation (3)) still applies in case of non-linear leakage.

5. Experiments

In this section, we will study how PKS distinguisher behaves in terms of both effectiveness and cost. In order to evaluate the effectiveness more comprehensively, we simulated both linear and non-linear scenarios. In each scenario, PKS distinguisher is evaluated as well as four representative MIA-like distinguishers, including histogram based MIA(MIA for short), KS test based MIA(MIA(KS)), KDE based MIA(MIA(KDE)) and KDE based MIA calculated using KL divergence(MIA(KDE,KL)). SR and GE[12] are chosen as quantitative evaluation metrics. In addition, we will discuss record the computational cost for PKS as well as all those four MIA-like distinguishers

5.1. Effectiveness

-Linear Leakage

The simulation chooses HW model to be a representative example of linear leakage models. The 1st S-box output of the 1st round in an AES-128 implementation is selected to be the target intermediate value.

Equation (3) is used to be the PKS distinguisher and $p = 0.25$. All interval numbers for MIA-like distinguishers are set to 9, which is a refined parameter for AES[11]. Both SR and GE are evaluated under four increasing noise levels ($\delta = 2/4/6/8$). All results are averaged over 100 times of attacks.

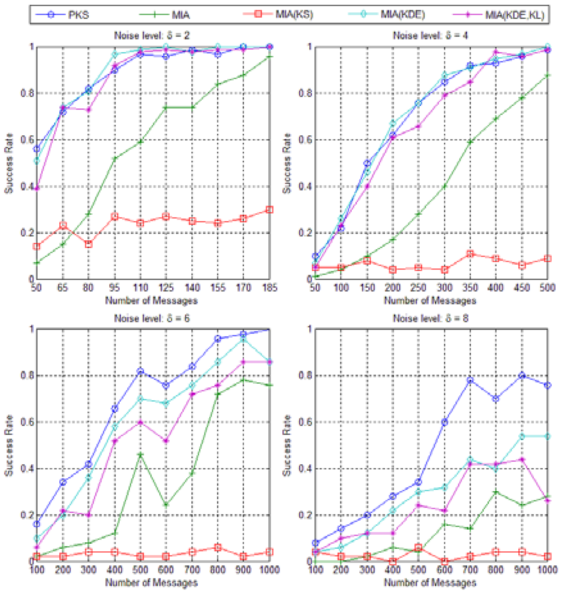

Figure 4. SRs - linear leakage

Figure 4 illustrates the SRs of five distinguishers under different levels of noise. The following observations are obtained.

- KDE based MIA-like distinguishers result in higher SRs than original MIA, which demonstrates that PDF estimation by KDE is more accurate than that by histogram[6][7][8].

- Directly applying KS test into the evaluation of mutual information could not lead to better SR. One reasonable explanation is that, if KS test is used to calculate mutual information, it needs to evaluate several conditional PDFs, each of which is with much smaller sample size. As a result, the overall error will inevitably increases.

- When the noise level is relatively low ($\delta = 2/4$), the SRs increase in the order MIA(KS) < MIA < PKS, MIA(KDE), MIA(KDE,KL) on the whole. Specifically, when $\delta = 2$, 95 power traces make the SRs of PKS,MIA(KDE),MIA(KDE,KL) exceed 90%, while the SRs of MIA and MIA(KS) are only 52% and 27%, respectively.

- When the noise level is relatively high ($\delta = 6/8$), the SRs increase in the order MIA(KS) < MIA < MIA(KDE,KL) < MIA(KDE) < PKS on the whole. Specifically, when $\delta = 8$, 700 power traces make the SRs of PKS, MIA(KDE), MIA(KDE,KL), MIA and MIA(KS) reach 2%, 15%, 42%, 43% and 78% respectively. This reflects that PKS has better adaptability in noisy environment.

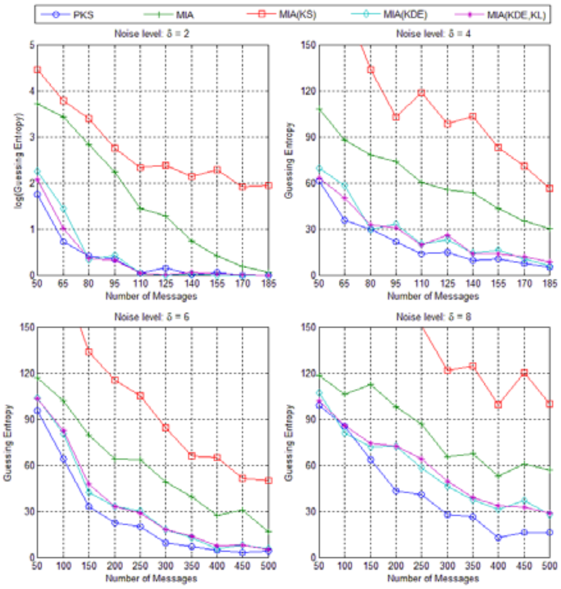

Figure 5. GEs - linear leakage

Figure 5 shows the GEs of five distinguishers under different levels of noise levels. According to Figure 6, we get observations similar to those derived from Figure 5. Under lower noise level ($\delta = 2$), PKS, MIA(KDE), MIA(KDE,KL) result in similar GEs. However, at a higher noise level, the advantage of PKS becomes more obvious. Specifically, when $\delta = 8$ and the number of messages is 500, the GE of PKS is 16, while those of MIA(KDE) and MIA(KDE,KL) are 29 and 30 respectively.

-Non-linear Leakage

In order to verify the applicability of PKS distinguisher for non-linear leakage, we use the same set of bit weights for STC89C52 micro-controller to simulate the sampled power traces. All other setups are identical to those used by experiments in linear scenario. The results are shown in Figure 6 and Figure 7.

Figure 6 and Figure 7 depict that, although SRs(resp. GEs) decrease(resp. increase) comparing to those in linear leakage, all five distinguishers work well. These results not only confirm the applicability of MIA-like distinguishers under non-linear leakage model, but also prove the advantages of PKS distinguisher over MIA-like distinguishers.

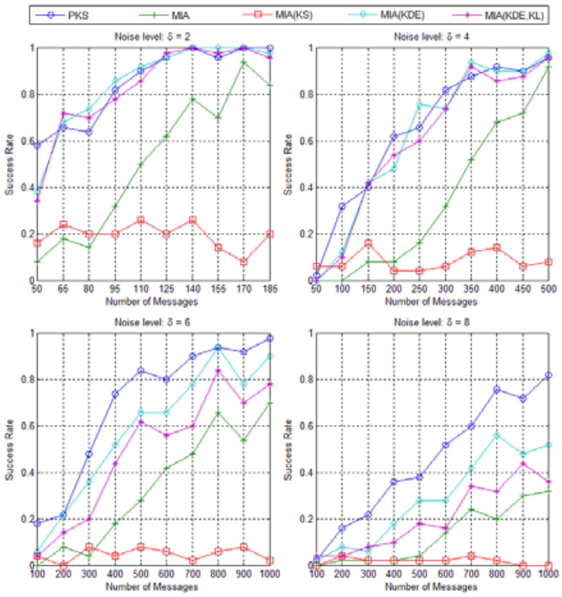

Figure 6. SRs - non-linear leakage

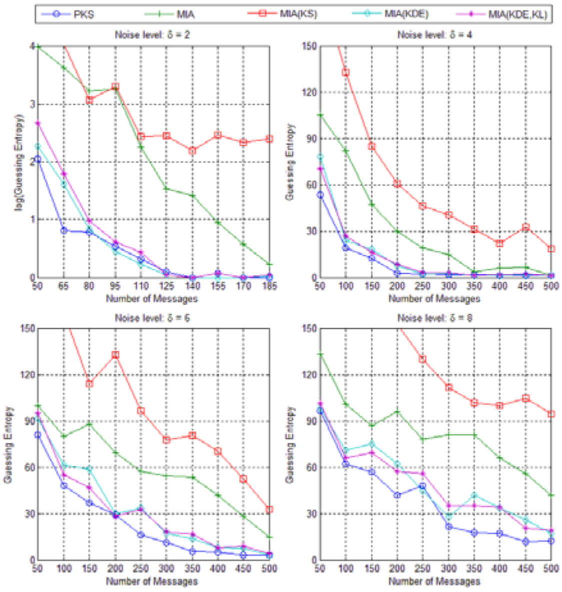

Figure 7. GEs - non-linear leakage

5.2. Computational Cost

We count the time costs of attacks on a PC using both PKS and all four MIA-like distinguishers with different number of messages. All results are averaged over 100 times of attacks, which are depicted in Figure 8.

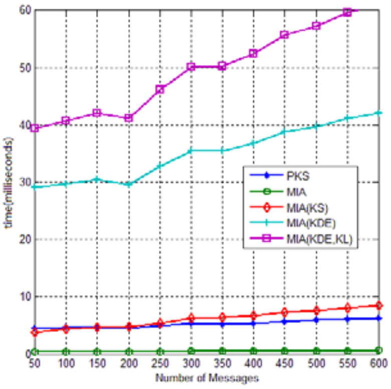

Figure 8. Computational costs of five distinguishers

Following observations can be retrieved from Figure 8. The calculation cost increase in the order MIA < PKS < MIA(KS) << MIA(KDE) < MIA(KDE,KL), and this trend becomes more obvious as the number of messages goes up. Estimation of joint PDF using KDE is time-consuming, so two distinguishers using KDE as PDF estimator are of higher cost in terms of execution time. Although PKS has bigger time overhead than MIA, it is still more efficient than MIA(KDE) and MIA(KDE,KL).

6. Conclusion

PKS distinguisher is proved to achieve remarkable technical merits at the cost of a very light-weight preprocessing, i.e. sorting operation on power values. This gave rise to at least two inspirations. Firstly, existing MIA-like distinguishers mostly likely have not made full use of side-channel leakage as they were claimed. In view of this, there seems to be some room for improvements of known MIA-like distinguishers in terms of utilization rate of side-channel leakage. Secondly, in practice-oriented cryptanalysis, we can resort to some simple yet effective solutions, and try to avoid being deeply involved in cumbersome technical details.

However, it is worth noting that distinguishers have been proved to show some preferences towards accuracy of underlying leakage models. With respect to this, evaluation results in one certain scenario may not be helpful for those in another. Our results do NOT make any assertions to the effectiveness of any side-channel distinguishers.

References

[1] Paul Kocher, Timing Attacks on Implementations of Diffie-Hellman, RSA, DSS, and Other Systems, CRYPTO 1996, LNCS 1109, pp. 104-113, 1996.

[2] Paul Kocher, Joshua Jaffe and Benjamin Jun, Differential Power Analysis, CRYPTO 99, LNCS 1666, pp. 388-397, 1999.

[3] Dakshi Agrawal, Bruce Archambeault, Josyula R. Rao and Pankaj Rohatgi, The EM Side-Channel(s),CHES 2002, LNCS 2523, pp. 29-45, 2002.

[4] Rita Mayer-Sommer, Smartly Analyzing the Simplicity and the Power of Simple Power Analysis on Smartcards, CHES 2000, LNCS 1965, pp. 78-92, 2000.

[5] Mathieu Renauld, Francois-Xavier Standaert, Nicolas Veyrat-Charvillon, Dina Kamel and Denis Flandre, A Formal Study of Power Variability Issues and Side-channel Attacks for Nanoscale Devices, Eurocrypt 2011, LNCS 6632, pp. 109-128, 2011.

[6] Benedikt Gierlichs, Lejla Batina and Pim Tuyls and Bart Preneel,Mutual Information Analysis - A Generic Side-Channel Distinguisher, CHES 2008, LNCS 5154, pp. 426-442, 2008.

[7] Emmanuel Prouff and Matthieu Rivain, Theoretical and Practical Aspects of Mutual Information Based Side Channel Analysis, ACNS 2009, LNCS 5536, pp. 499-518, 2009.

[8] Nicolas Veyrat-Charvillon, Francois-Xavier Standaert, Mutual Information Analysis: How, When and Why, CHES 2009, LNCS 5747, pp. 429-443, 2009.

[9] Thanh-Ha Le and Mael Berthier, Mutual Information Analysis under the View of Higher-Order Statistics,IWSEC 2010, LNCS 6434, pp. 285-300, 2010.

[10] Alexandre Venelli, Efficient Entgropy Estimation for Mutual Information Analysis Using B-Splines, WISTP 2010, LNCS 6033, pp. 17-30, 2010.

[11] Lejla Batina, Benedikt Gierlichs, Emmanuel Prouff, Matthieu Rivain, Francois-Xavier Standaert and Nicolas Veyrat-Charvillon, Mutual Information Analysis: a Comprehensive Study, Journal of Cryptology, Vol. 24, No. 2, 269-291

[12] Francois-Xavier Standaert, Tal G. Malkin and Moti Yung, A Unified Framework for the Analysis of Side-Channel Key Recovery Attacks, Eurocrypt 2009, LNCS 5479, pp 443-461, 2009.

[13] Suresh Chari, Josyula Rao and Pankaj Rohatgi, Template Attacks, CHES 2002, LNCS 2523, pp 51-62, 2002.

[14] Martin Bar, Hermann Drexler and Jurgen Pulkus, Improved Template Attack, COSADE 2010, pp. 81-89, 2010.

[15] Stefan Mangard, Elisabeth Oswald and Stefan Popp, Power Analysis Attacks: Revealing the Secrets of Smart Cards, Springer, Berlin, 2007.

[16] Jiye Liu, Yongbin Zhou, Yang Han, Jiantang LI, Shuguo Yang and Dengguo Feng, How to Characterisitc Side-Channel Leakage More Precisely? ISPEC 2011, LNCS 6672, pp 196-207, 2011.