The Journal of Systems and Software 120 (2016) 105–113

Contents lists available at ScienceDirect

The Journal of Systems and Software

journal homepage: www.elsevier.com/locate/jss

How many interesting points should be used in a template attack?

Hailong Zhang*, Yongbin Zhou

State Key Laboratory of Information Security, Institute of Information Engineering, Chinese Academy of Sciences, Minzhuang Road 89-A, Beijing, 100093, P.R. China

ARTICLE INFO

Article history:
Received 6 August 2015
Revised 16 July 2016
Accepted 19 July 2016
Available online 21 July 2016

Keywords:
Side channel attacks
Power analysis attacks
Template attack
Interesting points
Profiling efficiency

ABSTRACT

Considering that one can fully characterize and exploit the power leakages of the reference device in the process of recovering the secret key used by the target device, template attack (TA) is broadly accepted as the strongest power analysis attack from the perspective of information theory. In order to fully exploit the power leakages of the reference device, one usually has to concern the power leakages at different interesting points. Then, a natural question is how many interesting points should be used in a TA? We note that the number of interesting points one uses directly decides the profiling efficiency of TA. In light of this, we evaluate the optimal number of interesting points in simulated scenarios, and the evaluation results bring us an empirically useful formula. Then, in order to validate the empirical formula, we perform TA using power traces provided by DPA Contest v4.1. In the real scenario, the correlation method is used to select the interesting points, and the S-Box output of the 1st round AES encryption is chosen as the target intermediate value. Evaluation results show that the empirical formula is indeed correct and can be useful in practice.

1. Introduction

In practice, crypto algorithms are always implemented on crypto devices (e.g., microcontrollers, FPGA, ASIC, etc.). Different forms of side channel leakages exist when a crypto device is in operation. The side channel leakages of a crypto device can be used to recover the secret key, and that is the idea of side channel attacks. Typical side channel attacks include timing attacks (Kocher, 1996; Kelsey et al., 1998), electromagnetic attacks (Agrawal et al., 2003a; Gandolfi et al., 2001), power analysis attacks (Kocher et al., 1999; Akkar et al., 2000), and their combinations (Agrawal et al., 2003b; Souissi et al., 2012). Among different types of side channel attacks, power analysis attacks have received the most attention over the last two decades. The reasons are that power analysis attacks are relatively easy to implement, and the attack price is relatively low (Mangard et al., 2007).

The first successful power analysis attack was reported by Kocher et al. (1999). Since then, different forms of attacks, such as template attack (TA) (Chari et al., 2003), correlation power analysis (CPA) (Brier et al., 2004; Le et al., 2006), stochastic model based power analysis (SMPA) (Schindler et al., 2005; Lemke-Rust and Paar, 2007), and mutual information analysis (MIA) (Girelichs et al., 2008; Veyrat-Charvillon and Standaert, 2009) were proposed. Among them, TA is accepted as the strongest power analysis attack from the perspective of information theory. The reasons are that in TA a reference device identical to the target device can be used to accurately characterize the power leakages of the target device, and the characterized power leakages can be used to efficiently recover the secret key used by the target device.

Specifically, TA was proposed by Chari et al. (2003). In TA, the power leakages of the target device at different interesting points can be used to recover the secret key. The working procedure of TA consists of two phases, i.e., profiling and key-recovery. In profiling, mean vectors and covariance matrices are respectively used to characterize signals and noises at different interesting points, and one can obtain the so called templates. In order to accurately characterize signals and noises at different interesting points, a large number of power traces is usually needed in profiling. In key-recovery, a small number of power traces measured from the target device is used to recover the secret key. Under the assumption that noises at different interesting points follow the multivariate normal distribution, one can compute the match probability between the power leakages contained in a small number of power traces measured from the target device and the characterized templates. Among different key hypotheses, the key hypothesis that makes the match probability the largest is accepted as the secret key used by the target device.

Considering that it is the strongest form of power analysis attack, TA is widely used to practically evaluate the physical security of crypto devices, either unprotected or protected. However, problems still exist in TA, and they influence the effect of TA. In light of this, previous works (Rechberger and Oswald, 2004;

* Corresponding author.
E-mail addresses: zhanghailong@iie.ac.cn (H. Zhang), zhouyongbin@iie.ac.cn (Y. Zhou).

http://dx.doi.org/10.1016/j.jss.2016.07.028
0164-1212/

Agrawal et al., 2005; Oswald and Mangard, 2007; Gierlichs et al., 2006; Batina et al., 2008; Bär et al., 2010; Durvaux and Standaert, 2016; Archambeau et al., 2006; Elaabid and Guilley, 2010; Standaert and Archambeau, 2008) partially addressed problems exist in TA, which makes TA more applicable. First, in order to exploit the power leakages of the target device at different interesting points, one needs to use certain technique to effectively choose interesting points. In fact, the quality of the chosen interesting points directly decides the profiling efficiency of TA. Therefore, different interesting points chosen methods were proposed. Depending on their working principles, these methods can be divided into two groups. Methods in the first group choose the interesting points by using the data dependence property of the power leakages. Typical ones include the difference of means method (Rechberger and Oswald, 2004), the correlation method (Agrawal et al., 2005; Oswald and Mangard, 2007), and the T-Test method (Gierlichs et al., 2006; Batina et al., 2008; Bär et al., 2010; Durvaux and Standaert, 2016). On the other hand, some classification methods (e.g., principal component analysis (Archambeau et al., 2006; Elaabid and Guilley, 2010) and fisher linear discriminant analysis (Standaert and Archambeau, 2008)) are used to transform power traces, and the components that induce significant characterizations are exploited to recover the secret key.

Secondly, the optimal number of interesting points one should use in a certain scenario is still an open problem in TA. On one hand, when a small number of interesting points is chosen, the power leakages of the target device are not exploited efficiently, which means information loss and the key-recovery efficiency of TA is negatively influenced. On the other hand, when a large number of interesting points is chosen, the size of the covariance matrices is too large. In this case, numerical precision problems relate to the inversion of the covariance matrices of different templates will significantly lower the key-recovery efficiency of TA (Choudary and Kuhn, 2014; Lommé et al., 2013). In real scenarios, there exists four factors that may influence the optimal number of interesting points, i.e., the number of profiling traces and key-recovery traces, the signal-to-noise ratio, and the cross correlation between noises at different interesting points. Unfortunately, it is still not clear how different parameters influence the optimal number of interesting points, and one can just empirically choose a certain number of interesting points according to the engineering intuition, which is usually not optimal and therefore will influence the profiling efficiency of TA (Lerman et al., 2015).

In light of this, we evaluate the optimal number of interesting points in simulated scenarios. Specifically, we vary the number of power traces used in profiling and key-recovery, we vary the signal-to-noise ratio, and we vary the cross correlation between noises at different interesting points. We evaluate in each scenario the optimal number of interesting points. Based on the evaluation results, we can obtain an empirical formula. Then, in order to verify the validity of the empirical formula, we perform TA using the power traces provided by DPA Contest v4.1. In the real scenario, we use the correlation method to choose interesting points, and the S-Box output of the 1st round AES encryption is chosen as the target intermediate value. Empirical evaluation results show that the empirical formula reflects the real cases and can be useful in practice.

The organization of is paper is as follows. In Section 2, we present the working procedure of TA. In Section 3, we evaluate in simulated scenarios the optimal number of interesting points, and an empirically useful formula is obtained. In Section 4, we use power traces provided by DPA Contest v4.1 to verify that the empirical formula obtained from the simulated scenarios falls in line with the real cases. Finally, conclusions are given in Section 5.

2. Working procedure of template attack

The working procedure of TA consists of two phases, i.e., profiling and key-recovery. Firstly, the reference device identical to the target device can be used in profiling to accurately characterize the power leakages of the target device at different interesting points; then, in key-recovery, the secret key can be recovered utilizing the characterized power leakages about the target device. We respectively present the working procedure of profiling and key-recovery in this section.

2.1. Profiling

Because the reference device is under full control, power traces measured from the reference device can be used to characterize the power leakages of the target device. Under the assumption that the reference device is identical to the target device, its power leakages should be identical or highly similar to that of the target device. Therefore, the power leakages of the target device can be obtained.

In TA, the power leakages of the target device at different interesting points are exploited to recover the secret key used by the target device. Here, interesting points are those power samples that correspond to the processing of the target intermediate value v. The target intermediate value v is usually a sensitive intermediate value that depend on the secret key used by the target device.

It is assumed in TA that the power leakages at different interesting points follow the multivariate normal distribution. Here, we note that the power leakage of the target device at a single interesting point is usually assumed to be composed of signal and noise, while the signals at different interesting points are assumed to be mutually independent, the noises at different interesting points are usually assumed to be correlated. Therefore, in profiling, the mean vectors and the covariance matrices are used to respectively characterize the signals and the noises at different interesting points.

In order to relatively accurately characterize the signals and the noises at different interesting points, a large number of power traces is needed in profiling. However, the number of power traces available in profiling depends on the practical situation. Usually the number of profiling traces is limited. Under the known plaintext attack scenario, one can randomly feed n plaintexts $p_1, p_2, \ldots, p_n$ into the reference device. Using the leakage measurement setup, n power traces $I_1, I_2, \ldots, I_n$ can be measured and obtained during the operation of the reference device.

With the profiling traces, one needs to use a certain technique (shown in Section 1) to choose the interesting points. Here, we note that the reference device is usually assumed to under full control. Therefore, in profiling the secret key used by the reference device is assumed to be known, and the value of the target intermediate value v is known.

Before using the profiling traces $I_1, I_2, \ldots, I_n$ to characterize signals and noises at different interesting points, one needs to divide power traces into different groups according to the value of the target intermediate value v, i.e., power traces corresponding to the same value of v are placed into the same group. For example, assume the 1st S-Box output of the 1st round AES encryption is chosen as the target intermediate value v; then, one can divide n power traces into 256 groups.

If we denote power traces in the ith group as $I_1, I_2, \ldots, I_{n_i}$; then,

$$\mathbf{m}_i = \frac{1}{n_i}\sum_{j=1}^{n_i}\mathbf{t}_j, \mathbf{C}_i = \frac{1}{n_i - 1}\sum_{j=1}^{n_i}(\mathbf{t}_j - \mathbf{m}_i)^T(\mathbf{t}_j - \mathbf{m}_i), \quad (1)$$

where $\mathbf{t}_j$ denotes the power leakages contained in power trace I_j. We call $(\mathbf{m}_i, \mathbf{C}_i)$ a *template*. Obviously, the number of templates one can obtain is equal to the number of groups.

2.2. Key-recovery

In key-recovery, only a small number of power traces $I_1, \ldots, I_a$ measured from the target device can be obtained. For each power trace I_j, $j \in [1, a]$, one computes the match probability $P_{i,j}$ between the power leakages $\mathbf{t}_j$ contained in power trace I_j and the template $(\mathbf{m}_i, \mathbf{C}_i)$:

$$P_{i,j}(\mathbf{t}_j; (\mathbf{m}_i, \mathbf{C}_i)) = \frac{\exp(-\frac{1}{2}(\mathbf{t}_j - \mathbf{m}_i)^T \mathbf{C}_i^{-1}(\mathbf{t}_j - \mathbf{m}_i))}{\sqrt{(2\pi)^t \cdot \det(\mathbf{C}_i)}}. \quad (2)$$

Under the assumption that the power leakages obtained from different power traces are independent, the match probabilities $P_{i,1}, \ldots, P_{i,a}$ can be multiplied together, and P_i can be obtained:

$$P_i(\mathbf{t}_1, \ldots, \mathbf{t}_a; (\mathbf{m}_i, \mathbf{C}_i)) = \prod_{j=1}^{a} P_{i,j}(\mathbf{t}_j; (\mathbf{m}_i, \mathbf{C}_i)). \quad (3)$$

According to the maximum likelihood principle, among different match probabilities, the largest match probability P_{max} indicates the secret key used by the target device. Therefore, one should firstly compute the match probabilities between the power leakages and different templates; then, by comparing the match probabilities computed under different templates, the secret key used by the target device can be recovered.

3. Simulated evaluation

First of all, we would like to present the principle of the evaluation metric and the meaning of different parameters; then, we vary the values of different parameters, and we evaluate the optimal number of interesting points in different scenarios; finally, by analyzing the simulation results we propose an empirically useful formula.

3.1. Evaluation metric and parameters

In order to quantitatively evaluate the key-recovery efficiency of TA when different numbers of interesting points are used, we use the success rate (SR) as the evaluation metric. We note that SR was firstly proposed by Standaert et al. (2009) in EUROCRYPT 2009. Since then, SR was widely used to evaluate the key-recovery efficiency of different power analysis methods. Formally, the definition of SR is the probability that the secret key used by the target device can be successfully recovered given a certain number of power traces. In power analysis attacks, the divide-and-conquer strategy is employed, i.e., a large key can be divided into some small pieces, and each key piece can be recovered independently. Therefore, the probability that one can successfully recover the small key piece given a certain number of power traces is also known as the partial success rate (PSR). In our evaluations, only a small key piece is needed to be recovered once. Therefore, PSR is exactly the used evaluation metric.

On the other hand, we have shown in Section 1 that there are four parameters that may influence the optimal number of interesting points in real scenarios. Firstly, profiling traces are power traces measured from the reference device. One needs to use profiling traces to separate signals and noises at different interesting points. According to statistics (Coron et al., 2001; Ding et al., 2014; Fei et al., 2012), the larger the number of profiling traces, the more accurately the signals and the noises at different interesting points can be separated. Secondly, key-recovery traces are power traces measured from the target device. Because of the strong power of TA, only a small number of key-recovery traces is needed to successfully recover the secret key used by the target device. Thirdly, signal-to-noise ratio (also known as SNR) is defined as the ratio between the variance of signal at a single interesting point and the variance of noise at the same interesting point. Usually, the higher the value of SNR, the easier the secret key used by the target device can be recovered. In practice, one usually does not know exactly the SNR level of the target device. In TA, one can use profiling traces to estimate the SNR level of the target device (Durvaux et al., 2014). Finally, noises at different interesting points are usually not independent, one can characterize and exploit the cross correlation between noises at different interesting points in the process of recovering the secret key used by the target device. Usually, the higher the cross correlation between noises at different interesting points is, the more informative the interesting points are, and the higher the key-recovery efficiency of TA is.

Before we show the simulated evaluation results, we firstly present how to obtain the simulated power traces. We assume the power leakage at a single interesting point is composed of signal and noise, where signal corresponds to the power consumption of the target intermediate value v while noise is irrelevant to the power consumption of v. We assume the power consumption of each bit of v is different, and each bit of v leaks independently. This is actually the idea of the stochastic model and is practically verified to be correct (Doget et al., 2011; De Santis et al., 2013). We choose a number from the interval (1, 10) randomly to simulate the power consumption of a certain bit of the target intermediate value v (Whitnall et al., 2011). For example, if we choose the 1st S-Box output of the 1st round AES encryption as the target intermediate value; then we can randomly choose 8 numbers from the interval (1, 10), and each number is used to simulate the power consumption of one bit of the target intermediate value. Here, we note that based on previous experience (Whitnall et al., 2011), randomly choosing a number from the interval (1, 10) to simulate the power consumption of one bit of the target intermediate value can relatively accurately reflect the power leakages of signals in real cases. Then, for each possible value of v, we can compute and obtain a simulated signal sample, which is the sum of the power leakage of each bit of v. Considering that the plaintexts are randomly generated and values of v are equiprobable to occur (Batina et al., 2009), we can compute the variance of signal at a certain interesting point.

On the other hand, noises at different interesting points are usually assumed to follow the multivariate normal distribution. In our case, the correlation coefficient is used to measure the cross correlation between noises at different interesting points. In a certain SNR level scenario, when the variance of the signal at a certain interesting point is known, the variance of the noise at the interesting point can also be known. We use MATLAB to do the simulation evaluations, therefore we can use the function *mvnrnd* to generate noise samples at different interesting points. After we obtain noise samples at different interesting points, we can add a noise sample to a signal sample and obtain a simulated power sample. When the simulated power traces are obtained, we can evaluate the optimal number of interesting points in simulated scenarios.

3.2. Evaluation results

In the simulated scenarios, we can vary the number of profiling traces and the number of key-recovery traces, we can vary the SNR level, and we can vary the cross correlation between noises at different interesting points, then we can evaluate the optimal number of interesting points in different scenarios. The simulation results will show us the practical influence of different parameters on the optimal number of interesting points in different scenarios.

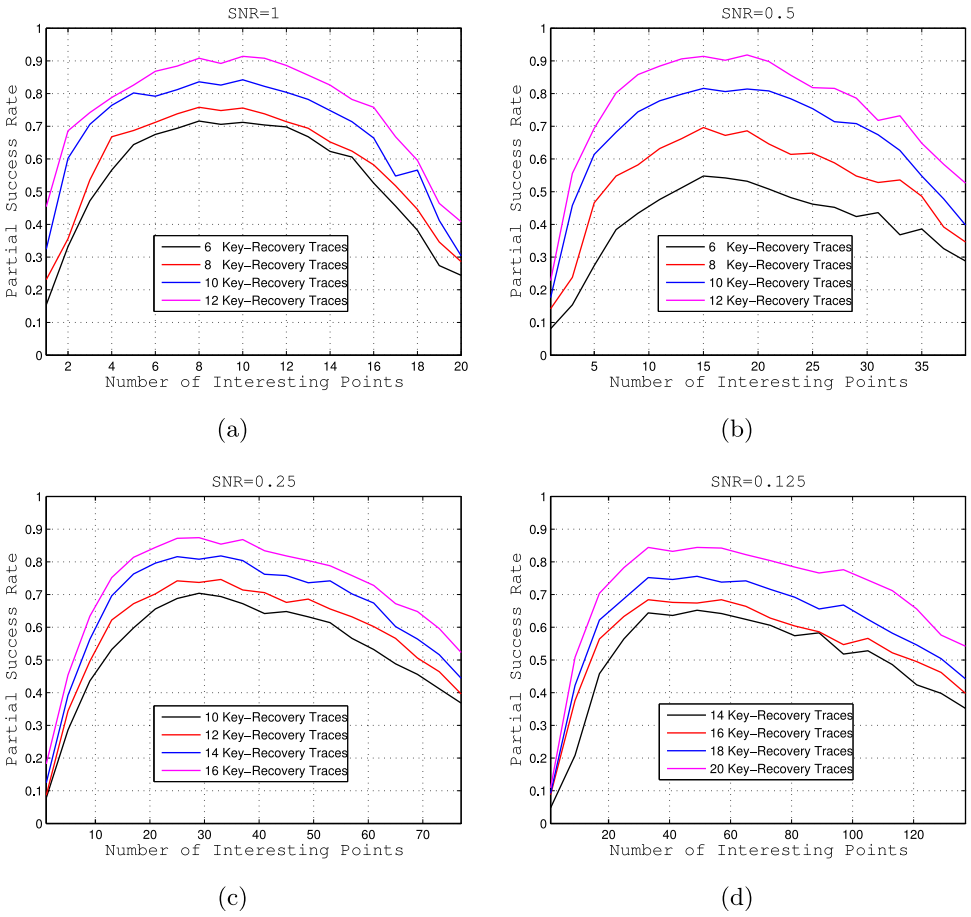

Fig. 1. Simulation results in four different scenarios. (a) 8000 profiling traces, cross correlation = 0.6, and SNR = 1. (b) 16,000 profiling traces, cross correlation = 0.6, and SNR = 0.5. (c) 32,000 profiling traces, cross correlation = 0.6, and SNR = 0.25. (d) 64,000 profiling traces, cross correlation = 0.6, and SNR = 0.125.

Firstly, we would like to evaluate the influence of the number of key-recovery traces on the optimal number of interesting points. With regard to this, we fix the number of profiling traces and the cross correlation between noises at different interesting points. We vary the number of key-recovery traces, and we evaluate its influence on the optimal number of interesting points in different SNR level scenarios.

It can be seen from Fig. 1 that the number of power traces used in key-recovery does not actually influence the optimal number of interesting points, it only influences the key-recovery efficiency of TA. The reason is that a larger number of key-recovery traces definitely brings more power leakages about the target device, which therefore helps improve the key-recovery efficiency of TA.

Then, we would like to evaluate the influence of the cross correlation between noises at different interesting points on the optimal number of interesting points. With regard to this, we evaluate the optimal number of interesting points in four different scenarios. In each scenario, we fix the number of profiling traces, the number of key-recovery traces, and the signal-to-noise ratio, we only vary the cross correlation between noises at different interesting points and we evaluate its influence on the optimal number of interesting points.

We note that depending on the previous experiences about the leakage characterization of crypto devices, the cross correlation between noises at different interesting points is usually around 0.3 to 0.6 (Rechberger and Oswald, 2004). Therefore, we vary the cross correlation between noises at different interesting points in this interval, and we evaluate its influence on the number of interesting points one should use. It can be seen from Fig. 2 that the cross correlation between noises at different interesting points does not actually influence the optimal number of interesting points one should use, it only influences the key-recovery efficiency of TA in a certain scenario.

In fact, we note that we evaluate the influence of the cross correlation between the noises at different interesting points on the optimal number of interesting points in other SNR level scenarios. The evaluation results show similar phenomenon as Fig. 2 shows. Therefore, we can come to the conclusion that the optimal number of interesting points in essence has nothing to do with the cross correlation parameter.

Overall, the above evaluation results show that either a larger number of key-recovery traces or the interesting points that induce a higher cross correlation helps improve the leakage exploitation of the target device and therefore improve the key-recovery efficiency of TA. So, on one hand, one should use as many key-recovery traces as possible; on the other hand, one should use the interesting points that induce a high cross correlation to make TA as powerful as possible in practice.

Thirdly, we would like to evaluate the influence of the number of profiling traces on the optimal number of interesting points. With regard to this, we fix the cross correlation between noises at different interesting points, we fix the number of key-recovery traces one uses, and we fix the SNR level. We only vary the number of profiling traces, and we evaluate its influence on the optimal number of interesting points.

It can be seen from Fig. 3 that the number of profiling traces one uses does essentially influence the optimal number of interesting points in TA. Qualitatively, the larger the number of power

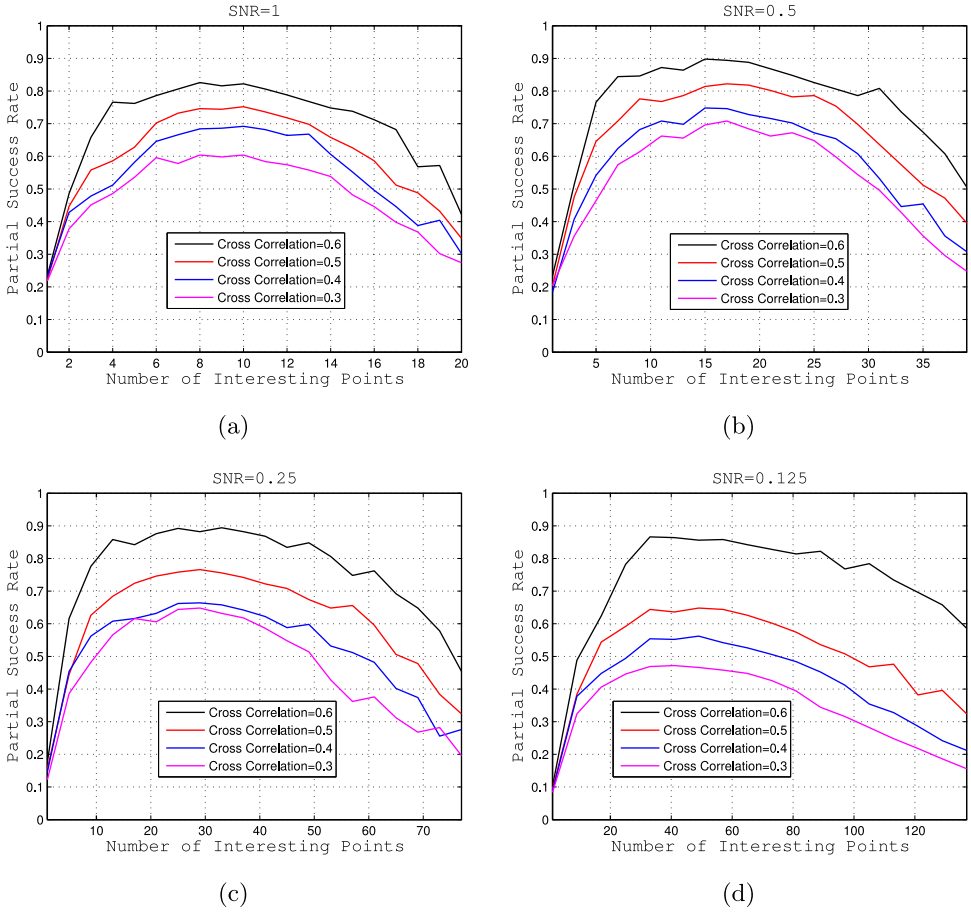

Fig. 2. Evaluation results in four scenarios. (a) 8000 profiling traces, 8 key-recovery traces, and SNR = 1. (b) 16,000 profiling traces, 12 key-recovery traces, and SNR = 0.5. (c) 32,000 profiling traces, 16 key-recovery traces, and SNR = 0.25. (d) 64,000 profiling traces, 20 key-recovery traces, and SNR = 0.125.

traces used in profiling, the larger the number of interesting points should be used. Quantitatively, the optimal number of interesting points increases linearly with the number of power traces used in profiling.

We note here that because of the variety of different parameters, it is impossible to show the influence of the values of different parameters on the optimal number of interesting points in all scenarios. However, large amounts of simulation results show that the number of key-recovery traces and the cross correlation between noises at different interesting points do not essentially influence the optimal number of interesting points, while the number of profiling traces and the SNR level will practically influence the optimal number of interesting points. We will quantitatively analyze this phenomenon in Section 3.3.

3.3. Summation and the empirical formula

According the simulation results, it is the number of profiling traces and the SNR level rather than the cross correlation or the number of key-recovery traces that influence the optimal number of interesting points in TA. The reason is that the accuracy of the covariance matrices of different templates is determined by the SNR level and the number of profiling traces, not by the number of key-recovery traces or the cross correlation between noises at different interesting points.

On one hand, in a certain SNR level scenario, the larger the number of profiling traces, the more accurately the covariance matrices of different templates are characterized. In this case, numerical precision problems related to the covariance matrices of different templates are less serious, and one can use a larger number of interesting points to more efficiently exploit the power leakages of the target device and therefore improve the key-recovery efficiency of TA. On the other hand, when the number of profiling traces is fixed, the higher the noise level, the more serious numerical precision problems relate to the covariance matrices of different templates are. In this case, a smaller number of interesting points should be used to effectively alleviate numerical precision problems relate to the covariance matrices of different templates and therefore the key-recovery efficiency of TA is negatively influenced.

According to the simulation results, we can obtain some interesting observations. In general, in a certain scenario, when the noise level is fixed, the optimal number of interesting points increases in proportion to the number of profiling traces. For example, when the SNR level equals to 1, each 2 additional interesting points should be used as each 2000 additional profiling traces are used. Also, the number of profiling traces and the optimal number of interesting points decreases in proportion to the SNR level. For example, as the SNR level decreases from 1 to 0.5, the number of profiling traces needs to be used increases from 4000 to 8000, and the number of interesting points needs to be used increases from 4 to 8. Based on the observations we can propose a formula, which is empirically useful to estimate the optimal number of interesting points. Assume the SNR level is $\frac{1}{2^c}$, where $c > 0$; and assume the number of profiling traces is N_{pro}, where $N_{pro} > 2000$; then the optimal number of interesting points N_{poi} approximately follows Eq. (4):

$$N_{poi} = \frac{N_{pro}}{2000 \times 2^c} \times (c+1). \tag{4}$$

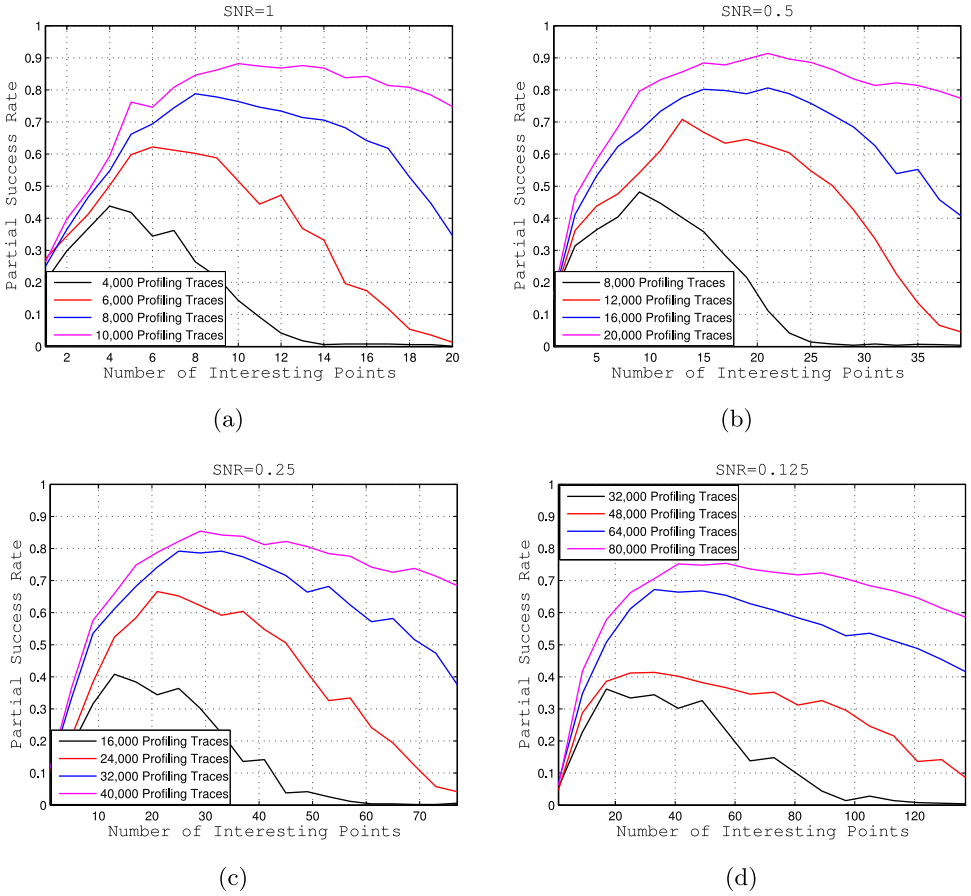

Fig. 3. Evaluation results in four different scenarios. (a) cross correlation = 0.6, 8 key-recovery traces, and SNR = 1. (b) cross correlation = 0.6, 10 key-recovery traces, and SNR = 0.5. (c) cross correlation = 0.6, 12 key-recovery traces, and SNR = 0.25. (d) cross correlation = 0.6, 14 key-recovery traces, and SNR = 0.125.

Here, one may wonder what if the value of $c < -1$. Because in this case the value of $N_{poi} < 0$, which is meaningless in practice. We note that nowadays crypto devices are always equipped with countermeasures (e.g., a classical countermeasure is to add noises to lower down the SNR level of the measured power leakages and therefore make power analysis attacks more difficult to succeed). Therefore, the SNR level of the measured power traces is usually below 1, which means the value of c is usually larger than zero, i.e., $c > 0$. Therefore, the value of N_{poi} is generally larger than zero, i.e., $N_{poi} > 0$, and the empirical formula shown in Eq. (4) is meaningful in practical scenarios.

4. Empirical verification

Now, the power traces provided by DPA Contest v4.1 are used to verify the validity of the empirical formula on the optimal number of interesting points. The power traces provided by DPA Contest v4.1 correspond to the power consumption of a low-entropy masking scheme protected AES-256 software implementation on an Atmel ATMega-163 smart-card. In total, 10 sets of power traces, each containing 10,000 power traces are provided. Power traces in the same set correspond to a certain secret key. One can download power traces provided by DPA Contest v4.1 from the DPA Contest website (http://www.dpacontest.org/v4/rsm-traces.php). We can obtain the information about the plaintext, the ciphertext, the mask offset, and the secret key corresponding to each power trace. In our case, we assume the mask offset is known. We do so to focus on the correctness of the empirical formula on the optimal number of interesting points, rather than to focus on how to get rid of the effect the mask offset poses on the key-recovery efficiency of TA.

The 1st S-Box output of the 1st round AES encryption is chosen as the target intermediate value, and the correlation method is used to choose interesting points, i.e., we compute the correlation coefficient between the hypothetical and the real power leakages at different power samples, those power samples that induce correlation peaks are chosen as the interesting points. 20 interesting points are chosen. Before verifying the validity of the empirical formula on the optimal number of the interesting points, we firstly evaluate the average cross correlation between the noises at differen interesting points and the average SNR level of the power traces. We note that the value of the average cross correlation between the noises at different interesting points and the value of the average SNR level of the power traces are important for us to verify the validity of the empirical formula.

We firstly divide 100,000 power traces into 256 groups according to the values of the 1st S-Box output of the 1st round AES encryption. For power traces in each group, their signal parts are equal, and only the noise parts are different. Using the power traces in the same group, we compute the correlation coefficient between noises at different interesting points. Then, by averaging the correlation coefficient computed with different groups of power traces, we can obtain the average cross correlation between noises at different interesting points. Evaluation result shows that the average cross correlation between noises at different interesting points is 0.3515, which falls in line with our previous experience about the cross correlation between noises at different interesting points.

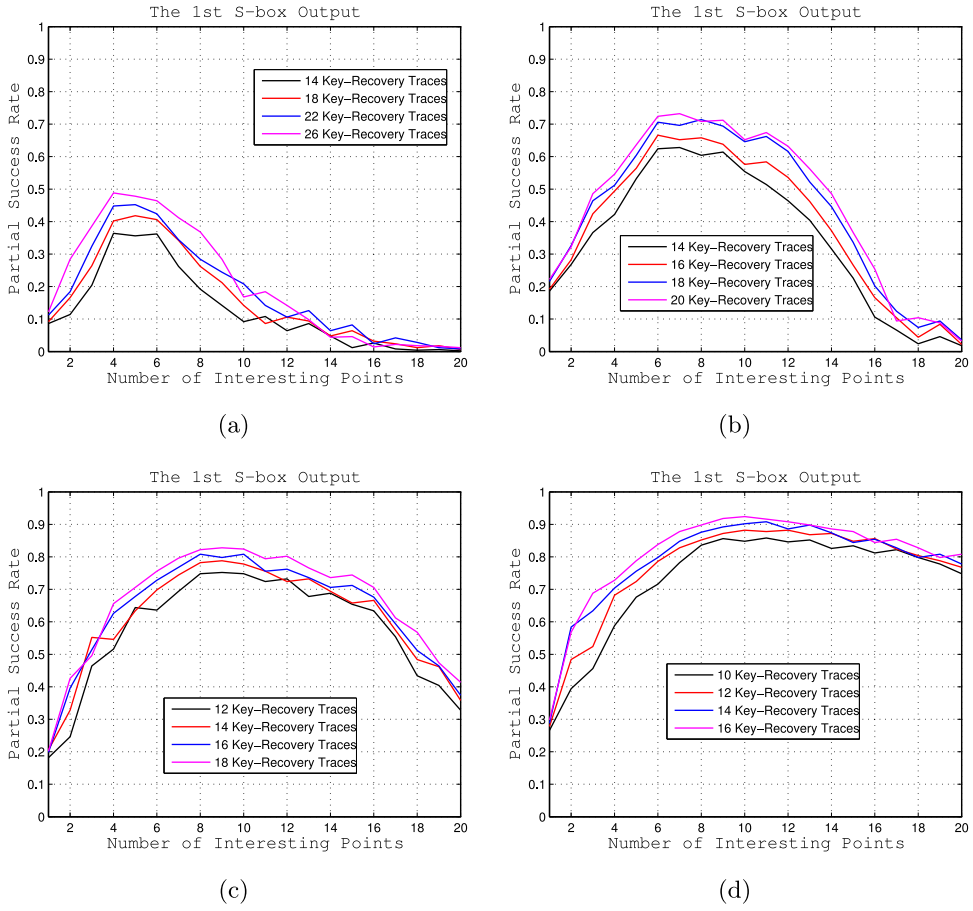

Fig. 4. Empirical evaluations in different scenarios. (a) 4000 profiling traces. (b) 6000 profiling traces. (c) 8000 profiling traces. (d) 10,000 profiling traces.

In order to evaluate the average SNR level of the measured power traces, we firstly compute the mean value and the variance for the power leakages at the chosen 20 interesting points using power traces in the same group. Then, the variance of the 256 mean values is assumed to be the variance of signals, and the mean of the 256 variances is assumed to be the variance of noises. By averaging the SNR levels computed at 20 interesting points, we can estimate the SNR level of the measured power traces. Evaluation result shows that the average SNR level of the power traces is 0.9856. It can be seen that the noise level of the measured power traces is relatively low. This is because the target is an AES software implementation on a smart-card, whose noise level is relatively low.

According to the empirical formula we know that in this scenario a relatively small number of interesting points is needed, and the optimal number of interesting points depend on the number of profiling traces. By varying the number of profiling traces and the number of key-recovery traces, we will practically verify this fact.

Fig. 4 shows that the optimal number of interesting points varies with the number of profiling traces, not the number of key-recovery traces. On one hand, the larger the number of profiling traces, the more accurately the covariance matrices of different templates are characterized, the less serious numerical precision problems relate to the covariance matrices of different templates are, and the larger the number of interesting points can be used. On the other hand, the number of key-recovery traces does not influence the optimal number of interesting points, it only influences the key-recovery efficiency of TA, i.e., the larger the number of key-recovery traces, the larger the amount of power leakages can be used, and the higher the key-recovery efficiency of TA is.

Also, it can be seen from Fig. 4 that the practical evaluation results fit the empirically obtained formula well. In the simulated evaluations, we have evaluated the number of interesting points one should use in the scenario that the SNR level is equal to 1. The evaluation results show that approximately 4 interesting points should be used when 4,000 profiling traces are used, and each 2 additional interesting points should be used as each 2000 additional profiling traces are used. Now, in the practical scenario, when the SNR level of the measured power traces is close to 1, the evaluation results fall in line with the simulation results, i.e., when 4,000 profiling traces are used, approximately 4 interesting points should be used, and each 2 additional interesting points should be used as each 2000 additional profiling traces are used.

Overall, evaluation results in real scenario show that the empirically obtained formula can relatively accurately reflect the practical cases, and therefore can be useful to guide evaluators to choose the optimal number of interesting points. Considering that the number of interesting points significantly influences the profiling efficiency of TA, and therefore it will influence our knowledge about the physical security of crypto devices in practical scenarios, this work will help the crypto device evaluators to make sure that the chosen number of interesting points is optimal, and it will not lower down the performance of TA in practice. Finally, we note that although in the real scenario the correlation method is used to choose interesting points and the 1st S-Box of the 1st round AES encryption, the methodology provided in this paper should be applicable to other scenarios. Indeed, the only thing that influences the optimal number of interesting points is the SNR level of the crypto device and the number of profiling traces. From this point of view, the method used to choose interesting points and the

chosen target intermediate value does not actually influence the optimal number of interesting points.

5. Conclusions and future work

In this paper, the optimal number of interesting points one should use in TA is evaluated in the simulated scenarios. Specifically, we vary the values of different parameters, and we evaluate the performance of TA in different scenarios. Then, by analyzing the key-recovery efficiency of TA in different scenarios, we obtain an empirically useful formula. In order to verify the correctness and the usefulness of the empirically obtained formula, power traces provided by DPA Contest v4.1 are used to perform a successful TA. Evaluation results show that the empirical formula can relatively accurately estimate the optimal number of interesting points. Therefore, we come to the conclusion that the empirical formula is indeed correct and can be useful in practice.

However, before using the empirical formula to decide the optimal number of interesting points, one needs to first know the SNR level and the number of profiling traces. Usually the SNR level of a crypto device is not clear, and the evaluator needs to use the profiling traces to estimate the SNR level of the crypto device. Also, we note that the empirical formula obtained in this paper is only applicable to the gaussian template scenario. In practice, there may exist the non-gaussian template scenario. What is the optimal number of interesting points in the non-gaussian template scenario is still not known, which is left as the future work.

Acknowledgements

This work was partially supported by the National Natural Science Foundation of China (Nos. 61472416 and 61272478), and the Strategic Priority Research Program of the Chinese Academy of Sciences (Nos. XDA06010701 and XDA06010703). We acknowledge their supports.

References

Agrawal, D., Archambeault, B., Rao, J.R., Rohatgi, P., 2003a. The EM side-channel(s). In: CHES 2002, LNCS, 2523, pp. 29–45.
Agrawal, D., Rao, J.R., Rohatgi, P., 2003b. Multi-channel attacks. In: CHES 2003, LNCS, 2779, pp. 2–16.
Agrawal, D., Rao, J.R., Rohatgi, P., Schramm, K., 2005. Templates as master keys. In: CHES 2005, LNCS, 3659, pp. 15–29.
Akkar, M.-L., Bevan, R., Dischamp, P., Moyart, D., 2000. Power analysis, what is now possible .... In: ASIACRYPT 2000, LNCS, 1976, pp. 489–502.
Archambeau, C., Peeters, E., Standaert, F.-X., Quisquater, J.-J., 2006. Template attacks in principal subspaces. In: CHES 2006, LNCS, 4249, pp. 1–14.
Bär, M., Drexler, H., Pulkus, J., 2010. Improved template attacks. COSADE.
Batina, L., Gierlichs, B., Lemke-Rust, K., 2008. Comparative evaluation of rank correlation based DPA on an AES prototype chip. ISC 2008, LNCS 5222, 341–354.
Batina, L., Gierlichs, B., Lemke-Rust, K., 2009. Differential cluster analysis. In: CHES 2009, LNCS, 5747, pp. 112–127.
Brier, E., Clavier, C., Olivier, F., 2004. Correlation power analysis with a leakage model. In: CHES 2004, LNCS, 3156, pp. 16–29.
Chari, S., Rao, J.R., Rohatgi, P., 2003. Template attacks. In: CHES 2002, LNCS, 2523, pp. 13–28.
Choudary, O., Kuhn, M.G., 2014. Efficient template attacks. In: CARDIS 2013, LNCS, 8419, pp. 253–270.
Coron, J.-S., Kocher, P., Naccache, D., 2001. Statistics and secret leakage. In: FC 2001, LNCS, 1962, pp. 157–173.
De Santis, F., Kasper, M., Mangard, S., Sigl, G., Stein, O., Stöttinger, M., 2013. On the relationship between correlation power analysis and the stochastic approach: an ASIC designer perspective. In: INDOCRYPT 2013, LNCS, 8250, pp. 215–226.
Ding, A.A., Zhang, L., Fei, Y., Luo, P., 2014. A statistical model for higher order DPA on masked devices. In: CHES 2014, LNCS, 8731, pp. 147–169.
Doget, J., Prouff, E., Rivain, M., Standaert, F.-X., 2011. Univariate side cnannel attacks and leakage modeling. J. Cryptogr. Eng. 1 (2), 123-14.
Durvaux, F., Standaert, F.-X., 2016. From improved leakage detection to the detection of points of interests in leakage traces. In: EUROCRYPT 2016, LNCS, 9665, pp. 240–262.
Durvaux, F., Standaert, F.-X., Veyrat-Charvillon, N., 2014. How to certify the leakage of a chip? In: EUROCRYPT 2014, LNCS, 8441, pp. 459–476.
Elaabid, M.A., Guilley, S., 2010. Practical improvements of profiled side-channel attacks on a hardware crypto-accelerator. In: AFRICACRYPT 2010, LNCS, 6055, pp. 243–260.
Fei, Y., Luo, Q., Ding, A.A., 2012. A statistical model for DPA with novel algorithmic confusion analysis. In: CHES 2012, LNCS, 7428, pp. 233–250.
Gandolfi, K., Mourtel, C., Olivier, F., 2001. Electromagnetic analysis: concrete results. In: CHES 2001, LNCS, 2162, pp. 251–261.
Gierlichs, B., Lemke-Rust, K., Paar, C., 2006. Template vs. stochastic methods - a performance analysis for side chennel cryptanalysis. In: CHES 2006, LNCS, 4249, pp. 15–29.
Girelichs, B., Batina, L., Tuyls, P., Preneel, B., 2008. Mutual information analysis. In: CHES 2008, LNCS, 5154, pp. 426–442.
Kelsey, J., Schneier, B., Wagner, D., Hall, C., 1998. Side channel cryptanalysis of product ciphers. In: ESORICS 1998, LNCS, 1485, pp. 97–110.
Kocher, P., Jaffe, J., Jun, B., 1999. Differential power analysis. In: CRYPTO 1999, LNCS, 1666, pp. 388–397.
Kocher, P.C., 1996. Timing attacks on implementations of Diffie–Hellman, RSA, DSS, and other systems. In: CRYPTO 1996, LNCS, 1109, pp. 104–113.
Le, T.-H., Clédiére, J., Canovas, C., Robisson, B., Serviére, C., Lacoume, J.-L., 2006. A proposition for correlation power analysis enhancement. In: CHES 2006, LNCS, 4249, pp. 174–186.
Lemke-Rust, K., Paar, C., 2007. Analyzing side channel leakage of masked implementations with stochastic methods. In: ESORICS 2007, LNCS, 4734, pp. 454–468.
Lerman, L., Poussier, R., Bontempi, G., Markowitch, O., Standaert, F.-X., 2015. Template attacks vs. machine learning revisited (and the curse of dimensionality in side-channel analysis). In: COSADE 2015, LNCS, 9064, pp. 20–33.
Lommé, V., Prouff, E., Roche, T., 2013. Behind the scene of side channel attacks. In: ASIACRYPT 2013, LNCS, 8269, pp. 506–525.
Mangard, S., Oswald, E., Popp, T., 2007. Power Analysis Attacks. Springer, Heidelberg.
Oswald, E., Mangard, S., 2007. Template attacks on masing-resistance is futile. In: CT-RSA 2007, LNCS, 4377, pp. 243–256.
Rechberger, C., Oswald, E., 2004. Practical template attacks. In: WISA 2004, LNCS, 3325, pp. 440–456.
Schindler, W., Lemke-Rust, K., Paar, C., 2005. A stochastic model for differential side channel cryptanalysis. In: CHES 2005, LNCS, 3659, pp. 30–46.
Souissi, Y., Bhasin, S., Guilley, S., Nassar, M., Danger, J.-L., 2012. Towards different flavors of combined side channel attacks. In: CT-RSA 2012, LNCS, 7178, pp. 245–259.
Standaert, F.-X., Archambeau, C., 2008. Using subspace-based template attacks to compare and combine power and electromagnetic information leakage. In: CHES 2008, LNCS, 5154, pp. 411–425.
Standaert, F.-X., Malkin, T.G., Yung, M., 2009. A unified framework for the analysis of side-channel key recovery attacks. In: EUROCRYPT 2009, LNCS, 5479, pp. 443–461.
Veyrat-Charvillon, N., Standaert, F.-X., 2009. Mutual information analysis: how, when and why? In: CHES 2009, LNCS, 5747, pp. 429–443.
Whitnall, C., Oswald, E., Standaert, F.-X., 2011. An exploration of the Kolomogov–Smirnov test as a competititor to mutual information analysis. In: CARDIS 2011, LNCS, 7079, pp. 234–251.

Hailong Zhang was born in 1986. He received the Ph.D degree in information security from Institute of Information Engineering, Chinese Academy of Sciences in January 2015. His main research interest includes side-channel attacks, especially power analysis attacks and the countermeasures and evaluations.

Yongbin Zhou was born in 1973. He received the Ph.D degree in computer science in March 2004. He is now a full professor of State Key Laboratory of Information Security, Institute of Information Engineering, Chinese Academy of Sciences. His research interests include theories and technologies of network and information Security.