How to Choose Interesting Points for Template Attacks More Effectively?

Guangjun Fan[1], Yongbin Zhou[2], Hailong Zhang[2], Dengguo Feng[1]

[1] State Key Laboratory of Computer Science, Institute of Software, Chinese Academy of Sciences
guangjunfan@163.com, feng@tca.iscas.ac.cn
[2] State Key Laboratory of Information Security, Institute of Information Engineering, Chinese Academy of Sciences
zhouyongbin@iie.ac.cn, zhanghailong@iie.ac.cn

Abstract. Template attacks are widely accepted to be the most powerful side-channel attacks from an information theoretic point of view. For template attacks to be practical, one needs to choose some special samples as the interesting points in actual power traces. Up to now, many different approaches were introduced for choosing interesting points for template attacks. However, it is *unknown* that whether or not the previous approaches of choosing interesting points will lead to the best classification performance of template attacks. In this work, we give a negative answer to this important question by introducing a practical new approach which has completely different basic principle compared with all the previous approaches. Our new approach chooses the point whose distribution of samples approximates to a normal distribution as the interesting point. Evaluation results exhibit that template attacks based on the interesting points chosen by our new approach can achieve obvious better classification performance compared with template attacks based on the interesting points chosen by the previous approaches. Therefore, our new approach of choosing interesting points should be used in practice to better understand the practical threats of template attacks.

Keywords: Side-Channel Attacks, Power Analysis Attacks, template attacks, Interesting Points.

1 Introduction

Side-channel attacks are one of the most important threats against modern cryptographic implementations. The basic idea of these attacks is to determine the key of a cryptographic device by exploiting its power consumption [11], its electromagnetic radiation [20], its execution time [19], and many more [21]. Traditional security notions (such as chosen-ciphertext security for public-key encryption schemes) do not provide any security guarantee against such attacks, and many implementations of provably secure cryptosystems were broken by side-channel attacks.

Power analysis attacks have received such a large amount of attention because they are very powerful and can be conducted relatively easily. Therefore, let us focus exclusively on power analysis attacks. As an important method of power analysis attacks, template attacks were firstly proposed by Chari et al. in 2002 [1] and belong to the category of profiled side-channel attacks. Under the assumption that one (an actual attacker or an evaluator) has a reference device identical or similar to the target device, and thus be well capable of characterizing power leakages of the target device, template attacks are widely accepted to be the strongest side-channel attacks from an information theoretic point of view [1]. We note that, template attacks are also important tools to evaluate the physical security of a cryptographic device.

Template attacks consist of two stages. The first stage is the profiling stage and the second stage is the extraction stage. In the profiling stage, one captures some actual power traces from a reference device identical or similar to the target device and builds templates for each key-dependent operation with the actual power traces. In the extraction stage, one can exploit a small number of actual power traces measured from the target device and the templates obtained from the profiling stage to classify the correct (sub)key.

1.1 Motivations

Note that for real-world implementation of cryptography devices, a side-channel leakage trace (i.e. an actual power trace for the case of power analysis attacks) usually contains multiple samples corresponding to the target intermediate values. The reason is that the key-dependent operations usually take more than one instruction cycles. 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 target cryptographic device.

For template attacks to be practical, it is paramount that not all the samples of an actual power trace are part of the templates. To reduce the number of samples and the size of the templates, one needs to choose some special points as the interesting points in actual power traces. Main previous approaches of choosing interesting points for template attacks can be divided into two kinds.

Approaches belong to the first kind try to choose the points which contain the most information about the characterized key-dependent operations as the interesting points with different principles. Classical template attacks [1] generally use the approaches belong to the first kind to choose interesting points. Moreover, many papers [2,3,5,10,12] suggested an accepted guideline for choosing interesting points for the approaches in the first kind. The accepted guideline is that one should only choose one point as the interesting point per clock cycle since more points in the same clock cycle do not provide more information. Disobeying this accepted guideline leads to poorer classification performance of template attacks even if a higher number of interesting points is chosen due to some numerical obstacles when one computes the inverse of the covariance matrices $\mathbf{C}_i$ (Please see Section 2.2 for more details.). Up to now, many different approaches of choosing interesting points which belong to the first kind were

introduced. These approaches are *Correlation Power Analysis based approach* (Chapter 6 in [11]) (CPA), *Sum Of Squared pairwise T-differences based approach* [10] (SOST), *Difference Of Means based approach* [1] (DOM), *Sum Of Squared Differences based approach* [10] (SOSD), *Variance based approach* [16] (VAR), *Signal-to-Noise Ratios based approach* (pp. 73 in [11]) (SNR), *Mutual Information Analysis based approach* [17] (MIA), and *Kolmogorov-Smirnov Analysis based approach* [18] (KSA). One uses these approaches to choose the points which contain the most information about the characterized key-dependent operations as the interesting points by computing the signal-strength estimate $\mathsf{SSE}(t)$ for each point P_t. For example, when one uses Correlation Power Analysis based approach to choose interesting points for template attacks, the signal-strength estimate $\mathsf{SSE}(t)$ is measured by the coefficient of correlation between the actual power consumptions and the hypothetical power consumptions of a point P_t. For these approaches, in each clock cycle, the point with the strongest signal-strength estimate $\mathsf{SSE}(t)$ is chosen as the interesting point.

Approaches belong to the second kind based on the principal components or Fisher's linear discriminant. *Principal Component Analysis based approach* [3] (PCA) and *Fisher's Linear Discriminant Analysis based approach* [9] (LDA) belong to the second kind. We note that, PCA-based template attacks is inefficient due to its high computational requirements [2] and may not improve the classification performance [7]. Therefore, PCA-based template attacks are not considered to be an approach which can be widely used to choose interesting points for template attacks. Moreover, LDA-based template attacks depends on the rare condition of equal covariances [4] (Please see Section 2.2 for more details.), which does not hold for most cryptographic devices. Therefore, it is not a better choice compared with PCA-based template attacks in most settings [4]. Due to these reasons, we ignore PCA-based template attacks as well as LDA-based template attacks and only consider the approaches of choosing interesting points for classical template attacks which are the *most* widely used profiled side-channel attacks in this paper.

However, up to now, it is still *unknown* that whether or not using the above approaches of choosing interesting points will lead to the best classification performance of template attacks. In other words, whether or not there exists other approaches which based on different basic principles will lead to better classification performance of template attacks is still *unclear*. If the answer to this question is negative, we can demonstrate that one can further improve the classification performance of template attacks by using the more advanced approach to choose interesting points rather than by designing some kind of improvements about the mathematical structures of the attacks. In this paper, we try to answer this important question.

1.2 Contributions

In this paper, we firstly present a new approach of choosing interesting points for template attacks which has completely different basic principle compared with all the previous approaches. The theoretical correctness of our new approach is

supported by an important mathematical property of the multivariate Gaussian distribution and the Pearson's chi-squared test for goodness of fit [25].

Furthermore, we experimentally verified that template attacks based on the interesting points chosen by our new approach can achieve obvious better classification performance compared with template attacks based on the interesting points chosen by the previous approaches. This gives a negative answer to the question that whether or not using the previous approaches of choosing interesting points will lead to the best classification performance of template attacks.

Moreover, the computational price of our new approach is low and practical. Therefore, our new approach of choosing interesting points for template attacks can be used in practice to better understand the practical threats of template attacks.

1.3 Related Work

Template attacks were firstly introduced in [1]. Answers to some basic and practical issues of template attacks were provided in [2], such as how to choose interesting points in an efficient way and how to preprocess noisy data. Efficient methods were proposed in [4] to avoid several possible numerical obstacles when implementing template attacks.

Hanley et al. [12] presented a variant of template attacks that can be applied to block ciphers when the plaintext and ciphertext used are unknown. In [8], template attacks were used to attack a masking protected implementation of a block cipher. Recently, a simple pre-processing technique of template attacks, normalizing the sample values using the means and variances was evaluated for various sizes of test data [7].

Gierlichs et al. [10] made a systematic comparison of template attacks and stochastic model based attacks [24]. How to best evaluate the profiling stage and the extraction stage of profiled side-channel attacks by using the information-theoretic and the security metric was shown in [22].

1.4 Organization of This Paper

The rest of this paper is organized as follows. In Section 2, we briefly introduce basic mathematical concepts and review template attacks. In Section 3, we introduce our new approach of choosing interesting points for template attacks. In Section 4, we experimentally verify the effectiveness of the new approach in improving the classification performance of template attacks. In Section 5, we conclude the whole paper.

2 Preliminaries

In this section, we first introduce some basic mathematical concepts which are used in this paper, then briefly review template attacks.

2.1 Basic Mathematical Concepts

We first introduce the Gamma function and the chi-squared distribution. Then, we briefly introduce the concept of the goodness of fit of a statistical model and the Pearson's chi-squared test for goodness of fit.

Definition 1. *The Gamma function is defined as follows:*

$$\Gamma(x) = \int_0^{\infty} e^{-t} t^{x-1} dt,$$

where $x > 0$.

Definition 2. *The probability density function of the chi-squared distribution with* k *degrees of freedom (denoted by* χ_k^2*) is*

$$f(x;k) = \begin{cases} \frac{1}{\Gamma(\frac{k}{2})2^{k/2}} e^{-x/2} x^{(k-2)/2}, & x > 0; \\ 0, & x \leq 0, \end{cases}$$

where $\Gamma(\cdot)$ *denotes the Gamma function.*

The *goodness of fit of a statistical model* describes how well it fits a set of observations (samples). Measures of goodness of fit typically summarize the discrepancy between the observed values and the values expected under the statistical model in question. Such measures of goodness of fit can be used in statistical hypothesis testing. The *Pearson's chi-squared test for goodness of fit* [25] is used to assess the goodness of fit establishes whether or not an observed frequency distribution differs from a theoretical distribution. In the following, we will briefly introduce the Pearson's chi-squared test for goodness of fit.

Assume that, there is a population X with the following theoretical distribution:

$$H_0 : \mathsf{Pr}[X = a_i] = f_i \quad (i = 1, \ldots, k),$$

where a_i, f_i $(i = 1, \ldots, k)$ are known and $a_1, \ldots, a_k$ are pairwise different, $f_i > 0$ $(i = 1, \ldots, k)$.

One obtains n samples (denoted by $X_1, X_2, \ldots, X_n$) from the population X and uses the Pearson's chi-squared test for goodness of fit to test whether or not the hypothesis H_0 holds. We use the symbol ω_i to denote the number of samples in $\{X_1, X_2, \ldots, X_n\}$ which equal to a_i. If the number n is large enough, it will has that $\omega_i/n \approx f_i$, namely $\omega_i \approx nf_i$. The value nf_i can be viewed as the theoretical value (TV for short) of the category "a_i". The value ω_i can be viewed as the empirical value (EV for short) of the category "a_i". Table 1 shows the theoretical value and the empirical value of the category "a_i".

Clearly, when the discrepancy of the last two lines of Table 1 is smaller, the hypothesis H_0 increasingly seems to be true. It is well known that the Pearson's goodness of fit χ^2 statistic (denoted by Z) is used to measure this kind of discrepancy and is shown as follows:

$$Z = \sum_{i=1}^{k} (nf_i - \omega_i)^2 / (nf_i). \tag{1}$$

Table 1. The theoretical value and the empirical value of each category

Category	a_1	a_2	$\cdots$	a_i	$\cdots$	a_k
TV	nf_1	nf_2	$\cdots$	nf_i	$\cdots$	nf_k
EV	ω_1	ω_2	$\cdots$	ω_i	$\cdots$	ω_k

The statistic Z can be exploited to test whether or not the hypothesis H_0 holds. For example, after choosing a constant Con under a given level, when $Z \leq Con$, one should accept the hypothesis H_0. When $Z > Con$, one should reject the hypothesis H_0. Now, let's consider a more general case. The following lemma about Z was given out by Pearson at 1900 [25] and the proof of Lemma 1 is beyond the scope of this paper.

Lemma 1. *If the hypothesis H_0 holds, when $n \to \infty$, the distribution of Z will approach to the chi-squared distribution with $k-1$ degrees of freedom, namely χ^2_{k-1}.*

Assume that one computes a specific value of Z (denoted by Z_0) by a group of specific data. Let

$$L(Z_0) = \mathsf{Pr}[Z \geq Z_0 | H_0] \approx 1 - K_{k-1}(Z_0), \tag{2}$$

where the symbol $K_{k-1}(\cdot)$ denotes the distribution function of χ^2_{k-1}. ***Clearly, when the probability $L(Z_0)$ is higher, the hypothesis H_0 increasingly seems to be true.*** Therefore, the probability $L(Z_0)$ can be used as a tool to test the hypothesis H_0.

If the theoretical distribution of the population X is continuous, the Pearson's chi-squared test for goodness of fit is also valid. In this case, assume that, one want to test the following hypothesis:

H_1: The distribution function of the population X is $F(x)$.

The distribution function $F(x)$ is continuous. To test the hypothesis H_1, one should set

$$-\infty = a_0 < a_1 < a_2 < \cdots < a_{k-1} < a_k = \infty,$$

and let $I_1 = (a_0, a_1], \cdots, I_i = (a_{i-1}, a_i], \cdots, I_k = (a_{k-1}, a_k)$. Moreover, one obtains n samples (denoted by $X_1, X_2, \ldots, X_n$) from the population X. Let ω_i denotes the cardinality of the set $\{X_j | X_j \in I_i,\ j \in \{1, 2, \ldots, n\}\}$ and

$$f_i = \mathsf{Pr}[x \in I_i, x \leftarrow X] = F(a_i) - F(a_{i-1}) \quad (i = 1, \ldots, k).$$

Then, one can also similarly compute $L(Z_0)$ (by equation (2)) to test the hypothesis H_1.

2.2 Template Attacks

Template attacks consist of two stages. The first stage is the profiling stage and the second stage is the extraction stage. We will introduce the two stages in the following.

The Profiling Stage Assume that there exist K different (sub)keys $key_i, i = 0, 1, \ldots, K-1$ which need to be classified. Also, there exist K different key-dependent operations O_i, $i = 0, 1, \ldots, K-1$. Usually, one will built K templates, one for each key-dependent operation O_i. One can exploit some methods to choose N interesting points $(P_0, P_1, \ldots, P_{N-1})$. Each template is composed of a mean vector and a covariance matrix. Specifically, the mean vector is used to estimate the data-dependent portion of side-channel leakages. It is the average signal vector $\mathbf{M}_i = (M_i[P_0], \ldots, M_i[P_{N-1}])$ for each one of the key-dependent operations. The covariance matrix is used to estimate the probability density of the noises at different interesting points. It is assumed that noises at different interesting points approximately follow the multivariate normal distribution. A N dimensional noise vector $\mathbf{n}_i(\mathbf{S})$ is extracted from each actual power trace $\mathbf{S} = (S[P_0], \ldots, S[P_{N-1}])$ representing the template's key dependency O_i as $\mathbf{n}_i(\mathbf{S}) = (S[P_0] - M_i[P_0], \ldots, S[P_{N-1}] - M_i[P_{N-1}])$. One computes the $(N \times N)$ covariance matrix $\mathbf{C}_i$ from these noise vectors. The probability density of the noises occurring under key-dependent operation O_i is given by the N dimensional multivariate Gaussian distribution $p_i(\cdot)$, where the probability of observing a noise vector $\mathbf{n}_i(\mathbf{S})$ is:

$$p_i(\mathbf{n}_i(\mathbf{S})) = \frac{1}{\sqrt{(2\pi)^N |\mathbf{C}_i|}} exp\Big(-\frac{1}{2}\mathbf{n}_i(\mathbf{S})\mathbf{C}_i^{-1}\mathbf{n}_i(\mathbf{S})^T\Big) \quad \mathbf{n}_i(\mathbf{S}) \in \mathbb{R}^N. \tag{3}$$

In equation (3), the symbol $|\mathbf{C}_i|$ denotes the determinant of $\mathbf{C}_i$ and the symbol $\mathbf{C}_i^{-1}$ denotes its inverse. We know that the matrix $\mathbf{C}_i$ is the estimation of the true covariance $\mathbf{\Sigma}_i$. The condition of equal covariances [4] means that the leakages from different key-dependent operations have the same true covariance $\mathbf{\Sigma} = \mathbf{\Sigma}_0 = \mathbf{\Sigma}_1 = \cdots = \mathbf{\Sigma}_{K-1}$. In most settings, the condition of equal covariances does not hold. Therefore, in this paper, we only consider the device in which the condition of equal covariances does not hold.

The Extraction Stage Assume that one obtains t actual power traces (denoted by $\mathbf{S}_1, \mathbf{S}_2, \ldots, \mathbf{S}_t$) from the target device in the extraction stage. When the actual power traces are statistically independent, one will apply maximum likelihood approach on the product of conditional probabilities (pp. 156 in [11]), i.e.

$$key_{ck} := argmax_{key_i}\Big\{ \prod_{j=1}^{t} \mathsf{Pr}[\mathbf{S}_j | key_i], i = 0, 1, \ldots, K-1 \Big\},$$

where $\mathsf{Pr}[\mathbf{S}_j | key_i] = p_{f(\mathbf{S}_j, key_i)}(n_{f(\mathbf{S}_j, key_i)}(\mathbf{S}_j))$. The key_{ck} is considered to be the correct (sub)key. The output of the function $f(\mathbf{S}_j, key_i)$ is the index of a key-dependent operation. For example, when the output of the first S-box in the first round of AES-128 is chosen as the target intermediate value, one builds templates for each output of the S-box. In this case, $f(\mathbf{S}_j, key_i) = Sbox(m_j \oplus key_i)$, where m_j is the input plaintext corresponding to the actual power trace $\mathbf{S}_j$.

3 Our New Approach to Choose Interesting Points For Template Attacks

Now, we begin to introduce our new approach of choosing interesting points for template attacks. Firstly, we show the following Lemma whose proof is in Appendix A.

Lemma 2. *The marginal distribution of multivariate Gaussian distribution is a normal distribution.*

The main idea of our new approach is as follows. In template attacks, it is assumed that the distribution of the noises of multiple interesting points follows the multivariate Gaussian distribution. Moreover, by Lemma 2, we know that the marginal distribution of the multivariate Gaussian distribution is a normal distribution. Therefore, in classical template attacks, if the distribution of samples of each interesting point increasingly to approximate a normal distribution, the multivariate Gaussian distribution statistical model will increasingly to be suitable to be exploited to build the templates for template attacks. Otherwise, if the points whose distributions of samples are not similar to normal distributions are chosen as the interesting points, the multivariate Gaussian distribution will not be suitable to be exploited to build the templates and the classification performance of template attacks will be poor. Therefore, for each clock cycle, our new approach chooses the point whose distribution of samples is more approximate to a normal distribution than other points in the same clock cycle as the interesting point.

The Pearson's chi-squared test for goodness of fit can be used as a tool to assess whether or not the distribution of samples of each point approximates to a normal distribution. Specifically speaking, assume that, for a fixed point P_t, one obtains n samples $(X_1, X_2, \ldots, X_n)$ for a fixed operation on fixed data and computes:

$$\hat{\mu} = \frac{1}{n} \cdot \sum_{i=1}^{n} X_i, \ s^2 = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (X_i - \hat{\mu})^2. \tag{4}$$

Note that, in template attacks, one can operate the reference device as many times as possible and samples a large number of actual power traces in the profiling stage. Therefore, the value of n can be large enough. When the value of n is large enough, one can assume that the theoretical distribution of samples of the point P_t is the normal distribution $\mathcal{N}(\hat{\mu}, s^2)$ and to test whether this hypothesis holds by exploiting the Pearson's chi-squared test for goodness of fit as follows. The distribution function of the normal distribution $\mathcal{N}(\hat{\mu}, s^2)$ is denoted by $F(x; \hat{\mu}, s^2)$. Let $a_0 = -\infty, a_1 = \hat{\mu} - 2s, a_2 = \hat{\mu} - 1.5s, \ldots, a_9 = \hat{\mu} + 2s, a_{10} = +\infty$ and $I_1 = (-\infty, \hat{\mu} - 2s], I_2 = (\hat{\mu} - 2s, \hat{\mu} - 1.5s], \ldots, I_{10} = (\hat{\mu} + 2s, +\infty)$. Then, one computes $Z_0 = \sum_{i=1}^{10} (nf_i - \omega_i)^2/(nf_i)$, where $f_i = F(a_i; \hat{\mu}, s^2) - F(a_{i-1}; \hat{\mu}, s^2)$ and $\omega_i = |\{X_j | X_j \in I_i, \ j \in \{1, 2, \ldots, n\}\}|$. After obtaining the statistic Z_0, one computes the value $L(Z_0)$ by using equation (2). When the value of n is large enough, if the n samples $(X_1, X_2, \ldots, X_n)$ fit the

normal distribution $\mathcal{N}(\hat{\mu}, s^2)$ well, the value $L(Z_0)$ will be high. Otherwise, the value $L(Z_0)$ will be low. Therefore, one can choose the interesting points based on the value $L(Z_0)$. For points in the same clock cycle, one computes the value $L(Z_0)$ of each point with the same actual power traces and chooses a point whose value $L(Z_0)$ is the highest one as the interesting point.

4 Experimental Evaluations

In this section, we will verify and compare the classification performance of template attacks based on the interesting points chosen by our new approach and the classification performance of template attacks based on the interesting points chosen by the previous approaches. Specifically speaking, our experiments are divided into two groups. In the first group, we tried to choose the interesting points by using different approaches. In the second group, we computed the classification performances of template attacks based on the interesting points chosen by different approaches.

For the implementation of a cryptographic algorithm with countermeasures, one usually first tries his best to use some methods to delete the countermeasures from actual power traces. If the countermeasures can be deleted, then one tries to recover the correct (sub)key using classical attack methods against unprotected implementation. For example, if one has actual power traces with random delays [15], he may first use the method proposed in [14] to remove the random delays from actual power traces and then uses classical attack methods to recover the correct (sub)key. The methods of deleting countermeasures from actual power traces are beyond the scope of this paper. Moreover, considering actual power traces without any countermeasures shows the upper bound of the physical security of the target cryptographic device. Therefore, we take unprotected AES-128 implementation as example.

The 1^{st} S-box outputs of the 1^{st} round of an unprotected AES-128 software implementation are chosen as the target intermediate values. The unprotected AES-128 software implementation is on an typical 8-bit microcontroller STC89C58RD+ whose operating frequency is 11MHz. The actual power traces are sampled with an Agilent DSA90404A digital oscilloscope and a differential probe by measurement over a 20Ω resistor in the ground line of the 8-bit microcontroller. The sampling rate was set to be 50MS/s. The average number of actual power traces during the sampling process was 10 times. For our device, the condition of equal covariances does not hold. This means that the differences between different covariance matrixes $\mathbf{C}_i$ are very evident (can easily be observed from visual inspection).

In order to choose interesting points and to test the classification performance of template attacks, we generated three sets of actual power traces which are respectively denoted by Set A, Set B, and Set C. The actual power traces in Set A were used in the profiling stage. The actual power traces in Set B were used in the extraction stage. The actual power traces in Set C were used to choose interesting points. The Set A captured 20,000 actual power traces which

were generated with a fixed main key and random plaintext inputs. The Set B captured 100,000 actual power traces which were generated with another fixed main key and random plaintext inputs. The Set C captured 110,000 actual power traces which were generated with a fixed main key and random plaintext inputs. Note that, we used the same device to generate the three sets of actual power traces, which provides a good setting for the focuses of our research.

4.1 Group 1

In all experiments, we chose 4 continual clock cycles about the target intermediate value (Note that, in our unprotected AES-128 software implementation, the target intermediate value only continued for 4 clock cycles.). In each clock cycle, there are 4 points. Therefore, there are 16 points (denoted by $P_0, P_1, \ldots, P_{15}$) totally[1]. Beside our new approach (denoted by CST), we also implemented all the other approaches of choosing interesting points for template attacks including CPA, SOST, DOM, SOSD, VAR, SNR, MIA, and KSA. All the approaches (CSF, CPA, SOST, DOM, SOSD, VAR, SNR, MIA, and KSA) used 110,000 actual power traces in Set C to choose interesting points. The leakage function of our device approximates the typical Hamming-Weight Model (pp. 40-41 in [11]). Therefore, we adopted this model for CPA, MIA, and KSA.

In order to get more accurate results, we conducted our new approach of choosing interesting points as follows. Due to the leakage function of our device approximates the typical Hamming-Weight Model, we chose 9 different values (denoted by $V_0, V_1, \ldots, V_8$) about the target intermediate value. The hamming weight of the 9 different values respectively are $0, 1, \ldots, 8$ (i.e. $HW(V_i) = i,\ i = 0, 1, \ldots, 8$). For each V_i $(i = 0, 1, \ldots, 8)$, we selected 400 actual power traces in which the target intermediate value equals to V_i from Set C. Therefore, for each value V_i $(i = 0, 1, \ldots, 8)$, there are 400 samples for each one of the 16 points $(P_0, P_1, \ldots, P_{15})$ and we computed the empirical mean value $\hat{\mu}$ and the empirical variance s^2 of the 400 samples for each one of the 16 points by equation (4). Then, for each V_i $(i = 0, 1, \ldots, 8)$, we tried to assess the goodness of fit establishes whether or not the actual distribution of samples of the point P_i $(i \in \{0, 1, \ldots, 15\})$ differs from its assumed theoretical distribution $\mathcal{N}(\hat{\mu}, s^2)$ by computing the value $L(Z_0)$ with the 400 samples like that in Section 3. For the value V_i $(i = 0, 1, \ldots, 8)$ and the point P_j $(j = 0, 1, \ldots, 15)$, we computed the value $L(Z_0)$ and rewrote it by $L_{(i,j)}(Z_0)$. Then, we computed the value $L_j(Z_0)$ $(j = 0, 1, \ldots, 15)$ for each one of the 16 points as follows:

$$L_j(Z_0) = \frac{1}{9} \cdot \sum_{i=0}^{8} L_{(i,j)}(Z_0),\ (j = 0, 1, \ldots, 15)$$

and chose the interesting points based on the values $L_0(Z_0), \ldots, L_{15}(Z_0)$. In one clock cycle, the point with the highest $L_j(Z_0)$ is chosen as the interesting point.

[1] The points $P_0, \ldots, P_3$ are in the first clock cycle. The points $P_4, \ldots, P_7$ are in the second clock cycle. The points $P_8, \ldots, P_{11}$ are in the third clock cycle. The points $P_{12}, \ldots, P_{15}$ are in the fourth clock cycle.

In Table 2, we show the interesting points chosen by different approaches using the 110,000 actual power traces in Set C. From Table 2, we find that our approach chooses different interesting points in the first three clock cycles compared with other approaches.

Table 2. The interesting points chosen by different approaches

Clock Cycle	1	2	3	4
CST	P_2	P_4	P_{11}	P_{12}
CPA	P_1	P_5	P_8	P_{12}
SOST	P_1	P_5	P_8	P_{12}
DOM	P_3	P_7	P_{10}	P_{12}
SOSD	P_3	P_7	P_{10}	P_{12}
VAR	P_3	P_7	P_{10}	P_{12}
SNR	P_3	P_7	P_{10}	P_{12}
MIA	P_1	P_5	P_8	P_{15}
KSA	P_1	P_5	P_8	P_{15}

4.2 Group 2

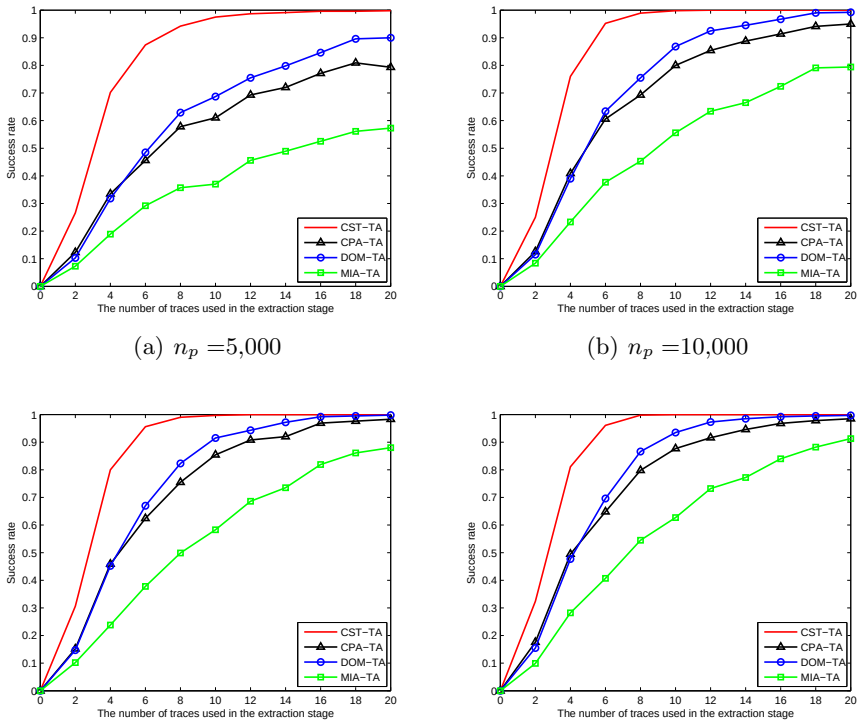
For simplicity, let n_p and n_e respectively denote the number of actual power traces used in the profiling stage and in the extraction stage. In this paper, we use the typical metric *success rate* [6] as the metric about the classification performance of template attacks.

In order to show the success rates of template attacks based on the interesting points chosen by different approaches under different attack scenarios, we conducted 4 groups of experiments. In these groups of experiments, the numbers of actual power traces used in the profiling stage are different. This implies that the level of accuracy of the templates in these groups of experiments are different. The higher number of actual power traces used in the profiling stage, the more accurate templates will be built. Moreover, in each groups of experiments, we still considered the cases that one can possess different numbers of actual power traces which can be used in the extraction stage.

Specifically speaking, in the 4 groups of experiments, we respectively chose 5,000, 10,000, 15,000, and 20,000 different actual power traces from Set A to build the 256 templates based on the interesting points chosen by different approaches in the profiling stage. Template attacks based on the interesting points chosen by approach A is denoted by the symbol "A-TA". We tested the success rates of template attacks based on the interesting points chosen by different approaches when one uses n_e actual power traces in the extraction stage as follows. We repeated the 9 attacks (CSF-TA, CPA-TA, SOST-TA, DOM-TA, SOSD-TA, SNR-TA, VAR-TA, MIA-TA, and KSA-TA) 1,000 times. For each time, we chose n_e actual power traces from Set B uniformly at random and the 9 attacks were conducted with the same n_e actual power traces. We respectively recorded how

many times the 9 attacks can successfully recover the correct subkey of the 1^{st} S-box.

From Table 2, we find that the CPA approach and the SOST approach provide the same result of choosing interesting points. The DOM approach, the SOSD approach, the VAR approach, and the SNR approach provide the same result of choosing interesting points. Moreover, the MIA approach and the KSA approach provide the same result of choosing interesting points. The approaches which provide the same result of choosing interesting points will lead to the same classification performance of template attacks. Therefore, in order to show the success rates more clearly, we only show the success rates of CSF-TA, CPA-TA, DOM-TA, and MIA-TA in Figure 1. The success rates of template attacks based on the interesting points chosen by different approaches when n_p equals to $5,000$ and n_e equals to $4,\ 8,\ 12,\ 16$, and 20 are shown in Table 3.



(a) n_p =5,000 (b) n_p =10,000

(c) n_p =15,000 (d) n_p =20,000

Fig. 1. The experiment results

From Figure 1 and Table 3, in all the attack scenarios, we find that template attacks based on the interesting points chosen by our new approach will achieve obvious higher success rates compared with template attacks based on the inter-

esting points chosen by the previous approaches. For example, when $n_p = 5,000$ and $n_e = 4$, the success rate of CST-TA equals to 0.70, while the success rate of DOM-TA equals to 0.32. What's more, when $n_p = 5,000$, CST-TA only needs 7 actual power traces in the extraction stage to achieve success rate higher than 0.9, while DOM-TA needs 20 actual power traces in the extraction stage to achieve success rate higher than 0.9 under the same attack scenario. Therefore, we believe that using our new approach to choose the interesting points can effectively improve the classification performance of template attacks.

Table 3. The success rates of template attacks when $n_p = 5,000$

n_e	4	8	12	16	20
CST-TA	0.70	0.94	0.99	1.00	1.00
CPA-TA	0.34	0.58	0.69	0.77	0.79
SOST-TA	0.34	0.58	0.69	0.77	0.79
DOM-TA	0.32	0.63	0.76	0.85	0.90
SOSD-TA	0.32	0.63	0.76	0.85	0.90
VAR-TA	0.32	0.63	0.76	0.85	0.90
SNR-TA	0.32	0.63	0.76	0.85	0.90
MIA-TA	0.19	0.36	0.46	0.53	0.57
KSA-TA	0.19	0.36	0.46	0.53	0.57

5 Conclusion

In this paper, we give a negative answer to the question that whether or not using the previous approaches of choosing interesting points will lead to the best classification performance of template attacks by introduction a new approach with completely different basic principle. Our new approach is based on the important mathematical property of the multivariate Gaussian distribution and exploits the Pearson's chi-squared test for goodness of fit.

Experiments verified that template attacks based on the interesting points chosen by our new approach will achieve obvious better classification performance compared with template attacks based on the interesting points chosen by the previous approaches. Moreover, the computational price of our new approach is low and practical. Therefore, our new approach of choosing interesting points can be used in practice to better understand the practical threats of template attacks. In the future, it is necessary to further verify our new approach in other devices such as ASIC and FPGA.

Acknowledgments This work was supported by the National Basic Research Program of China (No.2013CB338003), National Natural Science Foundation of China (Nos.61472416, 61272478), and National Key Scientific and Technological Project (No.2014ZX01032401-001).

References

1. Chari, S., Rao, J.R., Rohatgi, P.: Template Attacks. CHES2002, LNCS 2523, pp. 13-28, 2003.
2. Rechberger, C., Oswald, E.: Practical Template Attacks. WISA2004, LNCS 3325, pp. 440-456, 2004.
3. Archambeau, C., Peeters, E., Standaert, F.-X., Quisquater, J.-J.: Template Attacks in Principal Subspaces. CHES2006, LNCS 4249, pp. 1-14, 2006.
4. Choudary, O., Kuhn, M.G.: Efficient Template Attacks. CARDIS2013, LNCS 8419, pp. 253-270, 2013.
5. Bär, M., Drexler, H., Pulkus, J.: Improved Template Attacks. COSADE2010, 2010.
6. Standaert, F.-X., Malkin, T.G., Yung, M.: A Unified Framework for the Analysis of Side-Channel Key Recovery Attacks. EUROCRYPT2009, LNCS 5479, pp. 443-461, 2009.
7. Montminy, D.P., Baldwin, R.O., Temple, M.A., Laspe, E.D.: Improving cross-device attacks using zero-mean unit-variance mormalization. Journal of Cryptographic Engineering, Volume 3, Issue 2, pp. 99-110, June 2013.
8. Oswald, E., Mangard, S.: Template Attacks on Masking—Resistance Is Futile. CT-RSA2007, LNCS 4377, pp. 243-256, 2007.
9. Standaert, F.-X., Archambeau, C.: Using Subspace-Based Template Attacks to Compare and Combine Power and Electromagnetic Information Leakages. CHES2008, LNCS 5154, pp. 411-425, 2008.
10. Gierlichs, B., Lemke-Rust, K., Paar, C.: Templates vs. Stochastic Methods A Performance Analysis for Side Channel Cryptanalysis. CHES2006, LNCS4249, pp. 15-29, 2006.
11. Mangard, S., Oswald, E., Popp, T.: Power Analysis Attacks: Revealing the Secrets of Smart Cards. Springer 2007.
12. Hanley, N., Tunstall, M., Marnane, W.P.: Unknown Plaintext Template Attacks. WISA2009, LNCS 5932, pp. 148-162, 2009.
13. Jolliffe, I.: "Principal Component Analysis", John Wiley & Sons, Ltd, 2005.
14. Durvaux, F., Renauld, M., Standaert, F.-X. et al.: Efficient Removal of Random Delays from Embedded Software Implementations Using Hidden Markov Models. CARDIS2012, LNCS 7771, pp. 123-140, 2013.
15. Coron, J.-S., Kizhvatov, I.: Analysis and Improvement of the Random Delay Countermeasure of CHES 2009. CHES2010, LNCS 6225, pp. 95-109, 2010.
16. Mather, L., Oswald, E., Bandenburg, J., Wójcik, M.: Does My Device Leak Information? An a *priori* Statistical Power Analysis of Leakage Detection Tests. ASIACRYPT2013 Part I, LNCS 8269, pp. 486-505, 2013.
17. Gierlichs, B., Batina, L., Tuyls, P., Preneel, B.: Mutual Information Analysis. CHES2008, LNCS 5154, pp. 426-442, 2008.
18. Whitnall, C., Oswald, E., Mather, L.: An Exploration of the Kolmogorov-Smirnov Test as a Competitor to Mutual Information Analysis. CARDIS2011, LNCS 7079, pp. 234-251, 2011.
19. Kocher, P.C.: Timing Attacks on Implementations of Diffie-Hellman, RSA, DSS, and Other Systems. CRYPTO1996, LNCS 1109, pp. 104-113, 1996.
20. Gandolfi, K., Mourtel, C., Olivier, F.: Electromagnetic Analysis: Concrete Results. CHES2001, LNCS 2162, pp. 251-261, 2001.
21. European Network of Excellence (ECRYPT). The side channel cryptanalysis lounge. http://www.crypto.ruhr-uni-bochum.de/en sclounge.html.

22. Standaert, F.-X., Koeune, F., Schindler, W.: How to Compare Profiled Side-Channel Attacks? ACNS2009, LNCS 5536, pp. 485-498, 2009.
23. Cover, T.M., Thomas, J.A.: Elements of Information Theory. John Wiley & Sons, Chichester, 2006.
24. Schindler, W., Lemke, K., Paar, C.: A Stochastic Model for Differential Side Channel Cryptanalysis. CHES2005, LNCS 3659, pp. 30-46, 2005.
25. Pearson, K.: On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. Philosophical Magazine Series 5 50 (302): pp. 157-175, 1900.

Appendix A: The Proof of Lemma 2

Proof: For simplicity, we only consider the case when $N = 2$. For the case $N > 2$, this Lemma holds similarly.

Let (ξ, η) denote a 2 dimensional random vector. The continuous distribution function and the probability density function of the 2 dimensional random vector respectively are $F(x, y)$ and $p(x, y)$. Then, the marginal distribution functions are as follows:

$$F_1(x) = \int_{-\infty}^{x}\int_{-\infty}^{\infty} p(u, y)dudy,\ F_2(y) = \int_{-\infty}^{\infty}\int_{-\infty}^{y} p(x, u)dxdu.$$

The marginal density functions are as follows:

$$p_1(x) = \int_{-\infty}^{\infty} p(x, y)dy,\ p_2(y) = \int_{-\infty}^{\infty} p(x, y)dx.$$

For 2 dimensional multivariate Gaussian distribution, it has that

$$p(x, y) = \frac{1}{2\pi|\mathbf{C}|}exp\Big\{-\frac{1}{2}(x-a, y-b)\cdot\mathbf{C}^{-1}\cdot(x-a, y-b)^T\Big\},$$

where

$$\mathbf{C} = \begin{pmatrix} \sigma_1^2 & r\sigma_1\sigma_2 \\ r\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

and the values $a, b, \sigma_1, \sigma_2, r$ are constant, $\sigma_1 > 0, \sigma_2 > 0, |r| < 1$. The probability density function $p(x, y)$ can be rewritten as follows

$$p(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}}exp\Big\{-\frac{1}{2(1-r^2)}\cdot\Big[\frac{(x-a)^2}{\sigma_1^2}$$

$$-\frac{2r(x-a)(y-b)}{\sigma_1\sigma_2}+\frac{(y-b)^2}{\sigma_2^2}\Big]\Big\}.$$

Let

$$\frac{x-a}{\sigma_1} = u, \frac{y-b}{\sigma_2} = v$$

and it has that

$$p_1(x) = \int_{-\infty}^{\infty} p(x,y)dy$$

$$= \frac{1}{2\pi\sigma_1\sqrt{1-r^2}} \int_{-\infty}^{\infty} exp\Big\{ -\frac{1}{2(1-r^2)} \cdot [u^2 - 2ruv + v^2] \Big\} dv$$

$$= \frac{1}{\sqrt{2\pi}\sigma_1} e^{-u^2/2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi(1-r^2)}} \cdot exp\Big\{ -\frac{r^2u^2 - 2ruv + v^2}{2(1-r^2)} \Big\} dv$$

$$= \frac{1}{\sqrt{2\pi}\sigma_1} e^{-u^2/2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi(1-r^2)}} e^{-(v-ru)^2/2(1-r^2)} dv$$

$$= \frac{1}{\sqrt{2\pi}\sigma_1} e^{-u^2/2} = \frac{1}{\sqrt{2\pi}\sigma_1} e^{-(x-a)^2/2\sigma_1^2}.$$

Therefore, $p_1(x)$ is the probability density function of the normal distribution $\mathcal{N}(a, \sigma_1^2)$. Similarly, we can prove that

$$p_2(y) = \frac{1}{\sqrt{2\pi}\sigma_2} e^{-(x-b)^2/2\sigma_2^2}.$$

In this way, Lemma 2 is proven. □