Leakage Detection with Kolmogorov-Smirnov Test

Xinping Zhou[1], Kexin Qiao[1], Changhai Ou[2]

[1] Beijing Unionpay Card Technology Co., Ltd, China
(Bank Card Test Center, BCTC)
LastnameFirstname@bctest.com

[2] School of Computer Science and Engineering, Nanyang Technological University,
Singapore
chou@ntu.edu.sg

Abstract. Leakage detection seeking the evidence of sensitive data dependencies in the side-channel traces instead of trying to recover the sensitive data directly under the enormous efforts with numerous leakage models and state-of-the-art distinguishers can provide a fast preliminary security assessment on the cryptographic devices for designers and evaluators. Therefore, it is a popular topic in recent side-channel research of which the Welch's *t*-test-based Test Vector Leakage Assessment (TVLA) methodology is the most widely used one. However, the TVLA is not always the best option under all kinds of conditions (as we can see in the latter section of this paper). Kolmogorov-Smirnov test is a well-known nonparametric method for statistical analysis to determine whether the samples are from the same distribution by analyzing the cumulative distribution. It has been proposed into side-channel analysis as a successful distinguisher. This paper proposes—to our knowledge, for the first time—Kolmogorov-Smirnov test as a new method for leakage detection. Besides, we propose two implementations to speed up the KS leakage detection procedure. Experimental results on simulated leakage with various parameters and the practical traces verify that KS is an effective and robust leakage detection tool and the comprehensive comparison with TVLA shows that KS-based leakage detection can be a right-hand supplement to TVLA when performing the side-channel assessment.

Keywords: Side Channel Analysis, Kolmogorov-Smirnov Test, Leakage Detection, Cumulative Distribution Function, Histogram

1 Introduction

Side-channel attacks pose a serious threat to the secure devices implementing cryptographic algorithm since the devastating effectiveness and simplicity of such attacks began to become apparent with the work of Kocher et al. in the late 1990s [19]. Therefore, the evaluation of the vulnerability of cryptographic devices against side-channel attacks is an issue of increasing importance for designers and certification bodies. Leakage detection along this line has been attracting most attention recently in academia and industry [8,9,23,4,11,24,6]. It has shown its advantage as a convenient tool to fast conclude whether the device leaks by simply seeking evidence of sensitive data dependencies in the measured traces rather than trying to perform the sensitive data recovery attacks with numerous leakage models and state-of-the-art distinguishers.

The Test Vector Leakage Assessment (TVLA) [16,3] proposed by Cryptography Research at the 2011 Non-Invasive Attack Testing workshop is the most popular option among this trend and it has been well studied [31,10,32,35,36]. TVLA uses Welch's t-test to determine whether there is leakage by targeting the difference of expectation between two sets of traces of which one is associated with a fixed input and the other is associated with random input. If a significant difference is found the device is regarded as unsecure. It also can be extended into higher-order leakage detection by pre-processing the traces first then test the pre-processed traces [31]. The superiority of TVLA relies on the simple calculation of the parameters relating to the t-test.

KolmogorovSmirnov (KS) test is a nonparametric test of the equality of two distributions. In particular, the KolmogorovSmirnov statistic quantifies a distance between the empirical distribution functions of two samples to draw conclusion whether the two samples are from one distribution. It was suggested as an alternative distinguisher for Mutual Information Analysis [15] in [34] for side-channel community. Since then, several literatures demonstrating the efficiency of KS distinguisher have been published [37,20,18]. The basic principle of this distinguisher is that the samples partitioned based on the wrong key candidates are from the same distribution concluded by KS test and samples partitioned based on the true key candidate can be concluded from the different distribution by KS test.

1.1 Contribution

The contribution in this paper is threefold.

First, we confirm that this is the case KS can be used, not just to be a distinguisher for side-channel attack, but as an efficient information-theory based leakage detection tool. We demonstrate that the KS test can be used to find the sensitive data dependencies in the univariate traces and can be easily extended into multivariate leakage.

Second, we propose fast implementation based on histogram to speed up the KS leakage detection procedure.

Third, we carry out a series of experiments on both simulated leakage and real-world traces captured from the cryptographic device to conform the KS's efficiency on leakage detection. Besides, we perform a comprehensive comparison with the well-established TVLA (as a benchmark) by varying the leakage factors and find that KS test is more stable whilst TVLA suffers the fixed value used to generate the fixed traces used to generate the fixed traces and the order in masking scheme.

1.2 Outline

The rest of the paper proceeds as follows. Section 2 covers the preliminaries of notation used in this paper and the concept of Kolmogorov-Smirnov test. In Section 3 we describe how to detect leakage for both the univariate and multivariate settings using Kolmogorov-Smirnov test. Section 4 presents how to speed up the test procedure. In Section 5 we perform a series of experiments in the simulated context. Section 6 presents the results of experiments on the practical traces to verify the effectiveness of our proposed approach. Section 7 concludes the paper.

2 Preliminaries

In this section, we give a brief introduction on the notation and review the Kolmogorov-Smirnov test we put our focus on more closely in this paper.

2.1 Notations

We consider a 'standard DPA attack' scenario as defined in [21], and briefly explain the underlying idea as well as introduce the necessary terminology here. We assume that the power consumption $\mathbf{P} = \{P_1, ..., P_T\}$ of a cryptographic device (as measured at time points $\{1, ..., T\}$) depends, for at least some $\boldsymbol{\tau} \subset \{1, ..., T\}$, on some internal value (or state) $f_{k^*}(X)$ which we call the *target*: a function $f_{k^*} : \mathcal{X} \to \mathcal{Z}$ of some part of the known plaintext—a random variable $X \overset{R}{\in} \mathcal{X}$—which is dependent on some part of the secret key $k^* \in \mathcal{K}$. Consequently, we have that $P_t = L_t \circ f_{k^*}(X) + \varepsilon_t, t \in \boldsymbol{\tau}$, where $L_t : \mathcal{Z} \to \mathbb{R}$ describes the data-dependent leakage function at time t and ε_t comprises the remaining power consumption which can be modelled as independent random noise (this simplifying assumption is common in the literature—see, again, [21]). The attacker has N power measurements corresponding to encryptions of N known plaintexts $x_i \in \mathcal{X}$, $i = 1, \ldots, N$ and wishes to recover the secret key k^*. The attacker can accurately compute the internal values as they would be under each key hypothesis $\{f_k(x_i)\}_{i=1}^N$, $k \in \mathcal{K}$ and uses whatever information he possesses about the true leakage functions L_t to construct a prediction model (or models) $M_t : \mathcal{Z} \to \mathcal{M}_t$.

2.2 Kolmogorov-Smirnov Test

Kolmogorov-Smirnov (KS) test is a non-parametric hypothesis test of the equality of continuous or discontinuous one-dimensional probability distributions that can be used to compare a sample with a reference probability distribution or to compare two samples. The Kolmogorov-Smirnov statistic quantifies a distance between the empirical distribution function of the sample and the cumulative distribution function of the reference distribution Θ. Thus the null hypothesis of test H_0 and the alternative hypothesis H_{alt} are

H_0: the samples come from Θ vs. H_{alt}: the samples do not come from Θ

The empirical cumulative distribution function F_n for n independent and identically distributed (iid) ordered observations X_i is defined as

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I_{(-\infty,x]}(X_i) \tag{1}$$

where $I_{(-\infty,x]}(X_i)$ is the indicator function, equal to 1 if $X_i \leq x$ and equal to 0 otherwise. The Kolmogorov-Smirnov statistic for a given cumulative distribution function $F(x)$ is

$$D_n = \sup_x |F_n(x) - F(x)| \tag{2}$$

where $\sup_x |\cdot|$ is the supremum of the set of distances. If the sample comes from distribution $F(x)$, then D_n converges to 0 almost surely in the limit when n goes to infinity. The null hypothesis is not rejected if $D_n < D_{crit,n,\alpha}$ where $D_{crit,n,\alpha}$ is the critical value corresponding with a given level α.[3]

The one-sample Kolmogorov-Smirnov test can be extended for two-sample comparison to test whether they come from the one distribution. In this case, the hypothesis of Kolmogorov-Smirnov test is

$$H_0\text{: } F_A = F_B \text{ vs. } H_{alt} : F_A \neq F_B$$

where F_A and F_B are the two samples. The statistic for the two-sample test is

$$D_{n,m} = \sup_x |F_{A,n}(x) - F_{B,m}(x)| \tag{3}$$

where $F_{A,n}$ and $F_{B,m}$ are the empirical distribution functions of the first and the second one-dimensional sample respectively. An example of the calculation of $D_{n,m}$ is shown in Fig. 1 where the blue line and the red line represent cumulative probability of the two samples, and the black arrow is the KS statistic $D_{n,m}$ which measures the maximum distance of the two cumulative probability. For large samples, desired probability p value to accept the null hypothesis is

$$p = 2\sum_{j=1}^{\infty} (-1)^{j-1} e^{-2j^2 Z^2} \tag{4}$$

where Z is defined by

$$Z = D_{n,m}(\sqrt{J} + \frac{0.11}{\sqrt{J}} + 0.12) \tag{5}$$

and

$$J = \frac{nm}{n+m} \tag{6}$$

where n, m are the size of samples.

[3] For more details about how to calculate p value for one-sample Kolmogorov-Smirnov test, see [22].

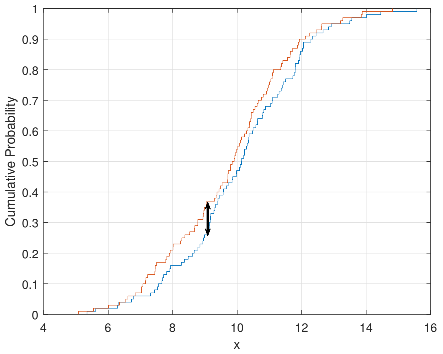

Fig. 1. An illustration of two-sample Kolmogorov-Smirnov test

The generalization of the one-dimensional Kolmogorov-Smirnov test to high dimensional probability distributions is a challenge. Peacock [26] proposed a bivariate KS for two two-dimensional samples test that uses four pairs of cumulative probability (instead of only one). For instance, the two-dimensional samples are $\mathbf{A} = (\mathbf{A}_1, \mathbf{A}_2) = \{A_{1,i}, A_{2,i}\}_{i=1}^{n}$ and $\mathbf{B} = (\mathbf{B}_1, \mathbf{B}_2) = \{B_{1,j}, B_{2,j}\}_{j=1}^{m}$, then the bivariate KS defined by Peacock is

$$D_{A,B} = \sup_{(x,y)\in(\mathbf{A}_1\cup\mathbf{B}_1,\mathbf{A}_2\cup\mathbf{B}_2)} |F_{\mathbf{A}_1,\mathbf{A}_2}(x,y) - F_{\mathbf{B}_1,\mathbf{B}_2}(x,y)| \tag{7}$$

The distribution-free property of the Kolmogorov-Smirnov test rests on being able to map any distribution function on to any other distribution function using a transformation that preserves the ordering of the data. In the one-dimensional case, the definition of empirical cumulative distribution function $F_n(x)$ is trivial since there only two ways to order the data which are $P(X \leq x)$ or $P(X > x)$. In fact, the two ways to order the one-dimensional data make no difference since they can be mapped from each other (i.e. $P(X \leq x) = 1 - P(X > x)$).

However, in the high-dimensional case the *directions* to order the tuples $(x, y, \ldots)$ are arbitrary and affect the definition of the empirical cumulative distribution function. For example, there is no way to map $P(X \leq x, Y \leq y)$ to $P(X \leq x, Y > y)$. For d dimensional data, there are 2^d independent ways to define cumulative distribution function. The four cumulative distribution functions for two-dimensional case are

$$F_{\mathbf{A}_1,\mathbf{A}_2}(x,y) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} I_{\mathbf{A}_{1,i}\leq x,\mathbf{A}_{2,j}\leq y} \tag{8}$$

$$F_{\mathbf{A}_1,\mathbf{A}_2}(x,y) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} I_{\mathbf{A}_{1,i}\leq x,\mathbf{A}_{2,j}> y} \tag{9}$$

$$F_{\mathbf{A}_1,\mathbf{A}_2}(x,y) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} I_{\mathbf{A}_{1,i}> x,\mathbf{A}_{2,j}> y} \tag{10}$$

$$F_{\mathbf{A}_1,\mathbf{A}_2}(x,y) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} I_{\mathbf{A}_{1,i}>x,\mathbf{A}_{2,j}\leq y} \tag{11}$$

We use symbols $F_{\mathbf{A}}^{--}, F_{\mathbf{A}}^{-+}, F_{\mathbf{A}}^{++}, F_{\mathbf{A}}^{+-}$ to represent (8) - (11) for short respectively. With the same manner, the definitions of higher-dimensional cumulative distribution function can be determined (e.g. one of three-dimensional cumulative distribution functions $F_{\mathbf{A}}^{---}$ can be defined by $F_{\mathbf{A}_1,\mathbf{A}_2,\mathbf{A}_3}(x,y,z) = \frac{1}{n}\sum\limits_{i=1}^{n}\sum\limits_{j=1}^{n}\sum\limits_{k=1}^{n} I_{\mathbf{A}_{1,i}\leq x,\mathbf{A}_{2,j}\leq y,\mathbf{A}_{3,k}\leq z}$). Then the KS statistic $D_{A,B}$ in (7) is given by the max difference distance among all types of cumulative distribution functions suggested by Peacock [26].

3 Methodology

Following the description about Kolmogorov-Smirnov test for comparison of two distributions in Section 2.2, we explain how to use it for leakage detection in this section.

3.1 General Approach

Suppose that n traces $\mathbf{P}^{(i)}(i \in \{1...n\})$ are captured from the device under test (DUT) when it is encrypting or decrypting the corresponding data $\mathbf{X}_{(i)}(i \in \{1...n\})$. Each of the n traces has T sample points $\{P_1^{(i)},...,P_T^{(i)}\}$ as described in Section 2.1. Note that the traces capture measurement and procedure are the same with the TVLA as described in detail in [31]. Then the traces can be categorized into two groups G_A and G_B according the intermediate value of each encryotion with some distinguisher function *df*. The G_A and G_B can be expressed as,

$$G_A = \{\mathbf{P}^{(i)}|df(\mathbf{X}_i) = x\}, G_B = \{\mathbf{P}^{(i)}|df(\mathbf{X}_i) \neq x\}$$

For instance, the *df* can be chosen as one bit of the intermediate value then the value of x in this case is 0 or 1. Then, the Kolmogorov-Smirnov test is performed on each time point of [1, T]. More precisely, for a time point $t \in [1,T]$ that can be regarded as a random variable, its samples F_A and F_B can be extracted from G_A and G_B at the corresponding point of length n_A and n_B respectively (where we assume that n_A=$|G_A|$ and n_B=$|G_B|$ representing the cardinality of the two groups). Then the leakage detection task is converted to determine whether these two samples come from a distribution. If they are from the same distribution which means it is difficult to recover the sensitive information by analysing these traces which concludes that the DUT passes the evaluation. The Kolmogorov-Smirnov test is used to test these two samples. The null hypothesis H_0 is set to be that they are from the same distribution which will conclude the DUT is secure enough under the given confidence level α while the alternative hypothesis H_{alt} is set to they are from the different distributions. The criterion to reject the null hypothesis only if the inequation (4) is fulfilled where the calculation of D is given by equation (3). Repeat the aforementioned procedure for all the time point t and conclude that the DUT is not secure when at least one null hypothesis is rejected.

Such a test that requires the participation of known key and the distinguisher function is so-called $specific$ test. Since this kind of test will suffer from different distinguisher functions and the information of the secret key which is contradictory with the black-box evaluation that none information should be known in advance. To avoid these drawbacks, *non-$specific$* (or *$fixed\ vs.$ random*) KS test can be performed for leakage detection. In order to conduct the *non-$specific$* KS test, the two groups G_A and G_B are derived when the traces collection rather than by the distinguisher function. First, a fixed data X is selected to be encrypted (or decrypted) for n_A times and the corresponding traces are captured to form the group G_A. Then, random data are used to feed the DUT for encrypting (or decrypting) for n_B times and then the corresponding traces are captured to form the group G_B. It is noticeable that the procedure of constructing G_A or G_B is not a consecutive one. The sequence of processing the fixed data or the random data should be random-cross (to avoid the irrelevant error caused by continuous processing of a fixed data—see, again, [31]). Once the G_A and G_B have been formed for the *non-$specific$* KS test, the Kolmogorov-Smirnov test is performed for each time point as the $specific$ KS test described before.

3.2 Multivariate Extensions

The above method can work sufficiently in the context that different points in a trace are treated as independent variables which can be happen in unprotected implementation of algorithm and the masked implementation on hardware which implies all the random shares processing concurrently. However, it is appealing to suppose that taking more than one data point into account might be beneficial as described in Section 2.1 that for at least some $\boldsymbol{\tau} \subset \{1, ..., T\}$ the samples points depend on some internal value (or state) $f_{k^*}(X)$. Besides, in the case of masked implementation on software the random shares are manipulated in sequence so that it is more efficient to detect the joint distributions of two or more trace points related to shares value.

Suppose the sensitive intermediate values $s = F_{k^*}(x)$, the principle behind masking is to split s into $d + 1$ shares $(r_0, ..., r_d \in \mathcal{Z})$ satisfying the relation[4]

$$s = r_0 \otimes r_1 \otimes ... \otimes r_d$$

where the $\otimes$ operation is the bitwise addition (or XOR) in the common case of Boolean masking. One of the shares, e.g. r_0, is sometimes referred to as the 'masked variable', with the other shares, $(r_1, ..., r_d)$ then viewed as the 'masks'. For the software implementation of masked scheme the leakage of the shares

[4] This relation exists implicitly even when it doesn't manifest directly in the cryptographic algorithm.

corresponding to the sensitive value s is

$$\mathbf{l} = (l_0, l_1, ..., l_d)$$

where

$$\begin{aligned} l_0 &= L_0 \circ (s \oplus r_1 \oplus \ldots \oplus r_d) + \varepsilon_0 \\ l_i &= L_i \circ (r_i) + \varepsilon_i, \qquad \text{for } 1 \leq i \leq d. \end{aligned}$$

Note that although it is difficult for the evaluator to precisely determine the location of l_i, we assume that ℓ time point candidates can be discovered for each share by some advanced methods such as points of interest extraction [29,7,13] or *a priori* knowledge about the masked implementation. Thus, the $(d+1)\ell$ time points should be analysed which implies that the size of all possible tuples would be ℓ^{d+1}. Since we focus on the general approach of leakage detection based on KS test and its efficiency in this paper, we just consider only one of the ℓ^{d+1} tuples.

For the *non-specific* test, the group G_A and G_B consist of the d-dimensional traces corresponding with the fixed input data and the random data respectively. And for the *specific* test, all the d-dimensional traces captured from the DUT with the random input are categorised to G_A and G_B by applying the distinguisher function on the sensitive value s based on the concept of the original differential power analysis [19].

Multivariate Test Kolmogorov-Smirnov test can be directly applied to test whether these multivariate samples $(l_0, l_1, ..., l_d)$ of G_A and G_B come from the same distribution according to equation (7) - (11) (recall Section 2.2). The calculation of the test statistic $D_{A,B}$ is a bit more complex here than in the univariate scenario since all the 2^d possible cases need to be taken into account. The final statistic is given by the max value of these $D_{A,B}s$.

Pre-processing The other option to deal with the multivariate leakage detection is to transform the multivariate samples into one-dimensional data first as the TVLA does [16,31] and then apply the univariate Kolmogorov-Smirnov test on the pre-processed one- dimensional data. The most popular transformation method (aka combined function) is probably the normalised product, shown in [28,33] to be the optimal choice in the idealised setting of a correlation attack [5] against Hamming weight leakage with Gaussian noise, which is defined as

$$l^* = \prod_{t=0}^{d} (l_t - \mu_t) \tag{12}$$

where l^* is the pre-processed data and μ_t denotes the mean of the samples at time point t. The G_A and G_B are pre-processed separately with (12) such that two univariate sample with size of n_A and n_B can be computed and univariate Kolmogorov-Smirnov test can be conducted.

4 Fast Implementation

In this section, we investigate how to implement the aforementioned methodology fast in practical evaluation. Since the evaluation might involve millions of traces, it is important to design an efficient algorithm to complete the process. Besides, since the statistic value is influenced by the number of traces designing an incremental algorithm can benefit the evaluation process. Inspired by the ideas to solve the fast TVLA problem [30] and the key rank estimation aspect in side-channel analysis [17,27], we introduce the histogram method to deal with the acceleration of the Kolmogorov-Smirnov test based leakage detection issue.

Assume that the evaluator captures n side-channel leakage traces each of which contains T points $\{P_1^{(i)}, ..., P_T^{(i)}\}$ $(i \in [1, n])$. The values of sample points are integral numbers and range from 0 to 2^{Q-1} since they are captured by oscilloscope (with number of quantization bits of Q). The fast implementation of Kolmogorov-Smirnov test-based leakage detection is divided into two categories implementation without pre-processing and implementation with pre-processing.

4.1 Implementation without Pre-processing

Suppose the multivariate test which is suitable for the software masking scheme and the situation exploiting joint leakage of multiple points in unprotected implementation be the bivariate test (other higher-dimensional test can be done in the same manner proposed in this section). In that case, two sample points of traces are used to conduct the leakage detection. The equation (7) can be rewritten by the following two equations

$$\varphi(x, y) = F_{\mathbf{A}_1,\mathbf{A}_2}(x, y) - F_{\mathbf{B}_1,\mathbf{B}_2}(x, y) \tag{13}$$

$$D_{A,B} = max|\varphi(x, y)|, \forall(x, y) \in (\mathbf{A}_1 \cup \mathbf{B}_1, \mathbf{A}_2 \cup \mathbf{B}_2) \tag{14}$$

The function $F(x, y)$ is defined as counting the number of samples that satisfy the conditions according to (7) - (11). In addition, we assume that the cardinality of group G_A and G_B are the same during the detection process. This constraint can be easily satisfied because the criterion used to category the whole traces is based on the random number (for *non-specific* test) or random intermediate value of cryptographic algorithm (for *specific* test). Therefore, the definition of $\varphi(x, y)$ in (13) becomes the mean of difference of the accumulative counts rather than the difference of mean accumulative counts. Consider the F^{+-} (equation (11)) scenario for instance, $\varphi(x, y)$ can be rewritten as

$$\begin{aligned}
\varphi(x, y) &= F_{\mathbf{A}_1,\mathbf{A}_2}(x, y) - F_{\mathbf{B}_1,\mathbf{B}_2}(x, y) \\
&= \frac{1}{n_A}\sum_{i=1}^{n_A}\sum_{j=1}^{n_A} I_{\mathbf{A}_{1,i}\le x, \mathbf{A}_{2,j}>y} - \frac{1}{n_B}\sum_{i=1}^{n_B}\sum_{j=1}^{n_B} I_{\mathbf{B}_{1,i}\le x, \mathbf{B}_{2,j}>y} \\
&= \frac{1}{n}(\sum_{i=1}^{n}\sum_{j=1}^{n} I_{\mathbf{A}_{1,i}\le x, \mathbf{A}_{2,j}>y} - \sum_{i=1}^{n}\sum_{j=1}^{n} I_{\mathbf{B}_{1,i}\le x, \mathbf{B}_{2,j}>y}) \\
&= \frac{1}{n}\phi(x, y)
\end{aligned} \tag{15}$$

where

$$\phi(x,y)=\sum_{i=1}^{n}\sum_{j=1}^{n}I_{\mathbf{A}_{1,i}\le x,\mathbf{A}_{2,j}>y}-\sum_{i=1}^{n}\sum_{j=1}^{n}I_{\mathbf{B}_{1,i}\le x,\mathbf{B}_{2,j}>y} \tag{16}$$

Hence, the problem is converted into compute the $\phi(x,y)$ function.

When a new observation $\mathbf{o}=(o_1,o_2)\in G_A\cup G_B$ is inserted into the set, if $\mathbf{o}\in G_A$ for all $x>o_1$ and at the same time $y\le o_2$, $\phi(x,y)$ is increased by 1. Otherwise, for all $x>o_1$ and at the same time $y\le o_2$, $\phi(x,y)$ is decreased by 1. The $\phi(x,y)$ for the rest of $x's$ and $y's$ are kept unchanged. In that way, all the $\phi(x,y)$ can be computed iteratively then the statistic value of KS can be computed as shown in Algorithm 1 which can be extended into univariate and higher-variate test logically in the same manner. Since the sample points x and y captured from oscilloscope are finite numbers ranging from 0 to 2^{Q-1}, such that the total computational complexity is $O(n*2^{Q*d})$ for the d-dimensional KS leakage detection[5]. It is comparable to the proposal based on brute force strategy suggested by Peacock [26], though optimized by Fasano and Franceschini [14] evaluating every $(x,y)\in(\mathbf{A}_1,\mathbf{A}_2)\cup(\mathbf{B}_1,\mathbf{B}_2)$ rather than every $(x,y)\in(\mathbf{A}_1\cup\mathbf{B}_1)\times(\mathbf{A}_2\cup\mathbf{B}_2)$, where $O(n^{d+1})$ is required.

Algorithm 1: Implementation without pre-processing

```
Input  : $n$ 2-dimensional traces $(P_1^{(i)}, P_2^{(i)})$, label vector $x_i \in \{0,1\}$
         $(i = 1,...,n)$
Output: $D_{A,B}$
for $i$=1 to $2^Q$ do
    for $j$=1 to $2^Q$ do
        $\phi(i,j) \leftarrow 0$;                          // initialization
    end
end
for $i$=1 to $n$ do
    for $j$=$P_1^{(i)}$ to $2^Q$ do
        for $k$=1 to $P_2^{(i)}$ do
            $\phi(j,k) \leftarrow \phi(j,k)$+1-2*$x_i$;           // recall (16)
        end
    end
end
$D_{A,B} \leftarrow 2*\max(|\phi(i,j)|)/n$
return $D_{A,B}$
```

4.2 Implementation with Pre-processing

The algorithm presented for the leakage detection without pre-processing takes advantage of the integral and finite numbers which can be used as the subscripts

[5] Q=8 for a typical oscilloscope.

of matrix in the algorithm. However, when dealing with the pre-processed traces such as (12) does, some or all elements of the consequential univariate trace are decimal numbers rather than integral numbers. In that case, Algorithm 1 is unable to be applied directly. Although the sample points in the traces are no longer integral numbers, the lower and upper bound of these sample points are limited. For example, in the bivariate setting l_0 in (12) ranges from 0 to 255 such that the mean value of this variable μ_0 also lays in the range of (0,255). Hence, $l_0 - \mu_0 \in [-\mu_0, 255 - \mu_0])$. The second variable has the same property so that $(l_0 - \mu_0)(l_1 - \mu_1) \in (-256^2, 256^2)$. Note that the lower bound and upper bound in the practical evaluation are much tighter than the theoretical result because the values of the samples at one time point are centralized in a much smaller range than [0,255].

The fast implementation of KS-based leakage detection on the traces with pre-processing is divided into two stages. The first stage is to pre-process the multivariate traces which can be implemented easily by computing the mean value of each variable and then calculating the normalised product in computational complexity of $O(n)$. Repeat that for the two groups G_A and G_B the univariate pre-processed traces can be computed. The second stage is to apply the univariate KS test on the two pre-processed traces. As explained before, the value of the samples in the traces are bounded thus the new histogram (with N_{bin} equally sized bins) can be applied to count the number of each sample value approximatively. Comparing the histograms belonging to the two groups can result in the computation of the statistic value $D_{n,n}$ in (3) (again, we assume that the cardinalities of G_A and G_B are the same). Algorithm 2 shows the detailed procedure of the second stages with the computational complexity of $O(n * N_{bin})$.

We now give the basic intention behind Algorithm 2. The first step is to determine the bounds of the samples (tighter than $(-256^2, 256^2)$) which are then used to calculate the binsize of the histogram. When the current observation l^* comes from group G_A, the counting number of those samples value larger than l^* should be increased by 1 according to (3) which leads to the histogram with larger subscripts than l^*'s subscript $round((l^* - LowerBound)/S_{bin})+1$ where the function $round(\cdot)$ represents the nearest integer.

4.3 Bounding the Error

Assume that the samples of two random variables X and Y are $\{x_1, ..., x_n\}$ and $\{y_1, ..., y_n\}$. For the KS test, the goal is to calculate the max distance of the cumulative distribution function. When the histogram proposed in 4.2 is applied the x_i and y_i are located in different places determined by the central value of histogram bins and the binsize. Thus, the overall max distance would not be exactly the same with the original one which can be explained by the following example.

Example 1. Let X be $\{0, 2.8, 1.3, 2.5, 1.8, 3.0, 2.7, 0.8, 0.4, 2.3\}$ and Y be $\{0.1, 1.4, 2.3, 0.3, 2.0, 1.9, 2.4, 1.5, 1.8, 0.7\}$, the max difference of cumulative distribution

Algorithm 2: Implementation with pre-processing

```
Input  : $n$ univariate pre-processed traces $\{l_A^{*(i)}\}$ for group $G_A$, $n$ univariate
         pre-processed traces $\{l_B^{*(i)}\}$ for group $G_B$ $(i = 1, ..., n)$, $N_{bin}$
Output: $D_{n,n}$
$UpperBound$=max(max($\{l_A^{*(i)}\}$),max($\{l_B^{*(i)}\}$)),$(i = 1, ..., n)$
$LowerBound$=min(min($\{l_A^{*(i)}\}$),min($\{l_B^{*(i)}\}$)),$(i = 1, ..., n)$
$S_{bin} \leftarrow (UpperBound - LowerBound)/N_{bin}$;                    // set binsize
for $i$=1 to $N_{bin}$ do
    $\phi(i) \leftarrow 0$;                                                // initialization
end
for $i$=1 to $n$ do
    $ID_A \leftarrow round((l_A^{*(i)} - LowerBound)/S_{bin}) + 1$
    $ID_B \leftarrow round((l_B^{*(i)} - LowerBound)/S_{bin}) + 1$;          // set subscript
    if $ID_A \leq ID_B$ then
        for $j$=$ID_A$ to $ID_B$ do
            $\phi(i) \leftarrow \phi(i) + 1$
        end
    else
        for $j$=$ID_B$ to $ID_A$ do
            $\phi(i) \leftarrow \phi(i) - 1$
        end
    end
end
$D_{n,n} \leftarrow \max(|\phi(i)|)/n$
return $D_{n,n}$
```

function is 0.4 at P($y \leq 2.4$)−P($x \leq 2.4$). When the bins number N_{bin} is set to be 3 which leads binsize S_{bin} to be 1. In this case, the histograms of X and Y are $H_X = \{3, 2, 5\}$ and $H_Y = \{3, 4, 3\}$ which means the max distance of cumulative distribution function is 0.2. However, for $N_{bin} = 5$ which leads to $S_{bin} = 0.6$ and the histograms are $H_X = \{2, 1, 2, 1, 4\}$ and $H_Y = \{2, 1, 2, 4, 1\}$, the corresponding max distance is 0.3 that is closer to the true value.

The error between the proposed histogram method and the original '$bruteforce$' method for the KS test can be bounded through some theoretical analysis. Suppose the probability density functions (PDF) of random variables X and Y are $\alpha(x)$ and $\beta(y)$ and assume that they are smooth (i.e. the derivative $\alpha'(x)$ and $\beta'(y)$ are bounded for all x and y). Again, let $\{x_1, ..., x_n\}$ and $\{y_1, ..., y_n\}$ be the samples of variables X and Y.

Proposition 1 *The mean error e between histogram method and the original 'brute force' method can be bounded by $|e| < 2B * S_{bin} * L$, where B is the boundary of the derivative of PDF satisfying $|\alpha'(x)| < B$ and $|\beta'(x)| < B$, S_{bin} is the binsize of histogram, L is the distance between smallest value and largest value among the samples.*

Proof. When the binsize is set to be S_{bin}, the samples are categorized into N_{bin} groups:

$$C_1 = [0, S_{bin}), C_2 = [S_{bin}, 2S_{bin}), ...C_N = [(N-1)S_{bin}, NS_{bin})$$

Let $\hat{\alpha}(x)$ be the PDF estimator of C_i.

$$\begin{aligned} \mathrm{E}(\hat{\alpha}(x)) &= \frac{1}{S_{bin}} P(x \in C_i) \\ &= \frac{1}{S_{bin}} \int_{(i-1)S_{bin}}^{iS_{bin}} \alpha(x) dx \\ &= \frac{F(iS_{bin}) - F((i-1)S_{bin})}{S_{bin}} \\ &= \alpha(t), (t \in ((i-1)S_{bin}, iS_{bin})) \end{aligned} \tag{17}$$

The last equality is satisfied because of the Lagrange Mean Value Theorem. The error e' between PDF is

$$\begin{aligned} e' &= |\mathrm{E}(\hat{\alpha}(x)) - \alpha(x)| \\ &= \alpha(t) - \alpha(x) \\ &= |\alpha'(s)(t-x)|, (again, s \in (t, x)) \\ &< B * S_{bin} \end{aligned} \tag{18}$$

The cumulative density function is defined as the integral of PDF, hence the error between these two kinds of CDF is denoted as:

$$\begin{aligned} |\int_{-\infty}^{u} \mathrm{E}(\hat{\alpha}(x)) dx - \int_{-\infty}^{u} \alpha(x) dx| &= |\int_{-\infty}^{u} (\mathrm{E}(\hat{\alpha}(x)) - \alpha(x)) dx| \\ &< u * B * S_{bin} \\ &< L * B * S_{bin} \end{aligned} \tag{19}$$

The second last inequation is done because the whole integral can be calculated within each of the S_{bin} size region. With the same procedure, the error of CDF on variable Y still can be bounded

$$|\int_{-\infty}^{u} \mathrm{E}(\hat{\beta}(y))dy - \int_{-\infty}^{u} \beta(y)dy| < L * B * S_{bin} \tag{20}$$

Therefore, the error e between the KS statistic that measures the largest difference of two CDF can be determined through (19) and (20),

$$||\int_{-\infty}^{u} \mathrm{E}(\hat{\alpha}(x))dx - \int_{-\infty}^{u} \mathrm{E}(\hat{\beta}(x))dx| - |\int_{-\infty}^{u} \alpha(x)dx - \int_{-\infty}^{u} \beta(x)dx|| < 2L*B*S_{bin}$$

□

5 Simulated Experiments

We here present the outcomes of several experiments on simulated leakages both in unprotected and masking contexts. As a baseline against which to compare the leakage detection performance of KS, we also tested the TVLA proposal. Our chosen evaluation metric is the mean p-values (among 200 trials for each experiment) to accept the null hypothesis which is concluding there is not a leak.

5.1 Unprotected Univariate Leakage

We first explore the performance of KS detection on the leakage of unprotected implementation. To do this, we simulate traces by adding Gaussian noise ε_G to the Hamming weight of intermediate value (recall Section 2.1).

$$l = HW(s) + \varepsilon_G$$

where $HW(\cdot)$ represents the hamming weight. The magnitude of noise is set to be $\frac{2}{SNR}$. Two groups of simulated traces G_A and G_B are generated where the traces in G_A correspond with the fixed intermediate value s and the traces in G_B correspond with the random intermediate values.

For the purpose of comprehensive comparison, we investigate the performance of the KS leakage detection under different realizations of these parameters in unprotected univariate leakage:

- The fixed value for simulating the $fixed$ leakage.
- The magnitude of noise.

Influence of noise First, we consider the influence of the noise. Since the noise is the parameter under test here, we fix the Hamming weight of fixed value to 3. We change the variance of noise and let the SNR be 2, 2^{-1}, 2^{-3}, 2^{-5}. Figure 2 shows the p-value results of KS based leakage detection and t-test based TVLA while SNR varies. We first observe that the TVLA outperforms KS leakage detection when SNR decreases.

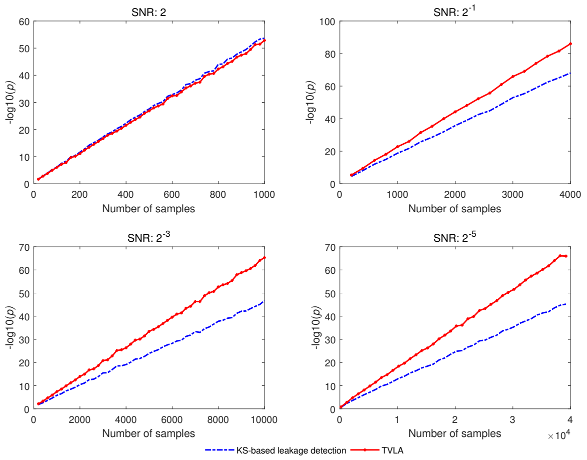

Fig. 2. Results of KS-based leakage detection and TVLA against unprotected leakage as SNR varies (Hamming weight: 3).

Influence of fixed value We then test the influence of fixed value. Since the fixed value is the parameter under test here, we fix the SNR to 2^{-3}. We change the hamming weight of the fixed value that is used to generate the fixed traces to be 2, 3, 4, and 5. The results are shown in Figure 3. We find that KS-based leakage detection can work stably for all hamming weights while the TVLA's performance falls a lot at hamming weight of 4. The main reason is that the TVLA takes the mean value of samples as the critical criterion to determine whether the two samples are from same distribution. Since the mean value of the informative-part samples from random traces will converge to 4 when number of traces increases through central limit theorem, the samples from random set and the samples from fixed set with fixed hamming weight 4 can not be distinguished by TVLA. However, KS as an information-theory tool measuring the characteristic of the CDF would not be influenced by the fixed value.

5.2 Masked Univariate Leakage

In this section, we consider the masking scheme that is for hardware implementation where the masks are manipulated in parallel. For the evaluation of the KS detection performance on the masked leakage (both in univariate and multivariate context in the later subsection), one more parameter is considered:

- The order of mask.

In order to simulate the univariate higher-order masking leakage, we use the following formula.

$$l = \sum_{i=0}^{d} HW(r_i) + \varepsilon_G \tag{21}$$

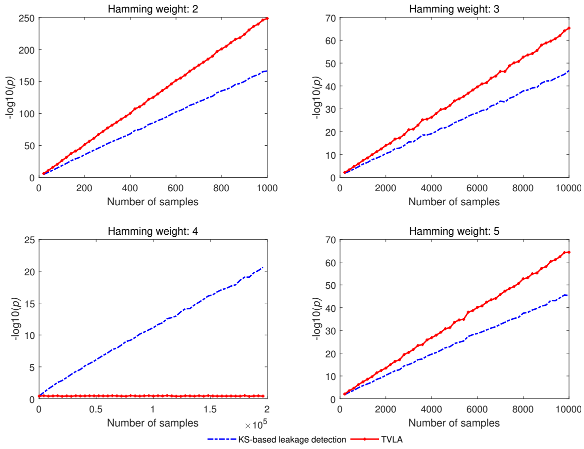

Fig. 3. Results of KS-based leakage detection and TVLA against unprotected leakage as hamming weight of fixed value varies (SNR: 2^{-3}).

where r_i $(i = 1, ..., d)$ represents the d random masks uniformly distributed in [0, 255], r_0 is referred as the 'masked variable' satisfying $r_0 = \overset{d}{\underset{i=1}{\oplus}} r_i \oplus s$. Note that the SNR here is slightly different with the unprotected simulation since the 'signal' consists of $d + 1$ valuable information. Ignore the difference introduced by the calculation of r_0 (we suppose the $d + 1$ variables are independent), the variance of signal can be given by $2(d + 1)$ approximately (relevant explanation can be found in Appendix A). Hence the variance of noise in this part is determined by $\frac{2(d+1)}{SNR}$. The pre-processing approach for the masked univariate leakage is:

$$l^* = (l - \mu_l)^d \tag{22}$$

where μ_l is the mean of the sample.

Influence of noise We first investigate the influence of noise on the performance of leakage detection in the masked univariate setting. As is done in the unprotected scenario, we fix the hamming weight of fixed value to 0, the order of mask to 3, an allow the SNR to be $\{4,\ 1,\ 2^{-1},\ 2^{-2}\}$. Figure 4 shows the experimentally observed performance of these two leakage detection tools given different numbers of traces. The results indicate that KS can be a leakage detection tool in masked leakage setting. Like the results in the unprotected scenario, the TVLA still outperforms KS-based leakage detection since the hamming weight is 0.

Influence of fixed value As before, we investigate the influence of fixed value on the performance of leakage detection in the masked univariate setting. The SNR is set to 4 and the order is set to 2. The experimental result is indicated in

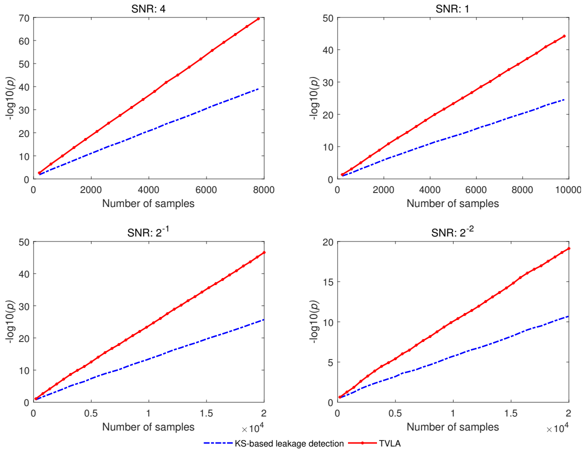

Fig. 4. Results of KS-based leakage detection and TVLA against masked univariate leakage as SNR varies (Hamming weight: 0, order of mask: 3).

Figure 5. The KS-based leakage detection works always more stably than TVLA when hamming weight of fixed value varies. Besides, when the hamming weight of fixed value is 4 KS-based leakage detection significantly outperforms TVLA.

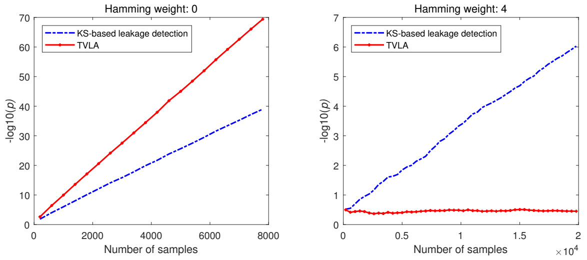

Fig. 5. Results of KS-based leakage detection and TVLA against masked univariate leakage as hamming weight of fixed value varies (Order of mask: 2, SNR: 4).

Influence of order Different with the unprotected scenario, we test a more parameter order of masks d on the performance of the leakage detection tools for the masked scenario. We fix the hamming weight to 0 (in consideration that TVLA might not work well in other settings), SNR to 16, and allow the d to vary in range of $\{3, 4, 6, 7\}$. The results are presented in Figure 6. When the order of masks increases, the advantage of KS-based leakage detection becomes more significant when compared with TVLA.

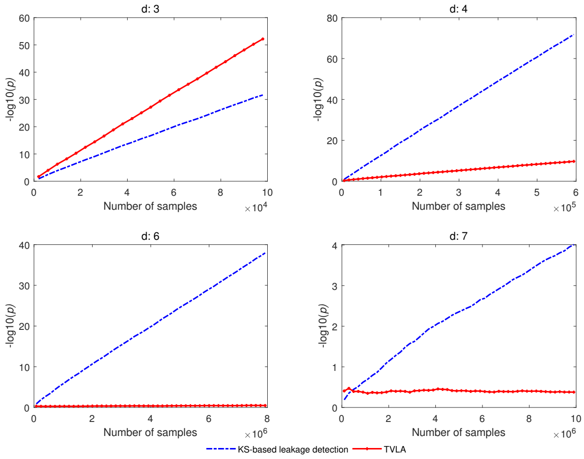

Fig. 6. Results of KS-based leakage detection and TVLA against masked univariate leakage as order of mask varies (Fixed value is 0, SNR is 16.)

6 Practical Experiments

We now move to consider the performance of the KS-based leakage detection in the practical setting. The targeted device is an AVR ATmega328P microcontroller with a clock frequency of 16MHz to execute uploaded programs. For the unprotected AES implementation, the encryption program is assembled using the library *AESLib* designed by Davy Landman for Arduino-specific port[1]. For the protected AES implementation, we changed the code 'AES-256 RSM' downloaded from the DPA contest v4 [2] which is designed for Atmel ATMega-163 smart-card to adapt for our microcontroller. The mask scheme implemented is the Rotating S-boxes Masking (RSM; for details, see [25]).The digital oscilloscope used to capture the current flow curve during encryption is 8-bit precision Lecroy waverunner 8104 at a sampling rate of 100MS/s.

6.1 Unprotected Scenario

We capture 400 thousand traces for each of fixed and random set respectively. The experiments is repeated for 200 times for each given number of traces. The results are shown in Figure 7. It can be learned from Figure 7 that the KS can still be a stable tool for leakage detection in practice and slightly outperforms TVLA when the number of traces increases.

6.2 Masked Scenario

For masked scenario we capture 500 thousand traces for each of fixed and random set respectively. The RSM scheme involves random masks and random offsets. We focus only on the two-dimensional leakage detection. It is a characteristic of

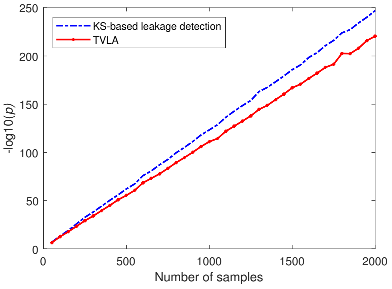

Fig. 7. Results of KS-based leakage detection and TVLA against unprotected traces in practical setting.

the RSM scheme that the output of the masked S-box and the masked value of next sub-plaintext have the same mask, so that their XOR result can remove the mask. In detail, the first part is $MSbox(x_i \oplus k \oplus r_{i+offset})$, and the second part is $x_{i+1} \oplus r_{i+1+offset}$ where $MSbox$ is the masked sbox, i is the index of the sub-plaintext, $offset$ is a random number, and r is a mask table. According to the description of the RSM algorithm, the first part can be expressed as

$$MSbox(x_i \oplus k \oplus r_{i+offset}) = Sbox(x_i \oplus k) \oplus r_{i+1+offset}$$

Hence, the XOR result of the two parts is $Sbox(x_i \oplus k) \oplus x_{i+1}$ which, although slightly different to the intermediates targeted in the simulated leakage scenario, can be computed for leakage detection. We don't promote this combination of intermediate values are the optimal choice, we simply make use of it as a target to demonstrate the performance of our proposed KS-based leakage detection. Unlike the univariate setting, the efficiency of multivariate KS is also tested here.

The results including performance of KS with normalised product, multivariate KS and TVLA are indicated in Figure 8. From Figure 8 it can be observed that the two kinds of KS-based leakage detection outperform TVLA in two setting of hamming weight of fixed value. The multivariate KS taking advantage of the original multivariate distribution of samples shows a distinct superiority when compared with the pre-processing-based KS leakage detection.

7 Conclusion and Future Perspectives

In this paper, we took the logical next step of extending Kolmogorov-Smirnov test which is a well-known nonparametric method for statistical analysis and has been widely studied in side-channel distinguisher application to the recently emerged leakage detection domain. In consideration of the side-channel leakage situation, we proposed two test methods both with fast implementations to significantly reduce the computation resource for performing practical leakage detection. We performed a range of experiments both on simulated leakage

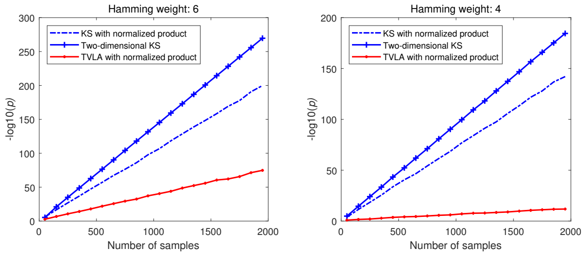

Fig. 8. Results of KS-based leakage detection and TVLA against masked traces in practical setting.

with various parameters and practical traces to verify its efficiency. Although the comprehensive comparison with TVLA reveals that KS-based leakage detection might not be the optimal under all conditions, its performance is strongly robust when the parameters vary especially in some cases where TVLA fails to detect leakage due to the self-characteristic of information-theory. On the other hand, the empirical results show that KS-based leakage detection would seem to be a right-hand supplement to TVLA for practitioners.

Since some recent literatures have proposed some ideas for how to speed up the KS test for statistical analysis[12,38]. Applications of these schemes for the multivariate KS test and the online KS leakage detection can be avenues for future work.

Besides, a unified framework of how to fairly and more comprehensively compare and evaluate the leakage detection methods would be an interesting topic along this line in the future work as per [35].

References

1. Arduino AESLib. `https://github.com/DavyLandman/AESLib`.
2. DPA Contest v4. `http://www.dpacontest.org/v4/`.
3. George Becker, J Cooper, Elke DeMulder, Gilbert Goodwill, Joshua Jaffe, G Kenworthy, T Kouzminov, A Leiserson, M Marson, Pankaj Rohatgi, et al. Test vector leakage assessment (tvla) methodology in practice. In *International Cryptographic Module Conference*, volume 1001, page 13, 2013.
4. Shivam Bhasin, Jean-Luc Danger, Sylvain Guilley, and Zakaria Najm. Nicv: normalized inter-class variance for detection of side-channel leakage. In *2014 International Symposium on Electromagnetic Compatibility, Tokyo*, pages 310–313. IEEE, 2014.
5. Eric Brier, Christophe Clavier, and Francis Olivier. Correlation power analysis with a leakage model. In Marc Joye and Jean-Jacques Quisquater, editors, *Cryptographic Hardware and Embedded Systems - CHES 2004: 6th International Workshop Cambridge, MA, USA, August 11-13, 2004. Proceedings*, volume 3156 of *Lecture Notes in Computer Science*, pages 16–29. Springer, 2004.
6. Olivier Bronchain, Tobias Schneider, and François-Xavier Standaert. Multi-tuple leakage detection and the dependent signal issue. *IACR Trans. Cryptogr. Hardw. Embed. Syst.*, 2019(2):318–345, 2019.

7. Eleonora Cagli, Cécile Dumas, and Emmanuel Prouff. Kernel discriminant analysis for information extraction in the presence of masking. In Kerstin Lemke-Rust and Michael Tunstall, editors, *Smart Card Research and Advanced Applications - 15th International Conference, CARDIS 2016, Cannes, France, November 7-9, 2016, Revised Selected Papers*, volume 10146 of *Lecture Notes in Computer Science*, pages 1–22. Springer, 2016.
8. Konstantinos Chatzikokolakis, Tom Chothia, and Apratim Guha. Statistical measurement of information leakage. In Javier Esparza and Rupak Majumdar, editors, *Tools and Algorithms for the Construction and Analysis of Systems, 16th International Conference, TACAS 2010, Held as Part of the Joint European Conferences on Theory and Practice of Software, ETAPS 2010, Paphos, Cyprus, March 20-28, 2010. Proceedings*, volume 6015 of *Lecture Notes in Computer Science*, pages 390–404. Springer, 2010.
9. Tom Chothia and Apratim Guha. A statistical test for information leaks using continuous mutual information. In *Proceedings of the 24th IEEE Computer Security Foundations Symposium, CSF 2011, Cernay-la-Ville, France, 27-29 June, 2011*, pages 177–190, 2011.
10. A. Adam Ding, Cong Chen, and Thomas Eisenbarth. Simpler, faster, and more robust t-test based leakage detection. In François-Xavier Standaert and Elisabeth Oswald, editors, *Constructive Side-Channel Analysis and Secure Design - 7th International Workshop, COSADE 2016, Graz, Austria, April 14-15, 2016, Revised Selected Papers*, volume 9689 of *Lecture Notes in Computer Science*, pages 163–183. Springer, 2016.
11. A. Adam Ding, Liwei Zhang, François Durvaux, François-Xavier Standaert, and Yunsi Fei. Towards sound and optimal leakage detection procedure. In Thomas Eisenbarth and Yannick Teglia, editors, *Smart Card Research and Advanced Applications - 16th International Conference, CARDIS 2017, Lugano, Switzerland, November 13-15, 2017, Revised Selected Papers*, volume 10728 of *Lecture Notes in Computer Science*, pages 105–122. Springer, 2017.
12. Denis Moreira dos Reis, Peter A. Flach, Stan Matwin, and Gustavo E. A. P. A. Batista. Fast unsupervised online drift detection using incremental kolmogorov-smirnov test. In Balaji Krishnapuram, Mohak Shah, Alexander J. Smola, Charu C. Aggarwal, Dou Shen, and Rajeev Rastogi, editors, *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, August 13-17, 2016*, pages 1545–1554. ACM, 2016.
13. François Durvaux and François-Xavier Standaert. From improved leakage detection to the detection of points of interests in leakage traces. In Marc Fischlin and Jean-Sébastien Coron, editors, *Advances in Cryptology - EUROCRYPT 2016 - 35th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Vienna, Austria, May 8-12, 2016, Proceedings, Part I*, volume 9665 of *Lecture Notes in Computer Science*, pages 240–262. Springer, 2016.
14. Giovanni Fasano and Alberto Franceschini. A multidimensional version of the kolmogorov–smirnov test. *Monthly Notices of the Royal Astronomical Society*, 225(1):155–170, 1987.
15. Benedikt Gierlichs, Lejla Batina, Pim Tuyls, and Bart Preneel. Mutual information analysis. In Elisabeth Oswald and Pankaj Rohatgi, editors, *Cryptographic Hardware and Embedded Systems - CHES 2008, 10th International Workshop, Washington, D.C., USA, August 10-13, 2008. Proceedings*, volume 5154 of *Lecture Notes in Computer Science*, pages 426–442. Springer, 2008.

16. Benjamin Jun Gilbert Goodwill, Josh Jaffe, Pankaj Rohatgi, et al. A testing methodology for side-channel resistance validation. In *NIST non-invasive attack testing workshop*, volume 7, pages 115–136, 2011.
17. Cezary Glowacz, Vincent Grosso, Romain Poussier, Joachim Schüth, and François-Xavier Standaert. Simpler and more efficient rank estimation for side-channel security assessment. In Gregor Leander, editor, *Fast Software Encryption - 22nd International Workshop, FSE 2015, Istanbul, Turkey, March 8-11, 2015, Revised Selected Papers*, volume 9054 of *Lecture Notes in Computer Science*, pages 117–129. Springer, 2015.
18. Annelie Heuser, Olivier Rioul, and Sylvain Guilley. A theoretical study of kolmogorov-smirnov distinguishers - side-channel analysis vs. differential cryptanalysis. In Emmanuel Prouff, editor, *Constructive Side-Channel Analysis and Secure Design - 5th International Workshop, COSADE 2014, Paris, France, April 13-15, 2014. Revised Selected Papers*, volume 8622 of *Lecture Notes in Computer Science*, pages 9–28. Springer, 2014.
19. Paul C. Kocher, Joshua Jaffe, and Benjamin Jun. Differential power analysis. In Michael J. Wiener, editor, *Advances in Cryptology - CRYPTO '99, 19th Annual International Cryptology Conference, Santa Barbara, California, USA, August 15-19, 1999, Proceedings*, volume 1666 of *Lecture Notes in Computer Science*, pages 388–397. Springer, 1999.
20. Houssem Maghrebi, Olivier Rioul, Sylvain Guilley, and Jean-Luc Danger. Comparison between side-channel analysis distinguishers. In Tat Wing Chim and Tsz Hon Yuen, editors, *Information and Communications Security - 14th International Conference, ICICS 2012, Hong Kong, China, October 29-31, 2012. Proceedings*, volume 7618 of *Lecture Notes in Computer Science*, pages 331–340. Springer, 2012.
21. Stefan Mangard, Elisabeth Oswald, and François-Xavier Standaert. One for all - all for one: unifying standard differential power analysis attacks. *IET Information Security*, 5(2):100–110, 2011.
22. Frank J Massey Jr. The kolmogorov-smirnov test for goodness of fit. *Journal of the American statistical Association*, 46(253):68–78, 1951.
23. Luke Mather, Elisabeth Oswald, Joe Bandenburg, and Marcin Wójcik. Does my device leak information? an a priori statistical power analysis of leakage detection tests. In Kazue Sako and Palash Sarkar, editors, *Advances in Cryptology - ASIACRYPT 2013 - 19th International Conference on the Theory and Application of Cryptology and Information Security, Bengaluru, India, December 1-5, 2013, Proceedings, Part I*, volume 8269 of *Lecture Notes in Computer Science*, pages 486–505. Springer, 2013.
24. Amir Moradi, Bastian Richter, Tobias Schneider, and François-Xavier Standaert. Leakage detection with the x2-test. *IACR Trans. Cryptogr. Hardw. Embed. Syst.*, 2018(1):209–237, 2018.
25. Maxime Nassar, Youssef Souissi, Sylvain Guilley, and Jean-Luc Danger. RSM: A small and fast countermeasure for aes, secure against 1st and 2nd-order zero-offset scas. In Wolfgang Rosenstiel and Lothar Thiele, editors, *2012 Design, Automation & Test in Europe Conference & Exhibition, DATE 2012, Dresden, Germany, March 12-16, 2012*, pages 1173–1178. IEEE, 2012.
26. JA Peacock. Two-dimensional goodness-of-fit testing in astronomy. *Monthly Notices of the Royal Astronomical Society*, 202(3):615–627, 1983.
27. Romain Poussier, François-Xavier Standaert, and Vincent Grosso. Simple key enumeration (and rank estimation) using histograms: An integrated approach. In Benedikt Gierlichs and Axel Y. Poschmann, editors, *Cryptographic Hardware and*

Embedded Systems - CHES 2016 - 18th International Conference, Santa Barbara, CA, USA, August 17-19, 2016, Proceedings, volume 9813 of *Lecture Notes in Computer Science*, pages 61–81. Springer, 2016.

28. Emmanuel Prouff, Matthieu Rivain, and Régis Bevan. Statistical analysis of second order differential power analysis. *IEEE Trans. Computers*, 58(6):799–811, 2009.
29. Oscar Reparaz, Benedikt Gierlichs, and Ingrid Verbauwhede. Selecting time samples for multivariate DPA attacks. In Emmanuel Prouff and Patrick Schaumont, editors, *Cryptographic Hardware and Embedded Systems - CHES 2012 - 14th International Workshop, Leuven, Belgium, September 9-12, 2012. Proceedings*, volume 7428 of *Lecture Notes in Computer Science*, pages 155–174. Springer, 2012.
30. Oscar Reparaz, Benedikt Gierlichs, and Ingrid Verbauwhede. Fast leakage assessment. In Wieland Fischer and Naofumi Homma, editors, *Cryptographic Hardware and Embedded Systems - CHES 2017 - 19th International Conference, Taipei, Taiwan, September 25-28, 2017, Proceedings*, volume 10529 of *Lecture Notes in Computer Science*, pages 387–399. Springer, 2017.
31. Tobias Schneider and Amir Moradi. Leakage assessment methodology - A clear roadmap for side-channel evaluations. In Tim Güneysu and Helena Handschuh, editors, *Cryptographic Hardware and Embedded Systems - CHES 2015 - 17th International Workshop, Saint-Malo, France, September 13-16, 2015, Proceedings*, volume 9293 of *Lecture Notes in Computer Science*, pages 495–513. Springer, 2015.
32. François-Xavier Standaert. How (not) to use welch's t-test in side-channel security evaluations. In Begül Bilgin and Jean-Bernard Fischer, editors, *Smart Card Research and Advanced Applications, 17th International Conference, CARDIS 2018, Montpellier, France, November 12-14, 2018, Revised Selected Papers.*, volume 11389 of *Lecture Notes in Computer Science*, pages 65–79. Springer, 2018.
33. François-Xavier Standaert, Nicolas Veyrat-Charvillon, Elisabeth Oswald, Benedikt Gierlichs, Marcel Medwed, Markus Kasper, and Stefan Mangard. The world is not enough: Another look on second-order DPA. In Masayuki Abe, editor, *Advances in Cryptology - ASIACRYPT 2010 - 16th International Conference on the Theory and Application of Cryptology and Information Security, Singapore, December 5-9, 2010. Proceedings*, volume 6477 of *Lecture Notes in Computer Science*, pages 112–129. Springer, 2010.
34. Nicolas Veyrat-Charvillon and François-Xavier Standaert. Mutual information analysis: How, when and why? In Christophe Clavier and Kris Gaj, editors, *Cryptographic Hardware and Embedded Systems - CHES 2009, 11th International Workshop, Lausanne, Switzerland, September 6-9, 2009, Proceedings*, volume 5747 of *Lecture Notes in Computer Science*, pages 429–443. Springer, 2009.
35. Carolyn Whitnall and Elisabeth Oswald. A cautionary note regarding the usage of leakage detection tests in security evaluation. *IACR Cryptology ePrint Archive*, 2019:703, 2019.
36. Carolyn Whitnall and Elisabeth Oswald. A critical analysis of ISO 17825 ('testing methods for the mitigation of non-invasive attack classes against cryptographic modules'). *IACR Cryptology ePrint Archive*, 2019:1013, 2019.
37. Carolyn Whitnall, Elisabeth Oswald, and Luke Mather. An exploration of the kolmogorov-smirnov test as a competitor to mutual information analysis. In Emmanuel Prouff, editor, *Smart Card Research and Advanced Applications - 10th IFIP WG 8.8/11.2 International Conference, CARDIS 2011, Leuven, Belgium, September 14-16, 2011, Revised Selected Papers*, volume 7079 of *Lecture Notes in Computer Science*, pages 234–251. Springer, 2011.
38. Yuanhui Xiao. A fast algorithm for two-dimensional kolmogorov–smirnov two sample tests. *Computational Statistics & Data Analysis*, 105:53–58, 2017.

A Sum of binomial distributions

Suppose that the two independent random integer numbers are x_0 and x_1 from [0, 255] with equal probability which can be expressed as $x_0 \sim U(0, 255)$ and $x_1 \sim U(0, 255)$. Therefore, the Hamming weight of x_0 $HW(x_0)$ and x_1 $HW(x_1)$ are from binomial distribution $B(8, \frac{1}{2})$. The sum Z of these two random variables are still from binomial distribution. Assume $X_0 = HW(x_0)$, $X_1 = HW(x_1)$ and $Z = X_0 + X_1$.

$$\begin{aligned}
P(X_0 + X_1 = k) &= \sum_{i=0}^{k} P(X_0 = i, X_1 = k - i) \\
&= \sum_{i=0}^{k} P(X_0 = i) P(X_1 = k - i) \\
&= \sum_{i=0}^{k} \binom{8}{i} (\frac{1}{2})^8 \binom{8}{k-i} (\frac{1}{2})^8 \\
&= (\frac{1}{2})^{16} \sum_{i=0}^{k} \binom{8}{i} \binom{8}{k-i} \\
&= \binom{16}{k} (\frac{1}{2})^{16}
\end{aligned}$$

Thus, $z \sim B(16, \frac{1}{2})$, and so is the sum of more random variables of such kind.