How to Estimate the Success Rate of Higher-Order Side-Channel Attacks

Victor Lomné[1], Emmanuel Prouff[1], Matthieu Rivain[2], Thomas Roche[1], and Adrian Thillard[1]

[1] ANSSI
firstname.name@ssi.gouv.fr

[2] CryptoExperts
matthieu.rivain@cryptoexperts.com

Abstract. The resistance of a cryptographic implementation with regards to side-channel analysis is often quantified by measuring the success rate of a given attack. This approach cannot always be followed in practice, especially when the implementation includes some countermeasures that may render the attack too costly for an evaluation purpose, but not costly enough from a security point of view. An evaluator then faces the issue of estimating the success rate of an attack he cannot mount. The present paper addresses this issue by presenting a methodology to estimate the success rate of higher-order side-channel attacks targeting implementations protected by masking. Specifically, we generalize the approach initially proposed at SAC 2008 in the context of first-order side-channel attacks. The principle is to approximate the distribution of an attack's score vector by a multivariate Gaussian distribution, whose parameters are derived by profiling the leakage. One can then accurately compute the expected attack success rate with respect to the number of leakage measurements. We apply this methodology to higher-order side-channel attacks based on the widely used correlation and likelihood distinguishers. Moreover, we validate our approach with simulations and practical attack experiments against masked AES implementations running on two different microcontrollers.

1 Introduction

Estimating the success rate of a side-channel attack –that uses a given number of leakage observations– is a central issue regarding the physical security evaluation of a cryptographic implementation. The empirical way is to perform the attack a certain number of times and to record the average number of successes. However, this approach is prohibitive against implementations protected by effective countermeasures since the attacks may become too costly to be performed several times (or even once). This does not mean that the implementation is *secure* though; this only means that the implementation is secure beyond the means of the evaluator (which may not compete with the means of a motivated attacker).

This situation is not satisfactory in practice where one desires that the computational cost of performing a security evaluation be fairly low and uncorrelated to the actual security of the target implementation.

In this paper, we propose a methodology to estimate the success rate of higher-order side-channel attacks targeting implementations protected by masking. Our methodology is based on the approach proposed by Rivain in [13] in the context of first-order side-channel attacks. The principle of this approach is to study the multivariate distribution of the score vector resulting from an attack. Specifically, Rivain suggests to approximate this distribution by a multivariate Gaussian distribution, which is sound in the context of *additive distinguishers* such as the correlation and the likelihood. We generalize this approach to higher-order side-channel analysis and we show how to derive the distribution parameters with respect to the leakage parameters. We show that using this methodology makes it possible to accurately estimate the success rate of a higher-order side-channel attack based on a simple profiling of the leakage parameters. Moreover, we demonstrate the soundness of our methodology by comparing its results to various attack experiments against masked AES implementations running on two different microcontrollers.

Related Works. In [10] and [17], the success rate of first-order side-channel analysis based on the correlation distinguisher is evaluated using Fisher's transformation. The obtained formulas are simple and illustrative, but they lack of accuracy. Indeed, it has been observed in [18] that the estimated success rates using this approach do not well match to the experimental ones. As explained in [18], this is mainly due to the incorrect assumption that the scores for the wrong key guesses are independent of the score for the good key guess. That is why, one should rather focus on the joint distribution of all scores as initially suggested in [13]. In the latter work, the author provide accurate formulae for the success rate of first-order side-channel attacks based on the correlation and likelihood distinguishers. A more recent work [6] further focuses on the mono-bit difference-of-means distingusisher as originally described by Kocher *et al.* [9].

Paper Organization. In Section 2, we provide some preliminaries about probability theory and the (multivariate) Gaussian distribution. Then Section 3 introduces our theoretical model for higher-order side-channel attacks and Section 4 describes the general methodology for estimating the success rate of such attacks based on additive distinguishers. In Sections 5 and 6, we apply the methodology to the correlation and the likelihood distinguishers respectively, and we show how to compute the score vector distribution parameters. Eventually, some attack simulations and practical attack experiments are reported in Sections 7 and 8 that demonstrate the soundness of our approach.

2 Preliminaries

Calligraphic letters, like $\mathcal{X}$, are used to denote finite sets (*e.g.* $\mathbb{F}_2^n$). The corresponding large letter X denotes a random variable over $\mathcal{X}$, while the lowercase letter x a value over $\mathcal{X}$. The probability of an event ev is denoted by $\mathrm{P}[ev]$. The expectation and the variance of a random variable X are respectively denoted by $\mathrm{E}[X]$ and $\mathrm{Var}[X]$. The covariance between two random variables X and Y is denoted by $\mathrm{Cov}[X, Y]$.

The Gaussian distribution of dimension T with T-size expectation vector $\boldsymbol{m}$ and $T \times T$ covariance matrix $\boldsymbol{\Sigma}$ is denoted by $\mathcal{N}(\boldsymbol{m}, \boldsymbol{\Sigma})$, and the corresponding probability density function (pdf) is denoted by $\phi_{\boldsymbol{m},\boldsymbol{\Sigma}}$. We recall that this pdf is defined for every $\boldsymbol{x} \in \mathbb{R}^T$ as

$$\phi_{\boldsymbol{m},\boldsymbol{\Sigma}}(\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^T |\boldsymbol{\Sigma}|}} \exp\left(-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{m})' \cdot \boldsymbol{\Sigma}^{-1} \cdot (\boldsymbol{x} - \boldsymbol{m})\right) , \tag{1}$$

where $(\boldsymbol{x} - \boldsymbol{m})'$ denotes the transpose of the vector $(\boldsymbol{x} - \boldsymbol{m})$ and $|\boldsymbol{\Sigma}|$ denotes the determinant of the matrix $\boldsymbol{\Sigma}$. The corresponding cumulative distribution function (cdf) is denoted $\Phi_{\boldsymbol{m},\boldsymbol{\Sigma}}$ and is defined for a pair of vectors $\boldsymbol{a} = (a_1, a_2, \ldots, a_T)$ and $\boldsymbol{b} = (b_1, b_2, \ldots, b_T)$ over $(\mathbb{R} \cup \{-\infty, +\infty\})^T$ by

$$\Phi_{\boldsymbol{m},\boldsymbol{\Sigma}}(\boldsymbol{a}, \boldsymbol{b}) = \int_{a_1}^{b_1} \int_{a_2}^{b_2} \cdots \int_{a_T}^{b_T} \phi_{\boldsymbol{m},\boldsymbol{\Sigma}}(\boldsymbol{x}) \; d\boldsymbol{x} \; . \tag{2}$$

If the dimension T equals 1, then the Gaussian distribution is said to be *univariate* and its covariance matrix is reduced to the variance of the single coordinate denoted σ^2. If T is greater than 1, the Gaussian distribution is said to be *multivariate*.

3 Higher-Order Side-Channel Model

We consider a cryptographic algorithm protected by *masking* and running on a leaking device. A (higher-order) side-channel attack exploits the leakage resulting from intermediate computations in order to recover (part of) the secret involved in the cryptographic algorithm. Let s denote such an intermediate variable satisfying:

$$s = \varphi(x, k^*) \; , \tag{3}$$

where x is (part of) the public input of the algorithm, k^* is (part of) the secret input of the algorithm, and φ is some function from $\mathcal{X} \times \mathcal{K}$ to $\mathcal{S}$.

For an implementation protected with masking, such a variable s is never stored nor handled in clear but in the form of several, say $d + 1$, *shares* s_0, s_1, $\ldots$, s_d satisfying the relation

$$s_0 \oplus s_1 \oplus \cdots \oplus s_d = s \tag{4}$$

for some operation $\oplus$. In the common case of Boolean masking this operation is the bitwise addition (or XOR), but it might be some other group addition law.

One of the share, say s_0, is sometimes referred to as *masked variable* and the other shares, s_1, s_2, ..., s_d as the *masks*. For masking approach to be sound, it is usually required that the masks are uniformly and independently generated. In that case, the $(d+1)$-tuple of shares can be modeled as a random vector $(S_0, S_1, \ldots, S_d)$ where $S_0 = s \oplus \bigoplus_{j=1}^{d} S_j$ and, for $j \geqslant 1$, the S_j are mutually independent random variables with uniform distribution over $\mathcal{S}$.

3.1 Leakage Model

During the execution of the algorithm, the processing of each share S_j produces some leakage L_j revealing some information about the share value. In what follows, we shall denote by $\boldsymbol{L}$ the leakage tuple:

$$\boldsymbol{L} = (L_0, L_1, \ldots, L_d) \ . \tag{5}$$

We shall sometimes use the alternative notation $\boldsymbol{L}_s$ or $\boldsymbol{L}_{x,k^*}$ to indicate that the leakage arises for the shared value $s = \varphi(x, k^*)$.

In this paper, we shall make the common assumption that given the values of the shares, the leakage has a Gaussian distribution. This assumption is referred here as the *Gaussian leakage assumption*, and it is formally stated by:

$$(L_j \mid S_j = s) \sim \mathcal{N}(\boldsymbol{m}_{j,s}, \boldsymbol{\Sigma}_{j,s}) \ , \tag{6}$$

for every $j \in \{0, 1, \ldots, d\}$ and for every $s \in \mathcal{S}$, where $\boldsymbol{m}_{j,s}$ are expectation vectors defined over $\mathbb{R}^T$ and $\boldsymbol{\Sigma}_{j,s}$ are (non-singular) covariance matrices defined over $\mathbb{R}^{T \times T}$. We shall further assume that the leakage L_j can be viewed as a deterministic function of S_j with an additive Gaussian noise:

$$L_j = f_j(S_j) + N_j \ . \tag{7}$$

This assumption, referred here as *Gaussian noise assumption*, is equivalent to the Gaussian leakage assumption with the additional requirement that the covariance matrices $\boldsymbol{\Sigma}_{j,s}$ are all equal to some matrix $\boldsymbol{\Sigma}_j$. We then have $f_j : s \mapsto \boldsymbol{m}_{j,s}$ and $N_j \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_j)$, where $\mathbf{0}$ denotes the null vector.

As a final assumption, we consider that for any fixed values of the shares, the leakage components are independent. That is, for every $(s_0, s_1, \ldots, s_d) \in \mathcal{S}^{d+1}$, the random variables $(L_j \mid S_j = s_j)$ are mutually independent. Under the Gaussian noise assumption, this simply means that the noises N_j are mutually independent, and that is why we shall refer to this assumption as the *independent noises assumption*.

Remark 1. For the sake of simplicity, we consider that all the leakages L_j have the same dimension T. Note that our analysis could be easily extended to the general case where each leakage L_j has its own dimension T_j.

3.2 Higher-Order Side-Channel Attacks

In a higher-order side-channel attack (HO-SCA), the adversary aims to extract information about k^* by monitoring the leakage of the shares. Specifically, the adversary observes several samples $\boldsymbol{\ell}_i \in \boldsymbol{\mathcal{L}}$ of the leakage $\boldsymbol{L}_{x_i,k^*}$, corresponding to some public input x_i that he may either choose or just know. According to the above leakage model, the leakage space $\boldsymbol{\mathcal{L}}$ is defined as $\boldsymbol{\mathcal{L}} = \mathbb{R}^{T\times(d+1)}$ and each leakage sample can be written as

$$\boldsymbol{\ell}_i = (\ell_{i,0}, \ell_{i,1}, \cdots, \ell_{i,d}) \ , \tag{8}$$

with $\ell_{i,j} \in \mathbb{R}^T$ for every j. Moreover, the Gaussian noise assumption implies that each leakage sample coordinate can be further written as

$$\ell_{i,j} = f_j(s_{i,j}) + n_{i,j} \ \ , \tag{9}$$

where $s_{i,1}$, $s_{i,2}$, $\ldots$, $s_{i,d}$ are d random mask values, where $s_{i,0} = \varphi(x_i, k^*) \oplus \bigoplus_{j=1}^{d} s_{i,j}$, and where $n_{i,0}$, $n_{i,1}$, $\ldots$, $n_{i,d}$ are samples of the Gaussian noises N_0, N_1, $\ldots$, N_d.

Once several, say q, leakage samples have been collected, the adversary makes use of a *distinguisher*, that is a function mapping the input-leakage samples $(x_1, \boldsymbol{\ell}_1)$, $(x_2, \boldsymbol{\ell}_2)$, $\ldots$, $(x_q, \boldsymbol{\ell}_q)$ to some *score vector* $\mathbf{d} = (d_k)_{k\in\mathcal{K}} \in \mathbb{R}^{|\mathcal{K}|}$. If the distinguisher is sound and if the leakage brings enough information on the shares, then the equality

$$k^* = \underset{k\in\mathcal{K}}{\operatorname{argmax}}\ d_k$$

should hold with a probability substantially greater than $\frac{1}{|\mathcal{K}|}$.

In what follows, we shall consider a natural equivalence relation between distinguishers. We say that two score vectors are *rank-equivalent* if for every $n \in \{1, 2, \ldots, |\mathcal{K}|\}$, the n coordinates with highest scores are the same for the two vectors. Two distinguishers d and d' are then said *equivalent*, denoted $\mathsf{d} \equiv \mathsf{d}'$ if for every $(x_i, \boldsymbol{\ell}_i)_i \in (\mathcal{X} \times \boldsymbol{\mathcal{L}})^q$, the score vectors $\mathsf{d}\big((x_i, \boldsymbol{\ell}_i)_i\big)$ and $\mathsf{d}'\big((x_i, \boldsymbol{\ell}_i)_i\big)$ are rank-equivalent.

In this paper, we focus on *additive distinguishers* which we formally define hereafter.

Definition 1. *A distinguisher* d *is* additive *if for every* $(x_1, x_2, \ldots, x_q) \in \mathcal{X}^q$*, there exists a family of functions* $\{g_{x,k} : \boldsymbol{\mathcal{L}} \to \mathbb{R}\ ;\ (x, k) \in \mathcal{X} \times \mathcal{K}\}$ *such that for every* $(\boldsymbol{\ell}_1, \boldsymbol{\ell}_2, \ldots, \boldsymbol{\ell}_q) \in \boldsymbol{\mathcal{L}}^q$ *we have*

$$\mathsf{d}\big((x_i, \boldsymbol{\ell}_i)_i\big) = (d_k)_{k\in\mathcal{K}} \quad \textit{with} \quad d_k = \frac{1}{q}\sum_{i=1}^{q} g_{x_i,k}(\boldsymbol{\ell}_i) \quad \textit{for every } k \in \mathcal{K}.$$

A distinguisher equivalent to an additive distinguisher as defined above is also said to be additive.

It was shown in [13] that the widely used first-order correlation and likelihood distinguishers are both additive distinguishers in the sense of the above definition. We will show in Sections 5 and 6 that their higher-order counterparts are also additive.

4 Estimating the Success Rate

In this section, we generalize the methodology introduced in [13] to HO-SCA as modelled in the previous section. Namely, we show how to get a sound estimation of the attack success rate by studying the multivariate probability distribution of the score vector for the case of additive distinguishers.

The success rate of a HODPA, denoted $\mathsf{Succ}^{\mathsf{d}}_{\boldsymbol{x},k^*}$, is defined with respect to some input vector $\boldsymbol{x} = (x_1, x_2, \ldots, x_q)$, some secret k^*, and some distinguisher d, as the probability:

$$\mathrm{P}\Big[k^* = \operatorname*{argmax}_{k\in\mathcal{K}} d_k \;\Big|\; \boldsymbol{\ell}_1 \xleftarrow{\$} \boldsymbol{L}_{x_1,k^*};\ \ldots;\ \boldsymbol{\ell}_q \xleftarrow{\$} \boldsymbol{L}_{x_q,k^*};\ (d_k)_{k\in\mathcal{K}} = \mathsf{d}\big((x_i, \boldsymbol{\ell}_i)_i\big)\Big] \ ,$$

where $\boldsymbol{\ell}_i \xleftarrow{\$} \boldsymbol{L}_{x_i,k^*}$ means randomly sampling $\boldsymbol{\ell}_i$ according to the distribution of $\boldsymbol{L}_{x_i,k^*}$.

Remark 2. For the sake of generality, we chose to fix the input vector $\boldsymbol{x}$ as a parameter of the attack so that we do not need to assume any specific strategy for the choice of the public inputs. However, we will investigate the particular setting where the x_i are uniformly distributed.

According to Definition 1, the score vector $(d_k)_{k\in\mathcal{K}}$ resulting from an additive distinguisher satisfies

$$d_k = \frac{1}{q}\sum_{i=1}^{q} g_{x_i,k}(\boldsymbol{\ell}_i) \ , \tag{10}$$

for some $g_{x,k} : \boldsymbol{\mathcal{L}} \to \mathbb{R}$. Then a simple application of the central limit theorem yields the following result, where we define the *occurrence ratio* τ_x of an element $x \in \mathcal{X}$ in the input vector $(x_1, x_2, \ldots, x_q)$ as

$$\tau_x = \frac{|\{i;\ x_i = x\}|}{q} \ . \tag{11}$$

Proposition 1. *The distribution of the score vector $(d_k)_{k\in\mathcal{K}}$ tends toward a multivariate Gaussian distribution as q grows, with expectation vector $(\mathrm{E}[d_k])_{k\in\mathcal{K}}$ satisfying*

$$\mathrm{E}[d_k] = \sum_{x\in\mathcal{X}} \tau_x \,\mathrm{E}[g_{x,k}(\boldsymbol{L}_{x,k^*})] \tag{12}$$

for every $k \in \mathcal{K}$, and with covariance matrix $(\mathrm{Cov}[d_{k_1}, d_{k_2}])_{(k_1,k_2)\in\mathcal{K}^2}$ satisfying

$$\mathrm{Cov}[d_{k_1}, d_{k_2}] = \frac{1}{q}\sum_{x\in\mathcal{X}} \tau_x \,\mathrm{Cov}[g_{x,k_1}(\boldsymbol{L}_{x,k^*}), g_{x,k_2}(\boldsymbol{L}_{x,k^*})] \tag{13}$$

for every $(k_1, k_2) \in \mathcal{K}^2$.

Proof. The first statement results by definition of additive distinguishers and the central limit theorem. Equations (12) and (13) directly holds by mutual independence between the leakage samples. □

The above proposition shows that for a sufficient number of leakage observations, the distribution of the score vector $\boldsymbol{d} = (d_k)_{k \in \mathcal{K}}$ can be soundly estimated by a multivariate Gaussian. As in [13], we now define the *comparison vector* as the $(|\mathcal{K}| - 1)$-size vector $\boldsymbol{c} = (c_k)_{k \in \mathcal{K}/\{k^*\}}$ whose coordinates satisfy

$$c_k = d_{k^*} - d_k \ , \tag{14}$$

for every $k \in \mathcal{K}/\{k^*\}$. The comparison vector is a linear transformation of the score vector by a $((|\mathcal{K}| - 1) \times |\mathcal{K}|)$-matrix P whose expression straightforwardly follows from (14). This implies that the distribution of the comparison vector can also be soundly estimated by a multivariate Gaussian distribution $\mathcal{N}(\boldsymbol{m_c}, \boldsymbol{\Sigma_c})$ where $\boldsymbol{m_c} = P \cdot \boldsymbol{m_d}$ and $\boldsymbol{\Sigma_c} = P \cdot \boldsymbol{\Sigma_d} \cdot P'$. Moreover, by definition of the comparison vector, an attack is successful (*i.e.* the correct secret k^* is ranked first in the score vector) if and only if all the coordinates of the comparison vector are positive. We deduce that the success rate $\mathsf{Succ}^{\mathsf{d}}_{\boldsymbol{x},k^*}$ of a distinguisher d satisfies

$$\mathsf{Succ}^{\mathsf{d}}_{\boldsymbol{x},k^*} \ = \ \mathrm{P}[\boldsymbol{c} > \boldsymbol{0}] \ \approx \ \Phi_{\boldsymbol{m_c},\boldsymbol{\Sigma_c}}(\boldsymbol{0}, \boldsymbol{\infty}) \tag{15}$$

where $\Phi_{\boldsymbol{m},\boldsymbol{\Sigma}}$ denotes the Gaussian cdf as defined in (2), $\boldsymbol{0}$ denotes the null vector, and $\boldsymbol{\infty}$ denotes the vector $(\infty, \infty, \ldots, \infty)$.

Remark 3. In [16], the authors propose to extend the notion of success rate to different orders. The o-th order success rate of a side-channel attack is defined as the probability that the target secret k^* is ranked among the o first key guesses by the score vector. The authors of [16] also suggest to consider the so-called *guessing entropy*, which is defined as the expected rank of the good key guess in the score vector [11,3]. As shown in [13], both the success rate of any order and the guessing entropy can be estimated using a similar approach as above.

Methodology. According to the above analysis, we propose the following methodology for an evaluator of some cryptographic algorithm to estimate the success rate of a HO-SCA against his masked implementation. We consider that the evaluator has access to the random masks generated during the computation, and is therefore able to predict the value of each share involved in the successive execution of the protected algorithm. The methodology is composed of three main steps:

1. Profile the leakage of every share using standard estimation techniques. Under the Gaussian leakage assumption, this estimation amounts to compute the sample means and the sample covariance matrices of the leakage $(L_i \mid S_i = s)$ for every share S_i and every possible value $s \in \mathcal{S}$ based on a set of collected leakage samples.
2. Use Proposition 1 to compute the expectation vector and covariance matrix of the score vector with respect to the leakage parameters.

3. Deduce the parameters of the comparison vector distribution and evaluate the success rate according to (15).

The precision of the obtained estimation is impacted by two main factors:

- the accuracy of the leakage parameter estimations, and
- the tightness of the Gaussian approximation arising in Proposition 1.

The accurate estimation of leakage parameters has been a widely investigated issue and efficient techniques are known to deal with it (see for instance [4,15,1,7]). Basically, the more noisy the leakage, the more samples must be used to get an accurate estimation. Note that in our approach, the evaluator only has to estimate first-order leakage parameters with respect to the share values. Practical aspects of leakage parameter estimation are further discussed in Section 8.

On the other hand, the Gaussian approximation is the main issue in our approach. One can fairly expect that if the considered implementation is not too weak, the convergence toward the Gaussian distribution should be rather fast compared to the number of leakage observations required to succeed the HO-SCA. In order to validate this intuition, we provide in Section 7 an empirical validation of the Gaussian approximation.

5 Application to the Correlation Distinguisher

In this section, we apply the general methodology described in Section 4 when the linear correlation coefficient is used as distinguisher [2]. For two samples $\boldsymbol{x} = (x_1, x_2, \ldots, x_q) \in \mathbb{R}^q$ and $\boldsymbol{y} = (y_1, y_2, \ldots, y_q) \in \mathbb{R}^q$, the linear coefficient is defined by

$$\rho(\boldsymbol{x}, \boldsymbol{y}) = \frac{\frac{1}{q}\sum_{i=1}^{q}(x_i - \overline{\boldsymbol{x}}) \cdot (y_i - \overline{\boldsymbol{y}})}{\sqrt{\frac{1}{q}\sum_i (x_i - \overline{\boldsymbol{x}})^2} \cdot \sqrt{\frac{1}{q}\sum_i (y_i - \overline{\boldsymbol{y}})^2}} \ , \tag{16}$$

where $\overline{\boldsymbol{x}}$ (resp. $\overline{\boldsymbol{y}}$) denotes the sample mean $q^{-1}\sum_i x_i$ (resp. $q^{-1}\sum_i y_i$).

In the context of HO-SCA, the correlation coefficient is used together with a *model function* $\mathtt{m} : \mathcal{X} \times \mathcal{K} \mapsto \mathbb{R}$ and a *combining function* $\mathtt{C} : \boldsymbol{\mathcal{L}} \mapsto \mathbb{R}$ (see for instance [12]). The combining function is involved to map a leakage sample into a univariate sample combining the leakages of the different shares. On the other hand, the model function computes some expected value for the combined leakage with respect to some input x and some guess k on the target secret. The correlation distinguisher $\mathsf{d}_{\mathsf{cor}}$ is then defined as

$$\mathsf{d}_{\mathsf{cor}}\big((x_i, \boldsymbol{\ell}_i)_i\big) = \rho\big((\mathtt{m}(x_i, k))_i, (\mathtt{C}(\boldsymbol{\ell}_i))_i\big) \ . \tag{17}$$

The following proposition extends the analysis conducted in [13] and states that the (higher-order) correlation distinguisher $\mathsf{d}_{\mathsf{cor}}$ is additive (see proof in appendix). This particularly implies that the methodology described in Section 4 can be applied to this distinguisher.

Proposition 2. *For any model function* $\mathtt{m} : \mathcal{X} \times \mathcal{K} \mapsto \mathbb{R}$ *and any combining function* $\mathtt{C} : \boldsymbol{\mathcal{L}} \mapsto \mathbb{R}$*, the correlation distinguisher* $\mathsf{d}_{\mathsf{cor}}$ *is additive. Moreover,* $\mathsf{d}_{\mathsf{cor}}$ *is equivalent to the distinguisher* $\mathsf{d}'_{\mathsf{cor}}$ *defined for every* $(x_i, \boldsymbol{\ell}_i)_i \in (\mathcal{X} \times \boldsymbol{\mathcal{L}})^q$ *by*

$$\mathsf{d}'_{\mathsf{cor}}\big((x_i, \boldsymbol{\ell}_i)_i\big) = \Big(\frac{1}{q}\sum_{i=1}^{q} g_{x_i,k}(\boldsymbol{\ell}_i)\Big)_{k\in\mathcal{K}} \ ,$$

where the function $g_{x,k} : \boldsymbol{\mathcal{L}} \to \mathbb{R}$ *satisfies*

$$g_{x,k}(\boldsymbol{\ell}) = \frac{1}{s_k}\left(\mathtt{m}(x,k) - \overline{\mathtt{m}}_k\right) \cdot \mathtt{C}(\boldsymbol{\ell}) \ , \tag{18}$$

for every $(x,k) \in \mathcal{X}\times\mathcal{K}$*, with* $\overline{\mathtt{m}}_k = \frac{1}{q}\sum_i \mathtt{m}(x_i,k)$ *and* $s_k = \sqrt{\frac{1}{q}\sum_i (\mathtt{m}(x_i,k) - \overline{\mathtt{m}}_k)^2}$.

Remark 4. If we focus on the uniform setting where the input vector $\boldsymbol{x} = (x_1, x_2, \ldots, x_q)$ is balanced (meaning that each value $x \in \mathcal{X}$ have an occurrence ratio of $\tau_x = \frac{1}{|\mathcal{X}|}$), then $\overline{\mathtt{m}}_k$ and s_k are constant with respect to k and $\mathsf{d}_{\mathsf{cor}}$ is equivalent to another simpler distinguisher:

$$\mathsf{d}''_{\mathsf{cor}} : \big((x_i, \boldsymbol{\ell}_i)_i\big) \mapsto \Big(\frac{1}{q}\sum\nolimits_i \mathtt{m}(x_i,k) \cdot \mathtt{C}(\boldsymbol{\ell_i})\Big)_{k\in\mathcal{K}} \ . \tag{19}$$

Application to the Normalized Product Combining. Let us now study the particular case of the higher-order correlation distinguisher based on the *centered product* combining function [12]. This combining function is defined for univariate share leakages (*i.e.* for $T = 1$ in the model of Section 3), namely its domain is $\boldsymbol{\mathcal{L}} = \mathbb{R}^{d+1}$. For every $(\ell_0, \ell_1, \ldots, \ell_d) \in \boldsymbol{\mathcal{L}}$, it is defined as

$$\mathtt{C}(\ell_0, \ell_1, \ldots, \ell_d) = \prod_{j=0}^{d} (\ell_j - \mu_j) \ , \tag{20}$$

where μ_j denotes the leakage expectation $\mathrm{E}\left[L_j\right]$.

Note that in practice, the adversary does not know the exact expectation μ_j but he can estimate it based on leakage samples. As argued in [12], the number of leakage samples required to succeed a HO-SCA is substantially greater than the number of leakage samples required to get precise estimations of the expectations μ_j. Therefore, we can soundly assume that the μ_j in (20) are the exact expectations $\mathrm{E}\left[L_j\right]$.

We recall that, according to the leakage model presented in Section 3.1, the jth leakage component L_j satisfies $L_j = f_j(S_j) + N_j$ where $f_j : s \mapsto m_{j,s}$ and $N_j \sim \mathcal{N}(0, \sigma_j^2)$. Since the noise N_j is centered in 0, we have $\mathrm{E}\left[f_j(S_j)\right] = \mathrm{E}\left[L_j\right] = \mu_j$. Moreover, we shall denote $\nu_j = \mathrm{Var}\left[f_j(S_j)\right]$. By uniformity of S_j over $\mathcal{S}$, we have:

$$\mu_j = \frac{1}{|\mathcal{S}|}\sum_{s\in\mathcal{S}} m_{j,s} \quad \text{and} \quad \nu_j = \frac{1}{|\mathcal{S}|}\sum_{s\in\mathcal{S}} (m_{j,s} - \mu_j)^2 \ . \tag{21}$$

In the following we shall further denote, for every $s \in \mathcal{S}$,

$$\alpha_s := \frac{1}{|\mathcal{S}|^d} \sum_{s_1 \in \mathcal{S}} \sum_{s_2 \in \mathcal{S}} \cdots \sum_{s_d \in \mathcal{S}} \prod_{j=0}^{d} (m_{j,s_j} - \mu_j) \tag{22}$$

and

$$\beta_s := \frac{1}{|\mathcal{S}|^d} \sum_{s_1 \in \mathcal{S}} \sum_{s_2 \in \mathcal{S}} \cdots \sum_{s_d \in \mathcal{S}} \prod_{j=0}^{d} (m_{j,s_j} - \mu_j)^2 \tag{23}$$

where $s_0 = s \oplus \bigoplus_{j=1}^{d} s_j$.

Note that both (22) and (23) can be expressed as a higher-order convolution product of the form

$$H(s) = \sum_{s_1} \sum_{s_2} \cdots \sum_{s_d} h_0(s \oplus s_1 \oplus s_2 \oplus \cdots \oplus s_d) \cdot h_1(s_1) \cdot h_2(s_2) \cdots h_d(s_d) \ . \tag{24}$$

We show in appendix how such a convolution can be efficiently computed for all values over $\mathcal{S}$ in $O(d \cdot |\mathcal{S}| \cdot \log |\mathcal{S}|)$ operations.

We then have the following corollary of Proposition 1 for the distinguisher $\mathsf{d}'_{\mathsf{cor}}$ with centered product combining function (see proof in appendix).

Corollary 1. *Let $k^* \in \mathcal{K}$, let $(x_1, x_2, \ldots, x_q) \in \mathcal{X}^q$ and let $\boldsymbol{\ell}_i \xleftarrow{\$} \boldsymbol{L}_{x_i,k^*}$ for every $i \in \{1, 2, \ldots, q\}$. Then the distribution of the score vector $(d'_k)_{k \in \mathcal{K}} = \mathsf{d}'_{\mathsf{cor}}\big((x_i, \boldsymbol{\ell}_i)_i\big)$ with centered product combining function tends toward a multivariate Gaussian distribution with expectation vector $(\mathrm{E}\,[d'_k])_{k \in \mathcal{K}}$ satisfying*

$$\mathrm{E}\,[d'_k] = \sum_{x \in \mathcal{X}} \tau_x \, \mathtt{M}(x, k) \, \alpha_{\varphi(x,k^*)} \ , \tag{25}$$

for every $k \in \mathcal{K}$, and with covariance matrix $(\mathrm{Cov}\left[d'_{k_1}, d'_{k_2}\right])_{(k_1,k_2) \in \mathcal{K}^2}$ satisfying

$$\begin{aligned} \mathrm{Cov}\left[d'_{k_1}, d'_{k_2}\right] = \frac{1}{q} \sum_{x \in \mathcal{X}} \tau_x \, \mathtt{M}(x, k_1) \, \mathtt{M}(x, k_2) \\ \times \left(\beta_{\varphi(x,k^*)} - \alpha^2_{\varphi(x,k^*)} + \prod_{j=0}^{d} (\nu_j + \sigma_j^2) - \prod_{j=0}^{d} \nu_j \right) \ , \end{aligned} \tag{26}$$

for every $(k_1, k_2) \in \mathcal{K}^2$, where

$$\mathtt{M} : (x, k) \mapsto \frac{\mathtt{m}(x, k) - \overline{\mathtt{m}}_k}{s_k} \ . \tag{27}$$

Remark 5. For the distinguisher $\mathsf{d}''_{\mathsf{cor}}$ defined in (19) and which is equivalent to the correlation distinguisher in the uniform setting (see Remark 4), we have the same result as in Corollary 1 but the function $\mathtt{M}$ is simply defined as the model function $\mathtt{m}$.

According to Corollary 1, the methodology presented in Section 4 can be applied to estimate the success rate of a HO-SCA based on the correlation distinguisher with centered product combining. The first step of the methodology shall provide estimations of the leakage functions $f_j : s \mapsto m_{j,s}$ (and hence of the corresponding μ_j and ν_j), while the second step shall simply consist in the evaluations of Formulae (25) and (26).

6 Application to the Likelihood Distinguisher

In this section, we apply the general methodology described in Section 4 when the likelihood is used as distinguisher [4]. The likelihood distinguisher, denoted $\mathsf{d}_{\mathsf{lik}}$, is usually applied after a *profiling step* whose goal is to provide an estimation $\hat{\mathsf{p}}_s$ of the pdf of the random variable $\boldsymbol{L}_s$ for every $s \in \mathcal{S}$. Then, for every sample $(x_i, \boldsymbol{\ell_i})_i \in (\mathcal{X} \times \boldsymbol{\mathcal{L}})^q$, the likelihood distinguisher is defined as

$$\mathsf{d}_{\mathsf{lik}}\big((x_i, \boldsymbol{\ell_i})_i\big) = \prod_{i=1}^{q} \hat{\mathsf{p}}_{\varphi(x_i,k)}(\boldsymbol{\ell_i}) \ . \tag{28}$$

In practice, one often makes use of the equivalent (averaged) log-likelihood distinguisher $\mathsf{d}'_{\mathsf{lik}}$ defined as

$$\mathsf{d}'_{\mathsf{lik}}\big((x_i, \boldsymbol{\ell_i})_i\big) = \frac{1}{q} \log \mathsf{d}_{\mathsf{lik}}\big((x_i, \boldsymbol{\ell_i})_i\big) = \frac{1}{q} \sum_{i=1}^{q} \log(\hat{\mathsf{p}}_{\varphi(x_i,k)}(\boldsymbol{\ell_i})) \ . \tag{29}$$

The log-likelihood distinguisher is usually preferred as it less susceptible to approximation errors than the likelihood. We straightforwardly get the following proposition.

Proposition 3. *The likelihood distinguisher* $\mathsf{d}_{\mathsf{lik}}$ *is additive and equivalent to the log-likelihood distinguisher* $\mathsf{d}'_{\mathsf{lik}}$*. Moreover, for every* $(x_i, \boldsymbol{\ell}_i)_i \in (\mathcal{X} \times \boldsymbol{\mathcal{L}})^q$*,* $\mathsf{d}'_{\mathsf{lik}}$ *satisfies*

$$\mathsf{d}'_{\mathsf{lik}}\big((x_i, \boldsymbol{\ell_i})_i\big) = \Big(\frac{1}{q} \sum_{i=1}^{q} g_{x_i,k}(\boldsymbol{\ell_i})\Big)_{k \in \mathcal{K}} \ , \tag{30}$$

where the function $g_{x,k} : \boldsymbol{\mathcal{L}} \to \mathbb{R}$ *satisfies*

$$g_{x,k}(\boldsymbol{\ell}) = \log(\hat{\mathsf{p}}_{\varphi(x,k)}(\boldsymbol{\ell})) \ , \tag{31}$$

for every $(x, k) \in \mathcal{X} \times \mathcal{K}$*.*

Under the Gaussian leakage assumption, it can be checked that the variable $\boldsymbol{L}_s$ has a Gaussian mixture distribution, with pdf p_s satisfying

$$\mathsf{p}_s : (\ell_0, \ell_1, \ldots, \ell_d) \mapsto \frac{1}{|\mathcal{S}|^d} \sum_{s_1 \in \mathcal{S}} \sum_{s_2 \in \mathcal{S}} \cdots \sum_{s_d \in \mathcal{S}} \prod_{j=0}^{d} \phi_{\boldsymbol{m}_{j,s_j}, \boldsymbol{\Sigma}_j}(\ell_j) \ , \tag{32}$$

where $s_0 = s \oplus \bigoplus_{j=1}^{d} s_j$. Note that for every $s \in \mathcal{S}$, the estimated pdf $\hat{\mathsf{p}}_s$ obtained from the profiling phase has a similar expression as p_s but with estimations $\hat{\boldsymbol{m}}_{j,s_j}$ and $\hat{\boldsymbol{\Sigma}}_j$ for the leakage means and covariance matrices.

Here again, it can be seen from (32) that for a given $\boldsymbol{\ell} \in \boldsymbol{\mathcal{L}}$ the probability $\mathsf{p}_s(\boldsymbol{\ell})$ is a higher-order convolution product as in (24). The set of probability values $\{\mathsf{p}_s(\boldsymbol{\ell})\ ;\ s \in \mathcal{S}\}$ can then be computed in $O(d \cdot |\mathcal{S}| \cdot \log |\mathcal{S}|)$ operations (see details in appendix).

Let us now consider the two functions:

$$\lambda(s_1, s_2) := \int_{\boldsymbol{\ell} \in \boldsymbol{\mathcal{L}}} \log(\hat{\mathsf{p}}_{s_1}(\boldsymbol{\ell}))\ \mathsf{p}_{s_2}(\boldsymbol{\ell})\ d\boldsymbol{\ell}\ , \tag{33}$$

and

$$\psi(s_1, s_2, s_3) := \int_{\boldsymbol{\ell} \in \boldsymbol{\mathcal{L}}} \log(\hat{\mathsf{p}}_{s_1}(\boldsymbol{\ell}))\ \log(\hat{\mathsf{p}}_{s_2}(\boldsymbol{\ell}))\ \mathsf{p}_{s_3}(\boldsymbol{\ell})\ d\boldsymbol{\ell}\ . \tag{34}$$

Then, by definition, we have

$$\Lambda(x, k, k^*) := \lambda(\varphi(x, k), \varphi(x, k^*)) = \mathrm{E}\left[g_{x,k}(\boldsymbol{L}_{x,k^*})\right]$$

and

$$\begin{aligned}\Psi(x, k_1, k_2, k^*) &:= \psi(\varphi(x, k_1), \varphi(x, k_2), \varphi(x, k^*)) \\ &= \mathrm{E}\left[g_{x,k_1}(\boldsymbol{L}_{x,k^*}) \cdot g_{x,k_2}(\boldsymbol{L}_{x,k^*})\right]\ .\end{aligned}$$

A direct application of Proposition 1 then yields the following corollary for the log-likelihood distinguisher.

Corollary 2. *Let $k^* \in \mathcal{K}$, let $(x_1, x_2, \ldots, x_q) \in \mathcal{X}^q$ and let $\boldsymbol{\ell}_i \xleftarrow{\$} \boldsymbol{L}_{x_i,k^*}$ for every $i \in \{1, 2, \ldots, q\}$. Then the distribution of the score vector $(d'_k)_{k \in \mathcal{K}} = \mathsf{d}'_{\mathsf{lik}}\big((x_i, \boldsymbol{\ell}_i)_i\big)$ tends toward a multivariate Gaussian distribution with expectation vector $(\mathrm{E}\left[d'_k\right])_{k \in \mathcal{K}}$ satisfying*

$$\mathrm{E}\left[d'_k\right] = \sum_{x \in \mathcal{X}} \tau_x\ \Lambda(x, k, k^*)\ , \tag{35}$$

for every $k \in \mathcal{K}$, and with covariance matrix $(\mathrm{Cov}\left[d'_{k_1}, d'_{k_2}\right])_{(k_1,k_2) \in \mathcal{K}^2}$ satisfying

$$\mathrm{Cov}\left[d'_{k_1}, d'_{k_2}\right] = \frac{1}{q} \sum_{x \in \mathcal{X}} \tau_x \big(\Psi(x, k_1, k_2, k^*) - \Lambda(x, k_1, k^*) \cdot \Lambda(x, k_2, k^*)\big)\ . \tag{36}$$

According to Corollary 2, the methodology presented in Section 4 can be applied to estimate the success rate of a HO-SCA based on the likelihood distinguisher.

7 Empirical Validation of the Gaussian Approximation

In Section 4, we have presented a methodology to estimate the success rate of higher-order side-channel attacks based on so-called additive distinguishers. The principle of this methodology is to approximate the distribution of the score vector by a multivariate Gaussian distribution whose parameters are derived from the leakage parameters. This Gaussian approximation is asymptotically sound by the central limit theorem. However, in the non-asymptotic context of a HO-SCA with a given number of leakage samples, it is fair to question whether this approximation is sound or not. In this section, we conduct an empirical study of the Gaussian approximation. For this purpose, we compare the success rates obtained from attack simulations, to the success rates obtained by applying the methodology of Section 4.

Since our purpose here is the sole validation of the Gaussian approximation, we do not focus on the leakage estimation issue, but we assume the exact leakage parameters $\{(m_{j,s}, \sigma_j^2)\ ;\ 0 \leqslant j \leqslant d,\ s \in \mathcal{S}\}$ are known (in a univariate setting). From these leakage parameters, and for a given HO-SCA based on some distinguisher $\mathsf{d} \in \{\mathsf{d}_{\mathsf{cor}}, \mathsf{d}_{\mathsf{lik}}\}$, we evaluate the success rate with the two following approaches:

- **Simulation success rate.** We perform several attack simulations and count the number of successes in order to get an estimation of the success rate. For each attack simulation, we randomly generate input-leakage samples $(x_1, \boldsymbol{\ell}_1)$, $(x_2, \boldsymbol{\ell}_2)$, ..., $(x_q, \boldsymbol{\ell}_q)$. Specifically, for every i, x_i is uniformly picked up and $\boldsymbol{\ell}_i$ is randomly sampled from the variable $\boldsymbol{L}_{x_i,k^*}$ according to the leakage parameters. Then we apply the distinguisher d to these samples, and we count a success whenever the good secret is ranked first.
- **Gaussian cdf evaluation.** We apply Corollaries 1 and 2 to compute the expectation vector and covariance matrix of the score vector with respect to the leakage parameters and taking $\tau_x = 1/|\mathcal{X}|$ as occurrence ratio for every $x \in \mathcal{X}$ (in accordance to the uniform distribution of the x_i). Then we compute the Gaussian cdf of the comparison vector to evaluate the success rate according to (15).

We plot hereafter the results obtained with these two approaches for different HO-SCA targeting an AES Sbox output:

$$\varphi(x, k^*) = \mathsf{SB}(x \oplus k^*)\ ,$$

where SB denote the AES Sbox function. For the leakage parameters, we used sample means and sample variances obtained by monitoring the leakage of two different devices running masked AES implementations (Device A and Device B, see Section 8 for details).

Figure 1 shows the results obtained for a second-order correlation attack with centered product combining function and Hamming weight model function (*i.e.* $\mathtt{m} = \mathtt{HW}$), for leakage parameters from Device A. Figure 2 plots the results of a second-order likelihood attack with the same leakage parameters, assuming

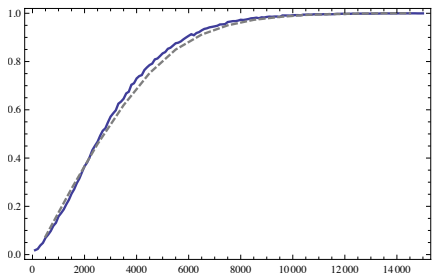

Fig. 1. Simulation SR (plain curve) *vs.* theoretical SR (dashed curve) for 2nd-order correlation attack.

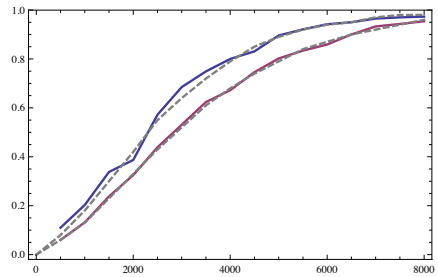

Fig. 2. Simulation SR (plain curves) *vs.* theoretical SR (dashed curves) for 2nd-order likelihood attacks.

a perfect profiling (*i.e.* $\hat{\mathsf{p}}_s = \mathsf{p}_s$ for every s) on the one hand and a slightly erroneous profiling on the other hand.[3] We observe that for both distinguishers, the experimental success rate curves and theoretical success rate curves clearly match. This validates the Gaussian approximation in these HO-SCA contexts.

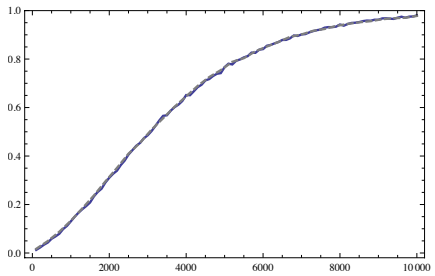

Fig. 3. Simulation SR (plain curve) *vs.* theoretical SR (dashed curve) for 3rd-order correlation attack.

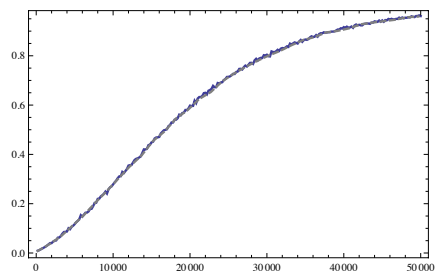

Fig. 4. Simulation SR (plain curve) *vs.* theoretical SR (dashed curve) for 4th-order correlation attack.

In order to test the Gaussian approximation to higher orders, we also performed third-order and fourth-order attacks, with leakage parameters from Device B. The results of the correlation attacks (centered product combining function and Hamming weight model function) are presented in Figure 3 and Figure 4 respectively. The figures for the higher-order likelihood attacks are provided in the full version of the paper. We see that the curves perfectly match, which further validates the Gaussian approximation in these higher-order contexts.

[3] Specifically, we introduce random errors in the $(m_{j,s})_{j,s}$ used in the estimated pdfs $\hat{\mathsf{p}}_s$.

8 Practical Experiments

In this section, we confront our methodology to practical attack experiments. We report the results of several higher-order correlation attacks against two different devices running masked AES implementations. We also apply our methodology to estimate the expected success rate of these attacks with respect to the inferred leakage parameters.

Experimental setup. Practical experiments were performed on two microcontrollers made in different CMOS technologies (130 and 350 nanometer processes, respectively called devices A and device B in the sequel). The side-channel traces were obtained by measuring the electromagnetic (EM) radiations emitted by the device during a masked AES-128 encryption handling one byte at a time. To this aim, an EM sensor was used (made of several coils of copper with diameter of 500μm), and was plugged into a low-noise amplifier. To sample the leakage measurements, a digital oscilloscope was used with a sampling rate of 10G samples per second for the device A and 2G samples per second for the device B, whereas microcontrollers were running at few dozen of MHz. As the microcontrollers clocks were not stable, we had to resynchronize the EM traces. This process is out of the scope of this work, but we would like to emphasize that resynchronization is always required in a practical context and it has a non negligible impact on the measurements noise.

In our attack context, the random values involved in the masking/sharing could be known by the evaluator and we used this ability to identify the time samples corresponding to the different manipulation of the different shares. This step allowed us to associate each share to a unique time sample (the one with maximal SNR) and to profile the leakage parameters.[4]

Estimation of the leakage parameters. To estimate the leakage functions $f_j : s \mapsto m_{j,s}$, we applied linear regression techniques on 200000 leakage samples. When applied on leakage samples $\ell_{1,j}, \ell_{2,j}, \ldots, \ell_{q,j}$, corresponding to successive share values $s_{1,j}, s_{2,j}, \ldots, s_{q,j}$, a linear regression *of degree* t returns an approximation of $f_j(s)$ as a degree-t polynomial in the bits of s (see [15,5] for more detail on linear regression in the context of side-channel attacks). We applied linear regression of degree 1 and 2 on Device A and B respectively. Once the f_j function estimated, we could easily get an estimation for the variance σ_j^2 of the noise N_j by computing the sample variance of $(\ell_{i,j} - f_j(s_{i,j}))_i$ for every j.

Methodology *versus* practice. In order to validate our methodology in practice, we performed higher-order correlation attacks with centered product combining function (see Section 5) and Hamming weight model function (*i.e.* $\mathtt{m} = \mathtt{HW}$). On the other hand, the success rate was estimated using the methodology described in Sections 4 and 5 by computing the parameters of the multivariate Gaussian distribution arising for the correlation distinguisher with respect to the inferred leakage parameters.

[4] The knowledge of the masks was however not used in the attack phase itself.

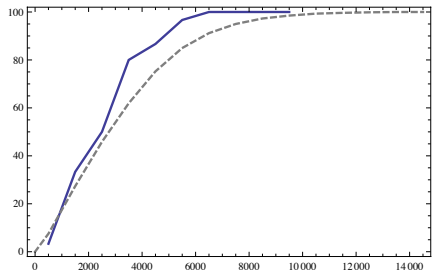

Fig. 5. Experimental SR (plain curve) *vs.* theoretical SR (dashed curve) for 2nd-order correlation attack on Device A.

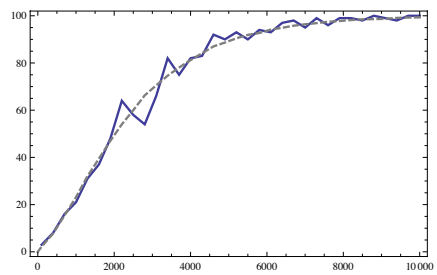

Fig. 6. Experimental SR (plain curve) *vs.* theoretical SR (dashed curve) for 2nd-order correlation attack on Device B.

Figures 5 and 6 plot the experimental success rates *versus* the theoretical success rates for the second-order correlation attacks against Device A and Device B. In order to validate our approach with respect to higher-order attacks in practice, we also compare the results obtained with our methodology to third-order and fourth-order attack results on Device B (see Figures 7 and 8). We observe a clear match between the experimental and theoretical success rate curves. These results demonstrate the soundness of the methodology in practice.

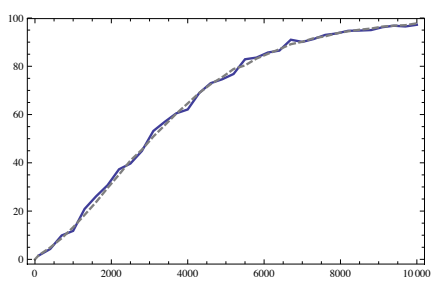

Fig. 7. Experimental SR (plain curve) *vs.* theoretical SR (dashed curve) for 3rd-order correlation attack on Device B.

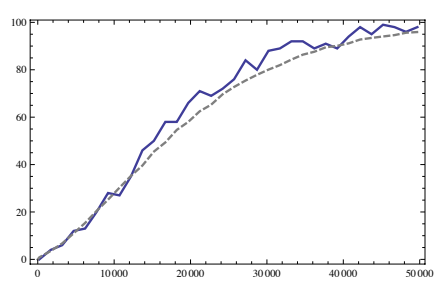

Fig. 8. Experimental SR (plain curve) *vs.* theoretical SR (dashed curve) for 4th-order correlation attack on Device B.

Further experiments are provided in the full version of the paper in order to observe the impact of the leakage profiling phase on our methodology.

9 Conclusion

In this work we have presented a methodology to evaluate the success rate of higher-order side-channel attacks. We have shown how to apply this methodology in the particular cases of attacks based on the correlation and likelihood distinguishers. The soundness of our approach has been validated by simulations and experiments performed on different microcontrollers. Using this methodology, an evaluator can estimate the side-channel resistance of his masked cryptographic implementation at the cost of inferring a few linear regression coefficients.

References

1. C. Archambeau, E. Peeters, F.-X. Standaert, and J.-J. Quisquater. Template Attacks in Principal Subspaces. In Goubin and Matsui [8], pages 1–14.
2. E. Brier, C. Clavier, and F. Olivier. Correlation Power Analysis with a Leakage Model. In M. Joye and J.-J. Quisquater, editors, *Cryptographic Hardware and Embedded Systems – CHES 2004*, volume 3156 of *Lecture Notes in Computer Science*, pages 16–29. Springer, 2004.
3. C. Cachin. *Entropy Measures and Unconditional Security in Cryptography*. PhD thesis, 1997.
4. S. Chari, J. Rao, and P. Rohatgi. Template Attacks. In B. Kaliski Jr., Ç. Koç, and C. Paar, editors, *Cryptographic Hardware and Embedded Systems – CHES 2002*, volume 2523 of *Lecture Notes in Computer Science*, pages 13–29. Springer, 2002.
5. J. Doget, E. Prouff, M. Rivain, and F.-X. Standaert. Univariate Side Channel Attacks and Leakage Modeling. *Journal of Cryptographic Engineering*, 1(2):123–144, 2011.
6. Y. Fei, Q. Luo, and A. A. Ding. A Statistical Model for DPA with Novel Algorithmic Confusion Analysis. In E. Prouff and P. Schaumont, editors, *CHES*, volume 7428 of *Lecture Notes in Computer Science*, pages 233–250. Springer, 2012.
7. B. Gierlichs, K. Lemke-Rust, and C. Paar. Templates vs. Stochastic Methods. In Goubin and Matsui [8], pages 15–29.
8. L. Goubin and M. Matsui, editors. *Cryptographic Hardware and Embedded Systems – CHES 2006*, volume 4249 of *Lecture Notes in Computer Science*. Springer, 2006.
9. P. Kocher, J. Jaffe, and B. Jun. Differential Power Analysis. In M. Wiener, editor, *Advances in Cryptology – CRYPTO '99*, volume 1666 of *Lecture Notes in Computer Science*, pages 388–397. Springer, 1999.
10. S. Mangard. Hardware Countermeasures against DPA – A Statistical Analysis of Their Effectiveness. In T. Okamoto, editor, *Topics in Cryptology – CT-RSA 2004*, volume 2964 of *Lecture Notes in Computer Science*, pages 222–235. Springer, 2004.
11. J. Massey. Guessing and Entropy. *IEEE International Symposium on Information Theory*, page 204, 1994.
12. E. Prouff, M. Rivain, and R. Bévan. Statistical Analysis of Second Order Differential Power Analysis. *IEEE Transactions on Computers*, 58(6):799–811, 2009.
13. M. Rivain. On the Exact Success Rate of Side Channel Analysis in the Gaussian Model. In R. M. Avanzi, L. Keliher, and F. Sica, editors, *Selected Areas in Cryptography – SAC 2008*, volume 5381 of *Lecture Notes in Computer Science*, pages 165–183. Springer, 2008.

14. M. Rivain, E. Prouff, and J. Doget. Higher-order Masking and Shuffling for Software Implementations of Block Ciphers. Cryptology ePrint Archive, 2009. http://eprint.iacr.org/.
15. W. Schindler, K. Lemke, and C. Paar. A Stochastic Model for Differential Side Channel Cryptanalysis. In J. Rao and B. Sunar, editors, *Cryptographic Hardware and Embedded Systems – CHES 2005*, volume 3659 of *Lecture Notes in Computer Science*. Springer, 2005.
16. F.-X. Standaert, T. G. Malkin, and M. Yung. A Formal Practice-Oriented Model For The Analysis of Side-Channel Attacks. Cryptology ePrint Archive, Report 2006/139, 2006.
17. F.-X. Standaert, E. Peeters, G. Rouvroy, and J.-J. Quisquater. An Overview of Power Analysis Attacks Against Field Programmable Gate Arrays. *IEEE*, 94(2):383–394, 2006.
18. A. Thillard, E. Prouff, and T. Roche. Success through Confidence: Evaluating the Effectiveness of a Side-Channel Attack. In G. Bertoni and J.-S. Coron, editors, *CHES*, volume 8086 of *Lecture Notes in Computer Science*, pages 21–36. Springer, 2013.

A Proof of Proposition 2

Proof. Let $(d_k)_{k\in\mathcal{K}} = \mathsf{d}_{\mathsf{cor}}\big((x_i,\boldsymbol{\ell}_i)_i\big)$ and $(d'_k)_{k\in\mathcal{K}} = \mathsf{d}'_{\mathsf{cor}}\big((x_i,\boldsymbol{\ell}_i)_i\big)$ for some input-leakage samples $(x_i,\boldsymbol{\ell_i})_{i\leqslant q} \in (\mathcal{X}\times\mathcal{L})^q$. We have:

$$d_k \;=\; \frac{1}{s_{\mathtt{C}}}\,\frac{\sum_{i=1}^{q}(\mathtt{m}(x_i,k)-\overline{\mathtt{m}}_k)\mathtt{C}(\boldsymbol{\ell_i})}{s_k} \;=\; \frac{1}{s_{\mathtt{C}}}\,d'_k \;,$$

where $s_{\mathtt{C}} = \sqrt{\frac{1}{q}\sum_i(\mathtt{C}(\boldsymbol{\ell_i})-\overline{\mathtt{C}})^2}$ with $\overline{\mathtt{C}} = \frac{1}{q}\sum_i \mathtt{C}(\boldsymbol{\ell_i})$.

Since $s_{\mathtt{C}}$ is strictly positive and constant with respect to the guess k, the score vectors $(d_k)_{k\in\mathcal{K}}$ and $(d'_k)_{k\in\mathcal{K}}$ are clearly rank-equivalent, implying that the distinguishers $\mathsf{d}_{\mathsf{cor}}$ and $\mathsf{d}'_{\mathsf{cor}}$ are equivalent. Moreover, after denoting by $g_{x,k}$ the function $\boldsymbol{\ell_i} \mapsto s_k^{-1}(\mathtt{m}(x,k)-\overline{\mathtt{m}}_k)\mathtt{C}(\boldsymbol{\ell_i})$, we get $d'_k = \frac{1}{q}\sum_{i=1}^{q} g_{x_i,k}(\boldsymbol{\ell_i})$, which implies that $\mathsf{d}'_{\mathsf{cor}}$ is additive. □

B Fast Evaluation of Higher-Order Convolution

Proposition 4. *Let d be a positive integer, and let $(\mathcal{S},\oplus)$ be a group of size $|\mathcal{S}| = 2^m$. Let $(h_j)_{0\leq j\leq d}$ be a family of functions from $\mathcal{S}$ into $\mathbb{R}$, such that $h_j(s)$ can be efficiently evaluated for every $s\in\mathcal{S}$ in $o(1)$ operations (one typically has a look-up table for every h_j). Consider the function $H:\mathcal{S}\to\mathbb{R}$ defined as*

$$H : s \mapsto \sum_{s_1\in\mathcal{S}}\sum_{s_2\in\mathcal{S}}\cdots\sum_{s_d\in\mathcal{S}} h_0(s\oplus s_1\oplus s_2\oplus\cdots\oplus s_d)\cdot h_1(s_1)\cdot h_2(s_2)\cdots h_d(s_d)\;.$$

Then, the whole set of outputs $\{H(s)\;;\; s\in\mathcal{S}\}$ can be computed in $O(d\cdot 2^m\cdot m)$ operations.

Proof. For every $s \in \mathcal{S}$, the function H satisfies

$$H(s) = \sum_{s_d \in \mathcal{S}} h_d(s_d) \cdots \sum_{s_2 \in \mathcal{S}} h_2(s_2) \sum_{s_1 \in \mathcal{S}} h_1(s_1) \cdot h_0(s \oplus s_1 \oplus s_2 \oplus \cdots s_d) \ .$$

Consider the convolution product of the form

$$h_1 \otimes h_0 : s \mapsto \sum_{t \in \mathcal{S}} h_1(t) \cdot h_0(s \oplus t) \ .$$

We have

$$\mathcal{WH}(h_1 \otimes h_0) = 2^{\frac{m}{2}} \ \mathcal{WH}(h_1) \cdot \mathcal{WH}(h_0) \ ,$$

where $\mathcal{WH}$ is the (normalized) Walsh-Hadamard transform (WHT). This convolution product can hence be efficiently computed from three evaluations of fast WHT that each takes $O(2^m \cdot m)$ operations.[5]

One can check that the sequence of functions $(H_i)_{0 \leq i \leq d}$ defined as

$$\begin{cases} H_0 = h_0 \\ H_i = h_i \otimes H_{i-1} \quad \text{for every } i \geqslant 1 \end{cases}$$

is such that $H_d = H$. One can then sequentially compute the set of outputs of H_1, H_2, $\ldots$, $H_d = H$ by evaluating d convolution products, which gives a total cost of $O(d \cdot 2^m \cdot m)$ operations. □

C Proof of Corollary 1

To prove the corollary, we first introduce the following lemma.

Lemma 1. *The expectation and variance of the random variable* $\mathtt{C}(\boldsymbol{L}_{x,k^*})$ *respectively satisfy*

$$\mathrm{E}\left[\mathtt{C}(\boldsymbol{L}_{x,k^*})\right] = \alpha_{\varphi(x,k^*)} \tag{37}$$

and

$$\mathrm{Var}\left[\mathtt{C}(\boldsymbol{L}_{x,k^*})\right] = \beta_{\varphi(x,k^*)} - \alpha^2_{\varphi(x,k^*)} + \prod_{j=0}^{d}(\nu_j + \sigma_j^2) - \prod_{j=0}^{d} \nu_j \ . \tag{38}$$

Proof. Since the N_j are independent and centered in 0, we have

$$\mathrm{E}\left[\mathtt{C}(\boldsymbol{L}_{x,k^*})\right] = \mathrm{E}\left[\mathtt{C}\big(f_0(S_0), f_1(S_1), \ldots, f_d(S_d)\big)^2\right] = \alpha_{\varphi(x,k^*)} \ ,$$

On the other hand, by definition of the variance, we have

$$\mathrm{Var}\left[\mathtt{C}(\boldsymbol{L}_{x,k^*})\right] = \mathrm{E}\left[\mathtt{C}(\boldsymbol{L}_{x,k^*})^2\right] - \mathrm{E}\left[\mathtt{C}(\boldsymbol{L}_{x,k^*})\right]^2 = \mathrm{E}\left[\mathtt{C}(\boldsymbol{L}_{x,k^*})^2\right] - \alpha^2_{\varphi(x,k^*)} \ .$$

[5] The WHT is involutive, hence we have $h_1 \otimes h_0 = 2^{\frac{m}{2}} \ \mathcal{WH}\big(\mathcal{WH}(h_1) \cdot \mathcal{WH}(h_0)\big)$.

Then, we have

$$\mathrm{E}\left[\mathtt{C}(\boldsymbol{L}_{x,k^*})^2\right] = \mathrm{E}\left[\prod_{j=0}^{d}\left(f_j(S_j)+N_j-\mu_j\right)^2\right] = \mathrm{E}\left[\prod_{j=0}^{d}\left((f_j(S_j)-\mu_j)^2+N_j^2\right)\right]$$

where the second holds since the N_j have zero means and are mutually independent and independent of the S_j. By developing the product, we get a sum of monomials, such that each monomial involves random variables that are mutually independent, except for one single monomial which is $\prod_{j=0}^{d}(f_j(S_j)-\mu_j)^2$. We can then develop the above equation as

$$\begin{aligned}\mathrm{E}\left[\mathtt{C}(\boldsymbol{L}_{x,k^*})^2\right] = \prod_{j=0}^{d}\left(\mathrm{E}\left[(f_j(S_j)-\mu_j)^2\right]+\mathrm{E}\left[N_j^2\right]\right) \\ -\prod_{j=0}^{d}\mathrm{E}\left[(f_j(S_j)-\mu_j)^2\right]+\mathrm{E}\left[\prod_{j=0}^{d}(f_j(S_j)-\mu_j)^2\right] \quad ,\end{aligned}$$

which gives

$$\mathrm{E}\left[\mathtt{C}(\boldsymbol{L}_{x,k^*})^2\right] = \prod_{j=0}^{d}(\nu_j+\sigma_j^2)-\prod_{j=0}^{d}\nu_j+\beta_{\varphi(x,k^*)}.$$

□

Proof of Corollary 1. Applying (12) and (13) to the functions $g_{x,k} : \boldsymbol{\ell} \mapsto \frac{1}{s_k}\left(\mathtt{m}(x,k)-\overline{\mathtt{m}}_k\right)\cdot\mathtt{C}(\boldsymbol{\ell})$ as defined in (18), we get

$$\mathrm{E}\left[d'_k\right] = \frac{1}{s_k}\sum_{x\in\mathcal{X}}\tau_x\left(\mathtt{m}(x,k)-\overline{\mathtt{m}}_k\right)\mathrm{E}\left[\mathtt{C}(\boldsymbol{L}_{x,k^*})\right] \quad ,$$

and

$$\mathrm{Cov}\left[d'_{k_1},d'_{k_2}\right] = \frac{1}{q}\frac{1}{s_{k_1}s_{k_2}}\sum_{x\in\mathcal{X}}\tau_x\left(\mathtt{m}(x,k_1)-\overline{\mathtt{m}}_{k_1}\right)\left(\mathtt{m}(x,k_2)-\overline{\mathtt{m}}_{k_2}\right)\mathrm{Var}\left[\mathtt{C}(\boldsymbol{L}_{x,k^*})\right] \quad ,$$

Then Lemma 1 directly yields the corollary statement. □