Masks will Fall Off
Higher-Order Optimal Distinguishers

Nicolas Bruneau[1,2], Sylvain Guilley[1,3], Annelie Heuser[1⋆] and Olivier Rioul[1]

[1] Télécom ParisTech, Institut Mines-Télécom, CNRS LTCI
Department Comelec, Paris, France
{firstname.lastname@telecom-paristech.fr}
[2] STMicroelectronics, AST division, Rousset, France
[3] Secure-IC S.A.S., Rennes, France

Abstract. Higher-order side-channel attacks are able to break the security of cryptographic implementations even if they are protected with masking countermeasures. In this paper, we derive the best possible distinguishers (High-Order Optimal Distinguishers or HOOD) against masking schemes under the assumption that the attacker can profile. Our exact derivation admits simple approximate expressions for high and low noise and shows to which extent the optimal distinguishers reduce to known attacks in the case where no profiling is possible. From these results, we can explain theoretically the empirical outcome of recent works on second-order distinguishers. In addition, we extend our analysis to any order and to the application to masked tables precomputation. Our results give some insight on which distinguishers have to be considered in the security analysis of cryptographic devices.

Keywords: Side-channel analysis, higher-order masking, masking tables, higher-order optimal distinguisher (HOOD), template attack.

1 Introduction

In order to secure embedded devices against side-channel attacks, masking schemes have been introduced. Recent works have shown provable protections with a security parameter d, such that each sensitive variable is secured with d random masks [4]. The computation is carried out in such a way that the knowledge of any tuple of d intermediate variables does not disclose any information on any sensitive variable. Accordingly, all distinguishers using up to d leakages will fail to recover the correct key. A successful attack would be a $(d+1)$th-order CPA, which uses combination functions to transform the measured leakage and the prediction on each share into a single value in order to compute Pearson correlation coefficients [4, 10, 13, 14, 18, 19, 23].

[⋆] Annelie Heuser is a Google European fellow in the field of privacy and is partially founded by this fellowship.

Using combination functions to fit to a known tool like CPA looks more like an engineering recipe than the optimal solution. Yet, it is shown by Prouff et al. [16] that, in the case of second-order attacks, using the normalized product combination function combined with an optimal prediction function is the most efficient solution among all known combination functions. Even more, Standaert et al. [20] showed that the information loss induced from the combination functions vanishes for high noise.

In [14] Oswald and Mangard introduced several template-based attacks on masking schemes. Among them is the so-called template-based DPA attack, which extends the traditional template attack from Chari et al. [5] to first-order masking schemes. Besides, their approach classifies measurements according to the key, and not according to sensitive variables.

A slightly different scenario has been analyzed by Tunstall et al. in [22], where the authors study the security of masking tables for software implementations as defined in [1]. The authors suggested a two-stage CPA: first, for each individual trace, extract the mask during the precomputation and, second, use this knowledge about the mask to reveal the secret key using a vertical attack.

In this article we tackle the questions *"what is the best possible distinguisher in case of profiling?"* and *"how far are they from known practical distinguishers?"*.

In particular, we derive optimal higher-order distinguishers against higher-order masking schemes in case profiling is possible. Here, optimality means maximizing the success rate. Starting from second-order optimal distinguisher we derive approximations for high and low noise and recover known attacks. In particular, we show to what extent the optimal second-order distinguisher can be translated into a second-order CPA attack using combination functions. Given these results for second-order we extend our analysis to $(d+1)$th-order distinguisher against dth-order masking schemes.

Additionally, we investigate the scenario of masking tables as in [22]. We derive the optimal attack against masking tables and again derive approximations for the scenario of high and low noise which results in new attacks and compare it to the two-stage CPA.

2 Preliminaries

2.1 Masking Countermeasure & Notations

Even though many different masking schemes have been investigated so far, which clearly differ in their strength, the principle of attacking is equivalent. A masking scheme is characterized by the number of random masks that are used per sensitive variable. In the following we consider a dth-order masked implementation where we assume that the masks are uniformly distributed over a space $\mathcal{M}$. Calligraphic letters (e.g., $\mathcal{X}$) denote sets, capital letters (e.g., X) denote random variables taking values in these sets, and the corresponding lowercase letters (e.g., x) denote their realizations. Let k^* denote the secret cryptographic key, k any possible key hypothesis from the keyspace $\mathcal{K}$, and T be the input or ciphertext of the cryptographic algorithm. The mapping $f : (\mathcal{T}, \mathcal{K}, \mathcal{M}) \to \mathbb{F}_2^n$ maps the input or ciphertext $t \in \mathcal{T}$, a key hypothesis $k \in \mathcal{K}$ and the mask $m \in \mathcal{M}$ to an internally processed variable in some space $\mathbb{F}_2^n$ that is assumed to relate to the measured leakage X, where n is the number of bits. Generally it is assumed that f is known to the attacker. The measured leakage X can then be written as

$$X = \varphi(f(T, k^*, M)) + N, \tag{1}$$

where N denotes an independent—not necessarily Gaussian—additive noise with zero mean and where φ a device-specific deterministic function. In this paper we start by assuming that φ is known to the attacker due to profiling to consider the most powerful attack. We then show to which extend and scenarios this assumption can be relaxed while still achieving the same efficiency of the attack.

Specifically, in a dth-order masking scheme the implementation is protected with d masks with corresponding leakages

$$X^{(\omega)} = \varphi^{(\omega)}(f^{(\omega)}(T^{(\omega)}, k^*, M^{(\omega)})) + N^{(\omega)}, \tag{2}$$

with $\omega \in \{0, \dots, d\}$ and $M^{(\omega)} \in \mathcal{M}^{(\omega)}$ where the $\mathcal{M}^{(\omega)}$ does not need to be equal in general. Accordingly, a dth-order masking scheme can be broken using $(d+1)$th-order distinguishers by targeting $d+1$ shares. For simplification we denote $Y(T^{(\omega)}, k, M^{(\omega)}) = \varphi^{(\omega)}(f^{(\omega)}(T^{(\omega)}, k, M^{(\omega)}))$.

Example 1 (First-order software masking). For example a first-order masking scheme ($d = 1$) might leak with

$$X^{(0)} = \mathsf{HW}[M] + N^{(0)}, \tag{3}$$

$$X^{(1)} = \mathsf{HW}[\mathtt{Sbox}[T \oplus k^*] \oplus M] + N^{(1)}, \tag{4}$$

with $\mathtt{Sbox} : \mathbb{F}_2^8 \to \mathbb{F}_2^8$ being the AES Substitution box and $T^{(1)} = T$ uniformly distributed over $\mathbb{F}_2^8$ (and $T^{(0)}$ is non-existent). Thus, $\varphi^{(0)}(\cdot) = \varphi^{(1)}(\cdot) = \mathsf{HW}[\cdot]$ (the Hamming weight function), $M^{(0)} = M^{(1)} = M$, $f^{(0)}(T, k, M) = M$ and $f^{(1)}(T, k, M) = \mathtt{Sbox}[T \oplus k] \oplus M$.

Example 2 (Tables pre-computation). Again when assuming a Hamming weight leakage model, a masking scheme using Sbox recomputation [11] might leak with

$$X^{(\omega)} = \mathsf{HW}[\omega \oplus M] + N^{(\omega)}, \qquad \forall \omega \in \{0, 1, \dots, 2^n - 1\} \cong \mathbb{F}_2^n \tag{5}$$

$$X^{(2^n)} = \mathsf{HW}[T \oplus k^* \oplus M] + N^{(2^n)}. \tag{6}$$

A detailed description will be given in Sect. 5.

Definition 1 (Perfect masking (dth-order) [2]). *Let us denote the random variables $F^{(\omega)}(t, k) = f^{(\omega)}(t, k, M^{(\omega)})$ for $\omega \in \{0, \dots, d\}$ and a fixed pair (t, k). A masking scheme is* perfect *at dth-order if the joint distribution of maximum d of $F^{(0)}(t, k), \dots, F^{(d)}(t, k)$ is identically distributed of any pair $(t, k) \in \mathcal{T} \times \mathcal{K}$.*

Note that dth-order security implies $1\text{st}, 2\text{nd}, \dots, (d-1)$th-order security.

Proposition 1. *If a masking scheme is perfect, then whatever function ψ, $\sum_{m^{(\omega)} \in \mathcal{M}^{(\omega)}} \psi(f^{(\omega)}(t,k,m^{(\omega)}))$ is constant for any pair (t,k) for any $0 \leq \omega \leq d$.*

Proof. Let $\tilde{t}, \tilde{k}$ be any value in $\mathcal{T}$ and $\mathcal{K}$, respectively. As the masking scheme is perfect up to dth-order (which implies 1st-order) the distribution of $f^{(\omega)}(t,k,M^{(\omega)})$ is equivalent to $f^{(\omega)}(\tilde{t},\tilde{k},M^{(\omega)})$, hence $\psi(f^{(\omega)}(t,k,M^{(\omega)}))$ and $\psi(f^{(\omega)}(\tilde{t},\tilde{k},M^{(\omega)}))$ have the same distribution. In particular, the sum of realizations is identical. □

In our setup we assume that the attacker is able to measure q i.i.d. measurements. All values indexed by $i \in \{1,\ldots,q\}$ are in bold face (e.g. $\mathbf{a} = (a_1,\ldots,a_q) \in \mathcal{A}^q$ for $a_i \in \mathcal{A}$). Values indexed by the intermediate variable index (ω) $(a^{(\omega)} \in \mathcal{A})$ are denoted by $a^{(\star)} = (a^{(0)},\ldots,a^{(d)}) \in \mathcal{A}^{d+1}$. Moreover, $a_i^{(\omega)} \in \mathcal{A}$ and $\mathbf{a}^{(\star)} \in \mathcal{A}^{q\times(d+1)}$.

Note that contrary to $\mathbf{a}$, the vectors along (ω) can be linked, e.g., $\bigoplus_{\omega=0}^{d} M^{(\omega)} = 0$ in Example 1 or $\forall\omega \in \{0,\ldots,2^n-1\} \cup \{2^n\}, M^{(\omega)} = M^{(0)}$ in Example 2. Thus the set of admissible masks, denoted by $\mathcal{M}^{(\star)}$, is a subset of the Cartesian product over all $\mathcal{M}^{(\omega)}$. Additionally, regarding the noise, we have that $\forall(i,\omega) \neq (i',\omega')$, $N_i^{(\omega)}$ is independent of $N_{i'}^{(\omega')}$.

We write $\mathbb{P}(m) = \mathbb{P}(M = m)$ for discrete probability distributions, p for densities, and when the random variable X is conditioned by the event $Y = y$, we use the notation $p_k(X|Y = y)$ to recall that y depends on a (fixed) key guess k. As the model is known by the attacker, we also have: $p_k(X|Y = y) = p_k(X|T = t, M = m)$ when $y = y(t,k,m)$. Indeed, owing to Eq. (2), Y is a *sufficient statistic* for X [8]. We then use $p_k(x|t,m)$ to denote $p(X = x|Y = \varphi(f(t,k,m)))$. We denote the scalar product between $\mathbf{x}$ and $\mathbf{y}$ by $\langle\mathbf{x}|\mathbf{y}\rangle = \sum_{i=1}^{q} x_i y_i$, the Euclidean norm as $\|\mathbf{x}\|_2 = \sqrt{\langle\mathbf{x}|\mathbf{x}\rangle}$ and the componentwise product with $\mathbf{x}\cdot\mathbf{y} = (x_1y_1,\ldots,x_qy_q)$. Given a function $g(k)$, we use the notation $\arg\max_k\ g(k)$ to denote the value of k that maximizes $g(k)$. Finally, $\mathbb{1}_E : E \to \{0,1\}$ denotes the indicator function of the set E.

2.2 Combination Functions for Higher-Order CPA Attacks

In order to conduct a second-order CPA attack, two kinds of *combination functions*, i.e., $c_X : \mathcal{X}^{d+1} \to \mathbb{R}$ and $c_Y : \mathcal{T}^{d+1} \to \mathbb{R}$, are required. However, this seems to be more inspired from an engineering perspective –an "act from necessity"– than a sound mathematical tool to maximize the success. That the use of combination functions comes with information loss was already pointed out by Mangard and Oswald and Mangard [14] and Standaert et al. [20]. The history and selection of combination functions is indeed epic, where the literature mostly concentrated on second-order CPA $(d = 1)$.

The most prominent function to combine the leakages is the *product* combining function $c_X^{\mathrm{prod}}(X^{(0)},X^{(1)}) = X^{(0)} \cdot X^{(1)}$ introduced by Chari et al. in [4] and the *absolute difference* $c_X^{\mathrm{diff}}(X^{(0)},X^{(1)}) = |X^{(0)} - X^{(1)}|$ by Messerges in [13]. Oswald and Mangard [14] proposed (for e.g. the setting given in Example 1) an even more exotic combination function and corresponding prediction function[4]:

$$c_X^{\sin}(X^{(0)}, X^{(1)}) = \sin\left((X^{(0)} - X^{(1)})^2\right) \tag{7}$$

$$c_Y^{\sin}(T) = -89.95\ \sin(\mathsf{HW}[Y])^3 - 7.82\ \sin(\mathsf{HW}[Y])^2 + 67.66\ \sin(\mathsf{HW}[Y]), \tag{8}$$

where $Y = Y(T, k, 0)$. Contrary to what was suggested in previous papers, Prouff et al. [16] showed that all these combination functions should be accompanied by $c_Y^{\mathrm{opt}}(T^{(0)}, T^{(1)}) = \mathbb{E}\{c_X^*(Y^{(0)}, Y^{(1)})|T^{(0)}, T^{(1)}\}$ to maximize the absolute value of correlation, where the expectation is taken over the mask M and c_X^* denotes the same combination function as c_X but defined as a map $\mathcal{Y}^{d+1} \to \mathbb{R}$. Moreover, the *normalized product function*, i.e., $c_X^{\text{n-prod}}(X^{(0)}, X^{(1)}) = (X^{(0)} - \mathbb{E}\{X^{(0)}\}) \cdot (X^{(1)} - \mathbb{E}\{X^{(1)}\})$ is shown to be the most efficient of all known combination functions when considering a Hamming weight leakage model.

3 Optimal Distinguisher for Second-Order Attacks

3.1 Motivation

As highlighted in Subsect. 2.2 the introduction of combination functions for second-order CPA is more a necessary evil than an optimized procedure to maximize success. In [20] the authors empirically showed that a combination function always goes hand in hand with information loss. However, the authors depicted that for large noise the second-order CPA with the normalized product function $c_X^{\text{n-prod}}(X^{(0)}, X^{(1)})$ becomes (nearly) equivalent to the maximum likelihood distinguisher applied to the joint distribution.

This observation might not be obvious in theory since correlation is only an appropriate statistical tool when the underlying noise is Gaussian. Unfortunately, when multiplying two Gaussian distributions, as it is done for $c_X^{\text{n-prod}}(X^{(0)}, X^{(1)})$, does clearly *not* result in a Gaussian distribution.

Thus, our aim is to precisely state the higher-order optimal distinguisher (HOOD) expression for second-order when the attacker has full information about underlying the leakage and determine when this knowledge can be lessened to relate the expression to second-order CPA. This will help to understand known empirical results [9, 14, 16, 20]. In particular, we investigate low and high noise scenarios to see which combination function from the pool described in Subsect. 2.2 would be a reasonable choice.

3.2 Explicit Derivations

In [14] the authors state various template attacks against first-order masking schemes. The most efficient is a straightforward extension of the classical template attack [5] over all pairs (t, k). Our approach goes in a similar direction: we utilize the joint distribution of both leakages $X^{(0)}$ and $X^{(1)}$ without using a combination function, which gives us the optimal second-order distinguisher maximizing the success rate.

[4] Note that these are the corrected formulas given in [16].

Theorem 2 (Second-order HOOD). *If the model (i.e., $\varphi^{(\omega)}$) is known to the attacker for all ω, then the* second-order HOOD *is*

$$\mathcal{D}^2_{opt}(\mathbf{x}^{(\star)}, \mathbf{t}^{(\star)}) = \arg\max_{k\in\mathcal{K}} \; p_k(\mathbf{x}^{(\star)}|\mathbf{t}^{(\star)}) \tag{9}$$

$$= \arg\max_{k\in\mathcal{K}} \prod_{i=1}^{q} \sum_{m^{(\star)}\in\mathcal{M}^{(\star)}} \mathbb{P}(m^{(\star)}) \prod_{\omega=0}^{1} p_k(x_i^{(\omega)}|t_i^{(\omega)}, m^{(\omega)}). \tag{10}$$

Note that as the attacker knows the model he is able to compute the required probability distributions and densities.

Proof. Let us denote the key guess of any second-order distinguisher as $\hat{k} = \arg\max_{k\in\mathcal{K}} \mathcal{D}^2(\mathbf{x}^{(\star)}, \mathbf{t}^{(\star)})$. Then, using a frequentist approach we start from the success probability $\mathbb{P}_S$ over all possible secret keys k

$$\mathbb{P}_S = \frac{1}{|\mathcal{K}|}\sum_{k\in\mathcal{K}} \mathbb{P}(\hat{k}=k) = \frac{1}{|\mathcal{K}|}\sum_{k\in\mathcal{K}}\sum_{\mathbf{t}^{(\star)}} \mathbb{P}(\mathbf{t}^{(\star)})\; \mathbb{P}(\hat{k}=k|\mathbf{T}^{(\star)}=\mathbf{t}^{(\star)}) \tag{11}$$

$$= \frac{1}{|\mathcal{K}|}\sum_{k\in\mathcal{K}}\sum_{\mathbf{t}^{(\star)}\in\mathcal{T}^{q\times(d+1)}} \mathbb{P}(\mathbf{t}^{(\star)}) \int_{\mathcal{X}^{q\times(d+1)}} p_k(\mathbf{x}^{(\star)}|\mathbf{t}^{(\star)})\mathbb{1}_{k=\hat{k}}\, \mathrm{d}\mathbf{x}^{(\star)} \tag{12}$$

$$= \frac{1}{|\mathcal{K}|}\sum_{k\in\mathcal{K}}\sum_{\mathbf{t}^{(\star)}\in\mathcal{T}^{q\times(d+1)}} \mathbb{P}(\mathbf{t}^{(\star)}) \int_{\mathcal{X}^{q\times(d+1)}} p_{\hat{k}}(\mathbf{x}^{(\star)}|\mathbf{t}^{(\star)})\, \mathrm{d}\mathbf{x}^{(\star)}. \tag{13}$$

In Eq. (11), we have to compute $\mathbb{P}(\hat{k}=k|\mathbf{t}^{(\star)})$ where $\hat{k} = \hat{k}(\mathbf{x}^{(\star)}, \mathbf{t}^{(\star)}) = \arg\max_{k\in\mathcal{K}}\mathcal{D}^2(\mathbf{x}^{(\star)}, \mathbf{t}^{(\star)})$ is a function of $\mathbf{x}^{(\star)}$ and $\mathbf{t}^{(\star)}$. This is therefore a probability on the random variable $\mathbf{X}^{(\star)}$ knowing $\mathbf{T}^{(\star)} = \mathbf{t}^{(\star)}$, which follows the density $p_k(\mathbf{x}^{(\star)}|\mathbf{t}^{(\star)})$. Like for every probability taken on a random variable with density, the required probability is the integral of density over the events. So $\mathbb{P}(\hat{k}(\mathbf{x}^{(\star)}, \mathbf{t}^{(\star)}) = k|\mathbf{t}^{(\star)}) = \int p_k(\mathbf{x}^{(\star)}|\mathbf{t}^{(\star)})\, \mathrm{d}\mathbf{x}^{(\star)}$, where the integral is taken over all $\mathbf{x}^{(\star)}$ such that $\hat{k}(\mathbf{x}^{(\star)}, \mathbf{t}^{(\star)}) = k$; this is the indicator function inside the integral in Eq. (12).

Now, $\mathbb{P}(\mathbf{t}^{(\star)})$ is independent of the key. Thus, for each given sequence $\mathbf{x}^{(\star)}, \mathbf{t}^{(\star)}$ maximizing the success rate amounts to choose $k = \hat{k}$ such that $p_k(\mathbf{x}^{(\star)}|\mathbf{t}^{(\star)})$ is maximized. Moreover,

$$p_k(\mathbf{x}^{(\star)}|\mathbf{t}^{(\star)}) = \prod_{i=1}^{q} p_k(x_i^{(\star)}|t_i^{(\star)}) \tag{14}$$

$$= \prod_{i=1}^{q} \sum_{m^{(\star)}\in\mathcal{M}^{(\star)}} \mathbb{P}(m^{(\star)})\; p_k(x_i^{(\star)}|t_i^{(\star)}, m^{(\star)}) \tag{15}$$

$$= \prod_{i=1}^{q} \sum_{m^{(\star)}\in\mathcal{M}^{(\star)}} \mathbb{P}(m^{(\star)}) \prod_{\omega=0}^{1} p_k(x_i^{(\omega)}|t_i^{(\omega)}, m^{(\omega)}). \tag{16}$$

We used from (15) to (16) that $N_i^{(\omega)}$ is i.i.d. across the values of $i = \{1, \ldots, q\}$ and independent for $\omega = \{0, 1\}$. Accordingly, $\arg\max_{k\in\mathcal{K}}$ of Eq. (16) forms the optimal distinguisher $\mathcal{D}^2_{opt}(\mathbf{x}^{(\star)}, \mathbf{t}^{(\star)})$. □

Remark 1. To simplify our notation, we assume in the following that the masks at each order are drawn from the same space $\mathcal{M}$, with uniform probability $\mathbb{P}(M = m) = 1/|\mathcal{M}|$ and only one text byte is manipulated with the masks, as in software implementations (cf. Ex. 1). That is, $\forall i\ t_i^{(0)} = t_i^{(1)} = t_i$ and, moreover, there is only one mask $m^{(0)} = m^{(1)} = m$. Accordingly, Eq. (10) simplifies to

$$\mathcal{D}^2_{opt}(\mathbf{x}^{(\star)}, \mathbf{t}) = \arg\max_{k\in\mathcal{K}} \prod_{i=1}^{q} \sum_{m\in\mathcal{M}} p_k(x_i^{(0)}|t_i, m) \cdot p(x_i^{(1)}|t_i, m). \tag{17}$$

However, all our results hereafter can be easily extended to the scenario without simplifications.

As it is most often assumed that the noise distribution at the manipulation of each share is Gaussian (e.g., [14,16]), we further deduce Eq. (17) for Gaussian noise.

Proposition 3 (Second-order HOOD for Gaussian noise). *Assuming that* $N^{(\omega)} \sim \mathcal{N}(O, {\sigma^{(\omega)}}^2)$ *then the second-order optimal distinguisher becomes*

$$\mathcal{D}^{2,G}_{opt}(\mathbf{x}^{(0)}, \mathbf{x}^{(1)}, \mathbf{t}) = \arg\max_{k\in\mathcal{K}} \prod_{i=1}^{q} \sum_{m\in\mathcal{M}} \exp\left\{-\frac{1}{2}\left(\frac{-2x_i^{(0)}y^{(0)}(t_i,k,m) + y^{(0)}(t_i,k,m)^2}{{\sigma^{(0)}}^2}\right.\right.$$
$$\left.\left. + \frac{-2x_i^{(1)}y^{(1)}(t_i,k,m) + y^{(1)}(t_i,k,m)^2}{{\sigma^{(1)}}^2}\right)\right\}. \tag{18}$$

Proof. In this case $p_k(x_i^{(\omega)}|t_i, m)$ is the 1D Gaussian density with mean $y^{(\omega)}(t_i, k, m)$ and variance $\sigma^{(\omega)}$. Removing all constants gives us the required formula. □

As a next step we give approximations for high noise and low noise.

Corollary 4 (Second-order HOOD for high Gaussian noise). *When considering that* $\mathbb{E}\{y(T, m, k)\} = \mathbb{E}\{\varphi(f(T, m, k))\}$ *is independent of the choice of* $k \in \mathcal{K}$ *(owing to Proposition 1)*[5]*, which is given in case of high noise since a large number of measurements* q *is considered, then the distinguishing rule simplifies to*

$$\mathcal{D}^{2,G,\sigma\uparrow}_{opt}(\mathbf{x}^{(\star)}, \mathbf{t}) = \arg\max_{k\in\mathcal{K}} \prod_{i=1}^{q} \sum_{m\in\mathcal{M}} \exp\left\{\frac{x_i^{(0)}y^{(0)}(t_i,k,m)}{{\sigma^{(0)}}^2} + \frac{x_i^{(1)}y^{(1)}(t_i,k,m)}{{\sigma^{(1)}}^2}\right\}. \tag{19}$$

Proof. In Eq. (18) we can now remove the terms $y^{(\omega)}(t_i, k, m)^2$ for $\omega = \{0, 1\}$ because, $\mathbb{E}\{y(T, m, k)\} = \mathbb{E}\{\varphi(f(T, m, k))\}$ is independent of $k \in \mathcal{K}$. This gives Eq. (19). □

[5] This assumption has been also made in [21].

Remark 2. We see that the exact optimal distinguishing expression (Eq. (18)) operates in the *direct scale*, such that the function φ (including its scaling factor) and thus the exact relationship between X and Y has to be known. Whereas the distinguisher for high noise (Eq. (19)) operates in the *proportional scale*[6], thus the relationship between X and Y has only to be known up to an irrelevant affine law. That is to say, the attacker shall know that $X^{(\omega)} = a^{(\omega)}\varphi^{(\omega)}(f^{(\omega)}(T^{(\omega)}, k, m)) + b^{(\omega)} + N^{(\omega)}$ with unknown $a^{(\omega)}, b^{(\omega)} \in \mathbb{R}$. For more information on direct and proportional scales we refer the reader to [24].

Remark 2 already gives a hint about a possible relationship between the second-order HOOD and second-order CPA for high noise, which we will discuss in the next subsection.

Proposition 5 (Second-order HOOD for low Gaussian noise). *Assuming that both shares have the same low noise variance $\sigma = \sigma^{(0)} = \sigma^{(1)}$ then the optimal distinguisher reduces at first order to*

$$\mathcal{D}_{opt}^{2,G,\sigma\downarrow}(\mathbf{x}^{(\star)}, \mathbf{t}) = \arg\min_{k\in\mathcal{K}} \sum_{i=1}^{q} \max_{m\in\mathbb{F}_2^n} (x_i^{(0)} - y^{(0)}(t_i, k, m))^2 + (x_i^{(1)} - y^{(1)}(t_i, k, m))^2. \tag{20}$$

Proof. Starting from Eq. (18) and using $y_i^{(\omega)} = y^{(\omega)}(t_i, k, m)$ we have

$$\hat{k} = \arg\max_{k\in\mathcal{K}} \prod_{i=1}^{q} \sum_{m\in\mathbb{F}_2^n} \exp\left\{ -\frac{1}{{\sigma^{(0)}}^2}(x_i^{(0)} - y_i^{(0)})^2 - \frac{1}{{\sigma^{(1)}}^2}(x_i^{(1)} - y_i^{(1)})^2 \right\}. \tag{21}$$

Now as $\sigma = \sigma^{(0)} = \sigma^{(1)}$ and as the sum over exponential reduces at first order to the minimum we have the first order approximation for $\sigma \to 0$

$$= \arg\max_{k\in\mathcal{K}} \prod_{i=1}^{q} \min_{m\in\mathbb{F}_2^n} \exp\left\{ -(x_i^{(0)} - y_i^{(0)}(t_i, k, m))^2 - (x_i^{(1)} - y_i^{(1)}(t_i, k, m))^2 \right\}.$$

Applying the logarithm that is strictly monotonous increasing yields

$$= \arg\max_{k\in\mathcal{K}} \sum_{i=1}^{q} \min_{m\in\mathbb{F}_2^n} \left(-(x_i^{(0)} - y_i^{(0)}(t_i, k, m))^2 - (x_i^{(1)} - y_i^{(1)}(t_i, k, m))^2 \right) \tag{22}$$

$$= \arg\min_{k\in\mathcal{K}} \sum_{i=1}^{q} \max_{m\in\mathbb{F}_2^n} \left((x_i^{(0)} - y_i^{(0)}(t_i, k, m))^2 + (x_i^{(1)} - y_i^{(1)}(t_i, k, m))^2 \right). \quad \square$$

Interestingly, one can see directly from Eq. (20) that the optimal distinguisher for low noise cannot be rewritten as correlation with any combination functions and moreover that it operates in the direct scale. Even more, the nature of distinguisher seems not very intuitive.

[6] But not *anti-proportional* scale, or in other words, the "sign" has to be known.

3.3 Comparison with Second-Order CPA

Proposition 6 (Relationship between second-order HOOD for high noise and second-order CPA). *The second-order HOOD for high noise can be approximated as*

$$\mathcal{D}_{opt}^{2,G,\sigma\uparrow} \approx \arg\max_k \langle \mathbf{x}^{(0)} \cdot \mathbf{x}^{(1)} | \sum_{m\in\mathcal{M}} y^{(0)}(\mathbf{t},k,m) y^{(1)}(\mathbf{t},k,m)\rangle, \tag{23}$$

which is all the more equivalent as the noise gets larger. Accordingly, as the noise is larger, the closer the optimal distinguishing rule to second-order CPA with

$$c_X^{n\text{-}prod}(X^{(0)},X^{(1)}) = (X^{(0)} - \mathbb{E}\{X^{(0)}\}) \cdot (X^{(1)} - \mathbb{E}\{X^{(1)}\}) \text{ and} \tag{24}$$

$$c_Y^{opt}(Y^{(\star)}) = \mathbb{E}\{c_X^*(Y^{(0)}(T,k,M), Y^{(1)}(T,k,M))|T\}. \tag{25}$$

Proof. We use the first order Taylor expansion $\exp\{\varepsilon\} = 1+\varepsilon+O(\varepsilon^2)$. Note that this approximation is all the better as ε is close to zero and thus as the argument of $\exp\{\cdot\}$ his high. Starting from Eq. (19) and using $y_i^{(\omega)} = y^{(\omega)}(t_i,k,m)$, we have

$$\mathcal{D}^{2,G,\sigma\uparrow}(\mathbf{x}^{(0)},\mathbf{x}^{(1)},\mathbf{t}) = \arg\max_k \prod_{i=1}^{q} \sum_{m\in\mathcal{M}} \exp\left\{\frac{1}{\sigma^{(0)^2}} x_i^{(0)} y_i^{(0)}\right\} \exp\left\{\frac{1}{\sigma^{(1)^2}} x_i^{(1)} y_i^{(1)}\right\}$$

$$\approx \arg\max_k \prod_{i=1}^{q} \sum_{m\in\mathcal{M}} \left(1+\frac{1}{\sigma^{(0)^2}} x_i^{(0)} y_i^{(0)}\right)\left(1+\frac{1}{\sigma^{(1)^2}} x_i^{(1)} y_i^{(1)}\right) \tag{26}$$

$$= \arg\max_k \prod_{i=1}^{q} \sum_{m\in\mathcal{M}} \left(1+\frac{1}{\sigma^{(0)^2}\sigma^{(1)^2}} x_i^{(0)} y_i^{(0)} x_i^{(1)} y_i^{(1)} + \frac{1}{\sigma^{(0)^2}} x_i^{(0)} y_i^{(0)} + \frac{1}{\sigma^{(1)^2}} x_i^{(1)} y_i^{(1)}\right). \tag{27}$$

In Eq. (27), owing to the perfect masking definition, the terms $\sum_{m\in\mathcal{M}} x_i^{(0)} y_i^{(0)}$ and $\sum_{m\in\mathcal{M}} x_i^{(1)} y_i^{(1)}$ are constant ($\text{const}^{(0)}$ and $\text{const}^{(1)}$). Additionally, as the logarithm function is increasing, we consider the logarithm of the product, and we use the approximation $\ln\{1+\varepsilon\} = \varepsilon + \mathcal{O}(\varepsilon^2)$, (reciprocal of the previous Taylor's expansion of the exponential function), which is again all the better as ε is close to zero and thus for high noise. Accordingly,

$$\mathcal{D}^{2,G,\sigma\uparrow}(\mathbf{x}^{(0)},\mathbf{x}^{(1)},\mathbf{t}) \approx \arg\max_k \ln \prod_{i=1}^{q} \sum_{m\in\mathcal{M}} \left(1+\frac{1}{\sigma^{(0)^2}\sigma^{(1)^2}} x_i^{(0)} y^{(0)}(t_i,k,m)\cdot\right.$$

$$\left. x_i^{(1)} y^{(1)}(t_i,k,m) + \frac{1}{\sigma^{(0)^2}} x_i^{(0)} y^{(0)}(t_i,k,m) + \frac{1}{\sigma^{(1)^2}} x_i^{(1)} y^{(1)}(t_i,k,m)\right) \tag{28}$$

$$= \arg\max_k \sum_{i=1}^{q} \ln\left\{1+\sum_{m\in\mathcal{M}} \frac{1}{\sigma^{(0)^2}\sigma^{(1)^2}} x_i^{(0)} y_i^{(0)} x_i^{(1)} y_i^{(1)} + \text{const}^{(0)} + \text{const}^{(1)}\right\}$$

$$\approx \arg\max_k \sum_{i=1}^{q} x_i^{(0)} x_i^{(1)} \sum_{m\in\mathcal{M}} y^{(0)}(t_i,k,m) y^{(1)}(t_i,k,m) + \text{const}^{(0)} + \text{const}^{(1)}$$

$$= \arg\max_k \langle \mathbf{x}^{(0)} \cdot \mathbf{x}^{(1)} | \sum_{m\in\mathcal{M}} y^{(0)}(\mathbf{t},k,m) y^{(1)}(\mathbf{t},k,m) \rangle. \tag{29}$$

Note that we can remove the $\text{const}^{(0)}, \text{const}^{(1)}$ as they do not depend on the key guess. For large number of measurements (resulting from large noise) the $\arg\max_{k\in\mathcal{K}}$ of the correlation coefficient can be simplified as

$$\arg\max_{k\in\mathcal{K}} \frac{\langle \mathbf{x} - \overline{\mathbf{x}} | \mathbf{y}(k) \rangle}{\|\mathbf{x} - \overline{\mathbf{x}}\|_2 \cdot \|\mathbf{y}(k) - \overline{\mathbf{y}(k)}\|_2} \approx \arg\max_{k\in\mathcal{K}} \langle \mathbf{x} - \overline{\mathbf{x}} | \mathbf{y}(k) \rangle. \tag{30}$$

Accordingly, if $\mathbf{x}^{(0)}$ and $\mathbf{x}^{(1)}$ are centered, then in Eq. (29) $c_X^{n\text{-}prod} = \mathbf{x}^{(0)} \cdot \mathbf{x}^{(1)}$, and $c_Y^{opt} = \sum_{m\in\mathcal{M}} y^{(0)}(\mathbf{t},k,m) y^{(1)}(\mathbf{t},k,m)$ is the *optimal prediction function.* □

As correlation is a measure in the proportional scale, we can relax our assumptions made about the knowledge of the attacker. More precisely, he does not need to know $y^{(0)}$ and $y^{(1)}$ exactly but any linear transformation $l^{(\omega)}(y^{(\omega)}) = ay^{(\omega)} + b$, as it is most often assumed in the literature [16, 20]. Yet, in Prop. 6 we do not recover the absolute value of the correlation, thus, for second-order CPA the "sign" must be known and taking the absolute value does not result in an equivalence for high noise, which is also empirically validated in our experiments in Subsect. 3.4.

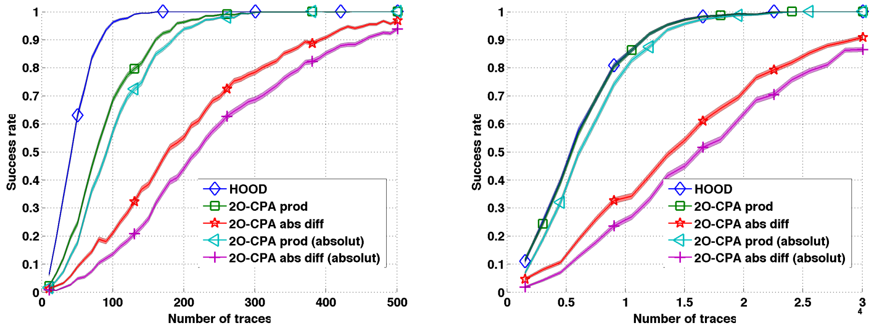
Remark 3. Prouff et al. illustrated in [16] that for large sigma the improved (i.e., centered) product combining function has the best efficiency among the known combination functions, which is inline with our findings in Prop. 6. Moreover, we can claim that the improved product combining function is the most efficient among *all* combining functions for high noise as it becomes equivalent to the optimal second-order distinguisher. Moreover, our study is not restricted to a particular HW or HD leakage model scenario as in the previous studies.

Remark 4. The determination of optimal combination functions is a vivid research topic. As already mentioned, the optimality of the centered product amongst all combination functions has been conjectured by Prouff et al. in [16]. Afterwards, mathematical arguments for optimality were given by Carlet et al. [3], and independently by Ding et al. in [7].

Remark 5. As underlined in [20], the function to be maximized in Eq. (23) is a straightforward generalization of Pearson's correlation coefficient to the case of three random variables: $X^{(0)}$, $X^{(1)}$, and $\mathbb{E}\{y^{(0)}(T,k,M) y^{(1)}(T,k,M) | T\}$, where the expectation is taken over M.

3.4 Experimental Validation

For our experimental validation we used simulations of a first-order masking scheme where each share is leaking in the Hamming weight model to be able to directly compare our results to previous publications conducting the same setting [16, 20] (see Example 1). We simulated the noise arising from a Gaussian



(a) $\sigma = 0.5$

(b) $\sigma = 4$

Fig. 1: Success rate for second-order attacks

distributions $N \sim \mathcal{N}(0, \sigma^2)$ for $\sigma = \sigma^{(0)} = \sigma^{(1)} \in \{0.5, 4\}$. To be reliable we conducted 500 independent experiments with uniformly distributed k^* to compute the empirical success rate. Moreover, when plotting the empirical success rate, we additionally highlight the standard deviation of the success rate by error bars. If the error bars do not overlap, we can unambiguously conclude that one distinguisher is better than the other [12].

For our simulations we calculated the second-order HOOD and second-order CPA with the normalized product and the absolute difference combination function as described in Subsect. 2.2. For a low value of σ, $\mathcal{D}^{2,G}_{opt}$ (HOOD) clearly outperforms 2nd-order CPA (2O-CPA) independent of the combination functions (see Fig. 1a), which is inline with our theoretical analysis and the empirical analysis in [20]. For high values of σ, the second-order HOOD and second-order CPA with the normalized product combining function become equivalently efficient (see Fig. 1b), which coincides with Prop. 6. Note that as said before, taking the absolute value of the correlation is not equivalent to HOOD, which is confirmed in Fig. 1b.

4 Higher-Order Optimal Distinguisher (HOOD) for Any Order

The claim in [16] that the normalized product combining function $c_X^{n\text{-}prod}$ in combination with c_Y^{opt} is optimal[7] was only done for $d = 1$. We now extent our investigation to $(d+1)$th-order distinguishers in order to analyze if the assumption can straightforwardly be generalized.

Theorem 7 (General HOOD). *When $\varphi^{(\omega)} : \mathbb{F}_2^n \to \mathbb{R}$ is known for all ω, $N_i^{(\omega)}$ i.i.d. across values of $i = \{1, \ldots, q\}$ and independent across the values of*

[7] Note again, that the authors used the absolute correlation coefficient of the correct key as a measure of optimality; not the success rate.

$\omega = \{0, \ldots, d\}$, *then the* general higher-order optimal distinguisher *is*

$$\mathcal{D}_{opt}^{d}(\mathbf{x}^{(\star)}, \mathbf{t}^{(\star)}) = \arg\max_{k\in\mathcal{K}} \prod_{i=1}^{q} \sum_{m^{(\star)}\in\mathcal{M}^{(\star)}} \mathbb{P}(m^{(\star)}) \prod_{\omega=0}^{d} p_k(x_i^{(\omega)}|t_i^{(\omega)}, m^{(\omega)}). \quad (31)$$

Proof. The proof is a straightforward extension of proof of Theorem 2. □

Proposition 8 (HOOD for Gaussian noise). *Under the same assumptions as in Theorem 7 and additionally assuming Gaussian noise, i.e.,* $N_i^{(\omega)} \sim \mathcal{N}(0, \sigma_\omega^2)$*, Eq.* (31) *becomes*

$$\mathcal{D}_{opt}^{d,G}(\mathbf{x}^{(\star)}, \mathbf{t}) = \arg\max_{k\in\mathcal{K}} \sum_{i=1}^{q} \log\left\{ \sum_{m^{(\star)}\in\mathcal{M}^{(\star)}} \exp\Big\{\sum_{\omega=0}^{d} \frac{1}{\sigma^{(\omega)^2}}\big(x_i^{(\omega)}y_i^{(\omega)} - \frac{1}{2}y_i^{(\omega)^2}\big)\Big\}\right\}. \quad (32)$$

Proof. As $p_k(x_i^{(\omega)}|t_i, m) = p_{k,N^{(\omega)}}(x_i^{(\omega)} - y^{(\omega)}(t_i, k, m))$ we have

$$\arg\max_{k\in\mathcal{K}} \prod_{i=1}^{q} \sum_{m^{(\star)}\in\mathcal{M}^{(\star)}} \mathbb{P}(m^{(\star)}) \prod_{\omega=0}^{d} p_k(x_i^{(\omega)}|t_i, m^{(\omega)}) \quad (33)$$

$$= \arg\max_{k\in\mathcal{K}} \prod_{i=1}^{q} \sum_{m^{(\star)}\in\mathcal{M}^{(\star)}} \mathbb{P}(m^{(\star)}) \prod_{\omega=0}^{d} \frac{1}{\sqrt{2\pi}\sigma^{(\omega)}} \exp\Big\{-\frac{1}{2\sigma^{(\omega)^2}}(x_i^{(\omega)} - y^{(\omega)}(t_i, k, m))^2\Big\}.$$

Now, removing all key-independent constants yields

$$\arg\max_{k\in\mathcal{K}} \prod_{i=1}^{q} \sum_{m\in\mathcal{M}} \prod_{\omega=0}^{d} \exp\left\{-\frac{1}{2\sigma^{(\omega)^2}}(x_i^{(\omega)} - y^{(\omega)}(t_i, k, m))^2\right\}$$

Now, as the product of $\exp\{\cdot\}$ is the $\exp\{\cdot\}$ of the sum and expanding the square and removing the key-independent factor $x_i^{(\omega)^2}$ gives the required equation. □

Proposition 9 (HOOD for high Gaussian noise). *For high Gaussian noise (low SNR) we can further approximate the HOOD to*

$$\mathcal{D}_{opt}^{d,G,\sigma\uparrow}(\mathbf{x}^{(\star)}, \mathbf{t}) = \arg\max_{k} \prod_{i=1}^{q} \sum_{m\in\mathcal{M}} \exp\left\{\sum_{\omega=0}^{d} \frac{1}{\sigma^{(\omega)^2}} x_i^{(\omega)} y^{(\omega)}\right\}, \quad (34)$$

and as $\sigma^{(\omega)}$ *becomes large Eq.* (34) *becomes closer to (d+1)th-order CPA with*

$$c_X^{n\text{-}prod}(X^{(\star)}) = \prod_{\omega=0}^{d}(X^{(\omega)} - \mathbb{E}\{X^{(\omega)}\}) \quad \textit{and} \quad c_Y^{opt}(Y^{(\star)}) = \mathbb{E}\{c_X^{*}(Y^{(\star)}(M, k))\}.$$

Proof. As in the case of $d = 1$, we use the first-order Taylor expansion $\exp\{\varepsilon\} = 1 + \varepsilon + O(\varepsilon^2)$. Starting from Eq. (32), we have

$$\mathcal{D}^{d,G,\sigma\uparrow}(\mathbf{x}^{(0)}, \mathbf{x}^{(1)}, \mathbf{t}) = \arg\max_k \prod_{i=1}^{m} \sum_{m\in\mathcal{M}} \prod_{\omega=0}^{d} \exp\left\{\frac{1}{2\sigma^{(\omega)^2}} x_i^{(\omega)} y_i^{(\omega)}\right\} \tag{35}$$

$$\approx \arg\max_k \prod_{i=1}^{m} \sum_{m\in\mathcal{M}} \prod_{\omega=0}^{d} \left(1 + \frac{1}{2\sigma^{(\omega)^2}} x_i^{(\omega)} y_i^{(\omega)}\right). \tag{36}$$

Now, in Eq. (36) when factorizing the product over ω all the terms not depending on all shares $1, \dots, d$ simultaneously, i.e., $\prod_{\omega=0}^{d} x_i^{(\omega)} y_i^{(\omega)}$, do not depend on the key due to the perfect masking definition. Moreover, following the same argumentation as for Prop. 6, we recover that if $\forall\omega$ $\mathbf{x}^{(\omega)}$ are centered, then $c_X^{n\text{-}prod} = \prod_{\omega=0}^{d} \mathbf{x}^{(\omega)}$, and $c_Y^{opt} = \sum_{m\in\mathcal{M}} \prod_{\omega=0}^{d} y^{(\omega)}(\mathbf{t}, k, m)$ is the *optimal prediction function* for higher-order CPA. □

Proposition 9 shows that the normalized production combination function combined with the optimal prediction function is therefore not only optimal for dth-order CPA in case of $d = 1$ but for any value of d.

Proposition 10 (HOOD for low Gaussian noise). *For low noise variance $\sigma = \sigma^{(0)} = \cdots = \sigma^{(d)}$ the optimal distinguisher (Eq.* (32)*) is simplified to*

$$\mathcal{D}_{opt}^{d,G,\sigma\downarrow}(\mathbf{x}^{(\star)}, \mathbf{t}) = \arg\min_{k\in\mathcal{K}} \sum_{i=1}^{q} \max_{m\in\mathcal{M}} \sum_{\omega=0}^{d} (x_i^{(\omega)} - y_i^{(\omega)})^2 \tag{37}$$

$$= \arg\min_{k\in\mathcal{K}} \sum_{i=1}^{q} \max_{m\in\mathcal{M}} \|x_i^{(\star)} - y_i^{(\star)}\|_2^2. \tag{38}$$

Proof. The proof is a straightforward extension of the proof for Prop. 5. □

5 HOOD for Precomputation Masking Tables

5.1 Classical Attacks

We now consider the attack of a masking scheme using Sbox recomputation as described in [11]. Appendix A provides a description of the underlying algorithm.

It is noteworthy that the traditional approach to reduce the multiplicity of leakage samples by a combination $c_X : \mathcal{X}^d \to \mathbb{R}$ actually **would fail** in the setup of masking tables.

Indeed, the combination functions are usually considered symmetric into its arguments, meaning that any swap of the inputs does not affect the combination. This (tacit) hypothesis has been made, for instance, for

- the absolute difference $c_X^{diff}(X^{(\star)}) = (|X^{(0)} - X^{(1)}| = |X^{(1)} - X^{(0)}|)$, and

- the centered product $c_X^{n\text{-}prod}(X^{(\star)}) = ((X^{(0)} - \mathbb{E}\{X^{(0)}\})(X^{(1)} - \mathbb{E}\{X^{(1)}\}) = (X^{(1)} - \mathbb{E}\{X^{(1)}\})(X^{(0)} - \mathbb{E}\{X^{(0)}\}))$.

We assume here that the attacker applies the combination function on the leakages occurring during the Sbox recomputation (see Alg. 1), i.e., the attacker gains 2^n leakages

$$\begin{aligned} X^{(0)} &= \varphi^{(0)}(M) + N^{(0)} \qquad (39)\\ X^{(1)} &= \varphi^{(1)}(M \oplus 1) + N^{(1)} \qquad (40)\\ &\vdots \\ X^{(2^n-1)} &= \varphi^{(2^n-1)}(M \oplus (2^n - 1)) + N^{(2^n-1)}, \qquad (41) \end{aligned}$$

and would apply e.g., $c_X^{diff}(X^{(\star)})$ or $c_X^{n\text{-}prod}(X^{(\star)})$. Additionally, he measures the leakage $X^{(2^n)} = \varphi^{(2^n)}(T \oplus k \oplus M) + N^{(2^n)}$ and finally combines it with the previous combined leakages as $\bar{c}_X(X^{(2^n)}, c_X(X^{(0)}, \ldots, X^{(2^n-1)}))$.

Following the methodology in [16] and assuming an equal leakage function on each share[8], i.e., $\varphi = \varphi^{(0)} = \cdots = \varphi^{(2^n)}$, the optimal function to combine the predictions would then be

$$\begin{aligned} c_Y^{opt} &= \mathbb{E}\{\bar{c}_X^*(c_X^*(\varphi(M), \varphi(M \oplus 1), \ldots, \varphi(M \oplus (2^n - 1))), \varphi(t \oplus k \oplus M))\} \qquad (42)\\ &= \frac{1}{2^n} \sum_{m \in \mathbb{F}_2^n} \bar{c}_X^*(c_X^*(\varphi(m), \varphi(m \oplus 1), \ldots, \varphi(m \oplus (2^n - 1))), \varphi(t \oplus k \oplus m))\\ &= \frac{1}{2^n} \sum_{m' \in \mathbb{F}_2^n} \bar{c}_X^*(c_X^*(\varphi(m' \oplus k), \ldots, \varphi(m' \oplus k \oplus (2^n - 1))), \varphi(t \oplus m')) \qquad (43)\\ &= \frac{1}{2^n} \sum_{m' \in \mathbb{F}_2^n} \bar{c}_X^*(c_X^*(\varphi(M'), \varphi(m' \oplus 1), \ldots, \varphi(M' \oplus (2^n - 1))), \varphi(t \oplus m')). \qquad (44) \end{aligned}$$

In Eq. (43), we change m for $m' = m \oplus k$ and in Eq. (44), the input terms at position ζ are replaced with those at position $\zeta \oplus k$ (because of the symmetry property of c). Accordingly, c_Y^{opt} does not depend on the key k and is even constant as the same operation can be done on $t \oplus k$, therefore higher-order CPA fails.

Of course, the Sbox precomputation masking scheme can be attacked by various attacks (e.g., the classic means, collision attacks, second-order attacks) that concentrate on specific stages of Alg. 1. However, a better attack would consist in using altogether all the leakages from the Sbox recomputation with one (or more) of the samples used during the computation proper (starting from line 8, when the key is involved). One example of such strategy has been exposed in [22], which we label as 2-stage CPA attack.

[8] This assumption is reasonable for software implementation, which is the adequate scenario for masking tables.

Definition 2 (2-stage CPA attack [22]).

$$2\times\mathsf{CPA}^{mt}(\mathbf{x},\mathbf{t}) = \arg\max_{k\in\mathcal{K}} \rho(\mathbf{x}^{(2^n)}, y^{(2^n)}(\mathbf{t},k,\hat{\mathbf{m}}), \tag{45}$$

where $\forall i$ $\hat{m}_i$ is the mask that maximizes the correlation between $x_i^{(\omega)}$ and $y_i^{(\omega)} = \omega \oplus m_i$ for $\omega \in [0, 2^n[$. This attack is a synergy between a horizontal and a vertical attack. For each trace (separately $\forall i$), the first attack in Eq. (45) consists in recovering the mask during the precomputation (lines 2 to 5 in Appendix A). Second, a regular CPA using a model in which both the plaintext t and the mask m are assumed as public knowledge is launched. Even if the mask $\hat{m}$ is not recovered correctly for each trace (since 2^n leakage samples during the precomputation can be seen as small), it can be expected that the value of the mask is recovered by the first horizontal attack probabilistically well enough for it to be biased, i.e., better guessed than random. This gives a rough idea of the proof of soundness for this attack.

Nonetheless, this attack is probably not the most efficient, as it uses separately the information available from the Sbox precomputation and from the leakage of the AES algorithm proper. The next subsection investigates the optimal attack and gives approximation for high and low noise.

5.2 HOOD for Precomputation Masking Tables

When using masking tables (Alg. 1) the attacker first has all leaking samples during the precomputation, i.e., $y_i^{(\omega)} = \varphi(\omega \oplus m)$ that are independent of i for $0 \leq \omega \leq (2^n - 1)$, and, second, the leakage arising from the combination of the mask m, plaintext t_i and the key, i.e., $y_i^{(2^n)} = \varphi(t_i \oplus k \oplus m)$. Thus, all terms for $\omega \neq 2^n$ do not depend on the key and the higher-order optimal distinguisher from Eq. (32) can be further deduced.

Theorem 11 (HOOD for masking tables). *When $\varphi : \mathbb{F}_2^n \to \mathbb{R}$ is known, $N_i^{(\omega)} \sim \mathcal{N}(0, \sigma_\omega^2)$ and i.i.d. across values of $i = \{1, \dots, q\}$ and independent across the values of $\omega = \{0, \dots, 2^n\}$, then the higher-order optimal distinguisher against masking tables takes the form*

$$\begin{aligned}
&\mathcal{D}_{opt}^{mt,G}(\mathbf{x}^{(\star)}, \mathbf{t}) = \\
&\quad \arg\max_{k\in\mathcal{K}} \sum_{i=1}^{q} \log \left\{ \sum_{m\in\mathbb{F}_2^n} \exp \left\{ \sum_{\omega\in\mathbb{F}_2^n} \frac{1}{\sigma^{(\omega)^2}} \left(x_i^{(\omega)} \varphi(\omega \oplus m) - \frac{1}{2}\varphi^2(\omega \oplus m) \right) \right.\right. \\
&\quad\quad \left.\left. + \frac{1}{\sigma^{(2^n)^2}} \left(x_i^{(2^n)} \varphi(t_i \oplus m \oplus k) - \frac{1}{2}\varphi^2(t_i \oplus m \oplus k) \right) \right\} \right\}.
\end{aligned} \tag{46}$$

Proof. Straightforward computation from Eq. (32) yields

$$\operatorname*{arg\,max}_{k\in\mathcal{K}} \prod_{i=1}^{q} \sum_{m\in\mathbb{F}_2^n} \prod_{\omega\in\mathbb{F}_2^n} \exp\left\{\frac{1}{\sigma^{(\omega)2}}\left(x_i^{(\omega)} y_i^{(\omega)} - \frac{1}{2} y_i^{(\omega)^2}\right)\right\} \tag{47}$$

$$= \operatorname*{arg\,max}_{k\in\mathcal{K}} \sum_{i=1}^{q} \log\left\{\sum_{m\in\mathbb{F}_2^n} \exp\left\{\sum_{\omega\in\mathbb{F}_2^n} \frac{1}{\sigma^{(\omega)2}}\left(x_i^{(\omega)} y_i^{(\omega)} - \frac{1}{2} y_i^{(\omega)^2}\right)\right\}\right\} \tag{48}$$

Now plugging the respective leakages as described in Subsect. 5.2 gives

$$\begin{aligned} &= \operatorname*{arg\,max}_{k\in\mathcal{K}} \sum_{i=1}^{q} \log\left\{\sum_{m\in\mathbb{F}_2^n} \exp\left\{\sum_{\omega\in\mathbb{F}_2^n} \frac{1}{\sigma^{(\omega)2}}\left(x_i^{(\omega)} \varphi(\omega\oplus m) - \frac{1}{2}\varphi^2(\omega\oplus m)\right)\right.\right. \\ &\quad \left.\left. + \frac{1}{\sigma^{(2^n)2}}\left(x_i^{(2^n)}\varphi(t_i\oplus m\oplus k) - \frac{1}{2}\varphi^2(t_i\oplus m\oplus k)\right)\right\}\right\}. \qquad \square \end{aligned} \tag{49}$$

Proposition 12 (HOOD for masking tables for low SNR). *For large Gaussian noise (or low SNR) the distinguisher becomes*

$$\mathcal{D}_{opt}^{mt,G,\sigma\uparrow}(\mathbf{x}^{(\star)}, \mathbf{t}) =$$

$$\operatorname*{arg\,max}_{k\in\mathcal{K}} \sum_{\omega\in\mathbb{F}_2^n} \frac{1}{\sigma^{(\omega)^2}} \sum_{i=1}^{q} \begin{pmatrix} x_i^{(\omega)} x_i^{(2^n)} \sum_m \varphi(\omega\oplus m)\varphi(t_i\oplus k\oplus m) \\ -\frac{1}{2} x_i^{(2^n)} \sum_m \varphi(t_i\oplus k\oplus m)\varphi(\omega\oplus m)^2 \\ -\frac{1}{2} x_i^{(\omega)} \sum_m \varphi(\omega\oplus m)\varphi(t_i\oplus k\oplus m)^2 \\ +\frac{1}{4}\sum_m \varphi(\omega\oplus m)^2\varphi(t_i\oplus k\oplus m)^2 \end{pmatrix}. \tag{50}$$

Proof. Due to the lack of space we neglect the term $\operatorname{arg\,max}_{k\in\mathcal{K}}$ in front of each line. Starting from Eq. (32) we use again the first-order Taylor expansion $\exp\{\varepsilon\} = 1 + \varepsilon + O(\varepsilon^2)$. So,

$$\prod_{i=1}^{q} \sum_{m\in\mathbb{F}_2^n} \prod_{\omega=0}^{2^n} \Big(1 + \frac{1}{\sigma^{(\omega)^2}}\big(x_i^{(\omega)} y_i^{(\omega)} - \frac{1}{2} y_i^{(\omega)^2}\big) + \frac{1}{2\sigma^{(\omega)^4}}\big(x_i^{(\omega)} y_i^{(\omega)} - \frac{1}{2} y_i^{(\omega)^2}\big)^2\Big).$$

Furthermore, an expansion at second-order gives

$$\begin{aligned} \prod_{i=1}^{q} \sum_{m\in\mathbb{F}_2^n} \Big(1 + \sum_{\omega=0}^{2^n} \frac{1}{\sigma^{(\omega)^2}}\big(x_i^{(\omega)} y_i^{(\omega)} - \frac{1}{2} y_i^{(\omega)^2}\big) + \frac{1}{2\sigma^{(\omega)^4}}\big(x_i^{(\omega)} y_i^{(\omega)} - \frac{1}{2} y_i^{(\omega)^2}\big)^2 \\ + \sum_{\omega\neq\omega'}^{2^n} \frac{1}{\sigma^{(\omega)^2}\sigma^{(\omega')^2}}\big(x_i^{(\omega)} y_i^{(\omega)} - \frac{1}{2} y_i^{(\omega)^2}\big)\big(x_i^{(\omega')} y_i^{(\omega')} - \frac{1}{2} y_i^{(\omega')^2}\big)\Big). \end{aligned} \tag{51}$$

From the perfect masking condition (see Prop. 1), the first-order term

$$\sum_{m\in\mathbb{F}_2^n} \sum_{\omega=0}^{2^n} \frac{1}{\sigma^{(\omega)^2}}\big(x_i^{(\omega)} y_i^{(\omega)} - \frac{1}{2} y_i^{(\omega)^2}\big) = \sum_{\omega=0}^{2^n} \frac{1}{\sigma^{(\omega)^2}}\big(x_i^{(\omega)} \sum_{m\in\mathbb{F}_2^n} y_i^{(\omega)} - \frac{1}{2} \sum_{m\in\mathbb{F}_2^n} y_i^{(\omega)^2}\big)$$

is constant as well as

$$
\begin{aligned}
&\sum_{m\in\mathbb{F}_2^n}\sum_{\omega=0}^{2^n}\frac{1}{2\sigma^{(\omega)^4}}\big(x_i^{(\omega)}y_i^{(\omega)}-\frac{1}{2}y_i^{(\omega)^2}\big)^2 && (52)\\
&=\sum_{\omega=0}^{2^n}\frac{1}{2\sigma^{(\omega)^4}}\big(x_i^{(\omega)^2}\sum_{m\in\mathbb{F}_2^n}y_i^{(\omega)^2}+\frac{1}{4}\sum_{m\in\mathbb{F}_2^n}y_i^{(\omega)^4}-x_i^{(\omega)}\sum_{m\in\mathbb{F}_2^n}y_i^{(\omega)^3}\big). && (53)
\end{aligned}
$$

The other terms in ω, ω' can be written as

$$
\begin{aligned}
2\sum_{\omega<\omega'}^{2^n}\frac{1}{\sigma^{(\omega)^2}\sigma^{(\omega')^2}}\Big(&x_i^{(\omega)}x_i^{(\omega')}\sum_{m\in\mathbb{F}_2^n}y_i^{(\omega)}y_i^{(\omega')}-\frac{1}{2}x_i^{(\omega')}\sum_{m\in\mathbb{F}_2^n}y_i^{(\omega')}y_i^{(\omega)^2}\\
&-\frac{1}{2}x_i^{(\omega)}\sum_{m\in\mathbb{F}_2^n}y_i^{(\omega)}y_i^{(\omega')^2}+\frac{1}{4}\sum_{m\in\mathbb{F}_2^n}y_i^{(\omega)^2}y_i^{(\omega')^2}\Big). \qquad (54)
\end{aligned}
$$

Moreover, all terms involving only combinations of $\omega < d = 2^n$ do not depend on the key, thus we can further simplify to the required equation

$$
\begin{aligned}
\sum_{\omega\in\mathbb{F}_2^n}\frac{1}{\sigma^{(\omega)^2}}\left(\sum_{i=1}^{q}x_i^{(\omega)}x_i^{(2^n)}\sum_{m\in\mathbb{F}_2^n}y^{(\omega)}y^{(2^n)}-\frac{1}{2}x_i^{(2^n)}\sum_{m\in\mathbb{F}_2^n}y^{(2^n)}y^{(\omega)^2}\right. \qquad (55)\\
\left.-\frac{1}{2}x_i^{(\omega)}\sum_{m\in\mathbb{F}_2^n}y^{(\omega)}y^{(2^n)^2}+\frac{1}{4}\sum_{m\in\mathbb{F}_2^n}y^{(\omega)^2}y^{(2^n)^2}\right). \qquad \square
\end{aligned}
$$

Proposition 13 (Relationship between HOOD and CPA for masking tables). *When all noise variances are equal, i.e., $\sigma = \sigma^{(\omega)}\ \forall\omega$, Eq. (50) further simplifies to*

$$
\begin{aligned}
\mathcal{D}_{opt}^{mt,G,\sigma\uparrow}(\mathbf{x}^{(\star)},\mathbf{t}) = \underset{k\in\mathcal{K}}{\arg\max}\sum_{\omega\in\mathbb{F}_2^n}\sum_{i=1}^{q}\Big(&x_i^{(\omega)}x_i^{(2^n)}\sum_{m\in\mathbb{F}_2^n}\varphi(\omega\oplus m)\varphi(t_i\oplus k\oplus m)\\
&-\frac{1}{2}x_i^{(\omega)}\sum_{m\in\mathbb{F}_2^n}\varphi(\omega\oplus m)\varphi^2(t_i\oplus k\oplus m)\Big), \qquad (56)
\end{aligned}
$$

which becomes close to a combination of higher-order CPAs, i.e.,

$$
\begin{aligned}
\mathcal{D}_{C\text{-}CPA}^{mt,\sigma\uparrow}(\mathbf{x}^{(\star)},\mathbf{t}) = \underset{k\in\mathcal{K}}{\arg\max}\sum_{\omega\in\mathbb{F}_2^n}&\rho(c_X^{n\text{-}prod}(\mathbf{x}^{(\omega)},\mathbf{x}^{(2^n)}),c_Y^{opt}(\mathbf{y}^{(\omega)},\mathbf{y}^{(2^n)})) \qquad (57)\\
&-\frac{1}{2}\rho(\mathbf{x}^{(\omega)},c_Y^{opt}(\mathbf{y}^{(\omega)},\mathbf{y}^{(2^n)^2})).
\end{aligned}
$$

Proof. If all the variances are equal we have

$$
\sum_{\omega\in\mathbb{F}_2^n}\frac{\varphi^2(\omega\oplus m)}{\sigma^{(\omega)}}=\frac{1}{\sigma}\sum_{\omega\in\mathbb{F}_2^n}\varphi^2(\omega\oplus m)=\frac{1}{\sigma}\sum_{\omega\in\mathbb{F}_2^n}\varphi^2(\omega). \qquad (58)
$$

So, regarding the second term in Eq. (50) we have

$$\sum_{\omega\in\mathbb{F}_2^n} \frac{1}{\sigma^{(\omega)2}} \sum_{i=1}^{q} x_i^{(2^n)} \sum_{m\in\mathbb{F}_2^n} \varphi(t_i\oplus k\oplus m)\varphi(\omega\oplus m)^2 \tag{59}$$

$$= \sum_{i=1}^{q} x_i^{(2^n)} \sum_{m\in\mathbb{F}_2^n} \varphi(t_i\oplus k\oplus m) \sum_{\omega\in\mathbb{F}_2^n} \frac{1}{\sigma^{(\omega)2}} \varphi(\omega\oplus m)^2 \tag{60}$$

$$= \sum_{i=1}^{q} x_i^{(2^n)} \sum_{m\in\mathbb{F}_2^n} \varphi(t_i\oplus k\oplus m) \sum_{\omega\in\mathbb{F}_2^n} \frac{1}{\sigma^2} \varphi(\omega)^2 \tag{61}$$

$$= \sum_{i=1}^{q} x_i^{(2^n)} \sum_{m\in\mathbb{F}_2^n} \varphi(t_i\oplus m) \sum_{\omega\in\mathbb{F}_2^n} \frac{1}{\sigma^2} \varphi(\omega)^2, \tag{62}$$

which clearly does not depend on the key k. The same goes for the fourth term, which proofs the first part. Now, rewriting Eq. (56) gives

$$\begin{aligned} \arg\max_{k\in\mathcal{K}} \sum_{\omega\in\mathbb{F}_2^n} &\langle \mathbf{x}^{(\omega)}\mathbf{x}^{(2^n)} \mid \sum_{m\in\mathbb{F}_2^n} \varphi(\omega\oplus m)\varphi(\mathbf{t}\oplus k\oplus m)\rangle \\ &-\langle \frac{1}{2}\mathbf{x}^{(\omega)} \mid \sum_{m\in\mathbb{F}_2^n} \varphi(\omega\oplus m)\varphi^2(\mathbf{t}\oplus k\oplus m)\rangle, \end{aligned} \tag{63}$$

and using the same argumentation as in the proof of Prop. 9 gives the required formula from the second part. □

Interestingly, instead of using one CPA to recover the mask and one to recover the secret key (see Def. 2) we recover that the best methodology is to attack each share $\omega < 2^n$ with $\omega = 2^n$ (minus a regulation term) and then use a combination of all attacks. Note again that we can make the same relaxations about the leakage model as done in Subsect. 3.3.

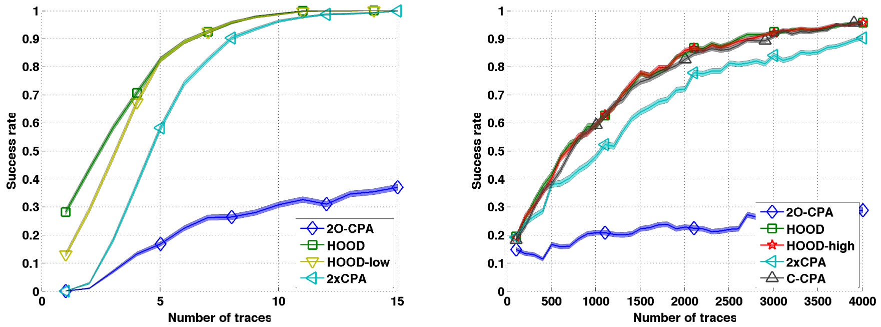
Remark 6. For low noise, we can straightforwardly use Prop. 10, which is validated in our empirical results.

5.3 Experimental Validation

To empirically validate our theoretical results we use simulations of a first order masking scheme with precomputation tables. We target the xor operation in the precomputation phase and the AddRoundKey of the algorithm (see line 3 and line 8 of Alg. 1 in Appendix A).

Thus, we have the same leakages as depicted in Examples 2, where for computationally reasons for all distinguishers we only target four bits ($n = 4$).

Remark 7. Targeting the AddRoundKey phase has some advantages. First, it allows to perform the evaluation on only four bits without the loss of generality of using a four bits Sbox. Second, in the Sbox precomputation algorithm of



(a) $\sigma = 0.5$

(b) $\sigma = 5$

Fig. 2: Success rate for masking tables

Coron [6] the output masks are different for each entry of the Sbox and could therefore not be combined with the mask of the precomputation table. However, as in our analysis the attacker can still take advantage of the 2^n leakages of the masked inputs of the Sbox combined with the leakage of the AddRoundKey operation.

Similarly to the previous experiments, T is uniformly distributed over $\mathbb{F}_2^4$ and the noise is arising from a Gaussian distribution $N \sim \mathcal{N}(0, \sigma^2)$ for $\sigma = \sigma^{(0)} = ... = \sigma^{(16)} \in \{0.5, 5\}$. Again to compute the success rate we conducted 500 independent experiments with uniformly distributed k^* and shaded the success rate with error bars.

Figure 2 shows the success rates. For low noise ($\sigma = 0.5$) the optimal distinguisher (HOOD) and its approximation for low noise (HOOD-low) perform similar and better than the 2*nd*-order CPA (2O-CPA) with normalized product combination function and the 2-stage CPA in Eq. 45 (2xCPA). Naturally, all distinguishers outperform 2nd-order CPA as it only utilizes two leakages $X^{(0)}$ and $X^{(256)}$. For higher noise ($\sigma = 5$) the HOOD and its approximation for high noise (HOOD-high) perform better than the 2-stage CPA (2xCPA) and 2nd-order CPA. Moreover it can be noticed that the distinguisher based on combinations of CPA (Eq. (57)) (C-CPA) and the optimal ones are equally efficient. Accordingly, we have empirically validated that our new distinguisher approximated from the HOOD is valid for high noise and more efficient than the two-stage CPA. In particular, it requires around 1000 traces less to reach $\hat{\mathbb{P}}_S = 90\%$ for $\sigma = 5$.

6 Conclusions and Perspectives

We have found the optimal distinguishers for higher-order masking, and especially, analyzed the application of second-order distinguisher and distinguisher against masking tables. This gives the first theoretical proof that for a high noise non-profiled second-order CPA becomes as efficient as the optimal distinguisher in terms of success rate. In particular, we explain that the normalized product combining function with the optimal prediction function [16] is sound and the optimal one among all (known and unknown) combination functions. We furthermore extended this result to $(d+1)$th-order distinguisher, which has not been analyzed before. For low noise, the optimal distinguisher does not reduce to any kind of correlation. In the application of masking tables we provide a new distinguisher based on correlation whose again is as efficient as the optimal distinguisher in case of high noise. Naturally, this new distinguisher outperforms all known (non-profiled) distinguisher for this application. Given all these results we theoretically and empirically show that for high noise the security analysis with non-profiled distinguisher is sufficient as it coincides with the optimal distinguisher.

These results raise various new perspectives. First of all, our methodology of starting from the optimal distinguisher and deriving approximated distinguisher could be applied to other scenarios. One application, for example, could be the scenario used in [17]. Moreover, future work should deal with the exact analysis of the impact of noise on the masking efficiency in a theoretical manner. This comes along with an analysis of the impact of the number of shares, in particular, with an investigation of the arguments done in [15, 23] about exponential attack complexity.

Acknowledgement

The authors would like to thank the anonymous reviewers for their valuable comments and suggestions to improve the quality of the paper

References

1. Mehdi-Laurent Akkar and Christophe Giraud. An Implementation of DES and AES Secure against Some Attacks. In LNCS, editor, *Proceedings of CHES'01*, volume 2162 of *LNCS*, pages 309–318. Springer, May 2001. Paris, France.
2. Johannes Blömer, Jorge Guajardo, and Volker Krummel. Provably Secure Masking of AES. In Helena Handschuh and M. Anwar Hasan, editors, *Selected Areas in Cryptography*, volume 3357 of *Lecture Notes in Computer Science*, pages 69–83. Springer, 2004.
3. Claude Carlet, Finley Freibert, Sylvain Guilley, Michael Kiermaier, Jon-Lark Kim, and Patrick Solé. Higher-order cis codes. *Information Theory, IEEE Transactions on*, 60(9):5283–5295, Sept 2014.
4. Suresh Chari, Charanjit S. Jutla, Josyula R. Rao, and Pankaj Rohatgi. Towards Sound Approaches to Counteract Power-Analysis Attacks. In *CRYPTO*, volume 1666 of *LNCS*. Springer, August 15-19 1999. Santa Barbara, CA, USA. ISBN: 3-540-66347-9.
5. Suresh Chari, Josyula R. Rao, and Pankaj Rohatgi. Template Attacks. In *CHES*, volume 2523 of *LNCS*, pages 13–28. Springer, August 2002. San Francisco Bay (Redwood City), USA.
6. Jean-Sébastien Coron. Higher order masking of look-up tables. In Phong Q. Nguyen and Elisabeth Oswald, editors, *EUROCRYPT*, volume 8441 of *Lecture Notes in Computer Science*, pages 441–458. Springer, 2014.

7. A. Adam Ding, Liwei Zhang, Yunsi Fei, and Pei Luo. A Statistical Model for Higher Order DPA on Masked Devices. Cryptology ePrint Archive, Report 2014/433, June 2014. `http://eprint.iacr.org/2014/433/` (to appear at CHES 2014).
8. R. A. Fisher. On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London, A*, 222:309–368, 1922.
9. Benedikt Gierlichs, Lejla Batina, Bart Preneel, and Ingrid Verbauwhede. Revisiting Higher-Order DPA Attacks: Multivariate Mutual Information Analysis. In *CT-RSA*, volume 5985 of *LNCS*, pages 221–234. Springer, March 1-5 2010. San Francisco, CA, USA.
10. Marc Joye, Pascal Paillier, and Berry Schoenmakers. On Second-Order Differential Power Analysis. In *CHES*, volume 3659 of *LNCS*, pages 293–308. Springer, August 29 – September 1st 2005. Edinburgh, UK.
11. Paul C. Kocher, Joshua Jaffe, and Benjamin Jun. Differential Power Analysis. In *Proceedings of CRYPTO'99*, volume 1666 of *LNCS*, pages 388–397. Springer-Verlag, 1999.
12. 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, *ICICS*, volume 7618 of *LNCS*, pages 331–340. Springer, October 29-31 2012. Hong Kong.
13. Thomas S. Messerges. Using Second-Order Power Analysis to Attack DPA Resistant Software. In *CHES*, volume 1965 of *LNCS*, pages 238–251. Springer-Verlag, August 17-18 2000. Worcester, MA, USA.
14. Elisabeth Oswald and Stefan Mangard. Template Attacks on Masking — Resistance Is Futile. In Masayuki Abe, editor, *CT-RSA*, volume 4377 of *Lecture Notes in Computer Science*, pages 243–256. Springer, 2007.
15. Emmanuel Prouff and Matthieu Rivain. Masking against Side Channel Attacks: a Formal Security Proof. In *EUROCRYPT*, volume 7881 of *LNCS*, pages 142–159. Springer, May 2013. Athens, Greece.
16. Emmanuel Prouff, Matthieu Rivain, and Régis Bevan. Statistical Analysis of Second Order Differential Power Analysis. *IEEE Trans. Computers*, 58(6):799–811, 2009.
17. Thomas Roche and Victor Lomné. Collision-Correlation Attack against Some 1st-Order Boolean Masking Schemes in the Context of Secure Devices. In Emmanuel Prouff, editor, *COSADE*, volume 7864 of *Lecture Notes in Computer Science*, pages 114–136. Springer, 2013.
18. Kai Schramm and Christof Paar. Higher Order Masking of the AES. In David Pointcheval, editor, *CT-RSA*, volume 3860 of *LNCS*, pages 208–225. Springer, 2006.
19. F. X Standaert, E. Peeters, and J-J Quisquater. On the masking countermeasure and higher-order power analysis attacks. In *Information Technology: Coding and Computing, 2005. ITCC 2005. International Conference on*, volume 1, pages 562–567 Vol. 1, 2005.
20. 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 *ASIACRYPT*, volume 6477 of *LNCS*, pages 112–129. Springer, December 5-9 2010.
21. Adrian Thillard, Emmanuel Prouff, and Thomas Roche. Success through Confidence: Evaluating the Effectiveness of a Side-Channel Attack. In Guido Bertoni and Jean-Sébastien Coron, editors, *CHES*, volume 8086 of *Lecture Notes in Computer Science*, pages 21–36. Springer, 2013.
22. Michael Tunstall, Carolyn Whitnall, and Elisabeth Oswald. Masking Tables - An Underestimated Security Risk. In Shiho Moriai, editor, *FSE*, volume 8424 of *Lecture Notes in Computer Science*, pages 425–444. Springer, 2013.

23. Jason Waddle and David Wagner. Towards Efficient Second-Order Power Analysis. In *CHES*, volume 3156 of *LNCS*, pages 1–15. Springer, 2004. Cambridge, MA, USA.

24. Carolyn Whitnall, Elisabeth Oswald, and François-Xavier Standaert. The Myth of Generic DPA...and the Magic of Learning. In Josh Benaloh, editor, *CT-RSA*, volume 8366 of *Lecture Notes in Computer Science*, pages 183–205. Springer, 2014.

A Algorithm of masking tables

```
   input  : $t$, one byte of plaintext, and $k$, one byte of key
   output : The application of AddRoundKey and SubBytes on $t$, i.e., $S(t \oplus k)$
1  $m \leftarrow_{\mathcal{R}} \mathbb{F}_2^n$, $m' \leftarrow_{\mathcal{R}} \mathbb{F}_2^n$ // Draw of random input and output masks ;
2  for $\omega \in \{0, 1, \ldots, 2^n - 1\}$ do // Sbox masking
3      $z \leftarrow \omega \oplus m$ // Masked input ;
4      $z' \leftarrow S[\omega] \oplus m'$ // Masked output ;
5      $S'[z] \leftarrow z'$ // Creating the masked Sbox entry ;
6  end
7  $t \leftarrow t \oplus m$ // Plaintext masking ;
8  $t \leftarrow t \oplus k$ // Masked AddRoundKey ;
9  $t \leftarrow S'[t]$ // Masked SubBytes ;
10 $t \leftarrow t \oplus m'$ // Demasking ;
11 return $t$
```

Algorithm 1: Beginning of a block cipher masked by Sbox precomputation

We have indicated the words length of all data as n, typically, $n = 8$ bit for AES. Two random masks m and m' are drawn initially from $\mathbb{F}_2^n$ and all the data manipulated by the algorithm will be exclusive-ored with one of the two masks.

Masking the plaintext is straightforward (see line 7). Key addition can be done safely as a second step, as the plaintext is already masked (see line 8). Passing through the Sbox is less obvious, as this operation is non-linear. Therefore, the Sbox is recomputed masked, as shown on lines 2 to 5: a new table S', that has also size $2^n \times n$ bits, is required for this purpose. In the Sbox precomputation step (lines 2 to 5), the key byte k is not manipulated. The leakage only concerns the mask.