Stochastic Collision Attack

Nicolas Bruneau, Claude Carlet, Sylvain Guilley, Annelie Heuser, Emmanuel Prouff, Olivier Rioul

Stochastic Collision Attack

Nicolas Bruneau, Claude Carlet, Sylvain Guilley, *Member, IEEE*, Annelie Heuser, Emmanuel Prouff, and Olivier Rioul

***Abstract*— On the one hand, collision attacks have been introduced in the context of side-channel analysis for attackers who exploit repeated code with the same data without having any knowledge of the leakage model. On the other hand, stochastic attacks have been introduced to recover leakage models of internally processed intermediate secret variables. Both techniques have shown advantages and intrinsic limitations. Most collision attacks, for instance, fail in exploiting all the leakages (e.g., only a subset of matching samples are analyzed), whereas stochastic attacks cannot involve linear regression with the full basis (while the latter basis is the most informative one). In this paper, we present an innovative attacking approach, which combines the flavors of stochastic and collision attacks. Importantly, our attack is derived from the optimal distinguisher, which maximizes the success rate when the model is known. Notably, we develop an original closed-form expression, which shows many benefits by using the full algebraic description of the leakage model. Using simulated data, we show in the unprotected case that, for low noise, the stochastic collision attack is superior to the state of the art, whereas asymptotically and thus, for higher noise, it becomes equivalent to the correlation-enhanced collision attack. Our so-called stochastic collision attack is extended to the scenario where the implementation is protected by masking. In this case, our new stochastic collision attack is more efficient in all scenarios and, remarkably, tends to the optimal distinguisher. We confirm the practicability of the stochastic collision attack thanks to experiments against a public data set (DPA contest v4). Furthermore, we derive the stochastic collision attack in case of zero-offset leakage that occurs in protected hardware implementations and use simulated data for comparison. Eventually, we underline the capability of the new distinguisher to improve its efficiency when the attack multiplicity increases.**

***Index Terms*— Side-channel analysis, collision attacks, optimal distinguisher, masking.**

I. Introduction

SIDE-CHANNEL attacks consist in exploiting some leakage occurring during the computation of a cryptographic algorithm. The best way to exploit leakages is known in general: whenever the context permits, the leakage is profiled and then a maximum likelihood distinguisher is applied [13], [45]. However, for some devices, like for banking or the ePassport, the attacker might not be able to hold an open copy of the device (i.e. a copy with full control), which prevents him from doing any profiling. Additionally, for some particular architectures, the leakage may be very specific to each device (owing for instance to technological dispersion). For scenarii like the two previous ones, attacks built on a maximum likelihood approach are impossible and a deeper analysis is thus needed to identify the most pertinent/efficient distinguishers, especially when the leakage nature cannot be captured by the classical models (as e.g. the Hamming weight).

A. Device-Specific Leakage

One such situation is when every device has a specific leakage model. This can happen for several reasons. For instance, in deep submicron (DSM) complementary metal-oxide-semiconductor (CMOS) technologies, the variability in fabrication can be large, hence the leakage model is unpredictable [34]. Such variability is, in practice, a drawback in terms of yield and reliability, since safety margins must be considered regarding the performances of each chip. Nonetheless, it can be turned into a constructive feature, for instance for physically unclonable functions (PUF). Such devices precisely require a deterministic behavior for every unique device, but an unpredictable behavior for a device which has not been characterized. Security-wise, process variation also has the nice advantage of making profiled side-channel attacks ineffective. Indeed, a leakage model learned from one design does not apply to another. While some "porting" of precharacterized template is possible when the acquisition conditions changed (e.g., using traces normalization [18], [30]), this situation is not possible when chips differ intrinsically. A second reason for the leakage to be specific to a chip is when countermeasures are applied, which aim at reducing the differences in the leakage due to the data. On hardware circuits, this technique is known as dual-rail precharge logic (see e.g., [43]). Ideally, the leakage of such circuits is the same for all data: we say that the countermeasure reduces the signal-to-noise ratio (SNR). But in practice, the SNR is not exactly equal to zero, owing to tiny technological unbalances which create leaks. Now, it is a priori difficult to predict theleakage model. For instance, it is difficult to know whether a bit (e.g., a gate) will leak positively or negatively. Similar protection goal can be reached on FPGAs using a double access to memory (see

Manuscript received November 12, 2016; revised February 23, 2017; accepted April 5, 2017. Date of publication April 24, 2017; date of current version June 14, 2017. This work was supported by the ANR CHIST-ERA project SECODE (Secure Codes to thwart Cyber-physical Attacks). The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Ozgur Sinanoglu. *(Corresponding author: Sylvain Guilley.)*

N. Bruneau and S. Guilley are with Secure-IC S.A.S., and also with LTCI, Telecom-ParisTech, Université Paris-Saclay, France (e-mail: sylvain.guilley@telecom-paristech.fr).

C. Carlet is with LAGA, UMR 7539, CNRS, University of Paris VIII and University of Paris XIII, France.

A. Heuser is with IRISA, France, and also with CNRS, France.

E. Prouff is with the Groupe SAFRAN, France.

O. Rioul is with LTCI, Telecom-ParisTech, Université Paris-Saclay, France.

e.g., [4], [44]), and also to software (see e.g., [14], [39]). Of course, balanced circuits implemented in DSM technology are even less amenable to template attacks [34].

B. State-of-the-Art About Unprofiled Attacks

In this situation where the leakage is specific to the chip under attack or simply unknown, only unprofiled strategies are suitable. One of them consists in assuming a hypothetical model, which is supposed to be close enough to reality. A subsequent correlation attack (also known as "correlation power analysis" or CPA [8]) is performed to extract the key. The difference of means (DoM) distinguisher has been introduced by Kocher *et al.* [24]: technically, it acts as a CPA, but as it operates on only one single bit, it is less sensitive to model errors. Another one is called mutual information analysis (MIA [20]), which selects the best key according to the entropy of a partitioning. However, for the attack to be sound, it is known that the partitioning must not be injective in the key. Hence, an educated guess of such leakage model is required.

In situations where the leakage model is known up to a parameterization, one can resort to the so-called linear regression analysis (LRA, see e.g., [17]). LRA is the non-profiled version of the stochastic approach (see e.g., [36]) that aims at inferring the model in a profiling phase. In case of profiling it is possible to use a *full basis* to accurately characterize the leakage [21]. However, for LRA one has to be careful to select an appropriate basis that is not too large. Otherwise, indeed, the distinguisher becomes not sound, since the method succeeds in recovering the functional dependency between the leakage and the hypothesis even when the key guess is wrong (see e.g., [17]). Consequently, even for LRA, the attacker has to make some assumptions about the algebraic properties of the underlying leakage model.

Collision attacks have the nice property that they allow to circumvent the modeling issue. They have been introduced by Schramm *et al.* [38] on DES and in [37] on AES, and consist in identifying same inputs to some repeated function (e.g., a substitution box (SBox)) through a side-channel. These inputs are *sensitive values*, that is intermediate values which are input to a subfunction of the cryptographic algorithm, and depend on a subkey and a part of the cleartext (or ciphertext). Besides, we can say that collision attacks are *unsupervised* side-channel attacks, since they do not require a training phase. Bogdanov presented generalized collisions in [6]: collisions can be found between different executions of AES. Bogdanov refined the collision detection method by binary and ternary voting in [7]. By design, these attacks are ignorant of the leakage model, and rely on only one assumption: repeated calls to a given code with the same arguments leak quite similarly. Unfortunately, the classical collision attack cannot cope with masking countermeasures. The work of Clavier *et al.* [15] follows the same path, but adapts to the context of side-channel, where collisions can be detected only after averaging traces. The pro of such attacks is that the collision detection probability is indeed improved despite noise, but the con is that many traces are needed to test each key hypothesis.

Moradi *et al.* [33] abstract the notion of collision by first estimating some instantaneous leakage moments corresponding to the manipulation of SBox outputs and in [31] to masking material. The pro of such an approach is that each trace contributes globally to the attack (it serves for all key hypothesis), whereby the leakage moments have to be estimated accurately enough for each class. The drawback in the case of masking countermeasures is that only independent characteristics (e.g., leakage distribution moments) of each SBox are extracted and then individually compared.

We overcome these limitations with a joint analysis of the twain SBoxes at once. In our approach, the collision is simply a constraint on a stochastic regression attack which assumes that SBoxes have globally the same leakage model.

C. Our Approach: Stochastic Collision Attack

In this paper, we present an innovative attacking approach, which combines the flavours of stochastic and collision attacks. Importantly, our attack is derived from the optimal distinguisher [22], which maximizes the success rate when the model is known. In particular, our attack does not need to make any assumption on the unknown leakage model, except that the deterministic leakage repeats when the code is reused. Hence we adhere to the seminal idea of Schramm *et al*. and of Bogdanov, in that we intend to exploit only collisions (namely *generalized collisions* between different executions of the block cipher). But in addition, we take advantage of the parameterization technique of LRA and the profiled stochastic approach. Used at its maximum power, namely with a full basis, the LRA is known not to be sound as a distinguisher. However, with our approach of using collisions, the same model is reused and we achieve soundness. Moreover, we extend our approach to masking countermeasures.

D. Contributions

We derive our novel stochastic collision attack and importantly give a closed form of it that is not straightforward. This closed-form is quite original and made possible by a stochastic characterization over a full basis. We conduct an exhaustive comparison of the distinguishers and highlight the efficiency of the stochastic collision attack by simulations and attacks on real traces. Our experiments show that the stochastic collision attack and the correlation-enhanced collision attack outperform the classical collision attack in any scenario. This is mainly due to the fact that the collision attack is a chosen plaintext attack that only uses a fraction of the traces, whereas the stochastic and the correlation-enhanced collision attacks are known plaintext attacks involving the complete set of traces. For low noise, the stochastic collision attack is superior to the correlation-enhanced collision attack, while they become closer when the noise increases. Even more, our mathematical derivation reveals that both distinguishers become asymptotically equivalent.

Interestingly, we show that it is worthwhile, in terms of success rate, if enough resources are available, to compute the stochastic collision attack of dimensionality L as large as

possible. Note that this feature does not apply on the state-of-the-art collision attacks as they are restricted only to $L = 2$ due to their underlying statistical method.

Furthermore, we derive the stochastic collision attack in case of masking when each SBox is masked with the same mask or the masks are related to each other as described in [12][1] and used for the DPA contest v4. In this case, the distinguisher does not reduce to a simple closed form expression and we show how to perform the optimization using the Expectation-Maximization (EM) algorithm (see e.g., [29], [40]). Remarkably, our simulations and experiments show that the stochastic collision distinguisher outperforms all state-of-the-art collision attacks for masking and that our attack converges to the optimal distinguisher. In particular, we compare our attack to the correlation-collision attack of Clavier *et al.* [15] and with the adaption for software implementations of Moradi [31]. Additionally, we adapt the attacks to the case of zero-offset leakage as present in masked hardware implementations.

E. Outlines

The new distinguisher is mathematically derived in Sec. II, and contrasted with the state-of-the-art collision attacks. Validation on simulated and real data is done in Sec. III. We derive our new distinguisher in case of masking with a theoretical and practical comparison to the state-of-the-art in Sec IV. Section V concludes and opens some perspectives. Appendix A provides a comparison between two variants of correlation-collision attacks and more details about the correlation-enhanced collision attack is given in appendix B.

II. Stochastic Collision Distinguisher

A. Preliminaries and Notation

In the sequel, we are interested in block ciphers which can be attacked by collisions. The access to substitution boxes (SBoxes) is especially leaky, because they consist in a memory look-up, therefore favorable case studies are block ciphers which reuse the same instance of SBox several times. This is the case for AES and PRESENT. We denote by n the fan-in (i.e., the number of inputs) of the SBox, that is $n = 8$ bits for AES and $n = 4$ bits for PRESENT. The implementation can be either hardware, or software (in which case the same memory is recalled sixteen times). For $\ell \in \{1, \dots, L\}$ (e.g. $L = 16$ for AES), we denote by $k^{*(\ell)} \in \mathbb{F}_2^n$ the ℓ-th secret key byte and by $k^{(\ell)}$ any possible key hypothesis thereof. The ℓ-th byte coordinate of the plaintext corresponding to the qth query ($1 \leq q \leq Q$) is denoted by $t_q^{(\ell)}$. The associated leakage is denoted by $x_q^{(\ell)}$.

For all the indices ℓ, the leakage models are assumed to be identical[2] but unknown, which leads to view $x_q^{(\ell)}$ as realizations of the following random variable:

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

[1]Extended version of [11] with more results in appendix.

[2]This assumption is reasonable when the code or the hardware is reused; this occurs in practice for the sake of cost overhead mitigation.

where the noise N is independent among ℓ as well as q and $\varphi = \psi \circ S$ is a composition of two deterministic functions. In particular, the SBox function $S : \mathbb{F}_2^n \to \mathbb{F}_2^n$ is algorithmic specific, whereas $\psi : \mathbb{F}_2^n \to \mathbb{R}$ is an unknown function related to the device architecture.

In the rest of the paper, we will additionally employ the following shortcut notations:

- $\vec{k}$ denotes the vector $(k^{(1)}, \cdots, k^{(L)})$,
- $\vec{x}^{(\cdot)}$ is the $Q \times L$ matrix whose rows correspond to L-variate leakages,
- to each matrix element $x_q^{(\ell)}$ of $\vec{x}^{(\cdot)}$ is associated a plaintext element $t_q^{(\ell)}$ and key element $k^{(\ell)}$ (note that, as usual, for each ℓ the key element $k^{(\ell)}$ is assumed to be the same for all the $t_q^{(\ell)}$ and $x_q^{(\ell)}$,
- $\vec{x}^{(\ell)}$ is a list of the Q leakages which are the consequence of the processing of $(t_q^{(\ell)})_{1 \leq q \leq Q}$,
- $\vec{x}_q$ is a list of the L leakages which are the consequence of the processing of $(t_q^{(\ell)})_{1 \leq \ell \leq L}$. It is the qth row of $\vec{x}^{(\cdot)}$. The corresponding vector of plaintext elements is denoted by $\vec{t}_q$.

For any vector $\vec{y} \in \mathbb{R}^L$, we will denote by $\|\vec{y}\|$ the *Euclidean norm* of $\vec{y}$. Moreover, we will denote by f_{μ,σ^2} the *Normal distribution* with mean μ and standard deviation σ. The notation will be simplified into f_{σ^2} when the mean is assumed to be null (note that for any y we have $f_{\mu,\sigma^2}(y) = f_{\sigma^2}(y-\mu)$). Eventually, for any function φ defined from a set E to a set F and any vector $\vec{y} \doteq (y_1, \cdots, y_L) \in E^L$, we will simply denote by $\varphi(\vec{y})$ the L-dimensional vector $(\varphi(y_1), \cdots, \varphi(y_L))$.

B. New Concept of Stochastic Collision Distinguisher

A *distinguisher* is a function which takes the known plaintexts and the measured leakages, and returns a key guess. A specific distinguisher is *optimal* when maximizing the success probability of the attack [22]. Besides considering univariate first-order distinguisher, two other scenarios have been considered so far: first, in [10] the authors derived optimal distinguishers when applied on masking countermeasures and second, [9] studied optimal distinguisher when dimension reduction is helpful. Thus, both previous studies considered multiple leaking points that are related to only one key byte and plaintext.

In the next proposition we derive the optimal distinguisher in the scenario of collision attacks, namely considering L independent leakage samples with L different key hypotheses and plaintexts. For the paper to be self-content, we recall in Fig. 1 the principle of a collision in a side-channel context.

Proposition 1 (Optimal Collision Based Distinguisher): The optimal collision based distinguisher, when the noise is Gaussian and isotropic in the L samples, namely $N^{(\cdot)} \sim \mathcal{N}(0, \mathrm{diag}(\sigma^2, \dots, \sigma^2))$, does not depend on σ^2, and can be expressed as:

$$\mathcal{D}_{\mathrm{opt}}(\vec{t}^{(\cdot)}, \vec{x}^{(\cdot)}) = \underset{\vec{k} \in (\mathbb{F}_2^n)^L}{\mathrm{argmin}} \sum_{q=1}^{Q} \left\| \vec{x}_q - \varphi(\vec{t}_q \oplus \vec{k}) \right\|^2. \tag{2}$$

Proof: The optimal distinguisher $\mathcal{D}_{\mathrm{opt}}$ is the ML (Maximum Likelihood). We write p for random variables' densities.

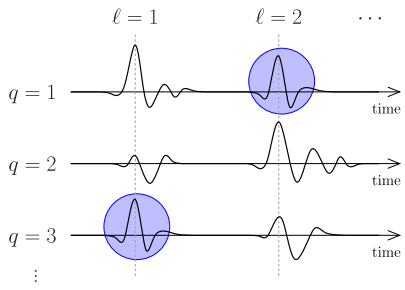

Fig. 1. Setup for a collision attack. The highlighted parts of the traces (blue circles) correspond to the following collision: $t_1^{(2)} \oplus k^{(2)} = t_3^{(1)} \oplus k^{(1)}$.

For example, p_N denotes the density of N. We can thus express the ML distinguisher as follows:

$$\begin{aligned}\mathcal{D}_{\text{opt}}(\vec{t}^{(\cdot)}, \vec{x}^{(\cdot)}) &= \underset{\vec{k}\in(\mathbb{F}_2^n)^L}{\text{argmax}}\, p(\vec{x}^{(\cdot)}|\vec{t}^{(\cdot)}, \vec{k}) \\ &= \underset{\vec{k}\in(\mathbb{F}_2^n)^L}{\text{argmax}} \prod_{q=1}^{Q} p_{N_q^{(\cdot)}}\left(\vec{x}_q - \varphi(\vec{t}_q \oplus \vec{k})\right) \qquad (3) \\ &= \underset{\vec{k}\in(\mathbb{F}_2^n)^L}{\text{argmax}} \sum_{q=1}^{Q} \log \prod_{\ell=1}^{L} f_{\sigma^2}(x_q^{(\ell)} - \varphi(t_q^{(\ell)} \oplus k^{(\ell)})), \\ & \qquad (4)\end{aligned}$$

which implies (2) by simply developing the normal distribution law $f_{\sigma^2}(\cdot)$, and by removing the key-independent terms. Basically, Eq. (3) results from the noise independence from trace to trace. Eq. (4) arises from the Gaussian distribution of $N^{(\cdot)}$, which is assumed *isotropic*, and from the fact that the logarithm is a strictly increasing function. ■

Remark 2: The hypothesis of noise isotropy in the sample space is often satisfied in practice. This can be justified as follows. Assuming that the thermal noise is white, the correlation window length is inversely proportional to the bandwidth of the measurement apparatus (typically 1 GHz). Therefore, the correlation between noisy samples does not extend beyond 1 ns. In practice, the Sbox calls are at least separated by one clock period, that is 10 to 100 ns. Therefore, the measurement noise samples are independent from one call to another. Still, if the L noise samples are correlated, the Proposition 1 would still hold, by trading the Euclidean distance by the Mahalanobis distance [27].

Unfortunately, Eq. (2) has one major drawback, as discussed for other optimal distinguishers in [9], [10], and [22] it can only be computed provided the leakage model (the function φ) is completely known. However, naturally, the strength of the state-of-the-art collision attacks is the unnecessariness of the knowledge of the leakage model. To cope with this problem we extend Proposition 1 to the case where the model becomes part of the *secrets* to be guessed as similarly done in LRA.

Definition 3 (Stochastic Collision Distinguisher):

$$\mathcal{D}_{\text{sto.coll}}(\vec{t}^{(\cdot)}, \vec{x}^{(\cdot)}) = \underset{\vec{k}\in(\mathbb{F}_2^n)^L}{\text{argmin}} \min_{\varphi:\mathbb{F}_2^n\to\mathbb{R}} \sum_{q=1}^{Q} \left\|\vec{x}_q - \varphi(\vec{t}_q \oplus \vec{k})\right\|^2,$$

where φ lives in the vector space of the functions $\mathbb{F}_2^n \to \mathbb{R}$.

Definition 3 involves a minimization over all leakage functions $\varphi : \mathbb{F}_2^n \to \mathbb{R}$, which appears as an unsolvable task if no information about the underlying algorithm and leakage model is known.[3] In order to solve this issue, we choose a basis decomposition for the unknown *univariate* leakage function φ. Note that, $\varphi : \mathbb{F}_2^n \to \mathbb{R}$ is a pseudo-Boolean function that can be seen as a 2^n-dimensional vector $\varphi(\cdot) \in \mathbb{R}^{2^n}$. Here $\mathbb{R}^{2^n}$ is the usual finite-dimensional Euclidean space equipped with the canonical scalar product $\langle \varphi \mid \varphi' \rangle = \sum_t \varphi(t)\varphi'(t)$. The key point is that for the stochastic collision attack we can make use of the *full (canonical) basis*, i.e. basis of highest degree, as the distinguisher involves (at least) two leakage samples.

Lemma 4: The stochastic collision attack of *highest degree* can be rewritten as follows:

$$\begin{aligned}&\mathcal{D}_{\text{sto.coll}}(\vec{t}^{(\cdot)}, \vec{x}^{(\cdot)}) \\ &\quad = \underset{\vec{k}\in(\mathbb{F}_2^n)^L}{\text{argmin}} \min_{\vec{a}\in\mathbb{R}^{2^n}} \sum_{q=1}^{Q} \left\|\vec{x}_q - \sum_{u\in\mathbb{F}_2^n} a_u \delta_u(\vec{t}_q \oplus \vec{k})\right\|^2, \qquad (5)\end{aligned}$$

where $\delta_u(t) = 1$ if $t = u$ and 0 otherwise.

Proof: In Definition 3 the leakage model space can be mathematically described using an orthonormal basis in $\mathbb{R}^{2^n}$ (see e.g., [36]). Using the canonical basis $(\delta_u)_{u\in\mathbb{F}_2^n}$, which is evidently orthonormal, simply gives $\varphi(t) = \sum_{u\in\mathbb{F}_2^n} \varphi(u)\delta_u(t)$. In Eq. (5), we denoted $\vec{a} = (\varphi(u))_{u\in\mathbb{F}_2^n}$. ■

Still, the stochastic collision distinguisher as presented in Eq. (5) is not satisfactory. Indeed, it is not a mathematical closed form, hence some optimization algorithm shall be run in order to minimize over $\vec{a} \in \mathbb{R}^{2^n}$. As the function inside the minimization (over $\vec{a}$) in Eq. (5) is actually convex, there is a global minimum, and a projection could hence be done as a first step of the optimization. However, the derivation of the global minimum implies a matrix inversion (see for instance [41, Sec. 4.2]), which is computationally intensive, and suffers drawbacks about accuracy and situations where the rank is not full. As a consequence of these observations, it would be preferable to derive the distinguisher by computing mathematically the value of the minimum (instead of directly running optimization algorithms). In the following subsection, we prove that this derivation can be done and we exhibit the closed form, which will be used in our attack simulations and experiments.

C. Main Result

This theorem is our main result:

Theorem 5: The stochastic collision attack defined in Eq. (5) is equal to:

$$\mathcal{D}_{\text{sto.coll}}(\vec{t}^{(\cdot)}, \vec{x}^{(\cdot)}) = \underset{\vec{k}\in(\mathbb{F}_2^n)^L}{\text{argmax}} \sum_{u\in\mathbb{F}_2^n} \frac{\left(\sum_\ell \sum_{q/t_q^{(\ell)}\oplus k^{(\ell)}=u} x_q^{(\ell)}\right)^2}{\sum_\ell \sum_{q/t_q^{(\ell)}\oplus k^{(\ell)}=u} 1}. \qquad (6)$$

[3] Note that we employ the notation $\text{argmin}_{k^{(\cdot)}} \min_\varphi$, which suggests that, first of all, a minimization on φ is performed for a set of vectorial keys $k^{(\cdot)}$, and then that the minimal value for all $k^{(\cdot)}$ is sought, and the corresponding key set is returned. But actually, the minimization is not forced to be in this order, as $k^{(\cdot)} \in (\mathbb{F}_2^n)^L$ and $\varphi : \mathbb{F}_2^n \to \mathbb{R}$ are independent variables.

In Eqn. (6), "$\sum_{q/t_q^{(\ell)}\oplus k^{(\ell)}=u}$" is short for "$\sum_{\substack{q\in\{1,\ldots,Q\}\\ \text{s.t. } t_q^{(\ell)}\oplus k^{(\ell)}=u}}$".

Proof: Let us denote by $G_{k^{(\ell)}}^{(\ell)}$ the $Q \times 2^n$ matrix, where $G_{k^{(\ell)}}^{(\ell)}[q,u] = \delta_u(t_q^{(\ell)} \oplus k^{(\ell)})$. We use $\|\cdot\|$ to define the norm-2 in $\mathbb{R}^Q$. Then, Eq. (5) can be expressed as:

$$\mathcal{D}_{\text{sto.coll}}(\vec{t}^{(\cdot)}, \vec{x}^{(\cdot)}) = \underset{\vec{k}\in(\mathbb{F}_2^n)^L}{\operatorname{argmin}} \min_{\vec{a}\in\mathbb{R}^{2^n}} \sum_{\ell=1}^{L} \left\| \vec{x}^{(\ell)} - G_{k^{(\ell)}}^{(\ell)} \vec{a} \right\|^2. \tag{7}$$

In this equation, $\vec{a}$ is seen as a column of 2^n real numbers and $G_{k^{(\ell)}}^{(\ell)}\vec{a}$ is a matrix vector product. Now, we can solve the minimization on $\vec{a} \in \mathbb{R}^{2^n}$ by minimizing the convex function:

$$\sum_\ell \left\| \vec{x}^{(\ell)} - G_{k^{(\ell)}}^{(\ell)} \vec{a} \right\|^2 = \sum_\ell \left\| \vec{x}^{(\ell)} \right\|^2 \tag{8}$$

$$+ \vec{a}^{\mathsf{T}} \left(\sum_\ell {G_{k^{(\ell)}}^{(\ell)}}^{\mathsf{T}} G_{k^{(\ell)}}^{(\ell)} \right) \vec{a} - 2\vec{a}^{\mathsf{T}} \left(\sum_\ell {G_{k^{(\ell)}}^{(\ell)}}^{\mathsf{T}} \vec{x}^{(\ell)} \right), \tag{9}$$

where $\vec{a}^{\mathsf{T}}$ is the vector (i.e., the line-matrix of dimension 1×2^n) equal to the transposed of $\vec{a}$. This quadratic form is minimal when its Jacobian is equal to zero [3]. Let us denote:

- $\Sigma_{k^{(\cdot)}}$ the square $2^n \times 2^n$ matrix $\sum_\ell {G_{k^{(\ell)}}^{(\ell)}}^{\mathsf{T}} G_{k^{(\ell)}}^{(\ell)}$ and
- $b_{k^{(\cdot)}}$ the column of 2^n elements $\sum_\ell {G_{k^{(\ell)}}^{(\ell)}}^{\mathsf{T}} \vec{x}^{(\ell)}$.

Let us assume $\Sigma_{k^{(\cdot)}}$ is invertible in the following (see Remark 6 otherwise), then we have $\vec{a} = \left(\Sigma_{k^{(\cdot)}}\right)^{-1} b_{k^{(\cdot)}}$. The optimal value of Eq. (9) is:

$$\sum_\ell \|\vec{x}^{(\ell)}\|^2 + b_{\vec{k}}^{\mathsf{T}} \cancel{\Sigma_{\vec{k}}^{-1}} \cancel{\Sigma_{\vec{k}}} \Sigma_{\vec{k}}^{-1} b_{\vec{k}} - 2 b_{\vec{k}}^{\mathsf{T}} \Sigma_{\vec{k}}^{-1} b_{\vec{k}} = \text{cst} - b_{\vec{k}}^{\mathsf{T}} \left(\Sigma_{\vec{k}}\right)^{-1} b_{\vec{k}}, \tag{10}$$

because $\Sigma_{\vec{k}}$ is symmetrical (hence its inverse is also). Let apart the constant, the term $-b_{\vec{k}}^{\mathsf{T}} \left(\Sigma_{\vec{k}}\right)^{-1} b_{\vec{k}}$ is negative, because $\left(\Sigma_{\vec{k}}\right)^{-1}$ is positive definite. Thus, in the sequel, we rather focus on maximizing the opposite of this value.

Let us now develop the expressions. For each coordinate of $b_{\vec{k}} = (b_{\vec{k}}[u])_{u\in\mathbb{F}_2^n}$ we have:

$$\begin{aligned} b_{\vec{k}}[u] &= \sum_{q=1}^{Q} \sum_\ell G_{k^{(\ell)}}^{(\ell)}[q,u] x_q^{(\ell)} \\ &= \sum_{q=1}^{Q} \sum_\ell \delta_u(t_q^{(\ell)} \oplus k^{(\ell)}) x_q^{(\ell)} \\ &= \sum_\ell \sum_{q/t_q^{(\ell)}\oplus k^{(\ell)}=u} x_q^{(\ell)}. \end{aligned}$$

Besides, the element $(u,v) \in \mathbb{F}_2^n \times \mathbb{F}_2^n$ of the square matrix $\Sigma_{\vec{k}}$ can be rewritten as:

$$\begin{aligned} \Sigma_{\vec{k}}[u,v] &= \sum_{q=1}^{Q} \sum_\ell \delta_u(t_q^{(\ell)} \oplus k^{(\ell)})\ \delta_v(t_q^{(\ell)} \oplus k^{(\ell)}) \\ &= \begin{cases} \sum_\ell \sum_{q/t_q^{(\ell)}\oplus k^{(\ell)}=u} 1 & \text{if } u = v, \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

Thus, $\Sigma_{\vec{k}}$ is diagonal as well as its inverse; every element of the diagonal shall simply be inversed. Therefore, the objective function (Eq. (10)) takes a closed form expression. Finally, the stochastic collision distinguisher rewrites:

$$\begin{aligned} \mathcal{D}_{\text{sto.coll}}(\vec{t}^{(\cdot)}, \vec{x}^{(\cdot)}) &= \underset{\vec{k}\in(\mathbb{F}_2^n)^L}{\operatorname{argmax}} \sum_{u\in\mathbb{F}_2^n} \left(\Sigma_{\vec{k}}[u,u]\right)^{-1} b_{\vec{k}}[u]^2 \\ &= \underset{\vec{k}\in(\mathbb{F}_2^n)^L}{\operatorname{argmax}} \sum_{u\in\mathbb{F}_2^n} \frac{\left(\sum_\ell \sum_{q/t_q^{(\ell)}\oplus k^{(\ell)}=u} x_q^{(\ell)}\right)^2}{\sum_\ell \sum_{q/t_q^{(\ell)}\oplus k^{(\ell)}=u} 1}. \end{aligned}$$

■

Remark 6: In case the matrix $\Sigma_{k^{(\cdot)}}$ is not invertible, there exists (at least) one value of $u \in \mathbb{F}_2^n$ such that $\sum_\ell \sum_{q/t_q^{(\ell)}\oplus k^{(\ell)}=u} 1 = 0$. Equivalently,

$$\left\{(q,\ell) \in \{1,\ldots,Q\} \times \{1,\ldots,L\} / t_q^{(\ell)} \oplus k^{(\ell)} = u\right\} = \varnothing.$$

Thus, both numerator and denominator corresponding to u in Eq. (6) are null. This means that the corresponding summand shall simply be dropped.

If there is only one value for ℓ, say $\ell \in \{1\}$, then Eq. (6) cannot distinguish keys. Indeed, by a variable change $u' \leftarrow u \oplus k$, we have:

$$\sum_{u\in\mathbb{F}_2^n} \frac{\left(\sum_{q/t_q\oplus k=u} x_q\right)^2}{\sum_{q/t_q\oplus k=u} 1} = \sum_{u'\in\mathbb{F}_2^n} \frac{\left(\sum_{q/t_q=u'} x_q\right)^2}{\sum_{q/t_q=u'} 1},$$

which clearly does not depend on k. Interestingly, this provides another proof of the fact that LRA is not sound when provided with a full basis [16].

This does not apply any longer when ℓ can take strictly more than one value. Indeed, the variable change $u' \leftarrow u \oplus k^{(1)}$ actually removes the key part in $u \oplus k^{(1)}$ (which becomes u'), but not in $u \oplus k^{(2)}$ which becomes $u' \oplus k^{(1)} \oplus k^{(2)}$.

D. Comparison With State-of-the Art Collision Attacks

1) Classical Collision Attack [37]: The collision attack, as published in the state-of-the-art, consists in distances to be minimized. In the case $L = 2$, the collision attacks minimize an Euclidean distance between the two leakages. This choice of distance is not justified in the state-of-the-art papers; still, it looks like a natural choice when the noise is Gaussian. So, when $L = 2$, the classical collision attack computes the distinguisher to test an hypothesis Δk on the sum of keys

$k^{(1)} \oplus k^{(2)}$:

$$\mathcal{D}_{\text{coll}}(\vec{t}^{(\cdot)}, \vec{x}^{(\cdot)}) = \underset{\Delta k \in \mathbb{F}_2^n}{\text{argmin}} \left(\sum_{\substack{1 \le q \le Q \\ t_q^{(1)} \oplus t_q^{(2)} = \Delta k}} 1 \right)^{-1} \sum_{\substack{1 \le q \le Q \\ t_q^{(1)} \oplus t_q^{(2)} = \Delta k}} \left(x_q^{(1)} - x_q^{(2)} \right)^2 . \tag{11}$$

Remark 7: In the state-of-the-art paper, for instance [15, Sec. 3.1], the Pearson correlation between $X^{(1)}$ and $X^{(2)}$, knowing $T^{(1)} \oplus k^{(1)} = T^{(2)} \oplus k^{(2)}$, is also suggested as a possible distinguisher. Now, the denominator of a correlation does not allow to distinguish, and the numerator is a covariance, which is proportional to the double-product term of Eq. (11) when the square is developed. Thus, in the sequel, we indifferently refer to the classical collision attack as a Euclidean distance or a correlation. See Appendix A for a more detailed discussion and a practical validation.

Remark 8: The stochastic collision distinguisher (Eq. (6)) uses in its computation all the $x_q^{(\ell)}$, for $\ell \in \{1, \ldots, L\}$ and $1 \le q \le Q$. Instead, the collision attack (Eq. (11)) uses only some values of $x_q^{(\ell)}$ in order to detect the collision. In particular, for each key guess only the traces which correspond to $t_q^{(1)} \oplus t_q^{(2)} = \Delta k$ are used, which is roughly $\frac{Q}{2^n}$ traces out of Q. This is one evidence why the stochastic distinguisher is more efficient in terms of success rate for a given number of traces (i.e., a given value of Q); Section III will confirm this remark. Notice that it applies even stronger if there are more than two leakages to combine.

In order to deal with the issue reported in Remark 8, and hence to exploit all traces for all key guesses, Moradi *et al.* proposed in [33] an enhancement of the classical collision attack.

2) Correlation-Enhanced Collision Attack: Let us also recall the correlation-enhanced collision attack [33]. It computes (see details in Appendix B):

$$\mathcal{D}_{\text{corr-coll}}(\vec{t}^{(\cdot)}, \vec{x}^{(\cdot)}) = \underset{k^{(1)}, k^{(2)}}{\text{argmax}} \sum_{u \in \mathbb{F}_2^n} \frac{\sum_{\substack{1 \le q \le Q \\ t_q^{(1)} \oplus k^{(1)} = u}} x_q^{(1)}}{\sum_{\substack{1 \le q \le Q \\ t_q^{(1)} \oplus k^{(1)} = u}} 1} \times \frac{\sum_{\substack{1 \le q \le Q \\ t_q^{(2)} \oplus k^{(2)} = u}} x_q^{(2)}}{\sum_{\substack{1 \le q \le Q \\ t_q^{(2)} \oplus k^{(2)} = u}} 1} . \tag{12}$$

As the correlation-enhanced collision attack first estimates averages of one SBox (conditioned by the plaintext) and then correlates it to the second one, a sufficient amount of traces is needed for each class in order to succeed. However, asymptotically the problem of estimation can be neglected. Indeed, we have this interesting result for large Q in the next lemma. These observations for small and large Q (which corresponds to low and high noise) are confirmed in our experiments in Sec. III.

Lemma 9: When the number of traces Q is large and the plaintexts are equidistributed, then $\mathcal{D}_{\text{corr-coll}}$ and $\mathcal{D}_{\text{sto.coll}}$ get equivalent.

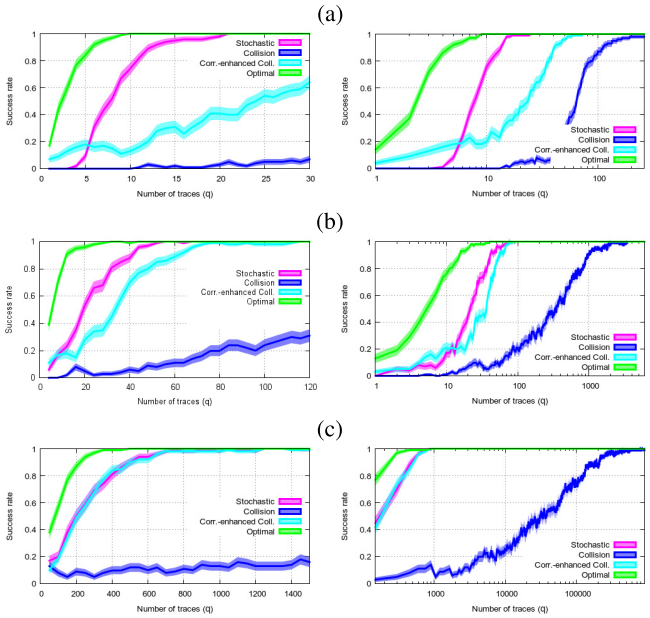

Fig. 2. Success rates for the four distinguishers $\mathcal{D}_{\text{sto.coll}}$ of Eq. (6), $\mathcal{D}_{\text{coll}}$ of Eq. (11), $\mathcal{D}_{\text{corr-coll}}$ of Eq. (12), and $\mathcal{D}_{\text{opt}}$ of Eq. (2) for (a) $\sigma = 0$, (b) $\sigma = 1$, (c) $\sigma = 4$. On the *left*, with a small number of traces, on the *right* with more traces (in logarithmic scale for q), so that all attacks succeed. In these figures, we have $n = 4$, so the SNR is simply σ^{-2}, that is (a) SNR $= +\infty$, (b) SNR $= 1$, (c) SNR $= 1/16$.

Proof: Owing to the uniform distribution of the plaintexts, we have (by the law of large numbers) that $\frac{1}{Q} \sum_{q/t_q = u} 1 \longrightarrow 2^{-n}$ when $Q \to +\infty$. Therefore, the denominator in both Eq. (6) ($\mathcal{D}_{\text{sto.coll}}$) and Eq. (12) ($\mathcal{D}_{\text{corr-coll}}$) can be neglected (asymptotically). Second, let us develop:

$$\begin{aligned} &\sum_{u \in \mathbb{F}_2^n} \left(\sum_{\ell} \sum_{q/t_q^{(\ell)} \oplus k^{(\ell)} = u} x_q^{(\ell)} \right)^2 \\ &= \sum_{\ell} \sum_{u \in \mathbb{F}_2^n} \sum_{q/t_q^{(1)} \oplus k^{(1)} = u} \left(x_q^{(\ell)} \right)^2 \\ &\quad + \sum_{\ell \neq \ell'} \sum_{u \in \mathbb{F}_2^n} \left(\sum_{q/t_q^{(\ell)} \oplus k^{(\ell)} = u} x_q^{(\ell)} \right) \left(\sum_{q/t_q^{(\ell')} \oplus k^{(\ell')} = u} x_q^{(\ell')} \right) . \end{aligned}$$

All the square terms do not depend on the key. Therefore, only cross-products remain, and each is exactly the expression of $\mathcal{D}_{\text{corr-coll}}$ for one pair. ∎

III. The New Distinguisher in Practice

A. Validation by Simulations for Different Noise Variances

The success rate SR of several attacks is plotted in Fig. 2 for different noise variances σ^2. The thickness of the curves represents the estimation error, namely plus/minus one standard deviation $(\text{SR}(1 - \text{SR})/N_{\text{attack}})^{1/2}$, where N_{attack} is the number of attack simulations performed to get the estimate SR. Indeed, the success rate is estimated as a counting, hence obeys a binomial distribution with parameters N_{attack} and SR [26, Sec. 3.4]. The more attacks are repeated, the more

narrow the error bars. In general, we carry out at least $N_{\text{attack}} = 100$ attacks, and sometimes 1000 to better distinguish between several curves which are close to each other.

The leakage model we used is $\varphi = w_H \circ S$, i.e., the composition of the Hamming weight and the PRESENT 4×4 SBox (that is, $n = 4$). As usual, the noise is assumed centered and of variance σ^2, therefore the SNR in Fig. 2 is equal to $n/(2\sigma)^2 = \sigma^{-2}$. On these figures, the performance of the optimal attack has been superimposed, which consists in a template attack on *one* SBox that nonetheless requires additionally the knowledge of the leakage model.

As we derived in theory (see Lemma 9) the stochastic collision attack outperforms the correlation-enhanced collision attack for lower noise (and thus smaller q), but they become equivalent for higher amount of traces. Moreover, we observe that the stochastic collision and the correlation-enhanced collision attack are always much more powerful than the classical collision attack, while not being far (in terms of performances) from the optimal attack. Actually, the more noise, the closer the stochastic collision attack and the optimal attack. On the contrary, as shown on the success rate with the logarithm of traces, the collision attack seems to get further to the stochastic collision attack when the noise variance increases. This can be accounted by the fact the classical collision attack is only using a fraction of the available traces (as discussed in Remark 8). Naturally, this amount gets higher the more traces are used (due to the higher noise variance) and thus the disadvantage of the classical collision attack becomes more compensated. This highlights that the classical collision attack is ad hoc, i.e., defined through an engineering idea but not a thorough analysis of the goal (maximizing the success probability).

B. Experiments on Real Traces

A similar analysis as in Fig. 2 is repeated on real traces, namely those publicly available from the DPA contest v4 dataset [42]. Those traces are collected from a masked implementation of AES. However, in this section we make the assumption that the masks are known to the adversary, which gives a similar context as in Subsect. III-A except that we have $n = 8$ instead of $n = 4$. The success rates are plotted in Fig. 3, in linear scale for q on the left-hand side, and in logarithmic scale on the right-hand side, so as to show that the classical collision attack indeed succeeds eventually (albeit with much more traces). We estimated an SNR of 2.2966 and 2.0652 from the leakage measurements for the first and second SBox, respectively, which shows that the variance of noise at both points in time are close.

C. Multi-Collisions

The state-of-the-art collision attacks are only suited for the two-leakage case ($L = 2$). On the contrary, the stochastic collision attack can apply to $L > 2$ leakages. It is interesting to assess in which respect this feature is an advantage. To do so, we compared the success rate to extract a pair of keys one by one,

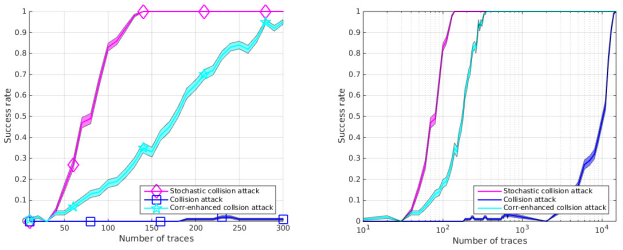

Fig. 3. DPA contest v4 (masks known). Left: linear, right: logarithmic scale for q.

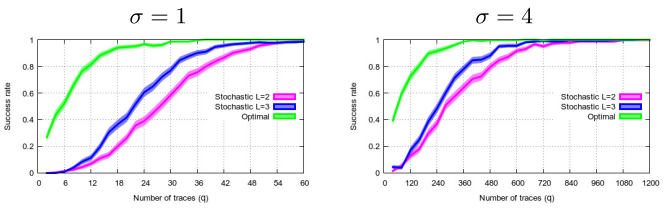

Fig. 4. Three-way vs. two-way collisions, applied to the recovery of one pair of keys.

- using two stochastic collision attacks for $L = 2$, ([4]) or
- using only one stochastic collision attack for $L = 3$.

Both strategies only allow to retrieve two keys out of three. Let us assume that $k^{*(1)}$ is known (or exhaustively searched for), and that $k^{*(2)}$ and $k^{*(3)}$ are unknown. A side-channel distinguisher thus estimates the pair $(\hat{k}^{*(2)}, \hat{k}^{*(3)})$ according to one of the two following strategies:

$$\begin{cases} (\hat{k}^{*(2)}, \hat{k}^{*(3)}) = (\mathcal{D}_{\text{sto.coll}}(\vec{t}^{(1,2)}, \vec{x}^{(1,2)}), \mathcal{D}_{\text{sto.coll}}(\vec{t}^{(1,3)}, \vec{x}^{(1,3)})), \\ (\hat{k}^{*(2)}, \hat{k}^{*(3)}) = \mathcal{D}_{\text{sto.coll}}(\vec{t}^{(1,2,3)}, \vec{x}^{(1,2,3)}). \end{cases}$$

There seems to be pros and cons to both strategies; hence a priori there is no clear intuition which option is the best. By guessing less keys (when $L = 2$, only one key is enumerated in $\mathbb{F}_2^n$), accordingly, there exist less rivals to the correct key. But an attack which exploits simultaneously $L = 3$ leakages collects more information than an attack which considers only $L = 2$ leakage samples.

To determine which tendency is preponderant, we launched simulations in the same setup as for Sec. III-A. The results are represented in Fig. 4 for two levels of noise: $\sigma \in \{1, 4\}$. Clearly, it appears better to guess two keys in one go. This result is one more reason to favor multivariate stochastic collision attacks over pairwise collision attacks. Recall that the state-of-the-art collision attacks are restricted to pairwise comparisons due to their underlying statistic tools. If computational power permits, it is all the more successful to carry out multivariate collision attacks as L is large.

Actually, to conclude, the distinguisher $\mathcal{D}_{\text{sto.coll}}$ being optimal (recall terms of Definition 3), when unknown parameters

[4]Such strategy can be improved by exploiting the natural redundancy in pairwise key search. Indeed, the $L = 2$ collision can concern leakages $\{1, 2\}$, $\{1, 3\}$, but also $\{2, 3\}$. Papers, such as [35], showed that a coding approach may constructively increase the efficiency of the attack. In this paper, we do not enter into this sophistication level.

are both the keys and the model, it is sensible that the second strategy has a smaller success probability.

IV. Stochastic Collision Attack in the Presence of Masking

In this section, we compare stochastic collision attacks with collision attacks in the case of masking, more specifically, in the case where the SBoxes are all masked with the same mask.[5] Indeed, this case is common in practice because recomputing a masked SBox is very lengthy, and masking every SBox with a different mask does not increase the (first) order of the countermeasure. With respect to the unprotected case, the model differs from Eq. (1), in that:

$$\forall \ell \in \{1, \ldots, L\}, \quad X^{(\ell)} = \psi(S(T^{(\ell)} \oplus k^{*(\ell)}) \oplus M) + N^{(\ell)}, \tag{13}$$

where the masks M are uniformly distributed in $\mathbb{F}_2^n$.

Furthermore, we extend and validate our attack to the scenario given by the DPA contest v4.1 in Subsect. IV-D in which the masks are not the same but related. Finally, we shortly discuss the case of zero-offset leakage given in hardware implementation in Subsect. IV-E.

A. Mathematical Expression of the Stochastic Collision Distinguisher

The optimal collision based distinguisher in case of masking is given in the following proposition.

Proposition 10 (Optimal Collision Based Distinguisher for Masking): The optimal collision based distinguisher when the model is masked at first order and when the noise is Gaussian and isotropic in the sample space, namely $N^{(\cdot)} \sim \mathcal{N}(0, \mathrm{diag}(\sigma^2, \ldots, \sigma^2))$ can be expressed as:

$$\mathcal{D}_{\mathrm{opt}}^{\mathrm{mask}}(\vec{t}^{(\cdot)}, \vec{x}^{(\cdot)}) = \underset{\vec{k} \in (\mathbb{F}_2^n)^L}{\mathrm{argmax}} \sum_{q=1}^{Q} \log \sum_{m \in \mathbb{F}_2^n} \exp \frac{-\left|\left|\vec{x}_q - \psi(S(\vec{t}_q \oplus \vec{k}) \oplus m)\right|\right|^2}{2\sigma^2}. \tag{14}$$

Proof: The proof is similar to that of Proposition 1, by noticing that:

$$\begin{aligned}\mathcal{D}_{\mathrm{opt}}^{\mathrm{mask}}(\vec{t}^{(\cdot)}, \vec{x}^{(\cdot)}) &= \underset{\vec{k} \in (\mathbb{F}_2^n)^L}{\mathrm{argmax}}\, p(\vec{x}^{(\cdot)} | \vec{t}^{(\cdot)}, \vec{k}) \\ &= \underset{\vec{k} \in (\mathbb{F}_2^n)^L}{\mathrm{argmax}} \prod_{q=1}^{Q} p(\vec{x}_q | \vec{t}_q, \vec{k}) \\ &= \underset{\vec{k} \in (\mathbb{F}_2^n)^L}{\mathrm{argmax}} \sum_{q=1}^{Q} \log \sum_{m \in \mathbb{F}_2^n} \\ &\quad \times \mathbb{P}(M = m)\, p(\vec{x}_q | \vec{t}_q, \vec{k}, m). \end{aligned} \tag{15}$$

Now, as the masks are uniformly distributed, the constant quantity $\mathbb{P}(M = m) = 1/2^n$ can be removed. Nonetheless, the logarithm of a sum of exponentials cannot be further simplified, which is why Eq. (14) is more complex than Eq. (2). ∎

Now, to cope with the scenario in which ψ is unknown to the attacker, we build the stochastic collision distinguisher in case of masking:

Definition 11 (Stochastic Collision Distinguisher for Masking):

$$\begin{aligned}\mathcal{D}_{\mathrm{sto.coll}}^{\mathrm{mask}}(\vec{t}^{(\cdot)}, \vec{x}^{(\cdot)}) &= \underset{\vec{k} \in (\mathbb{F}_2^n)^L}{\mathrm{argmax}} \max_{\psi : \mathbb{F}_2^n \to \mathbb{R}} \sum_{q=1}^{Q} \log \backslash \\ &\quad \sum_{m \in \mathbb{F}_2^n} \exp \frac{-\left|\left|\vec{x}_q - \psi(S(\vec{t}_q \oplus \vec{k}) \oplus m)\right|\right|^2}{2\sigma^2}. \end{aligned} \tag{16}$$

Lemma 12: By decomposing ψ in the canonical basis $(\delta_u)_{u \in \mathbb{F}_2^n}$, we get the following equivalent expressions:

$$\begin{aligned}\mathcal{D}_{\mathrm{sto.coll}}^{\mathrm{mask}}(\vec{t}^{(\cdot)}, \vec{x}^{(\cdot)}) &= \underset{\vec{k} \in (\mathbb{F}_2^n)^L}{\mathrm{argmax}} \max_{a \in \mathbb{R}^{2^n}} \sum_{q=1}^{Q} \log \backslash \\ &\quad \sum_{m \in \mathbb{F}_2^n} \exp \frac{-\left\|\vec{x}_q - \vec{a}_{S(\vec{t}_q \oplus \vec{k}) \oplus m}\right\|^2}{2\sigma^2}, \end{aligned} \tag{17}$$

where $\vec{a}_{S(\vec{t}_q \oplus \vec{k}) \oplus m}$ denotes the vector $(a_{S(t_q^{(\ell)} \oplus k^{(\ell)}) \oplus m})_{1 \le \ell \le L}$.

Proof: Similar to the proof of Lemma 4. ∎

In Eq. (17), the maximization on $a \in \mathbb{R}^{2^n}$ is a difficult problem, because the objective function:

$$\sum_{m \in \mathbb{F}_2^n} p(\vec{x}_q | \vec{t}_q, \vec{k}, m) = \sum_{m \in \mathbb{F}_2^n} \exp \frac{-\left\|\vec{x}_q - \vec{a}_{S(\vec{t}_q \oplus \vec{k}) \oplus m}\right\|^2}{2\sigma^2} \tag{18}$$

is a mixture of Gaussians. Therefore, it is not convex, but can instead have several local maxima.

Fortunately, some iterative algorithms (see for instance [29], [40]) exist that allow to numerically solve Eq. (17), which we detail next. We leverage on the Expectation Maximization (EM): it is a general framework to optimize in the context of Gaussian mixtures. It has already been used in side-channel analysis to extract the parameters of templates during profiling stage [25]. However, apparently for the first time, we exploit EM in the online attacking phase.

In order to simplify the presentation, we assume the following:

- The Q traces $\vec{x}^{(\cdot)} = (x_q^{(\ell)})_{1 \le q \le Q, 1 \le \ell \le L}$ corresponding to the Q plaintexts $\vec{t}^{(\cdot)} = (t_q^{(\ell)})_{1 \le q \le Q, 1 \le \ell \le L}$ have been drawn.
- We fix a key guess $\vec{k} \in (\mathbb{F}_2^n)^L$.
- We consider that the mask M is a random variable, uniformly distributed on $\mathbb{F}_2^n$.
- We model the unknown weights $a \in \mathbb{R}^{2^n}$ as a random variable.

Thus, the objective function (recall Eq. (18)) can rewrite:

$$p(\vec{x}^{(\cdot)} | a),$$

[5] This situation occurs in products implementing the so-called table recomputation countermeasure proposed by Kocher [23] – see also [28] for a description of this countermeasure when applied to AES.

where the lefthand side of Eq. (18) actually displays $p(\vec{x}^{(\cdot)}|a) = \sum_m \mathbb{P}(m)p(\vec{x}^{(\cdot)}|a,m)$. The EM algorithm is a method to get a series of values $a_0, a_1, \ldots$, such that $p(\vec{x}^{(\cdot)}|a_0) \leq p(\vec{x}^{(\cdot)}|a_1) \leq \ldots$. The model is given in the following

Lemma 13 (EM): Let $a' \in \mathbb{R}^{2^n}$. We define $a'' \in \mathbb{R}^{2^n}$ as:

$$a'' = \underset{a\in\mathbb{R}^{2^n}}{\operatorname{argmin}} \sum_m p(\vec{x}^{(\cdot)}|m,a')[-\log p(\vec{x}^{(\cdot)}|m,a)]. \quad (19)$$

Then, we have $p(\vec{x}^{(\cdot)}|a') \leq p(\vec{x}^{(\cdot)}|a'')$.

Proof: Let $a, a' \in \mathbb{R}^{2^n}$. We consider the quantity $Q(a,a')$ defined as:

$$Q(a,a') = \sum_m \mathbb{P}(m)p(\vec{x}^{(\cdot)}|m,a')\log\frac{p(\vec{x}^{(\cdot)}|m,a)}{p(\vec{x}^{(\cdot)}|m,a')}.$$

Then, using that for all $z \geq 0$, $\log z \leq z-1$, we have that:

$$Q(a,a') \leq \sum_m \mathbb{P}(m)p(\vec{x}^{(\cdot)}|m,a) - \sum_m \mathbb{P}(m)p(\vec{x}^{(\cdot)}|m,a'),$$

and the right-hand side equals $p(\vec{x}^{(\cdot)}|a) - p(\vec{x}^{(\cdot)}|a')$. Thus, in particular, for

$$a'' = \underset{a\in\mathbb{R}^{2^n}}{\operatorname{argmax}} Q(a,a') = \underset{a\in\mathbb{R}^{2^n}}{\operatorname{argmax}} \sum_m p(\vec{x}^{(\cdot)}|m,a') \times \log p(\vec{x}^{(\cdot)}|m,a),$$

we have that $p(\vec{x}^{(\cdot)}|a'') - p(\vec{x}^{(\cdot)}|a') \geq Q(a'',a') \geq 0$. Indeed, by definition of a'', $\forall a, Q(a'',a') \geq Q(a,a')$, hence in particular for $a = a'$, we have $Q(a'',a') \geq Q(a',a') = 0$. ∎

So, the minimization of the objective function $p(\vec{x}^{(\cdot)}|a)$ can be achieved according to this procedure:

1) Choose a random initial value $a_0 \in \mathbb{R}^{2^n}$;
2) Set an iteration counter i, starting from 0, and compute a_{i+1} as the value given in Eq. (19), where $a' = a_i$ and the obtained $a'' = a_{i+1}$. Notice that the Eq. (19) gives its name to the EM algorithm, since it is a minimization of an expectation over the masks M. When $p(\vec{x}^{(\cdot)}|a_{i+1}) - p(\vec{x}^{(\cdot)}|a_i) \leq \varepsilon$, where ε is a small value, then break the loop and return a_{i+1}.

The convergence of this procedure can be guaranteed [46].

Notice that the optimization problem of $p(\vec{x}^{(\cdot)}|a)$ is turned into a series of optimization problems (Eq. (19)). However, the problem of Eq. (19) is simple because in our case, the objective function $a \mapsto \sum_m p(\vec{x}^{(\cdot)}|m,a')[-\log p(\vec{x}^{(\cdot)}|m,a)]$ is concave.

Lemma 14 (Minimization Within EM Iterations): Let $a' \in \mathbb{R}^{2^n}$ and $\alpha_q(m) = p(\vec{x}_q|m,a')$. Then, $a'' = (a''_u)_{u\in\mathbb{F}_2^n}$ given in Eq. (19) has its coordinates equal to:

$$a''_u = \frac{\sum_{q,m,\ell/S(t_q^{(\ell)}\oplus k^{(\ell)})\oplus m=u} \alpha_q(m)x_q^{(\ell)}}{\left(\sum_{q,m,\ell/S(t_q^{(\ell)}\oplus k^{(\ell)})\oplus m=u} \alpha_q(m)\right)}. \quad (20)$$

Proof: Let $a' \in \mathbb{R}^{2^n}$ and $\alpha_q(m) = p(\vec{x}^{(\cdot)}|m,a')$. The function to minimize is

$$\begin{aligned} a \mapsto & -\sum_m p(\vec{x}^{(\cdot)}|m,a')\cdot\log p(\vec{x}^{(\cdot)}|m,a) \\ &= -\sum_{m\in\mathbb{F}_2^n} \alpha_q(m)\cdot\log\prod_{q=1}^{Q}\prod_{\ell=1}^{L} f_{\sigma^2}(x_q^{(\ell)} - a_{S(t_q^{(\ell)}\oplus k^{(\ell)})\oplus m}) \\ &= \mathrm{cst}_1 + \underbrace{\mathrm{cst}_2}_{>0} \times \sum_{q,m,\ell} \frac{\alpha_q(m)}{2}\left(x_q^{(\ell)} - a_{S(t_q^{(\ell)}\oplus k^{(\ell)})\oplus m}\right)^2, \end{aligned}$$

which is concave (as the sum of $2^n QL$ concave functions). Now, for all $u \in \mathbb{F}_2^n$ and after denoting $E_u = \{(q,m,\ell);\ S(t_q^{(\ell)}\oplus k^{(\ell)})\oplus m = u\}$, we have that:

$$\begin{aligned} &\frac{\partial}{\partial a_u}\sum_{q,m,\ell}\frac{\alpha_q(m)}{2}\left(x_q^{(\ell)} - a_{S(t_q^{(\ell)}\oplus k^{(\ell)})\oplus m}\right)^2 \\ &= \sum_{q,m,\ell}\alpha_q(m)\delta_u(S(t_q^{(\ell)}\oplus k^{(\ell)})\oplus m)\frac{\partial}{\partial a_u}\frac{1}{2}\left(x_q^{(\ell)} - a_u\right)^2 \\ &= -\sum_{(q,m,\ell)\in E_u}\alpha_q(m)\left(x_q^{(\ell)} - a_u\right) \\ &= a_u\left(\sum_{(q,m,\ell)\in E_u}\alpha_q(m)\right) - \left(\sum_{(q,m,\ell)\in E_u}\alpha_q(m)x_q^{(\ell)}\right). \end{aligned}$$

Clearly, this derivative is equal to zero if and only if Eq. (20) is satisfied. ∎

The whole procedure using the EM approach is summarized in Alg. 1. We notice that Eq. (20) is not well defined ("0/0") if there exists one $u \in \mathbb{F}_2^n$, such that for all q, m, ℓ, we have $S(t_q^{(\ell)} \oplus k^{(\ell)}) \oplus m \neq u$. But this has no impact, since in this case, one simply does not evaluate this value of a_u (i.e., the corresponding line 16 is skipped in Alg. 1). Indeed, this ill-defined value of a_u is not required in the sequel (that is at line 8).

B. State-of-the-Art Collision Attacks

The classical collision attack described in [15] still applies successfully in the masked context described by Eq. (13). Actually, we notice that this distinguisher ($\mathcal{D}_{\text{coll}}^{\text{mask}}$) does not need to be adapted as explained by the authors. Here, as underlined in Remark 7, we consider that the Euclidean distance and the correlation are alike. Indeed, as the mask is shared by all the SBoxes, the collisions happen under the same conditions as in the case without masking.

Correlation-enhanced collision ($\mathcal{D}_{\text{corr-coll}}$, coined in [33] and presented in Eq. (12)) fails if the masking is perfect. Indeed, in a perfect masking scheme [5], the average of each leakage ℓ ($1 \leq \ell \leq L$) of Eqn. (13)) depends neither on $T^{(\ell)}$ nor on $k^{*(\ell)}$. Therefore, we also introduce the adaptation of correlation-enhanced collision to the case of masking without any first-order leakage. Such adaptation is done by Moradi [31]. As explained by the author in [31, Sec. 4.4], the collision can be done only provided a bivariate combination is done at each

Algorithm 1 Stochastic Collision Distinguisher in the Case of Masking in Gaussian Noise Using the EM Approach

input : $\vec{t}^{(\cdot)}$ and $\vec{x}^{(\cdot)}$, a series of L plaintexts and leakages,
σ, the noise standard deviation.
output: $\mathcal{D}^{\text{mask}}_{\text{sto.coll}}(\vec{t}^{(\cdot)}, \vec{x}^{(\cdot)}) \in (\mathbb{F}_2^n)^L$, as defined in Eq. (17)

// Attack using stochastic-collision powered with EM for defeating masking
for $\vec{k}^{(\cdot)} \in (\mathbb{F}_2^n)^L$ **do**
  // The initial value of a_u must not be constant, otherwise, obviously, the attack fails!
  for $u \in \mathbb{F}_2^n$ **do**
    $a_u \leftarrow w_H(u)$ // It is better if this choice is close to the real model $\psi : \mathbb{F}_2^n \to \mathbb{R}$
  end
  for iteration $\in \{1, \ldots, +\infty\}$ **do**
    // Expectation step
    for $q \in \{1, \ldots, Q\}$ **do**
      for $m \in \mathbb{F}_2^n$ **do**
        $\beta_q(m) \leftarrow \exp\left(-\frac{1}{2\sigma^2}\sum_{\ell=1}^{L}\left(x_q^{(\ell)} - a_{S(t_q^{(\ell)} \oplus k^{(\ell)}) \oplus m}\right)^2\right)$
      end
      $\Sigma\beta_q \leftarrow \sum_{m \in \mathbb{F}_2^n} \beta_q(m)$
      for $m \in \mathbb{F}_2^n$ **do**
        $\alpha_q(m) \leftarrow \beta_q(m) / \Sigma\beta_q$
        // Normalization (Lemma 14)
      end
    end
    // Minimization step
    for $u \in \mathbb{F}_2^n$ **do**
      $a_u \leftarrow \left(\sum_{q,m,\ell / S(t_q^{(\ell)} \oplus k^{(\ell)}) \oplus m = u} \alpha_q(m)\right)^{-1} \times \left(\sum_{q,m,\ell / S(t_q^{(\ell)} \oplus k^{(\ell)}) \oplus m = u} \alpha_q(m) x_q^{(\ell)}\right)$
      // Eq. (20)
    end
    break if a_u has not changed (up to given tolerance)
  end
  Log-Likelihood$(\vec{k}^{(\cdot)}) \leftarrow \sum_{q=1}^{Q} \log \Sigma\beta_q$
end
return $\text{argmax}_{\vec{k}^{(\cdot)}}$ Log-Likelihood$(\vec{k}^{(\cdot)})$
// Notice: here, we assume for instance that $k^{(0)}$ is known or brute-forced

SBox before correlation-enhanced collision. It is thus needed to add L leakages to Eq. (13). For $L = 2$, the correlation-enhanced collision is 4-variate attack which exploits, for $\ell \in \{1, 2\}$:

$$\begin{cases} X^{(\ell)} = \psi(S(T^{(\ell)} \oplus k^{*(\ell)}) \oplus M) + N^{(\ell)}, & \text{(e.g., SBox output)} \\ X'^{(\ell)} = \psi(M) + N'^{(\ell)}, & \text{(e.g., mask leakage)} \end{cases}$$

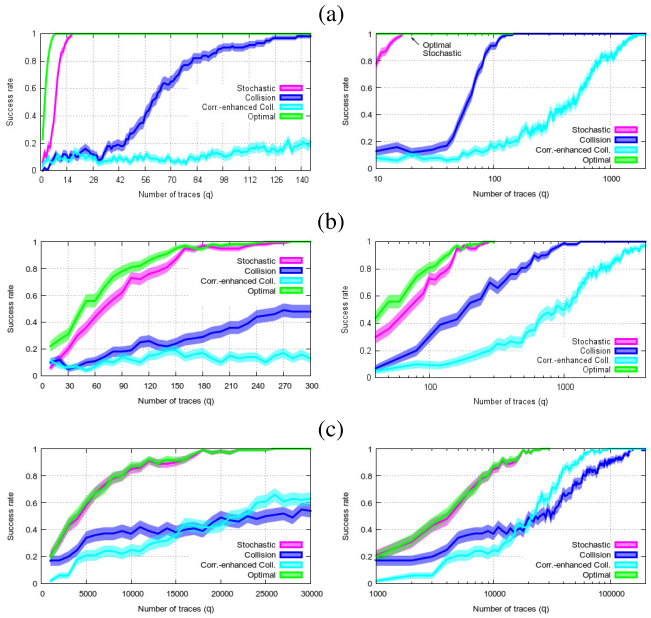

Fig. 5. Success rates for the four distinguishers $\mathcal{D}^{\text{mask}}_{\text{sto.coll}}$ of Eq. (17), $\mathcal{D}_{\text{coll}}$ of Eq. (11), $\mathcal{D}^{\text{mask}}_{\text{corr-coll}}$ of Eq. (21), and $\mathcal{D}^{\text{mask}}_{\text{opt}}$ (Eq. (14)), for (a) $\sigma = 0.1$, (b) $\sigma = 1$, and (c) $\sigma = 4$. On the *left*, with a small number of traces, on the *right* with more traces (in logarithmic scale for q), so that all attacks succeed. In these figures, we have $n = 4$, so the SNR is simply σ^{-2}, that is (a) SNR $= 100$, (b) SNR $= 1$, and (c) SNR $= 1/16$.

where $N^{(1)}$, $N^{(2)}$, $N'^{(1)}$ and $N'^{(2)}$ are independent (and assumed i.i.d. following the same law $\mathcal{N}(0, \sigma^2)$ for simplification). In this scenario, the correlation-enhanced collision of lowest order, which we denote by $\mathcal{D}^{\text{mask}}_{\text{corr-coll}}$, is given (by analogy with Eq. (12)) by:

$$\begin{aligned} &\mathcal{D}^{\text{mask}}_{\text{corr-coll}}(\vec{t}^{(\cdot)}, (\vec{x}^{(\cdot)}, \vec{x}'^{(\cdot)})) \\ &\quad = \underset{k^{(1)}, k^{(2)}}{\text{argmax}} \sum_{u \in \mathbb{F}_2^n} \frac{\sum_{\substack{1 \le q \le Q \\ \text{s.t. } t_q^{(1)} \oplus k^{(1)} = u}} x_q^{(1)} x_q'^{(1)}}{\sum_{\substack{1 \le q \le Q \\ \text{s.t. } t_q^{(1)} \oplus k^{(1)} = u}} 1} \\ &\qquad \times \frac{\sum_{\substack{1 \le q \le Q \\ \text{s.t. } t_q^{(2)} \oplus k^{(2)} = u}} x_q^{(2)} x_q'^{(2)}}{\sum_{\substack{1 \le q \le Q \\ \text{s.t. } t_q^{(2)} \oplus k^{(2)} = u}} 1}. \end{aligned} \tag{21}$$

In Eq. (21), the leakages could be centered, meaning that $x_q^{(\ell)}$ is replaced by $x_q^{(\ell)} - \frac{1}{Q}\sum_{q=1}^{Q} x_q^{(\ell)}$. But the distinguisher $\mathcal{D}^{\text{mask}}_{\text{corr-coll}}$ would not change, as the centering only introduces, after expansion, terms which do not depend on the key hypotheses $k^{(1)}$ and $k^{(2)}$.

C. Validation by Simulations for Different Noise Variances

We give in Fig. 5, for $L = 2$, a validation of $\mathcal{D}^{\text{mask}}_{\text{sto.coll}}$ (Eq. (17)), compared to the optimal $\mathcal{D}^{\text{mask}}_{\text{opt}}$ (Eq. (14)) and to the state-of-the-art $\mathcal{D}_{\text{coll}}$ (Eq. (11)) and $\mathcal{D}^{\text{mask}}_{\text{corr-coll}}$ (Eq. (21)). These results can be directly compared to that of the various collision attacks without masking (Fig. 2), except that one cannot choose $\sigma = 0$ for $\mathcal{D}^{\text{mask}}_{\text{opt}}$ (unless a division by zero occurs). Thus, we replaced the $\sigma = 0$ (case (a) of Fig. 2) by $\sigma = 0.1$ (case (a) of Fig. 5).

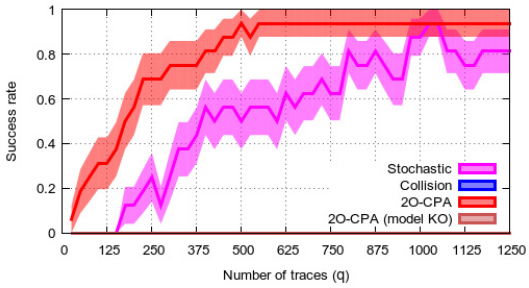

Fig. 6. Attack results on DPA contest v4.1 standardized traces (Sbox 0 and 5).

One can see that $\mathcal{D}^{\text{mask}}_{\text{corr-coll}}$ performs better than $\mathcal{D}_{\text{coll}}$ for large noises, which was not noticed before our work. However, despite $\mathcal{D}^{\text{mask}}_{\text{corr-coll}}$ exploits 4 leakages (it is 4-variate), it is less efficient than $\mathcal{D}^{\text{mask}}_{\text{sto.coll}}$, which exploits only 2 leakages (it is bi-variate). Obviously, $\mathcal{D}^{\text{mask}}_{\text{opt}}$ is the best distinguisher, but most importantly, it represents the threshold of possible attacks: no attack can have a success rate above that of the optimal distinguisher (by definition, the optimal distinguisher maximizes the success probability). Remarkably, the stochastic collision attack $\mathcal{D}_{\text{sto.coll}}$ converges to the optimal attack, whereas both state-of-the-art attacks become equivalent and stay inferior even for high noise. Note that, as for large q the chosen plaintext attack setting becomes equivalent to the known plaintext attack setting, naturally, both state-of-the-art attack become closer the more traces and thus more plaintext are involved.

Remark 15: Note that, $\mathcal{D}^{\text{mask}}_{\text{corr-coll}}$ and $\mathcal{D}^{\text{mask}}_{\text{sto.coll}}$ have opposed approaches to cope with masking. More precisely, $\mathcal{D}^{\text{mask}}_{\text{corr-coll}}$ tries to remove the effect of the mask by multiplying the leakage of the masked SBox and the mask alone, as it is done in the classical second-order DPA. On the contrary, $\mathcal{D}^{\text{mask}}_{\text{sto.coll}}$ handles the mask within the estimation algorithm (see Alg. 1), and iteratively attempts to refine the model despite masking.

Remark 16: It is noteworthy that the classical collision attack $\mathcal{D}_{\text{coll}}$ recovers the key with fewer traces when there is masking than when there is not. For example, it can be seen in the right part of Fig. 2(c) (resp. of Fig. 5(c)) that $\mathcal{D}_{\text{coll}}$ needs $120,000$ (resp. $60,000$) traces to reach 80% of success rate without masking (resp. with masking). This is related to the confusion coefficient [19]. Unfortunately, due to the lack of space we cannot go into further details.

D. Experiments on Real Traces

In the DPA contest v4.1, the masks are not uniformly distributed over $\mathbb{F}_2^n$, but over a code (we refer to [12] for more details). Besides, the masks are not the same for all sboxes: If mask $m^{(1)} = m_i$ $(0 \le i \le 15)$ for sbox 1, then the mask for sbox 2 is $m^{(1)} = m_{i+1 \mod 16}$. It is straightforward to check that the stochastic collision approach still works. The classical collision attack also works, albeit with an adaptation. This is mainly due to the fact that $m_i \oplus m_{i+\ell}$ is not balanced as observed in [32].

The results are shown in Fig. 6. The signal to noise ratio $\|\psi\|_2^2/\sigma^2 = \|a\|_2^2/\sigma^2$ is 2.94 for Sbox 5 and 2.61 for Sbox 10. Traces have been standardized prior to attack: that is, the mean is removed and a scaling by the standard deviation is applied, as done in [30]. We have a finite amount of $200,000$ traces, hence with $Q = 1250$, only 16 attacks can be launched. Assuming we know the model (except the masks), we would compute a second-order CPA. This has already been done by combining an XOR and an SBox call in [2, Fig. 4(b)], where ≈ 300 traces are needed for 80% success rate. Here, we compute a second-order attack by comparing it with a SBox 1 and 5 as the two sources of leakage. This CPA requires about 400 traces to reach 80% success rate. The stochastic collision attack requires about twice more traces. However, despite the stochastic collision attack, CPA requires the knowledge of the ψ on a proportional scale.

E. Extension to Zero-Offset Collision Attacks

In a hardware setup, the various SBoxes can be instantiated in parallel. Therefore, there is only one measurement (and one noise), and the leakage model is slightly different from Eq. (1):

$$X = \sum_{\ell=1}^{L} \varphi(T^{(\ell)} \oplus k^{*(\ell)}) + N. \tag{22}$$

Such leakage is called Zero-Offset (ZO). An adaptation of $\mathcal{D}_{\text{sto.coll}}$ (Eq. (6)) is denoted $\mathcal{D}^{\text{ZO}}_{\text{sto.coll}}$ and given in Theorem 17. As shown in Corollary 18, the expression can be simplified for a large number of traces and balanced plaintexts as $\mathcal{D}^{\text{ZO,coll}}_{\text{sto.coll}}$, which as the same expression as $\mathcal{D}_{\text{sto.coll}}$. Note that, the adaptation of $\mathcal{D}_{\text{coll}}$ (Eq. (11)) is not straightforward. Indeed, the difference in the square is always equal to zero. Therefore, we fall back to the variant $\mathcal{D}_{\text{coll (with corr)}}$ defined in Eq. (25).

Theorem 17: Let G is the $Q \times 2^n$ matrix whose element at position (q,u) is $G[q,u] = \sum_\ell \delta_u(t_q^{(\ell)} \oplus k^{(\ell)})$. The stochastic collision attack defined by analogy of $\mathcal{D}^{\text{ZO}}_{\text{opt}}$ in the case of zero-offset, where the leakage function φ is guessed in parallel to the key (recall argumentation of Sec. II-B), is

$$\mathcal{D}^{\text{ZO}}_{\text{sto.coll}}(\vec{t}^{(\cdot)}, \vec{x}) = \underset{\vec{k}^{(\cdot)} \in (\mathbb{F}_2^n)^L}{\text{argmax}}\ x^{\mathsf{T}} G G^{+} x, \tag{23}$$

where G^+ is the Moore-Penrose pseudo-inverse of G [1].

Proof: The expression of $\mathcal{D}^{\text{ZO}}_{\text{sto.coll}}$ is

$$\begin{aligned}
&\underset{\vec{k}^{(\cdot)} \in (\mathbb{F}_2^n)^L}{\text{argmin}}\ \min_{\varphi:\mathbb{F}_2^n \to \mathbb{R}} \sum_{q=1}^{Q} \left(x_q - \sum_{\ell=1}^{L} \varphi(t_q^{(\ell)} \oplus k^{(\ell)}) \right)^2 \\
&= \underset{\vec{k}^{(\cdot)} \in (\mathbb{F}_2^n)^L}{\text{argmin}}\ \min_{a \in \mathbb{R}^{2^n}} \sum_{q=1}^{Q} \left(x_q - \sum_{\ell=1}^{L} \sum_{u \in \mathbb{F}_2^n} a_u \delta_u(t_q^{(\ell)} \oplus k^{(\ell)}) \right)^2 \\
&= \underset{\vec{k}^{(\cdot)} \in (\mathbb{F}_2^n)^L}{\text{argmin}}\ \min_{a \in \mathbb{R}^{2^n}} \sum_{q=1}^{Q} \left(x_q - \sum_{u \in \mathbb{F}_2^n} a_u \underbrace{\textstyle\sum_{\ell=1}^{L} \delta_u(t_q^{(\ell)} \oplus k^{(\ell)})}_{G[q,u]} \right)^2 .
\end{aligned}$$

The Euclidean norm that $\mathcal{D}^{\text{ZO}}_{\text{sto.coll}}$ minimizes over $a \in \mathbb{R}^{2^n}$ is $\|\vec{x} - Ga\|^2$, where G is the $Q \times 2^n$ matrix whose element at position (q,u) is $G[q,u] = \sum_\ell \delta_u(t_q^{(\ell)} \oplus k^{(\ell)})$. It is known that

the solution that minimizes the Euclidean norm is $a = G^+\vec{x}$. Hence the minimization over $\vec{k}^{(\cdot)} \in (\mathbb{F}_2^n)^L$ consists of:

$$\begin{aligned}&\left\|(I - GG^+)\vec{x}\right\|^2\\ &= \vec{x}^\mathsf{T}\vec{x} - \vec{x}^\mathsf{T}(GG^+)^\mathsf{T}\vec{x} - \vec{x}^\mathsf{T}(GG^+)\vec{x} + \vec{x}^\mathsf{T}(GG^+)^\mathsf{T}(GG^+)\vec{x}\\ &= \vec{x}^\mathsf{T}\vec{x} - \vec{x}^\mathsf{T}GG^+\vec{x}.\end{aligned}$$

Indeed, the $Q \times Q$ matrix GG^+ is symmetrical, and one has the remarkable identity $GG^+G = G$. As $\vec{x}^\mathsf{T}\vec{x}$ does not depend on the key,

$$\begin{aligned}\mathcal{D}^{\mathrm{ZO}}_{\mathrm{sto.coll}}(\vec{t}^{(\cdot)}, \vec{x}) &= \underset{\vec{k}^{(\cdot)} \in (\mathbb{F}_2^n)^L}{\operatorname{argmin}} \ \vec{x}^\mathsf{T}\vec{x} - \vec{x}^\mathsf{T}GG^+\vec{x}\\ &= \underset{\vec{k}^{(\cdot)} \in (\mathbb{F}_2^n)^L}{\operatorname{argmax}} \ \vec{x}^\mathsf{T}GG^+\vec{x}.\end{aligned}$$

■

Corollary 18: The stochastic collision attack in the case of zero-offset leakage $\mathcal{D}^{\mathrm{ZO}}_{\mathrm{sto.coll}}$ is well approximated by a formula sibling to $\mathcal{D}_{\mathrm{sto.coll}}$:

$$\mathcal{D}^{\mathrm{ZO.approx}}_{\mathrm{sto.coll}}(\vec{t}^{(\cdot)}, \vec{x}) = \underset{\vec{k}^{(\cdot)} \in (\mathbb{F}_2^n)^L}{\operatorname{argmax}} \sum_{u \in \mathbb{F}_2^n} \frac{\left(\sum_\ell \sum_{q/t_q^{(\ell)} \oplus k^{(\ell)} = u} x_q\right)^2}{\sum_\ell \sum_{q/t_q^{(\ell)} \oplus k^{(\ell)} = u} 1}. \tag{24}$$

One recognizes $\mathcal{D}_{\mathrm{sto.coll}}$ (Eq. (6)) where $x_q^{(\ell)}$ is replaced by x_q (indeed, in the ZO context, the leakage is monovariate).

Proof: By the law of large numbers, we have

$$\frac{1}{Q}\Sigma_{k^{(\cdot)}}[u, v] \xrightarrow[Q \to +\infty]{} \sum_{\ell, \ell'} p(T^{(\ell)} \oplus k^{(\ell)} = u \wedge T^{(\ell')} \oplus k^{(\ell')} = v).$$

Now,

- when $\ell = \ell'$, $p(T^{(\ell)} \oplus k^{(\ell)} = u \wedge T^{(\ell)} \oplus k^{(\ell)} = v) = \begin{cases} 2^{-n} & \text{if } u = v,\\ 0 & \text{otherwise;}\end{cases}$
- when $\ell \neq \ell'$, $p(T^{(\ell)} \oplus k^{(\ell)} = u \wedge T^{(\ell')} \oplus k^{(\ell')} = v) = p(T^{(\ell)} = k^{(\ell)} \oplus u) \times p(T^{(\ell')} = k^{(\ell')} \oplus v) = 2^{-2n}$ because $T^{(\ell)}$ and $T^{(\ell')}$ are independent.

Therefore, we have that:

$$\begin{aligned}&\frac{1}{Q}\Sigma_{k^{(\cdot)}}\\ &\xrightarrow[Q \to +\infty]{} L \times 2^{-n}\begin{pmatrix}1 & 0 & \cdots & 0\\ 0 & 1 & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & 1\end{pmatrix}\\ &\quad + L(L-1) \times 2^{-2n}\begin{pmatrix}1 & 1 & \cdots & 1\\ 1 & 1 & \cdots & 1\\ \vdots & \vdots & \ddots & \vdots\\ 1 & 1 & \cdots & 1\end{pmatrix}\\ &= 2^{-2n}L\begin{pmatrix}2^n+(L-1) & (L-1) & \cdots & (L-1)\\ (L-1) & 2^n+(L-1) & \cdots & (L-1)\\ \vdots & \vdots & \ddots & \vdots\\ (L-1) & (L-1) & \cdots & 2^n+(L-1)\end{pmatrix}.\end{aligned}$$

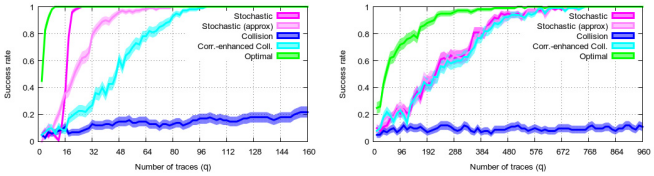

Fig. 7. Success rates for the four collision attacks in the context of zero-offset leakage for $\sigma = 0.1$ and $\sigma = 4$. In these figures, we have $n = 4$, so the SNR is simply σ^{-2}, that is SNR $= 100$ and SNR $= 1/16$.

Clearly, as in practice one has $2^n \gg (L-1)$ (e.g., $n = 4$ or 8, and $L = 2$ or 3), the extra-diagonal terms can be dropped from the matrix without altering it significantly.

So, for small values of Q, we simplify $\Sigma_{k^{(\cdot)}}$ as a diagonal matrix, whose element at position (u, u) is $\sum_{q, \ell / t_q^{(\ell)} \oplus k^{(\ell)} = u} 1$. Thence, the approximate distinguisher $\mathcal{D}^{\mathrm{ZO.approx}}_{\mathrm{sto.coll}}$ takes on the announced form. ■

The success rates of the four distinguishers are represented in Fig. 7, under similar conditions as those obtained for a bivariate leakage ($L = 2$, and two distinct leakage samples — recall Fig. 2). First of all, we notice that all four distinguishers succeed even in the ZO (monovariate collision) context, and that the optimal distinguisher $\mathcal{D}^{\mathrm{ZO}}_{\mathrm{opt}}$ is, as expected, the best. Second, we see that the stochastic collision is still achieving better than other state-of-the-art variants $\mathcal{D}^{\mathrm{ZO}}_{\mathrm{corr\text{-}coll}}$ and $\mathcal{D}^{\mathrm{ZO}}_{\mathrm{coll\ (with\ corr)}}$. Although the distinguisher $\mathcal{D}^{\mathrm{ZO}}_{\mathrm{sto.coll}}$ can be computed for any number $Q \geq 1$ of traces, its performance is fairly bad for strictly less than $Q = 2^n$ traces. The reason is that not all the possible texts have been encountered, hence a poor stochastic characterization of the leakage function. So, in our case, the success rate is ≈ 0 for $Q < 2^n$, and the approximation $\mathcal{D}^{\mathrm{ZO.approx}}_{\mathrm{sto.coll}}$ is thus better in this tiny region, which is especially visible for extremely high SNRs ($\sigma = 0.1$). This is actually *the only limitation* we observed about stochastic collision attacks: they are not reliable when the system of observations and unknown variables (model and key) is underdetermined. But clearly, when $Q \geq 2^n$, the accurate $\mathcal{D}^{\mathrm{ZO}}_{\mathrm{sto.coll}}$ overcomes the approximate $\mathcal{D}^{\mathrm{ZO.approx}}_{\mathrm{sto.coll}}$ in terms of success rate.

For $\sigma \geq 1$, many traces ($Q \gtrapprox 100$) are required to recover the key, hence $\mathcal{D}^{\mathrm{ZO}}_{\mathrm{sto.coll}}$ is very well approximated by $\mathcal{D}^{\mathrm{ZO.approx}}_{\mathrm{sto.coll}}$. Besides, Lemma 9 applies, hence $\mathcal{D}^{\mathrm{ZO.approx}}_{\mathrm{sto.coll}}$ is in turn equivalent to $\mathcal{D}^{\mathrm{ZO}}_{\mathrm{corr\text{-}coll}}$. Only $\mathcal{D}^{\mathrm{ZO}}_{\mathrm{coll\ (with\ corr)}}$ lags behind, because it is impeded by the fact traces are specialized to test one key hypothesis out of 2^n (recall Remark 8).

Eventually, we notice that in the absence of masking, even in high noise setups, the non-optimal distinguishers never perform as well as the optimal distinguisher (as was already shown for the L-variate collisions in Fig. 2).

V. Conclusions

In this paper, we mathematically derived the *stochastic collision attack* from the optimal distinguisher, while combining the flavors from collision attacks and stochastic modelling. The classical collision attack is inefficient in that it discards traces, which do not feature collisions. Similar to the correlation-enhanced collision attack, our approach consists in

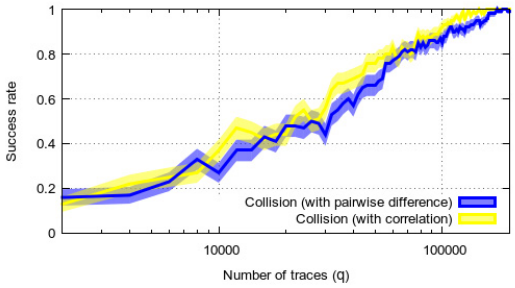

Fig. 8. Comparison between $\mathcal{D}_{\text{coll}}$ and $\mathcal{D}_{\text{coll (with corr)}}$.

taking advantage of all traces with a view to build a leakage model while distinguishing keys. Compared to stochastic attacks, our new methodology can apply linear regressions with the full basis which puts aside the problem of soundly choosing a sound sub-basis and allows, asymptotically, for perfectly recovering the algebraic description of the deterministic part of the leakage. Both in simulation and in real life, we have shown that the stochastic collision attack outperforms the state-of-the-art collision attacks, except in the unprotected case with (very) low SNR where the correlation-enhanced collision attack has similar efficiency, which confirms that in this unprotected low signal-to-noise scenario the state-of-the-art correlation-enhanced collision attack was already an appropriate choice. Furthermore, we extended our attack to the scenario of masking countermeasures in software and hardware and confirmed its practicability in simulations and experiments.

Appendix A
Comparison Between Variants of CHES 2011 [15] Collision Attack

In this appendix, we provide a comparison between two variants of collision-correlation power analysis on first-order protected AES, where the same mask is used at two SBoxes. Namely, we compare the collision based on pairwise square of differences ($\mathcal{D}_{\text{coll}}$, defined in Eq. (11)) with the collision based on correlation, in the same setup as in Sec. IV. According to Remark 7, the correlation sibling variant of $\mathcal{D}_{\text{coll}}$ is:

$$\mathcal{D}_{\text{coll (with corr)}}(\vec{t}^{(\cdot)}, \vec{x}^{(\cdot)}) = \underset{k^{(\cdot)} \in (\mathbb{F}_2^n)^2}{\text{argmax}} \frac{\sum_{\substack{1 \le q \le Q \\ \text{s.t. } t_q^{(1)} \oplus t_q^{(2)} = k^{(1)} \oplus k^{(2)}}} x_q^{(1)} \times x_q^{(2)}}{\sum_{\substack{1 \le q \le Q \\ \text{s.t. } t_q^{(1)} \oplus t_q^{(2)} = k^{(1)} \oplus k^{(2)}}} 1}. \quad (25)$$

Notice that both $\mathcal{D}_{\text{coll}}$ and $\mathcal{D}_{\text{coll (with corr)}}$ use only a fraction 2^{-n} of the traces for each key guess. This is an inherent limitation of the correlation-collision attacks of CHES 2011 [15]. As explained in Remark 7, both attacks are supposed to be equivalent, which can be checked in Fig. 8 (curve in yellow and dark blue) for a noise of variance $\sigma = 4$.

Appendix B
Correlation-Enhanced Collision Attack

The correlation-enhanced collision attack [33] consists in estimating the average value of the leakage of an SBox for a given text. Then, assuming the two SBoxes leak the same, their average value is expected to match when the correct key is XORed to each corresponding plaintext byte. Moradi *et al.* propose to use a correlation coefficient to estimate the similarity of leakage of each SBox. The attack first estimates moments on one SBox (whose key byte must be known), and then matches it online on the second SBox. Thus, a bivariate attack reveals (like $\mathcal{D}_{\text{coll}}$ and $\mathcal{D}_{\text{sto.coll}}$) only the XOR between the two key bytes.

Formally, let us denote the mean of leakage of SBox ℓ, $1 \le \ell \le L$, conditioned by the text byte $u \in \mathbb{F}_2^n$, as:

$$\hat{\mathbb{E}}(X^{(\ell)} | T^{(\ell)} = u) = \left(\sum_{\substack{1 \le q \le Q \\ \text{s.t. } t_q^{(\ell)} \oplus k^{(\ell)} = u}} 1 \right)^{-1} \left(\sum_{\substack{1 \le q \le Q \\ \text{s.t. } t_q^{(\ell)} \oplus k^{(\ell)} = u}} x_q^{(\ell)} \right).$$

Assume a fictive random variable U uniformly distributed on $\mathbb{F}_2^n$. For a random variable $A(U)$, which depends on U, let us also denote $\langle A(U) \rangle = \frac{1}{2^n} \sum_{u \in \mathbb{F}_2^n} A(u)$. So, the $\mathcal{D}_{\text{corr-coll}}$ distinguisher is computed as:

$$\mathcal{D}_{\text{corr-coll}}(\vec{t}^{(\cdot)}, \vec{x}^{(\cdot)}) = \underset{k^{(1)}, k^{(2)}}{\text{argmax}} \rho_U(\hat{\mathbb{E}}(X^{(1)} | T^{(1)} = U \oplus k^{(1)}), \hat{\mathbb{E}}(X^{(2)} | T^{(2)} = U \oplus k^{(2)})),$$

where the Pearson correlation coefficient is

$$\rho_U(A(U), B(U)) = \frac{\langle A(U)B(U) \rangle - \langle A(U) \rangle \langle B(U) \rangle}{\sqrt{\langle A(U)^2 \rangle - \langle A(U) \rangle^2} \sqrt{\langle B(U)^2 \rangle - \langle B(U) \rangle^2}}.$$

Now, the average $\langle \hat{\mathbb{E}}(X^{(\ell)} | T^{(\ell)} = U \oplus k^{(\ell)}) \rangle = \langle \hat{\mathbb{E}}(X^{(\ell)} | T^{(\ell)} = U) \rangle$ does not depend on a key. Thus, in the computation of $\mathcal{D}_{\text{corr-coll}}$, only the expectation of the product is key-dependent. Hence, $\mathcal{D}_{\text{corr-coll}}$ may be rewritten:

$$\begin{aligned} &\mathcal{D}_{\text{corr-coll}}(\vec{t}^{(\cdot)}, \vec{x}^{(\cdot)}) \\ &= \underset{k^{(1)}, k^{(2)}}{\text{argmax}} \sum_{u \in \mathbb{F}_2^n} \hat{\mathbb{E}}(X^{(1)} | T^{(1)} = u \oplus k^{(1)}) \\ &\qquad \times \hat{\mathbb{E}}(X^{(2)} | T^{(2)} = u \oplus k^{(2)}) \\ &= \underset{k^{(1)}, k^{(2)}}{\text{argmax}} \sum_{u \in \mathbb{F}_2^n} \frac{\sum_{\substack{1 \le q \le Q \\ \text{s.t. } t_q^{(1)} \oplus k^{(1)} = u}} x_q^{(1)}}{\sum_{\substack{1 \le q \le Q \\ \text{s.t. } t_q^{(1)} \oplus k^{(1)} = u}} 1} \times \frac{\sum_{\substack{1 \le q \le Q \\ \text{s.t. } t_q^{(2)} \oplus k^{(2)} = u}} x_q^{(2)}}{\sum_{\substack{1 \le q \le Q \\ \text{s.t. } t_q^{(2)} \oplus k^{(2)} = u}} 1}. \end{aligned}$$

References

[1] S. Barnett, *Matrices: Methods and Applications* (Oxford Applied Mathematics and Computing Science Series). Jun. 1990.

[2] P. Belgarric *et al.*, "Time-frequency analysis for second-order attacks," in *Smart Card Research and Advanced Applications* (Lecture Notes in Computer Science), vol. 8419, A. Francillon and P. Rohatgi, Eds. Springer, 2013, pp. 108–122.

[3] G. S. G. Beveridge and R. S. Schechter, *Optimization: Theory and Practice*. New York, NY, USA: McGraw-Hill, 1970.

[4] S. Bhasin, J.-L. Danger, S. Guilley, and W. He, "Exploiting FPGA block memories for protected cryptographic implementations," *ACM Trans. Reconfigurable Technol. Syst.*, vol. 8, no. 3, pp. 16:1–16:16, May 2015.

[5] J. Blömer, J. Guajardo, and V. Krummel, "Provably secure masking of AES," in *Selected Areas in Cryptography* (Lecture Notes in Computer Science), vol. 3357, H. Handschuh and M. A. Hasan, Eds. Springer, 2004, pp. 69–83.

[6] A. Bogdanov, "Improved side-channel collision attacks on AES," in *Selected Areas in Cryptography* (Lecture Notes in Computer Science), vol. 4876, C. M. Adams, A. Miri, and M. Wiener, Eds. Springer, 2007, pp. 84–95.

[7] A. Bogdanov, "Multiple-differential side-channel collision attacks on AES," in *Cryptographic Hardware and Embedded Systems* (Lecture Notes in Computer Science), vol. 5154, E. Oswald and P. Rohatgi, Eds. Springer, 2008, pp. 30–44.

[8] É. Brier, C. Clavier, and F. Olivier, "Correlation power analysis with a leakage model," in *Cryptographic Hardware and Embedded Systems* (Lecture Notes in Computer Science), vol. 3156. Springer, Aug. 2004, pp. 16–29.

[9] N. Bruneau, S. Guilley, A. Heuser, D. Marion, and O. Rioul, "Less is more—Dimensionality reduction from a theoretical perspective," in *Cryptographic Hardware and Embedded Systems* (Lecture Notes in Computer Science), vol. 9293, T. Güneysu and H. Handschuh, Eds. Springer, Sep. 2015, pp. 22–41.

[10] N. Bruneau, S. Guilley, A. Heuser, and O. Rioul, "*Masks will fall off* — Higher-order optimal distinguishers," in *Proc. 20th Int. Conf. Theory Appl. Cryptol. Inf. Secur. (ASIACRYPT)*, vol. 8874. Kaohsiung, Taiwan, Dec. 2014, pp. 344–365.

[11] C. Carlet and S. Guilley, "Side-Channel Indistinguishability," in *Proc. HASP*, New York, NY, USA, Jun. 2013, pp. 9:1–9:8.

[12] C. Carlet and S. Guilley. (Jul. 19, 2014). *Side-Channel Indistinguishability*. [Online]. Available: http://hal.archives-ouvertes.fr/hal-00826618

[13] S. Chari, J. R. Rao, and P. Rohatgi, "Template attacks," in *Cryptographic Hardware and Embedded Systems* (Lecture Notes in Computer Science), vol. 2523. Springer, Aug. 2002, pp. 13–28.

[14] C. Chen, T. Eisenbarth, A. Shahverdi, and X. Ye, "Balanced encoding to mitigate power analysis: A case study," in *Smart Card Research and Advanced Applications* (Lecture Notes in Computer Science), vol. 8968. Springer, Nov. 2014.

[15] C. Clavier, B. Feix, G. Gagnerot, M. Roussellet, and V. Verneuil, "Improved collision-correlation power analysis on first order protected AES," in *Cryptographic Hardware and Embedded Systems* (Lecture Notes in Computer Science), vol. 6917, B. Preneel and T. Takagi, Eds. Springer, 2011, pp. 49–62.

[16] G. Dabosville, J. Doget, and E. Prouff, "A new second-order side channel attack based on linear regression," *IEEE Trans. Comput.*, vol. 62, no. 8, pp. 1629–1640, Aug. 2013.

[17] J. Doget, E. Prouff, M. Rivain, and F.-X. Standaert, "Univariate side channel attacks and leakage modeling," *J. Cryptogr. Eng.*, vol. 1, no. 2, pp. 123–144, 2011.

[18] M. A. Elaabid and S. Guilley, "Portability of templates," *J. Cryptogr. Eng.*, vol. 2, no. 1, pp. 63–74, 2012, doi: 10.1007/s13389-012-0030-6.

[19] Y. Fei, Q. Luo, and A. A. Ding, "A statistical model for DPA with novel algorithmic confusion analysis," in *Cryptographic Hardware and Embedded Systems* (Lecture Notes in Computer Science), vol. 7428, E. Prouff and P. Schaumont, Eds. Springer, 2012, pp. 233–250.

[20] B. Gierlichs, L. Batina, P. Tuyls, and B. Preneel, "Mutual information analysis," in *Cryptographic Hardware and Embedded Systems* (Lecture Notes in Computer Science), vol. 5154. Springer, Aug. 2008, pp. 426–442.

[21] A. Heuser, M. Kasper, W. Schindler, and M. Stöttinger, "A new difference method for side-channel analysis with high-dimensional leakage models," in *Topics in Cryptology* (Lecture Notes in Computer Science), vol. 7178, O. Dunkelman, Ed. Springer, 2012, pp. 365–382.

[22] A. Heuser, O. Rioul, and S. Guilley, "Good is not good enough—Deriving optimal distinguishers from communication theory," in *Cryptographic Hardware and Embedded Systems* (Lecture Notes in Computer Science), vol. 8731, L. Batina and M. Robshaw, Eds. Springer, Sep. 2014, pp. 55–74.

[23] P. C. Kocher, "Timing attacks on implementations of Diffie–Hellman, RSA, DSS, and other systems," in *Advances in Cryptology* (Lecture Notes in Computer Science), vol. 1109. Springer-Verlag, 1996, pp. 104–113.

[24] P. Kocher, J. Jaffe, and B. Jun, "Differential power analysis," in *Advances in Cryptology* (Lecture Notes in Computer Science), vol. 1666, M. J. Wiener, Ed. Springer, 1999, pp. 388–397.

[25] K. Lemke-Rust and C. Paar, "Gaussian mixture models for higher-order side channel analysis," in *Cryptographic Hardware and Embedded Systems* (Lecture Notes in Computer Science), vol. 4727, P. Paillier and I. Verbauwhede, Eds. Springer, 2007, pp. 14–27.

[26] H. Maghrebi, O. Rioul, S. Guilley, and J.-L. Danger, "Comparison between side-channel analysis distinguishers," in *Information and Communications Security* (Lecture Notes in Computer Science), vol. 7618, T. W. Chim and T. H. Yuen, Eds. Springer, 2012, pp. 331–340.

[27] P. C. Mahalanobis, "On the generalised distance in statistics," *Proc. Nat. Inst. Sci. India*, vol. 2, no. 1, pp. 49–55, 1936.

[28] S. Mangard, E. Oswald, and T. Popp, *Power Analysis Attacks: Revealing the Secrets of Smart Cards*. Springer, Dec. 2006. [Online]. Available: http://www.dpabook.org/

[29] G. McLachlan and T. Krishnan, *The EM Algorithm and Extensions* (Wiley Series in Probability and Statistics), 2nd ed. Hoboken, NJ, USA: Wiley, Apr. 2008.

[30] D. P. Montminy, R. O. Baldwin, M. A. Temple, and E. D. Laspe, "Improving cross-device attacks using zero-mean unit-variance normalization," *J. Cryptogr. Eng.*, vol. 3, no. 2, pp. 99–110, 2013.

[31] A. Moradi, "Statistical tools flavor side-channel collision attacks," in *Advances in Cryptology* (Lecture Notes in Computer Science), vol. 7237, D. Pointcheval and T. Johansson, Eds. Springer, 2012, pp. 428–445.

[32] A. Moradi, S. Guilley, and A. Heuser, "Detecting hidden leakages," in *Applied Cryptography and Network Security*, vol. 8479, I. Boureanu, P. Owesarski, and S. Vaudenay, Eds. Springer, Jun. 2014.

[33] A. Moradi, O. Mischke, and T. Eisenbarth, "Correlation-enhanced power analysis collision attack," in *Cryptographic Hardware and Embedded Systems* (Lecture Notes in Computer Science), vol. 6225. Springer, Aug. 2010, pp. 125–139.

[34] M. Renauld, F.-X. Standaert, N. Veyrat-Charvillon, D. Kamel, and D. Flandre, "A formal study of power variability issues and side-channel attacks for nanoscale devices," in *Advances in Cryptology* (Lecture Notes in Computer Science), vol. 6632, K. G. Paterson, Ed. Springer, May 2011, pp. 109–128.

[35] T. Roche and V. Lomné, "Collision-correlation attack against some 1^{st}-order Boolean masking schemes in the context of secure devices," in *Constructive Side-Channel Analysis and Secure Design* (Lecture Notes in Computer Science), vol. 7864, E. Prouff, Ed. Springer, 2013, pp. 114–136.

[36] W. Schindler, K. Lemke, and C. Paar, "A stochastic model for differential side channel cryptanalysis," in *Cryptographic Hardware and Embedded Systems* (Lecture Notes in Computer Science), vol. 3659. Springer, Sep. 2005, pp. 30–46.

[37] K. Schramm, G. Leander, P. Felke, and C. Paar, "A collision-attack on AES: Combining side channel- and differential-attack," in *Cryptographic Hardware and Embedded Systems* (Lecture Notes in Computer Science), vol. 3156, M. Joye and J. J. Quisquater, Eds. Springer, Aug. 2004, pp. 163–175.

[38] K. Schramm, T. Wollinger, and C. Paar, "A new class of collision attacks and its application to DES," in *Fast Software Encryption* (Lecture Notes in Computer Science), vol. 2887, T. Johansson, Ed. Springer, Feb. 2003, pp. 206–222.

[39] V. Servant, N. Debande, H. Maghrebi, and J. Bringer, "Study of a novel software constant weight implementation," in *Smart Card Research and Advanced Applications* (Lecture Notes in Computer Science). Springer, Nov. 2014.

[40] J. A. Snyman, *Practical Mathematical Optimization: An Introduction to Basic Optimization Theory and Classical and New Gradient-Based Algorithms* (Applied Optimization). New York, NY, USA: Springer, 2005.

[41] F.-X. Standaert, F. Koeune, and W. Schindler, "How to compare profiled side-channel attacks?" in *Applied Cryptography and Network Security* (Lecture Notes in Computer Science), vol. 5536. Heidelberg, Germany: Springer, 2009, pp. 485–498.

[42] TELECOM ParisTech SEN Research Group. *DPA Contest (4th Edition), 2013–2014*. [Online]. Available: http://www.DPAcontest.org/v4/

[43] K. Tiri *et al.*, "Prototype IC with WDDL and differential routing—DPA resistance assessment," in *Cryptographic Hardware and Embedded Systems* (Lecture Notes in Computer Science), vol. 3659. Springer, Aug./Sep. 2005, pp. 354–365.

[44] R. Velegalati and J.-P. Kaps, "Techniques to enable the use of block RAMs on FPGAs with dynamic and differential logic," in *Proc. IEEE Int. Conf. Electron., Circuits, Syst. (ICECS)*, Dec. 2010, pp. 1251–1254.

[45] N. Veyrat-Charvillon, B. Gérard, and F.-X. Standaert, "Soft analytical side-channel attacks," in *Advances in Cryptology* (Lecture Notes in Computer Science), vol. 8873, P. Sarkar and T. Iwata, Eds. Springer, Dec. 2014, pp. 282–296.
[46] C. F. J. Wu, "On the convergence properties of the EM algorithm," *Ann. Statist.*, vol. 11, no. 1, pp. 95–103, 1983.

Nicolas Bruneau received the M.S. degree in communication security from the University of Lyon in 2012. He is currently pursuing the Ph.D. degree with Télécom ParisTech, with a focus on Multivariate Multitarget High-Order Side-Channel Attacks. He was a Research and Development Engineer with STMicroelectronics, Rousset. He is currently a Research and Development Engineer with Secure-IC, where he is involved in the innovative Virtualyzr security evaluation tool in the Threat Protection business line. His research interests include cryptographic engineering and secure embedded systems. In particular, he has been conducting research toward evaluating the security of protected implementations for more than four years.

Claude Carlet received the Ph.D. degree from the University of Paris 6, France, in 1990, and the Habilitation degree to Direct theses from the University of Amiens, France, in 1994. He was an Associate Professor with the Department of Computer Science, University of Amiens, from 1990 to 1994, and a Professor with the Department of Computer Science, University of Caen, France, from 1994 to 2000, and the Department of Mathematics, University of Paris 8, France, from 2000 to 2017. He has participated as the author of chapters or as editor to 11 books. He has authored or coauthored 100 international journal papers, 60 papers in international proceedings, and 20 shorter international papers. He has supervised 13 Ph.D. students and is currently supervising five. His research interests include Boolean functions (bent, correlation-immune, algebraic immune, and SAC), vectorial functions (bent-PN and APN), cryptography (in particular, stream ciphers, block ciphers, and side-channel attacks), finite fields, and coding theory (in relationship with the domains mentioned earlier). He has been a member of 70 program committees of international conferences and workshops (seven as a Co-Chair). He has been in charge of the French research group Codage-Cryptographie C2 (gathering the researchers in these domains) during ten years. He has been a Plenary Invited Speaker in 20 international conferences and an Invited Speaker in 25 other international conferences and workshops. He has been an Associate Editor of the IEEE TRANSACTIONS ON INFORMATION THEORY. He is currently the Editor-in-Chief of the journal *Cryptography and Communications* (Springer) and Editor of the four journals DCC (Springer), American Institute of Mathematical Sciences (AMC), IJCM-TCOM (Taylor & Francis), and IJOCT (Inderscience Publishers).

Sylvain Guilley (M'08) received the degree from the École Polytechnique (X97), Telecom ParisTech, in 2002, the M.Sc. degree in quantum physics from ENS, Paris 6 University, in 2010, the Ph.D. degree in digital electronics from Telecom ParisTech in 2007, and the HDR degree in mathematical cryptography from Paris 7 University in 2012. He is currently the Director of the Business Line Think Ahead with Secure-IC, an international SME headquartered in Rennes, France, with business branches in Paris, Singapore, and Tokyo. He is also an Ingénieur en Chef des Mines and a Full Professor with Télécom ParisTech, Université Paris-Saclay, France. He has been conducting research toward defining provable secure architectures for trusted computing for more than ten years. He authored over 200 scientific publications and patents related to security and embedded systems. He is a member of IACR and a Senior Member of the Cryptarchi Club. He is an active member in the ISO/IEC sub-committee SC27, dealing with information security. He is a Rapporteur for a study period on white box cryptography, and leads RISQ (http://www.risq.fr/), a French-wide national project involving 15 partners working on quantum-safe cryptography. He is an editor of standard projects on physically unclonable functions (ISO 20897) and side-channel calibration (ISO 20085).

Annelie Heuser received the Diploma (M.S.) degree in mathematics from Technische Universität, Darmstadt, Germany, with a special focus on computer science, and the Ph.D. degree from Telecom ParisTech in 2016. She was a Post-Doctoral Researcher with Telecom ParisTech. She was a Researcher with the Center of Advanced Security Research Darmstadt, Germany. She is currently a Researcher with the French National Center for Scientific Research, IRISA, Rennes, France. She has authored over 20 publications at international scientific conferences and journals. Her main research interests lie in the area of side-channel analysis, machine learning, hardware security, and malware analysis.

Emmanuel Prouff received the Ph.D. degree in applied mathematics from the Université de Caen Basse-Normandie and the Institut National de Recherche en Informatique et Automatique, Paris. He was an Expert in embedded systems security with the Agence Nationale de la Sécurité des Systèmes d'Information and a Cryptography and Security Research Manager with Oberthur Technologies. He is currently the Lead of the Cryptography and Security Team with Safran Identity and Security and an Associate Member of the POLSYS Team with the University Pierre et Marie Curie (Paris 6). He is also involved in physical attacks and dedicated countermeasures for cryptographic protocols and design of secure protocols for embedded systems.

Olivier Rioul graduated from École Polytechnique, Paris, France, in 1987 and from École Nationale Supérieure des Télécommunications, Paris, in 1989. He received the Ph.D. degree from École Nationale Supérieure des Télécommunications, Paris, in 1993. He is currently a Full Professor with the Department of Communication and Electronics, Télécom ParisTech, Université Paris-Saclay, France, and also a Professor Chargé de Cours with the Department of Applied Mathematics, École Polytechnique, Université Paris-Saclay. His research interests are in applied mathematics and include various, sometimes unconventional, applications of information theory, such as inequalities in statistics, hardware security, and experimental psychology.