Noname manuscript No.
(will be inserted by the editor)

Univariate Side Channel Attacks and Leakage Modeling

Julien Doget · Emmanuel Prouff · Matthieu Rivain · François-Xavier Standaert

the date of receipt and acceptance should be inserted later

Abstract Differential power analysis is a powerful cryptanalytic technique that exploits information leaking from physical implementations of cryptographic algorithms. During the two last decades numerous variations of the original principle have been published. In particular, the univariate case, where a single instantaneous leakage is exploited, has attracted much research effort. In this paper, we argue that several univariate attacks among the most frequently used by the community are not only asymptotically equivalent, but can also be rewritten one in function of the other, only by changing the leakage model used by the adversary. In particular, we prove that most univariate attacks proposed in the literature can be expressed as correlation power analyses with different leakage models. This result emphasizes the major role plays by the model choice on the attack efficiency. In a second point of this paper we hence also discuss and evaluate side channel attacks that involve no leakage model but rely on some general assumptions about the leakage. Our experiments show that such attacks, named robust, are a valuable alternative to the univariate differential power analyses. They only loose bit of efficiency in case a perfect model is available to the adversary, and gain a lot in case such information is not available.

Keywords Side Channel Attack · Correlation · Regression · Model

Research associate of the Belgian Fund for Scientific Research (FNRS - F.R.S.).

J. Doget · E. Prouff
Oberthur Technologies, 71-73 rue des Hautes Pâtures, F-92 726 Nanterre, France
E-mail: {j.doget, e.prouff}@oberthur.com

J. Doget
Université Catholique de Louvain-la-Neuve, UCL Crypto Group, B-1348 Louvain-la-Neuve, Belgium

J.Doget
Université Paris 8, Département de Mathématiques, 2, rue de la Liberté, F-93 526 Saint-Denis, France

F.-X. Standaert
Université Catholique de Louvain-la-Neuve, UCL Crypto Group, B-1348 Louvain-la-Neuve, Belgium
E-mail: fstandae@uclouvain.be

M. Rivain
CryptoExperts, Paris, France
E-mail: matthieu.rivain@cryptoexperts.com

1 Introduction

The goal of a Differential Power Analysis (DPA) is to take advantage of the key-dependent physical leakages provided by a cryptographic device, in order to recover secret information (key bytes, typically). Most of these attacks exploit the leakages by comparing them with key-dependent models that are available for the target device. Since the seminal work of Kocher *et al.* in the late 1990's [1], a large variety of statistical tests, also called distinguishers, have been introduced for this purpose. Namely, the original attack (that we will always refer to as DPA for convenience) was described using a Difference-of-Means test. Following works, including the all-or-nothing multiple-bit DPA [2], the generalized multiple-bit DPA [2], the Correlation Power Analysis (CPA) [3], the Partitioning Power Analysis (PPA) [4] and the enhanced DPA of Knudsen and Bévan [5], systematically proposed ways to enhance the Difference-of-Means test. Their goal was to better take advantage of the available information, *e.g.* , by allowing the adversary to incorporate more precise leakage models in the statistics. Hence, and in view of the large variety of distinguishers available in the literature, a natural question is to determine the exact relations between them and the conditions upon which one of them would be more efficient.

Closely related to this question, Mangard *et al.* showed in [6] that for a category of attacks, denoted as standard univariate DPA, a number of distinguishers (namely, those using a Difference-of-Means test or a Pearson's correlation coefficient or Gaussian templates) are in fact asymptotically equivalent, given that they are provided with the same *a priori* information about the leakages (*i.e.* if they use the same model). More precisely, [6] shows that these distinguishers only differ in terms that become key-independent once properly estimated. While this result is limited to first-order (aka univariate) attacks, it clearly underlines that the selection (or construction) of a proper leakage model in Side Channel Attacks (SCA) is at least as important as the selection of a good distinguisher.

A natural extension of Mangard *et al.* 's work is to study whether their statement holds in non-asymptotic contexts (*i.e.* when the number of measurements is reasonably small). Such a study is of particular importance since it corresponds to a practical issue from both the attacker and the security designer side. Indeed the latter ones often need to precisely determine which of the numerous existing attacks is the most suitable one in a given context, or reciprocally, which context is the most appropriate one for a given attack.

The results in this paper can be seen as a complement to the previous analyses and are in two parts. We first focus on the aforementioned list of non-profiled side channel distinguishers. We prove that they not only are asymptotically equivalent but also, that they can be explicitly re-written one in function of another, by only changing the leakage model. In other words, we show that all these distinguishers exploit essentially the same statistics and that any difference can be expressed as a change of model. This provides us with a unified framework to study and compare the attacks. Moreover, this emphasizes how strong the impact of the model choice on the attack efficiency is. Since a good leakage model is not always available to the attacker, we study in the second part of this paper, side channel attacks introduced in [7] which do not relate on a model choice and can be performed with a few general assumptions about the leakage. Those attacks are presented and analysed in the unified framework introduced in the first two sections of the paper. Our results show that such *robust side channel attacks*[1] are only slightly less efficient than a correlation power analysis performed with a perfect leakage model (which is a very favourable context for the CPA). At the opposite when no perfect leakage model is available, robust side channel attacks are more efficient than a correlation power analysis. Moreover in this case, they can deal with situations in which a correlation power analysis would fail.

[1] The term *robust* is related to the statistical notion of *robustness* that is the property of being insensitive to small deviations from assumptions.

2 Background

Let $\mathsf{E}_K(p)$ denote the output of the encryption of a plaintext p parameterized by a master key K. Let v_k be an intermediate result occurring during the processing of $\mathsf{E}_K(p)$, which can be expressed as a deterministic function of the plaintext p and a guessable part k of the secret key K (*e.g.* , an S-box output in a Substitution-Permutation Network (SPN) cipher). We shall refer to v_k as *sensitive variable* in the following. We consider an adversary, who has access to a physical implementation of $\mathsf{E}_K(\cdot)$ and, who observes the side channel leakage of N successive encryptions of plaintexts p_i. Each encryption $\mathsf{E}_K(p_i)$ gives rise to a value $v_{k,i}$ of the sensitive variable. The computation of this intermediate result by the device generates some physical leakage $\ell_{k,i}$. We denote by V_k and L the random variables over the sample $(v_{k,i})_i$ and $(\ell_{k,i})_i$ respectively. We assume the leakage L to be composed of two parts: a deterministic part $\delta(\cdot)$ and an independent noise B such that

$$L = \delta(V_k) + B \ , \tag{1}$$

which implies

$$\ell_{k,i} = \delta\left(v_{k,i}\right) + b_i \ ,$$

where b_i denotes the leakage noise value in the i^{th} leakage measurement.

Assumption 1 (Independent Noise) *The noise B is independent of the sensitive variable V_k.*

To mount an attack, the adversary measures leakages $(\ell_{k,i})_i$ from the targeted device using a sample $(p_i)_i$ of plaintexts. Then, he computes the hypothetic value $v_{\hat{k},i}$ of the sensitive variable $v_{k,i}$ for every p_i and for every possible $\hat{k}$. A *leakage model function* m is subsequently applied to map the hypothetic sensitive values toward estimated leakage values $m_{\hat{k},i} = \mathsf{m}(v_{\hat{k},i})$. Eventually, the adversary uses a distinguisher to compare the different model samples $(m_{\hat{k},i})_i$ with the actual leakage sample $(\ell_{k,i})_i$. If the attack is successful, the best comparison result (*i.e.* , the highest – or lowest – value of the distinguisher) should be obtained for the model sample corresponding to the correct subkey candidate $\hat{k} = k$. This procedure can then be repeated for different subkeys in order to eventually recover the full master key.

We sum-up hereafter the different steps of a standard univariate SCA:

1. Perform N measurements $(\ell_{k,i})_i$ on the cryptographic device using a sample $(p_i)_i$ of N plaintexts.
2. Choose a function m to model the deterministic part of the leakage.
3. For every key hypothesis $\hat{k}$, compute the model values $m_{\hat{k},i}$ from the plaintexts p_i's and the model function m.
4. Choose a statistical distinguisher Δ.

5. For every key hypothesis $\hat{k}$, compute the *distinguishing value* $\Delta_{\hat{k}}$ defined by:

$$\Delta_{\hat{k}} = \Delta\left((\ell_{k,i})_i, (m_{\hat{k},i})_i\right) \ .$$

This results in a *score vector* $(\Delta_{\hat{k}})_{\hat{k}}$.
6. Output as the o most likely key candidates the o key hypotheses that maximize – or minimize – $\Delta_{\hat{k}}$.

As it can be seen in the previous list, a standard univariate SCA on a given sensitive variable v_k is only characterized by the model function m and the distinguisher Δ. For this reason we shall use in the following the notation (m, Δ)-*SCA* to differentiate one such an attack from another.

In the rest of the paper we aim to compare different distinguishers targeting the same intermediate variable. For this purpose, we introduce hereafter the notion of *reduction between two SCAs*:

Definition 1 (SCA-reduction) A (m, Δ)-SCA is said to be *SCA-reducible* to a (m', Δ')-SCA if there exists a function f such that $\mathsf{m} = \mathsf{f} \circ \mathsf{m}'$ and for every pair $(k, \hat{k})$ and every samples $(\ell_{k,i})_i$ and $(v_{\hat{k},i})_i$, there exists a strictly monotonous function g such that:

$$\Delta\left((\ell_{k,i})_i, (m_{\hat{k},i})_i\right) = \mathsf{g} \circ \Delta'\left((\ell_{k,i})_i, (m'_{\hat{k},i})_i\right) \ ,$$

where $m_{\hat{k},i} = \mathsf{m}(v_{\hat{k},i})$ and $m'_{\hat{k},i} = \mathsf{m}'(v_{\hat{k},i})$.

Definition 2 (SCA-equivalence) Let A be a (m, Δ)-SCA and let B be a (m', Δ')-SCA. A is said to be *SCA-equivalent* to B if and only if A is SCA-reducible to B and B is SCA-reducible to A.

It is clear from the general attack description recalled above that two major choices are left to the adversary when the latter one wishes to perform a standard SCA attack on a given sensitive variable computed on some device:

- the choice of the distinguisher,
- the choice of the model.

In this paper, we will study both questions and will show that they are linked. We will first show that most of univariate SCA distinguishers that have been proposed in the literature give rise to attacks reducible to CPA under Definition 1. Namely, they lead to similar results up to a change of model. We will then discuss the importance of the model for the attack soundness and we will investigate attacks that do not require any *a priori* choice of a model.

2.1 Notations

Let X be a random variable and let x and Ω be respectively an element and a subset of the definition set $\mathscr{X}$ of X. In the rest of the paper, we shall denote by $\mathrm{P_r}(X = x)$ and $\mathrm{P_r}(X \in \Omega)$ the probabilities associated with the events $(X = x)$ and $(X \in \Omega)$ respectively. We shall moreover denote by $\mathbb{E}(X)$ the expectation of X. Estimations of the expectation and of the probability over a sample $(x_i)_i$ of values taken by X shall be denoted by $\widehat{\mathbb{E}}(X)$ and $\widehat{\mathrm{P}}_{\mathrm{r}}(X = x)$ respectively. For instance, if N denotes the size of the sample $(\ell_{k,i})_i$, notations $\widehat{\mathbb{E}}(L)$ and $\widehat{\mathrm{P}}_{\mathrm{r}}(L = \ell)$ shall refer to the mean value $\frac{1}{N}\sum_i \ell_{k,i}$ of the leakage sample and to ratio $\frac{\#\{i;\ell_{k,i}=\ell\}}{N}$. Eventually, we shall say that a sample $(x_i)_i$ of a random variable X is a *balanced sample* if it contains each value of $\mathscr{X}$ a same number of times. Clearly, the size N of such a sample is a multiple of the cardinality of $\mathscr{X}$.

The random variable related to the observations $v_{\hat{k},i}$ and $m_{\hat{k},i}$ will be denoted by $V_{\hat{k}}$ and $M_{\hat{k}}$ respectively. Throughout this paper we will hence have $M_{\hat{k}} = \mathsf{m}(V_{\hat{k}})$.

3 Reduction Between Various Side Channel Attacks

In this section, we first describe the focused distinguishers and then we give reduction relations between them.

3.1 Distinguisher Descriptions

The first (m, Δ)-SCA was introduced by Kocher *et al.* in [1], and was called *Differential Power Analysis*. It targets a single bit of the sensitive variable v_k and shall be therefore referred to as *single-bit DPA* in the rest of the paper. Since this bit usually depends on all bits of the subkey, the single-bit DPA may allow to unambiguously discriminate the correct subkey. However, for some kinds of algebraic relationships between the manipulated data and the subkey, several key candidates (including the correct one) may result in the same distinguishing value and the attack fails (this phenomenon is referred to as *ghost peaks* in [3]). To exploit more information from the leakage related to the manipulation of v_k and to succeed when single-bit DPA does not, the attack was extended to several bits by Messerges in [8] in two ways: the *all-or-nothing DPA* and the *generalized DPA*. The original single-bit DPA of Kocher and its extensions by Messerges can all be defined in a similar way as follows:

Definition 3 (Differential Power Analysis (DPA)) A *DPA* is a (m, Δ)-SCA, which involves a distinguisher Δ defined as a *Difference of Means (DoM)* between two leakage partitions defined according to the image set $\mathrm{Im}(\mathsf{m})$.

Depending on the definition of the leakage model function m, we recognize the classical presentations of the three DPA attacks listed above:

- In a *single-bit DPA*, the image set $\mathrm{Im}(\mathsf{m})$ is reduced to two elements w_0 and w_1 and for every $\hat{k}$ we have:

$$\Delta_{\hat{k}} = \widehat{\mathbb{E}}\left(L \mid M_{\hat{k}} = w_0\right) - \widehat{\mathbb{E}}\left(L \mid M_{\hat{k}} = w_1\right) \ . \quad (2)$$

– In an *all-or-nothing DPA*, the image set $\mathrm{Im}(\mathsf{m})$ can have a cardinality greater than 2. Two elements ω_0 and ω_1 are chosen in $\mathrm{Im}(\mathsf{m})$ and for every $\hat{k}$ we have:

$$\Delta_{\hat{k}} = \widehat{\mathbb{E}}\left(L \mid M_{\hat{k}} = \omega_0\right) - \widehat{\mathbb{E}}\left(L \mid M_{\hat{k}} = \omega_1\right) \ . \tag{3}$$

– In a *generalized DPA*, two subsets Ω_0 and Ω_1 of $\mathrm{Im}(\mathsf{m})$ are chosen and for every $\hat{k}$ we have:

$$\Delta_{\hat{k}} = \widehat{\mathbb{E}}\left(L \mid M_{\hat{k}} \in \Omega_0\right) - \widehat{\mathbb{E}}\left(L \mid M_{\hat{k}} \in \Omega_1\right) \ . \tag{4}$$

Distinguishers $\Delta_{\hat{k}}$ defined in (2) - (4) shall be denoted by $\mathsf{SB\text{-}DPA}(\hat{k})$, $\mathsf{AON\text{-}DPA}(\hat{k})$ and $\mathsf{G\text{-}DPA}(\hat{k})$ respectively, where $\hat{k}$ is the key hypothesis.

Example 1 Typical choices for the model functions in (2) - (4) are as follows [1, 8]:

Single-bit DPA: m is the function that maps the vector $v_{\hat{k}}$ to one of its bit-coordinates and we hence have $\mathrm{Im}(\mathsf{m}) = \{\omega_0, \omega_1\} = \{0, 1\}$.

All-or-nothing DPA: m is the Hamming weight and thus we have $\{\omega_0, \omega_1\} = \{0, n\}$ (n being the bit-size of $v_{\hat{k}}$).

Generalized DPA: m is the Hamming weight and thus we have $\{\Omega_0, \Omega_1\} = \left\{\{1, \ldots, \lfloor \frac{n}{2} \rfloor\}, \{\lceil \frac{n}{2} \rceil, \ldots, n\}\right\}$.

However, different choices for m, (ω_0, ω_1) and (Ω_0, Ω_1) may be arbitrary made by the attacker, which explains why we do not fix a particular choice in this paper.

After Messerges' works, two extensions of the DPA have been proposed respectively by Le *et al.* in [4] and by Brier *et al.* in [3].

The generalization proposed in [4] starts from (4) and enables to involve more than 2 subsets to eventually compute a weighted sum of means instead of a simple DoM. We recall hereafter its definition:

Definition 4 (Partition Power Analysis (PPA)) A *PPA* is a (m, Δ)-SCA, which involves a distinguisher Δ defined for every $\hat{k}$ by:

$$\Delta_{\hat{k}} = \sum_{\omega_i \in \mathrm{Im}(\mathsf{m})} \alpha_i \cdot \widehat{\mathbb{E}}\left(L \mid M_{\hat{k}} = \omega_i\right) \ , \tag{5}$$

where the α_i's are constant coefficients in $\mathbb{R}$.

A distinguisher $\Delta_{\hat{k}}$ defined such as in (5) shall be denoted $\mathsf{PPA}_{(\alpha_i)_i}(\hat{k})$. Moreover, when we shall need to exhibit the model m in the PPA, we shall use the notation $\mathsf{PPA}_{(\alpha_i)_i,\mathsf{m}}(\hat{k})$ for the distinguisher. As discussed in [4], the tricky part when specifying a PPA attack is the choice of the most suitable coefficients α_i's.

The generalization of the DPA proposed in [9] involves the *linear correlation coefficient*. We recall hereafter the definition of this attack:

Definition 5 (Correlation Power Analysis (CPA)) A *CPA* is a (m, Δ)-SCA, which involves the Pearson's correlation coefficient ρ as distinguisher. Namely, for every $\hat{k}$, we have:

$$\Delta_{\hat{k}} = \widehat{\rho}\left(L, M_{\hat{k}}\right) = \frac{\widehat{\mathrm{cov}}\left(L, M_{\hat{k}}\right)}{\widehat{\sigma}\left(L\right) \cdot \widehat{\sigma}\left(M_{\hat{k}}\right)} \ , \tag{6}$$

where $\widehat{\sigma}\left(L\right)$ and $\widehat{\sigma}\left(M_{\hat{k}}\right)$ denote the standard deviations of the samples $\left(\ell_{k,i}\right)_i$ and $\left(m_{\hat{k},i}\right)_i$ respectively and where their covariance is denoted by $\widehat{\mathrm{cov}}\left(L, M_{\hat{k}}\right)$ which is $\widehat{\mathbb{E}}\left(LM_{\hat{k}}\right) - \widehat{\mathbb{E}}\left(L\right)\widehat{\mathbb{E}}\left(M_{\hat{k}}\right)$.

A distinguisher $\Delta_{\hat{k}}$ defined such as in (6) shall be denoted by $\mathsf{CPA}(\hat{k})$. Moreover, when we shall need to exhibit the model m used in the CPA, we shall use the notation $\mathsf{CPA}_{\mathsf{m}}(\hat{k})$ for the distinguisher.

The attacks listed above have been applied in many papers, *e.g.* , [8, 10, 11] and have even been sometimes experimentally compared one to another [6, 12]. However, none of those works have enabled to draw definitive conclusions about the similarities and the differences of the attacks. Next sections aim to overcome this lack. The study shall be conducted under the following assumption:

Assumption 2 (Target Uniformity) *The predicted variable sample $\left(v_{\hat{k},i}\right)_i$ is balanced for every key hypothesis $\hat{k}$.*

Remark 1 Assumption 2 is realistic in the SCA context. Indeed, the $\left(v_{\hat{k},i}\right)_i$'s result from the evaluation of a *balanced* cryptographic primitive (*e.g.* , an S-box or a linear operation over a small vector space), and we can fairly assume when N is large enough that $\left(v_{\hat{k},i}\right)_i$ is a balanced sample.

Remark 2 Since m is defined over the definition set of the values $v_{\hat{k}}$ and since the distribution over $(v_{\hat{k}})_i$ is balanced independent of $\hat{k}$, then Assumption 2 implies that the mean and the standard deviation of $M_{\hat{k}} = \mathsf{m}(V_{\hat{k}})$ are always estimated from a balanced sample. As a consequence, those estimations are constant with respect to the key hypothesis $\hat{k}$ and exactly correspond to the mean $\mathbb{E}\left(M_{\hat{k}}\right)$ and the standard deviation $\sigma\left(M_{\hat{k}}\right)$ of $M_{\hat{k}}$.

In what follows, we state the SCA-reductions between DPA, PPA and CPA (Sections 3.2 and 3.3). We show that each of those attacks can be reformulated to reveal a correlation coefficient computation and that they only differ in the involved model function. A direct consequence of this statement is that comparing those attacks simply amounts to compare the accuracy/soundness of the underlying models. Afterward, we address attacks that consist in summing distinguishers and we show that they are also SCA-reducible to a CPA (Section 3.4). These results emphasize the importance of making a good choice for the model according to the attack context specificities, which is eventually discussed (Section 3.5).

3.2 From DPA to PPA

As the PPA is a generalization of the DPA that is based on the same statistical tool (namely a DoM test), we can reasonably expect that all the DPA presented in Section 3.1 can be rewritten in terms of a PPA. We give in the following proposition a formal proof for this intuition. Note that our proof is constructive and we exhibit how to reformulate any DPA in terms of a PPA.

Proposition 1 *Let* $\mathsf{DPA}(\hat{k})$ *be one of the DPA defined in* (2) - (4). *There exist coefficients* $(\alpha_i)_i$ *such that* $\mathsf{DPA}(\hat{k}) = \mathsf{PPA}_{(\alpha_i)_i}(\hat{k})$.

Proof. Let us first focus on the $\mathsf{SB\text{-}DPA}(\hat{k})$ distinguisher and let us denote by α_0 and α_1 respectively the coefficients 1 and -1. Relation (2) can be rewritten:

$$\mathsf{SB\text{-}DPA}(\hat{k}) = \alpha_0 \widehat{\mathbb{E}}\left(L \mid M_{\hat{k}} = w_0\right) + \alpha_1 \widehat{\mathbb{E}}\left(L \mid M_{\hat{k}} = w_1\right) \ . \tag{7}$$

The right part of (7) defines a PPA distinguisher $\mathsf{PPA}(\hat{k})$ involving the same 2-valued model m as $\mathsf{SB\text{-}DPA}(\hat{k})$ and the pair of coefficients (α_0, α_1). The same reasoning holds for an all-or-nothing DPA and its distinguisher $\mathsf{AON\text{-}DPA}(\hat{k})$ defined in (3), by stating $\alpha_0 = 1$, $\alpha_1 = -1$ and $\alpha_i = 0$ for every $\omega_i \in \mathrm{Im}(\mathsf{m}) \backslash \{\omega_0, \omega_1\}$.

Let us now focus on the generalized DPA and its distinguisher $\mathsf{G\text{-}DPA}(\hat{k})$. It can be easily checked that (4) can be rewritten:

$$\begin{aligned}\mathsf{G\text{-}DPA}(\hat{k}) = &\sum_{\omega \in \Omega_0} \frac{\widehat{\mathrm{P}}_{\mathrm{r}}\left(M_{\hat{k}} = \omega\right)}{\widehat{\mathrm{P}}_{\mathrm{r}}\left(M_{\hat{k}} \in \Omega_0\right)} \widehat{\mathbb{E}}\left(L \mid M_{\hat{k}} = \omega\right) \\ &- \sum_{\omega \in \Omega_1} \frac{\widehat{\mathrm{P}}_{\mathrm{r}}\left(M_{\hat{k}} = \omega\right)}{\widehat{\mathrm{P}}_{\mathrm{r}}\left(M_{\hat{k}} \in \Omega_1\right)} \widehat{\mathbb{E}}\left(L \mid M_{\hat{k}} = \omega\right) \\ &+ \sum_{\omega \in \mathrm{Im}(\mathsf{m}) \backslash \Omega_0 \cup \Omega_1} 0 \cdot \widehat{\mathbb{E}}\left(L \mid M_{\hat{k}} = \omega\right) \ . \end{aligned} \tag{8}$$

Let us denote by $(\omega_i)_i$ the elements in $\mathrm{Im}(\mathsf{m})$ and let $(\alpha_i)_i$ be a family coefficients defined such that:

$$\alpha_i = \begin{cases} \frac{\widehat{\mathrm{P}}_{\mathrm{r}}\left(M_{\hat{k}} = \omega_i\right)}{\widehat{\mathrm{P}}_{\mathrm{r}}\left(M_{\hat{k}} \in \Omega_0\right)} & \text{if } \omega_i \in \Omega_0, \\ -\frac{\widehat{\mathrm{P}}_{\mathrm{r}}\left(M_{\hat{k}} = \omega_i\right)}{\widehat{\mathrm{P}}_{\mathrm{r}}\left(M_{\hat{k}} \in \Omega_1\right)} & \text{if } \omega_i \in \Omega_1, \\ 0 & \text{otherwise.} \end{cases}$$

Under Assumption 2, coefficients α_i are constant (namely independent of the sample size and of the key hypothesis). After replacing the coefficients in (8) by those α_i's, we recognize in (8) the definition of a PPA distinguisher involving the same model m as $\mathsf{G\text{-}DPA}(\hat{k})$ and the family $(\alpha_i)_i$ as coefficients. ◇

As a direct consequence of Proposition 1, we get the following corollary:

Corollary 1 *Under Assumption 2, a DPA is SCA-reducible to a PPA.*

In the next section, we compare the PPA with the CPA.

3.3 From PPA to CPA

It is already well known in statistics that a linear correlation coefficient can be written as a weighted sum of means over a partition of a probability space. As a straightforward consequence and as mentioned by Le *et al.* in [4], a CPA can be viewed as a particular case of a PPA (*i.e.* , a CPA is SCA-reducible to a PPA). What we prove in this section is that a PPA can be re-stated as a CPA. Eventually, we argue that both attacks are SCA-equivalent under Assumption 2.

Proposition 2 *Let* $\mathsf{PPA}_{(\alpha_i)_i}(\hat{k})$ *be a PPA distinguisher defined with respect to a model function* m *and a family of coefficients* $(\alpha_i)_i$. *Then, there exists a function* f *and two constant coefficients a and b such that* $\mathsf{PPA}_{(\alpha_i)_i}(\hat{k}) = a \cdot \mathsf{CPA}(\hat{k}) + b$, *where* $\mathsf{CPA}(\hat{k})$ *is a CPA distinguisher involving the model function* $\mathsf{f} \circ \mathsf{m}$.

Proof. We recall that, in the definition of $\mathsf{PPA}_{(\alpha_i)_i}(\hat{k})$ (see (5)), every $\omega_i \in \mathrm{Im}(\mathsf{m})$ is associated with the coefficient α_i. From those ω_i's and α_i's we define a function f on $\mathrm{Im}(\mathsf{m})$ by:

$$\mathsf{f}(\omega_i) = \frac{\alpha_i}{\widehat{\mathrm{P}}_{\mathrm{r}}\left(M_{\hat{k}} = \omega_i\right)} \ . \tag{9}$$

Under Assumption 2, probabilities $\widehat{\mathrm{P}}_{\mathrm{r}}\left(M_{\hat{k}} = \omega_i\right)$, and thus coefficients $\mathsf{f}(\omega_i)$, are constant (namely independent of the sample size and of the key hypothesis $\hat{k}$). With those new notations, (5) can be rewritten as:

$$\begin{aligned}&\mathsf{PPA}_{(\alpha_i)_i,\mathsf{m}}(\hat{k}) = \\ &\qquad \sum_{\omega_i \in \mathrm{Im}(\mathsf{m})} \mathsf{f}(\omega_i) \cdot \widehat{\mathrm{P}}_{\mathrm{r}}\left(M_{\hat{k}} = \omega_i\right) \cdot \widehat{\mathbb{E}}\left(L \mid M_{\hat{k}} = \omega_i\right) \ . \end{aligned} \tag{10}$$

We therefore get the following relation:

$$\begin{aligned}&\mathsf{PPA}_{(\alpha_i)_i,\mathsf{m}}(\hat{k}) = \\ &\qquad \sum_{\alpha \in \mathrm{Im}(\mathsf{f})} \alpha \cdot \widehat{\mathrm{P}}_{\mathrm{r}}\left(M_{\hat{k}} \in \mathsf{f}^{-1}(\alpha)\right) \cdot \widehat{\mathbb{E}}\left(L \mid M_{\hat{k}} \in \mathsf{f}^{-1}(\alpha)\right) \end{aligned} \tag{11}$$

i.e.

$$\begin{aligned}&\mathsf{PPA}_{(\alpha_i)_i,\mathsf{m}}(\hat{k}) = \\ &\qquad \sum_{\alpha \in \mathrm{Im}(\mathsf{f})} \widehat{\mathrm{P}}_{\mathrm{r}}\left(\mathsf{f}(M_{\hat{k}}) = \alpha\right) \cdot \widehat{\mathbb{E}}\left(\alpha \cdot L \mid \mathsf{f}(M_{\hat{k}}) = \alpha\right) \ . \end{aligned} \tag{12}$$

After denoting by $M'_{\hat{k}}$ the random variable $\mathsf{f}(M_{\hat{k}})$ and thanks to the law of total expectation, we eventually deduce:

$$\mathsf{PPA}_{(\alpha_i)_i,\mathsf{m}}(\hat{k}) = \widehat{\mathbb{E}}\left(LM'_{\hat{k}}\right) \ . \tag{13}$$

On the other hand, we have:

$$\mathsf{CPA}_{\mathsf{m}'}(\hat{k}) = \frac{1}{\widehat{\sigma}(L)\,\widehat{\sigma}\left(M'_{\hat{k}}\right)} \cdot \widehat{\mathbb{E}}\left(LM'_{\hat{k}}\right) - \frac{\widehat{\mathbb{E}}(L)\widehat{\mathbb{E}}\left(M'_{\hat{k}}\right)}{\widehat{\sigma}(L)\,\widehat{\sigma}\left(M'_{\hat{k}}\right)} \ ,$$

where m' denote the function $\mathsf{f}\circ\mathsf{m}$. Under Assumption 2, values $\widehat{\mathbb{E}}(L)$, $\widehat{\sigma}(L)$, $\widehat{\mathbb{E}}\left(M_{\hat{k}}\right)$ and $\widehat{\sigma}\left(M_{\hat{k}}\right)$ are constant with respect to $\hat{k}$. This implies that the CPA distinguisher $\mathsf{CPA}(\hat{k})$ associated with the model function $\mathsf{f}\circ\mathsf{m}$ satisfies the following equality:

$$\widehat{\mathbb{E}}\left(LM'_{\hat{k}}\right) = a\cdot\mathsf{CPA}_{\mathsf{m}'}(\hat{k}) + b \ , \tag{14}$$

where a and b are two constant values satisfying

$$a = \widehat{\sigma}(L)\,\widehat{\sigma}\left(M'_{\hat{k}}\right) \ \text{ and } \ b = \frac{\widehat{\mathbb{E}}(L)\widehat{\mathbb{E}}\left(M'_{\hat{k}}\right)}{\widehat{\sigma}(L)\,\widehat{\sigma}\left(M'_{\hat{k}}\right)} \ .$$

From (13) and (14) we deduce that there exist two constant terms a and b and a model transformation f such that

$$\mathsf{PPA}_{(\alpha_i)_i,\mathsf{m}}(\hat{k}) = a\cdot\mathsf{CPA}_{\mathsf{m}'}(\hat{k}) + b \ , \tag{15}$$

with $\mathsf{m}' = \mathsf{f}\circ\mathsf{m}$. ◇

As a straightforward consequence of Proposition 2 we get the following corollary:

Corollary 2 *Under Assumption 2, a PPA is SCA-equivalent to a CPA.*

Proposition 2 implies that a PPA and a CPA only differ in the model, which is involved to correlate the leakage signal. As a consequence, if a PPA with model m and coefficients α_i's is more efficient than a CPA with model m', this simply means that the model $\mathsf{f}\circ\mathsf{m}$ (for f defined as in Prop. 2) is more linearly related to the deterministic leakage function $\delta(\cdot)$ than m'. In such a case, the CPA must be performed with the most accurate model between both, namely $\mathsf{f}\circ\mathsf{m}$. In other terms, the problem of finding the most pertinent coefficients α_i's for the PPA is equivalent to the problem of finding the model with maximum linear correlation with the deterministic leakage function.

3.4 Summing Distinguishers

In previous sections, we have established the SCA-reduction of DPA and PPA to CPA. Namely, we have shown that for every DPA or PPA with model m, there exists a new model $\mathsf{m}' = \mathsf{f}\circ\mathsf{m}$ such that a CPA with m' leads to a similar key-guess classification. This shows that, when performing such attacks, the real issue is the choice of the model and not the choice of the distinguisher. To deal with this issue when the best model is not known, an approach could consist in applying one of the distinguishers recalled in previous sections to a family of models $(\mathsf{m}_i)_i$ and to sum the results to define a new distinguisher. Actually, this distinguisher is still affinely reducible to a CPA-distinguisher involving a model defined with respect to $(\mathsf{m}_i)_i$ and the "new" attack is thus no more than a CPA attack with a new model. This comes down as a consequence of the following lemma:

Lemma 1 *Let $\mathsf{CPA}_{\mathsf{m}_1}(\hat{k})$ and $\mathsf{CPA}_{\mathsf{m}_2}(\hat{k})$ be two CPA distinguishing values defined for the same samples $(\ell_{k,i})_i$ and $(v_{\hat{k},i})_i$, and with two different model functions m_1 and m_2 respectively. Then, after denoted by m_3 the function $\frac{\mathsf{m}_1}{\widehat{\sigma}\left(M_{\hat{k},1}\right)} + \frac{\mathsf{m}_2}{\widehat{\sigma}\left(M_{\hat{k},2}\right)}$, we have:*

$$\mathsf{CPA}_{\mathsf{m}_1}(\hat{k}) + \mathsf{CPA}_{\mathsf{m}_2}(\hat{k}) = a\ \mathsf{CPA}_{\mathsf{m}_3}(\hat{k}) \ ,$$

where $M_{\hat{k},1}$, $M_{\hat{k},2}$ and $M_{\hat{k},3}$ denote the model variables associated with the model functions m_1, m_2 and m_3 respectively, and where $a = \widehat{\sigma}\left(M_{\hat{k},3}\right)$.

The idea consisting in summing several distinguishers to define a new one has been for instance applied by Bévan and Knudsen in [5] to enhance the original Kocher's DPA. The authors propose to perform a single-bit DPA for each bit of the sensitive variable $V_{\hat{k}}$ and then to sum the results. We call this attack a *Multiple-DPA* attack hereafter and we denote the involved distinguisher by $\mathsf{M\text{-}DPA}(\hat{k})$. It is defined as follows:

$$\mathsf{M\text{-}DPA}(\hat{k}) = \sum_{j=0}^{t} \mathsf{SB\text{-}DPA}(\hat{k})_j \ . \tag{16}$$

where t is any integer lower than or equal to the dimension of v_k viewed as a binary-vector and where $\mathsf{SB\text{-}DPA}(\hat{k})_j$ denotes the single-bit DPA with a model function m_j defined w.r.t. two real values $\omega_{0,j}$ and $\mathrm{c}\omega_{1,j}$ by $\mathsf{m}_j(v_{\hat{k}}) = (1-v_{\hat{k}}[j])\cdot\omega_{0,j} + v_{\hat{k}}[j]\cdot\omega_{1,j}$. As argued at the beginning of this section (and as a consequence of Propositions 1 and 2 and Lemma 1), this attack is SCA-reducible to a CPA. We state this in the following proposition and, for completeness, we exhibit in its proof the way how to define the CPA-distinguisher of this reduced CPA.

Proposition 3 *A M-DPA attack is SCA-reducible to a CPA under Assumption 2.*

Proof. Let us focus on Relation (16). Due to Proposition 1, for every j the single-bit DPA distinguisher $\mathsf{SB\text{-}DPA}(\hat{k})_j$ is affinely reducible to the CPA-distinguisher $\mathsf{CPA}(\hat{k})_j$ involving the model function $\mathsf{f}_j \circ \mathsf{m}_j$ where f_j is defined on $\mathrm{Im}(\mathsf{m}_j) = \{\omega_{0,j}, \omega_{1,j}\}$ by $\mathsf{f}_j(\omega_{0,j}) = 1/\widehat{\mathrm{P}}_{\mathrm{r}}\left(\mathsf{m}_j(V_{\hat{k}}) = \omega_{0,j}\right)$ and $\mathsf{f}_j(\omega_{1,j}) = -1/\widehat{\mathrm{P}}_{\mathrm{r}}\left(\mathsf{m}_j(V_{\hat{k}}) = \omega_{1,j}\right)$. Let $M_{\hat{k},j}$ denote the random variable $\mathsf{f}_j \circ \mathsf{m}_j(V_{\hat{k}})$. As a consequence of Proposition 1, we have:

$$\mathsf{SB\text{-}DPA}(\hat{k})_j = \frac{\mathsf{CPA}(\hat{k})_{\mathsf{m}_j} + b}{a} \ ,$$

with $a = \frac{1}{\widehat{\sigma}(L)\widehat{\sigma}\left(M_{\hat{k},j}\right)}$ and $b = \frac{\widehat{\mathbb{E}}(L)\widehat{\mathbb{E}}\left(M_{\hat{k},j}\right)}{\widehat{\sigma}(L)\widehat{\sigma}\left(M_{\hat{k},j}\right)}$. It can be checked that under Assumption 2, a and b are constant with respect to j and $\hat{k}$. We therefore deduce that (16) is equivalent with:

$$\mathsf{M\text{-}DPA}(\hat{k}) = \frac{1}{a}\sum_{j=0}^{t} \mathsf{CPA}(\hat{k})_{\mathsf{m}_j} + \frac{t \cdot b}{a} \ .$$

Lemma 1 then implies the following equality:

$$\mathsf{M\text{-}DPA}(\hat{k}) = \frac{\widehat{\sigma}\left(M^{\star}_{\hat{k}}\right)}{a}\mathsf{CPA}(\hat{k})_{\mathsf{m}^{\star}} + \frac{t \cdot b}{a} \ , \tag{17}$$

with $\mathsf{m}^{\star}$ being the function $\sum_{j=1}^{t} \frac{\mathsf{f} \circ \mathsf{m}_j}{\widehat{\sigma}\left(M_{\hat{k},j}\right)}$ and where $M^{\star}_{\hat{k}}$ denotes the model variable associated with $\mathsf{m}^{\star}$. ◇

3.5 On the Choice of the Model

In previous sections we argued that most of existing linear power analysis attacks are reducible to CPAs that only differ in the model they involve. As a first important consequence, one of those attacks is more efficient than another if and only if the corresponding SCA-reduced CPA involves a better model. This naturally raises the question of defining the model that optimizes the CPA efficiency. It has been proven in [13] that the model function $\mathsf{m} : v \mapsto \mathbb{E}\left(L \mid V_{\hat{k}} = v\right)$ maximizes the amplitude of the correlation coefficient (6) when the good key is tested and hence optimizes the attack efficiency (as argued in [14]). In the context of univariate SCA with leakage satisfying (1), this function is the deterministic leakage function $\delta(\cdot)$. Note that any model $\mathsf{m}(\cdot) = a\ \delta(\cdot) + b$ where $a \neq 0$, b are constants will also maximize the amplitude of the correlation. As a particular observation, when all the bits of the targeted variable v_k impact the leakage expectation, the result in [13] implies that the model must take into account all the bits of v_k and that attacks exploiting only a limited number of bits (such as *e.g.* , the single-bit DPA) are sub-optimal. It is worth noticing that if the model is perfect (*i.e.* , if $\mathsf{m}(\cdot) = \delta(\cdot)$), then under the *Gaussian Noise Assumption* (*i.e.* , the noise B in (1) is drawn from a gaussian distribution), the CPA is equivalent to a maximum likelihood attack [6], which is known to be optimal for key-recovery. However, computing $\mathsf{m} : v \mapsto \mathbb{E}\left(L \mid V_{\hat{k}} = v\right)$ with no *a priori* knowledge about L is not possible when no profiling stage is enabled. This implies that the adversary model is often not perfect and the resulting attacks are thus most of the time sub-optimal. In the next section, we investigate a family of side channel attacks that make weaker assumptions on the device behavior than the CPA-like attacks do. To succeed, those attacks, termed *robust*, do not require a good affine estimation of the deterministic part $\delta(\cdot)$ of the device leakage. Actually, they only require some general assumptions on the algebraic properties of $\delta(\cdot)$ (*e.g.* , the output value of the function is any linear combination of the bits of the input value).

4 Robust Side Channel Attacks

In this section, we investigate robust side channel attacks that are able to succeed with only a very limited knowledge on how the device leaks information. The starting point is to replace the requirement that the deterministic part of the leakage $\delta(\cdot)$ is greatly correlated to the attack model m by the weaker requirement that $\delta(\cdot)$ belongs to a set of functions sharing some algebraic properties.

Before presenting the attacks and in order to determine the kind of algebraic properties of $\delta(\cdot)$ they focus on, let us have a closer look at this function. As any real function defined over $\mathbb{F}_{2^n}$, it can be represented by a polynomial in $\mathbb{R}[x_0, \ldots, x_{n-1}]$, where the degree of every x_i in every monomial is at most 1 (because ${x_i}^m = x_i$ for every $x_i \in \mathbb{F}_2$ and $m \in \mathbb{N}^*$). Namely, there exists a *multivariate degree* (or a *degree* for short) $d \leq n$ and a set of real coefficients $(\alpha_u)_{u \subseteq \{0,\ldots,n-1\}}$ such that for every $x \in \mathbb{F}_{2^n}$ we have:

$$\delta(x) = \alpha_{-1} + \sum_{i=0}^{n-1} \alpha_i x_i + \sum_{i_1,i_2=0}^{n-1} \alpha_{i_1,i_2} x_{i_1} x_{i_2} + \cdots + \sum_{i_1,\ldots,i_d=0}^{n-1} \alpha_{i_1,\ldots,i_d} x_{i_1} x_{i_2} \cdots x_{i_d} \ . \tag{18}$$

In view of (18), a side channel adversary could use his knowledge of the device technology to make an assumption on the degree d of $\delta(\cdot)$ viewed as a polynomial with coefficients in $\mathbb{R}$. This amounts to make the following assumption on the device.

Assumption 3 (Leakage Interpolation Degree) *The multivariate degree of the deterministic part $\delta(\cdot)$ of the leakage is upper bounded by d, for some d lower than or equal to n.*

In practice and for most of devices such as smart cards, the coefficients α_{-1}, α_0, …, α_{n-1} are significantly greater than the others. This implies that the value of $\delta(x)$ is very

close to the value of the linear part in (18), the other non-linear terms playing a minor role [15]. In this case, it makes sense for the adversary to make Assumption 3 for $d = 1$. It is sometimes referred as the *Independent Bit Leakage* (IBL) Hypothesis in the literature [16] since it amounts to assume that the leakages related to the manipulation of two different bit-coordinates of V_k are independent. This assumption fits well with the physical reality of numerous electronic devices. Indeed, the power consumption and electromagnetic emissions both result from logical transitions occurring on the circuit wires. Thus assume that every bit of a processed variable contributes independently to the overall instantaneous leakage is therefore realistic.

From an attacker point of view, assuming the IBL hypothesis is often a good strategy in practice since it enables to define an attack which, without being optimal, has an adequate efficiency. However, from the security designer perspective the IBL hypothesis may be considered as too restrictive. In this case indeed, the security analysis must include the largest class of adversaries as possible and proving resistance under the IBL hypothesis is therefore no longer sufficient. This is all the more true that for some new devices (*e.g.* , based on architectures using 65 nm manufacturing technology), it has been observed [16–18] that the coefficients of the quadratic terms in (18) are not negligible compared to those of the linear terms: the leakages related to the manipulation of two different bit-coordinates of V_k are no longer independent. In this case, Assumption 3 for $d \geqslant 2$ shall yield a better representation of the reality.

To sum up our discussion, even if making the Assumption 3 for $d = 1$ may be sufficient for an attacker to perform a succesfull attack, one (typically a device designer) must choose d as large as possible if the purpose is to test a device resistance in the worst case scenario.

In the next two sections we present two side channel attack that are able to successfully recover the expected k with no other assumption on the deterministic part of the leakage than Assumption 3 for some limited value of d. In particular, their efficiency does not rest on the adversary ability to find a model m which is a good affine approximation of $\delta(\cdot)$ as it was the case for CPA-like attacks. The two attacks are described in the particular case where Assumption 3 is done for $d = 1$. This situation is indeed sufficient for most of practical attack contexts and it has the advantage to allow for a simple description of the attacks outlines. Eventually, in Section 4.3 we briefly explain how they can be simply extended to deal with Assumption 3 for $d > 1$ (*i.e.* , when neglecting the terms of degree greater than 1 leads to attack failure).

4.1 Absolute Sum DPA

It may first be noted that the multi-bit DPA recalled in Section 3.4 is not a "robust" extension of the binary single-bit DPA. Indeed, if we take a closer look at (16), we can check that the sign of each single-bit DPA distinguisher in the sum depends on the choice of the values $\omega_{0,j}$ and $\omega_{1,j}$. Hence, depending on the models m_j's chosen for the attack, the sum of the values returned by the single-bit DPA distinguishers when the good key is tested may be very close to zero, which may result in a wrong-key discrimination. As already pointed out in [19], a straightforward solution to circumvent this issue consists in replacing the sum in (16) by a sum of absolute values – or a sum of squares – of single-bit DPA distinguishers. This leads us to define the following AS-DPA distinguisher:

$$\mathsf{AS\text{-}DPA}(\hat{k}) = \sum_{i=0}^{t} \left|\mathsf{DPA}(\hat{k})_i\right| \quad . \tag{19}$$

Contrary to what happens for $\mathsf{M\text{-}DPA}(\hat{k})$, the value of each element in the sum in $\mathsf{AS\text{-}DPA}(\hat{k})$ stays unchanged if we replace a family of bijective model functions $(\mathsf{m}_j)_j$ by another one. We can therefore chose any m which shows that our new AS-DPA attack is "robust". In Appendix A, we give an example illustrating the differences between our new attack and the M-DPA or the CPA.

4.2 Linear Regression

In [7], Schindler *et al.* describe an efficient profiling method for SCA. Assuming that the attacker knows the subkey k, they explain how to recover the leakage function δ (*i.e.* , the α_j coefficients under the IBL assumption) using linear regression. As mentioned by the authors, their approach could also allow for the recovering of k (but no details nor experiments are provided). We develop hereafter the ideas introduced in [7] to get a robust SCA. Let $(v_k[n-1], \ldots, v_k[0])$ be the binary decomposition of the variable v_k targeted by the attack and let $(\ell_{k,i})_i$ and $(v_{\hat{k},i})_i$ be respectively a family of N leakage measurements and the corresponding hypotheses on the leakage deterministic part. The core idea is to compute, for each key candidate $\hat{k}$, a set of coefficients $\hat{\alpha}_{-1}$, $\hat{\alpha}_0$, ..., $\hat{\alpha}_{n-1}$ such that the families $(\ell_{k,i})_i$ and $(\hat{\alpha}_{-1} + \sum_{j=0}^{n} \hat{\alpha}_j v_{\hat{k},i}[j])_i$ are as close as possible for a well-chosen *distance*. Under Assumption 2, this process should result in a minimal distance when the good key candidate $\hat{k} = k$ is tested. As pointed out in [7], the *Euclidean distance* (or equivalently *the least-square distance*) is a sound distance choice and it is actually optimal when the noise in (1) has a Gaussian distribution [20]. Moreover, in this case the coefficients $\hat{\alpha}_j$ can be efficiently computed by performing a *linear regression*.

Let $\mathbf{L}$ be the $N \times 1$ matrix $(\ell_{k,1}, \ell_{k,2}, \ldots, \ell_{k,N})$ composed of the N leakage measurements. To proceed the linear regression for a key candidate $\hat{k}$, the following $N \times (n+1)$ matrix is first constructed:

$$\mathbf{M} = \begin{pmatrix} 1 & v_{\hat{k},1}[0] & \cdots & v_{\hat{k},1}[n-1] \\ 1 & v_{\hat{k},2}[0] & \cdots & v_{\hat{k},2}[n-1] \\ \vdots & \vdots & \ddots & \vdots \\ 1 & v_{\hat{k},i}[0] & \cdots & v_{\hat{k},i}[n-1] \\ \vdots & \vdots & \ddots & \vdots \\ 1 & v_{\hat{k},N}[0] & \cdots & v_{\hat{k},N}[n-1] \end{pmatrix} .$$

Notation *In the linear regression terminology, the Boolean coordinate functions $v_{\hat{k}}[j] : i \mapsto v_{\hat{k},i}[j]$ (j being the coordinate index) play the role of* basis functions*.*

In a second time, the *ordinary least square method* is applied, resulting in the construction of the coefficients $\hat{\alpha}_j$ of the column vector $\alpha_{\hat{k}}$ defined such that:

$$\alpha_{\hat{k}} = {}^t(\hat{\alpha}_{-1}, \hat{\alpha}_0, \ldots, \hat{\alpha}_{n-1}) = \left({}^t\mathbf{M} \cdot \mathbf{M}\right)^{-1} \cdot {}^t\mathbf{M} \cdot \mathbf{L} \ .$$

Eventually, the Euclidean distance, denoted by $\|\cdot\|^2$, between the hypotheses $\mathbf{M} \cdot \alpha_{\hat{k}}$ and the leakage vector $\mathbf{L}$ is computed. This results in the construction of a distinguishing value $\Delta_{\hat{k}}$ defined such that:

$$\Delta_{\hat{k}} = \|\mathbf{L} - \mathbf{M} \cdot \alpha_{\hat{k}}\|^2 \ .$$

By definition, the linear regression outputs the vector $\alpha_{\hat{k}}$ that optimally minimizes $\|\mathbf{L} - \mathbf{M} \cdot \alpha_{\hat{k}}\|^2$ according to the chosen basis functions.

Under Assumption 3 with $d = 1$, the distinguishing value $\Delta_{\hat{k}}$ is expected to be minimal for the good hypothesis $\hat{k} = k$. Contrary to the attacks analyzed in Section 3, which involve a fixed model function, regression attacks output a different model function $\mathsf{m}_{\hat{k}} : (v_{\hat{k}}[n-1], \ldots, v_{\hat{k}}[0]) \mapsto (v_{\hat{k}}[n-1], \ldots, v_{\hat{k}}[0]) \cdot \alpha_{\hat{k}}$ for each key candidate $\hat{k}$. For the key discrimination step, an Euclidean distance is processed in place of a correlation coefficient with the leakage sample.

Remark 3 In the literature, *goodness of fit* is the common way to describe how well a model fits a set of observations. Different measures of goodness of fit can be used depending on the context. The *coefficient of determination* or the *Akaike information criterion* are examples of such a measure. In this paper, we privileged the following coefficient of determination:

$$R^2(\hat{k}) = \frac{\|\mathbf{L} - \mathbf{M} \cdot \alpha_{\hat{k}}\|^2}{\mathsf{var}(L)} = \frac{\mathbb{E}\left(\left(\mathbf{L} - \mathbf{M} \cdot \alpha_{\hat{k}}\right)^2\right)}{\mathsf{var}(L)} \ . \tag{20}$$

It first permits to have a value in the range $[0, 1]$. Moreover, it is closely related to the correlation coefficient. Note that in your specific case, all models result from a linear regression with the same basis functions set and with the same observations. This implies that in this particular case the main known estimators are equivalent to the Euclidian distance estimator.

4.3 Extension of the Attacks to Non-linear Contexts

The choice of the coordinate functions $v_{\hat{k}}[j]$ as a basis for the linear regression is due to Assumption 3 assuming $d = 1$. If we relax our assumption and assume that the leakage also depends on some monomials $v_k[j_1]v_k[j_2]\cdots v_k[j_r]$, with $d \geqslant r \geqslant 2$, then the corresponding hypothesis-related monomials $v_{\hat{k}}[j_1]v_{\hat{k}}[j_2]\cdots v_{\hat{k}}[j_r]$ can be added to the initial basis $(v_{\hat{k}}[j])_j$. In this case, the regression detailed in previous section can be straightforwardly adapted to apply on the new (extended) basis. The new regression is still a linear one, but with a polynomial (and not simply linear) basis.

In the same manner, the outlines of the generalization process can be extended to the AS-DPA by using cross-products of SB-DPA. Namely, if we assume that the leakage depends on some monomials $v_k[j_1]v_k[j_2]\cdots v_k[j_r]$, with $d \geqslant r \geqslant 2$, then the corresponding SB-DPA cross-product $|\mathsf{DPA}(\hat{k})_{j_1} \times \mathsf{DPA}(\hat{k})_{j_2} \times \cdots \times \mathsf{DPA}(\hat{k})_{j_r}|$ can be added to the initial AS-DPA .

5 Attack Simulations and Experiments

In previous sections, we have shown that common univariate SCAs based on a restrictive model are equivalent to a CPA. At the opposite, we have exhibited two pertinent ways of attacking where some constraints on the model can be relaxed. It involves as a distinguisher either $\mathsf{AS\text{-}DPA}(\hat{k})$ or linear regression techniques. In the following we aim to confront our theoretical analyses with simulations in realistic scenarios. Simulation parameters are described below.

Attacks Target. The 8-bit output of the AES s-box, denoted by S, is targeted. Namely the variable V_k in (1) satisfies:

$$V_k = S(P \oplus k) \ , \tag{21}$$

where P corresponds to an 8-bit value known by the adversary.

Attack Types. We list below the attacks we have performed:

1. Single-bit DPA (SB-DPA)
2. All-Or-Nothing DPA (AON-DPA)
3. Generalized DPA (G-DPA)
4. Correlation Power Analysis (CPA)
5. Partition Power Analysis (PPA)
6. Absolute-Sum DPA (AS-DPA)
7. Regression Attack with $(v_{\hat{k}}[i])_{0 \leqslant i \leqslant 7}$ as basis functions (this corresponds to Assumption 3 with $d = 1$).

Model Choice. We recall that AON-DPA, G-DPA, CPA and PPA require the choice of a model function m, whereas SB-DPA, AS-DPA and the regression attack do not. In our simulation, we have assumed that the definition of the function $\delta(\cdot)$ in (1) is not known by the adversary and we thus systematically used the Hamming weight function when a model was required to perform the attack. Namely, in AON-DPA, G-DPA, CPA and PPA the model m satisfies:

$$\mathsf{m}(V_{\hat{k}}) = \mathrm{HW}(V_{\hat{k}}) = \sum_i V_{\hat{k}}[i] \quad . \tag{22}$$

This model choice is very classical and has been experimentally validated in several papers *e.g.* , [15]. Once the model function has been specified, parameters (ω_0, ω_1) in AON-DPA and (Ω_0, Ω_1) in G-DPA still need to be chosen in order to determine the distinguishers defined in (3) and (4) respectively. We chose

$$(\omega_0, \omega_1) = (\min_{V_{\hat{k}}} \mathsf{m}(V_{\hat{k}}), \max_{V_{\hat{k}}} \mathsf{m}(V_{\hat{k}})) = (0, 8)$$

and if we denote by $\mathrm{med}_X f(X)$ the *median* of the sample $f(X)$ with respect to X, we chose

$$\begin{aligned}(\Omega_0, \Omega_1) = \\ ([\min_{V_{\hat{k}}} \mathsf{m}(V_{\hat{k}}); \operatorname*{med}_{V_{\hat{k}}} \mathsf{m}(V_{\hat{k}})[,] \operatorname*{med}_{V_{\hat{k}}} \mathsf{m}(V_{\hat{k}}); \max_{V_{\hat{k}}} \mathsf{m}(V_{\hat{k}})]) = \\ ([0;4[,]4;8]) \quad .\end{aligned} \tag{23}$$

Note that this choice is optimal and exactly corresponds to the attacks performed by Messerges in his original papers [2, 8]. Additionally, we chose the coefficients α_i of the PPA distinguisher such that (13) is satisfied for the model function m defined in (22) (*i.e.* , $\mathsf{PPA}_{(\alpha_i)_i}(\hat{k}) = \widehat{\mathbb{E}}\left(L \cdot \mathrm{HW}(V_{\hat{k}})\right)$).

Leakage Simulations. Leakages have been simulated in accordance with (1), with the noise variable B being a Gaussian random variable with mean 0 and standard deviation σ. As explained in the following sections, we launched our attack simulations for different definitions of the function $\delta(\cdot)$ in (1), leading to two different scenarios:

- *Scenario* 1*:* we chose $\delta(\cdot)$ in (1) to be the Hamming weight function. Namely, the leakage variable L satisfies:

  $$L = \mathrm{HW}(V_k) + B \quad , \tag{24}$$

  In our attack settings, this first scenario is ideally suited for AON-DPA, G-DPA, CPA and PPA since the model function m used by the adversary exactly corresponds to the deterministic function $\delta(\cdot)$. It will be referred as the *perfect model* scenario.

- *Scenario* 2*:* we chose $\delta(\cdot)$ to be a linear combination of the $V_{\hat{k}}[i]$'s with randomly generated coefficients. Namely the leakage variable L satisfies:

  $$L = \alpha_{-1} + \sum_{i=0}^{7} \alpha_i \cdot V_k[i] + B \quad , \tag{25}$$

  with coefficients $(\alpha_i)_{-1 \leqslant i \leqslant 7}$ uniformly picked in $[-1, 1]$. This scenario is used to observe the distinguishers behavior when the deterministic part of the leakage differs from the model used by the adversary. We restricted ourselves to functions $\delta(\cdot)$ that are linear combinations in $\mathbb{R}$ of the bit-coordinates of the targeted value $V_{\hat{k}}$ *i.e.* as in Assumption 3 with $d = 1$. It will be referred as the *random linear leakage* scenario.

Remark 4 We do not restrict ourselves to Assumption 2. Namely we do not ensure that the size of the plaintext sample is a multiple of 256. Nevertheless plaintexts are drawn from a uniform distribution.

Attack Efficiency. In the following, an attack is said to be *successful* if the good key is output by the attack, that is if the key corresponding to the first element in the score vector is the key used in the simulated cryptographic device. An attack is said to be *more efficient than* another if it needs less messages to achieve the same success rate. Success rate is measured over $1,000$ tries.

We report and analyze in next two sections our attack simulations results for Scenario 1 (Section 5.1) and Scenario 2 (Section 5.2).

5.1 Attack Results in the Perfect Model Scenario

In this section we assume that L satisfies (24). In Fig. 1, the number of messages needed to achieve a success rate of 90% is recorded for each attack mentionned before[2]. Note that a success rate threshold has been fixed at 90% but in this configuration each attack can reach 100%.

Curves in Fig. 1 can be split in two parts depending on the noise standard deviation: the *oversampling* part, where a huge number of observations are needed to deal with the important noise effects and the *undersampling* part, where a small number of observations is sufficient. The two situations are analyzed separately in the following. In both cases, the most relevant observations are listed and discussed.

[2] We inform the reader that the curves are plotted fitted with a fourth degree polynomial to ease the reading of the figure. Fitted curves permit to observe the general behavior. Raw data can be found in Appendix C.

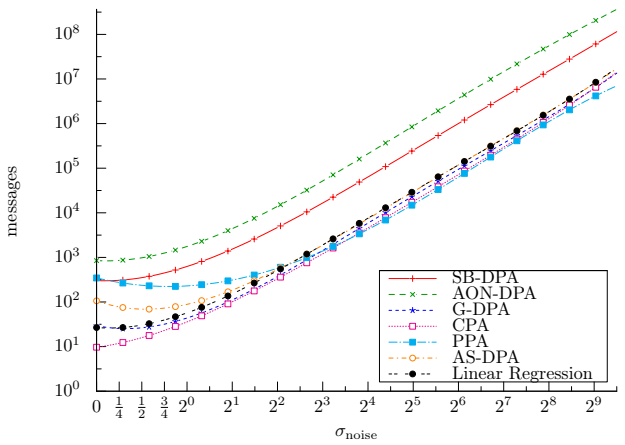


Fig. 1: Evolution of the number of messages (y-axis logscaled) to achieve a success rate of 90% according to the noise standard deviation (x-axis logscaled) – Fitted curves.

Oversampling. When the noise standard deviation is strictly greater than 2^3, each distinguisher needs a large number of messages (greater than 500) to reach a success rate of 90%. In this case the curves have the same shape for each distinguisher, which is compliant with the asymptotical results in [6]. Our observations are detailled below:

- The efficiency curves of each attack have the same gradient. This suggests us that the noise similarly impacts the efficiency of the attacks.
- The curves corresponding to G-DPA, CPA, PPA, AS-DPA and the regression attack are stacked. Note that with the logscaling that implies that those attacks share *approximatively* the same efficiency and that none of them is emerging as better candidate than the others. In fact, in the perfect model scenario, the distinguishers corresponding to these attacks are equivalent to a maximum likelihood test and the attacks therefore perform in a similar (and optimal) way [6]. This pinpoints the equivalence between the distinguishers when the model function used in the model-based attacks (*i.e.* , AON-DPA, G-DPA, CPA and PPA) is optimal (*i.e.* , perfectly corresponds to the function $\delta(\cdot)$ in 1).
- As expected, SB-DPA and AON-DPA are less powerful than the others (around 100 and 30 times less efficient than G-DPA, CPA, PPA, AS-DPA and the regression attack for the SB-DPA and the AON-DPA respectively). Indeed, by nature they do not exploit all the information contained in the leakage signal: in SB-DPA only one output bit is targeted over the 8 output bits of the AES, whereas the AON-DPA only exploits a limited part of the leakage measurements.

Remark 5 The good result of G-DPA can be surprizing as the involved model is not the Hamming weight model. The G-DPA model only takes two values -1 and 1 depending on the Hamming weight of the sensitive variable is lower than 4 or not. In fact the linear correlation between the G-DPA model and the Hamming weight model is high (greater than 0.9). That implies an efficiency ratio of 1.2 (0.08 in a $\log_{10}$ scale) according to [21]. This explains why G-DPA's curve appears stacked with CPA's curve.

Undersampling. When the noise standard deviation is lower than 2^3, the number of messages needed to perform an attack is quite small (lower than 500). In this case, the statistical stability of the involved distinguisher plays a role. To better understand how the different attacks perform in this context we redrew in Fig. 2 the curves with a thiner resolution than in Fig. 1. We detail our observations below:

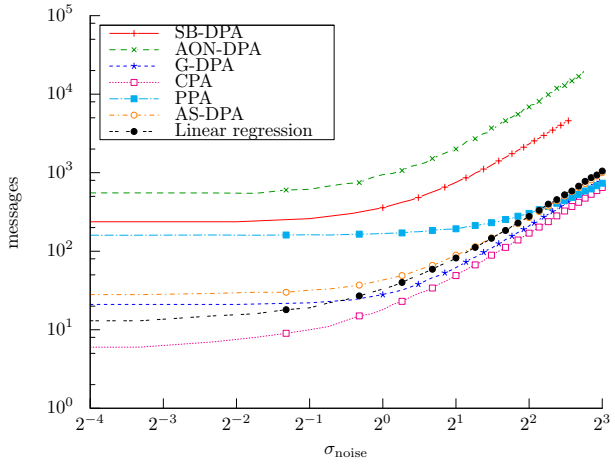


Fig. 2: Evolution of the number of messages (y-axis logscaled) to achieve a success rate of 90% according to the noise standard deviation (x-axis logscaled) – Higher resolution.

- An important efficiency difference occurs between the CPA, the DPAs and the PPA. For example with a noise standard deviation of 1, CPA needs only 30 messages to reach a success rate of 90%, whereas PPA needs 280 messages to achieve the same threshold.
- CPA is the most efficient attack. This confirms that Pearson's coefficient is the good tool to measure a linear correlation.
- In comparison, the PPA is much less efficient than the CPA (and even also than the DPAs). This result was actually expected. Indeed, *centering* the leakage and the model random variables (*i.e.* computing $\widehat{\mathbb{E}}\left(L\cdot \mathsf{m}(V_{\hat{k}})\right) - \widehat{\mathbb{E}}(L)\widehat{\mathbb{E}}\left(\mathsf{m}(V_{\hat{k}})\right)$ instead of $\widehat{\mathbb{E}}\left(L\cdot \mathsf{m}(V_{\hat{k}})\right)$ in the PPA attack) and then *normalizing* the centered mean by the standard deviations of the random variables (*i.e.* dividing

$\widehat{\mathbb{E}}\left(L \cdot \mathsf{m}(V_{\hat{k}})\right) - \widehat{\mathbb{E}}(L)\widehat{\mathbb{E}}\left(\mathsf{m}(V_{\hat{k}})\right)$ by $\widehat{\sigma}(L)$ and $\widehat{\sigma}\left(\mathsf{m}(V_{\hat{k}})\right)$ thus getting the CPA distinguisher $\mathsf{CPA}(\hat{k})$) is useful to reduce the linear dependency estimation errors when the number of observations is small (*i.e.* undersampling), which is the case when the attacks are performed for a small amount of noise.

– G-DPA, CPA and PPA are more efficient than AS-DPA and regression attacks. It may be noted that this situation is the opposite of the one occuring in the oversampling case.

Eventually, our results corroborate our theoretical analysis: the SB-DPA and the AON-DPA are less efficient than the other simulated attacks independent of the noise amount in the leakage. This highlights the fact that targeting a subspace of the model (*i.e.* , a single bit over eight or targeting 2 values over 256) is suboptimal when the adversary uses a model that well corresponds to the function $\delta(\cdot)$ (G-DPA, CPA and PPA) or when an AS-DPA or a regression attack is performed. Whatever the signal-to-noise ratio, CPA is always the best attack. However its efficiency is very close to that of G-DPA and PPA when the noise standard deviation reaches the threshold 4. Actually CPA is mainly better than the other tested attacks when the leakage is not very noisy (*i.e.* , when the noise standard deviation is between 0 and 4). Eventually, it can be noted that the efficiency of AS-DPA and linear regression attack tends to be close to that of the CPA while the perfect model scenario is optimally suited for CPA.

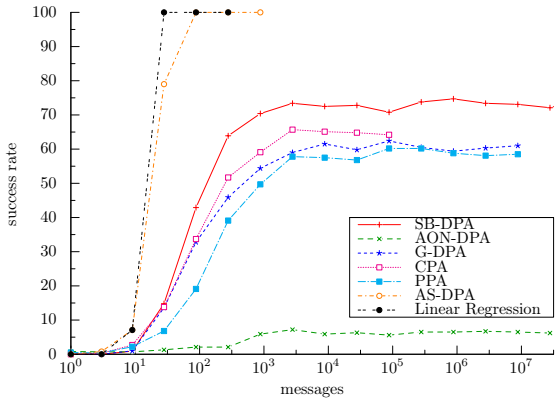
5.2 Attack Results in the Random Linear Leakage Scenario

In this section we assume that L satisfies (25). In Fig. 3, we recorded the success rate for different numbers of messages and for different values of noise standard deviation.

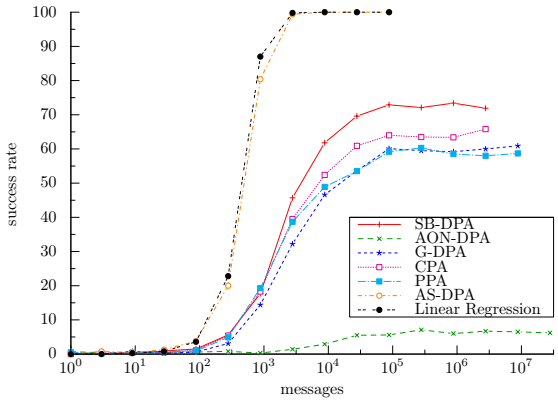
Observations are reported below. As in the perfect model scenario we can split our observations in two parts.

Oversampling. When the number of messages available is greater than approximately $10^5 \times \sigma^2$, the curves have the same shape for each distinguisher but contrary to what happened in the perfect model scenario, all the attacks do not reach a success rate of 100%.

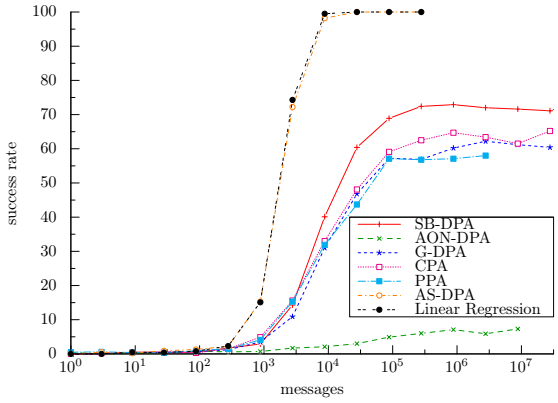
– The maximum success rate achieved by the model-based attacks is lower than 75% (*e.g.* , CPA achieves a success rate of 62% while G-DPA and PPA are still less efficient with a success rate limit of 58%) independent of the noise standard deviation. In other terms, for some linear functions $\delta(\cdot)$, those attacks do not succeed in discriminating the good key candidate when the Hamming weight function is involved as model. In Appendix B, we give a theoretical explanation of the CPA ineffectiveness for some linear functions $\delta(\cdot)$ and we argue that it



(a) No noise



(b) Mid noise (4.00)



(c) High noise (8.00)

Fig. 3: Evolution of the success rate (1,000 tries) for different numbers of messages and according to some critic noise standard deviations – whole data can be found in Appendix C.

is related to the algebraic properties of the s-box S that is targeted.
- At the opposite, the regression attack and the AS-DPA always succeeds in recovering the key and, actually, in a more efficient way than other attacks. Moreover, as it can be observed in Fig. 3b–3c, this assessment is confirmed independent of the noise standard deviation.
- AON-DPA only reaches a maximal success rate of 6% which is very low compared to the others. A possible explanation for the AON-DPA poor effectiveness resides in the fact that the design of the sets Ω_0 and Ω_1 under the hypothesis $\mathsf{m} = \mathrm{HW}$ is not relevant when $\delta(\cdot)$ is far away from the Hamming weight function
- At the opposite SB-DPA reaches a maximal success rate of 72% which is better than CPA. This observation is not surprising since SB-DPA targets only one bit (independently of the model choice) over eight, which lowers the impact of the model choice on the remaining seven bits.

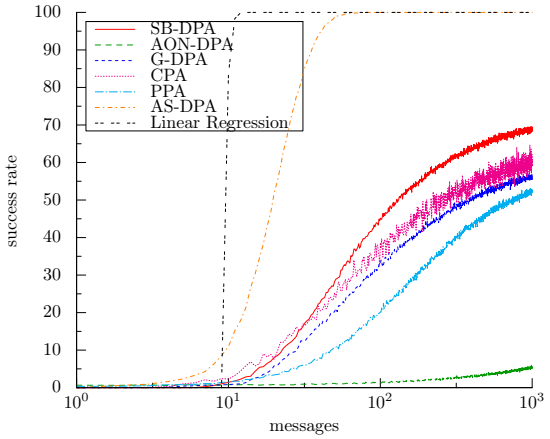
Undersampling. Let us focus on the critic values when a small number of messages is involved in the attack (lower than 500). In this case, the statistical stability of the involved distinguisher plays a role. To better understand how the different attacks perform in this context we redrew in Fig. 4 the curves with a thiner resolution than in Fig. 3.

Our observations are detailled below:

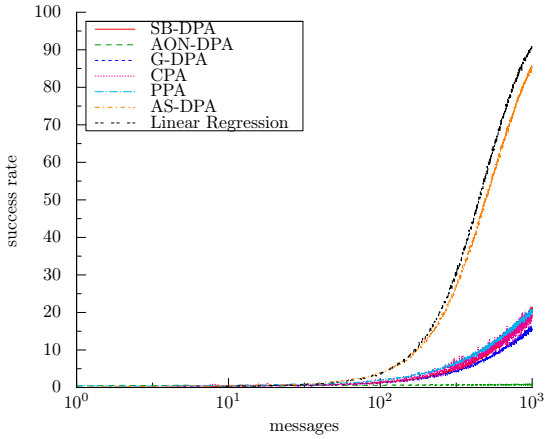
- In this situation, each distinguisher has the same ranking as in oversampling.
- G-DPA, CPA and PPA are relatively less efficient than in the perfect model scenario. Indeed, in the latter model scenario they are more efficient than AS-DPA and regression attack which is not the case here.
- SB-DPA and AON-DPA still have a different behavior than other model based attacks due to the use of a suboptimal model (with respect to the attacker choice in (22)).

The impact of the noise on the attacks efficiency in our linear random model scenario is very close to what we observed in the perfect model context. Namely the maximal success rate is the same whatever the noise deviation but more messages are needed to achieve it. In fact, we confirm the theoretical analysis in [21], where the author shows that doubling the noise deviation just increases the number of needed messages by $\sqrt{N}$ to reach the same success rate.

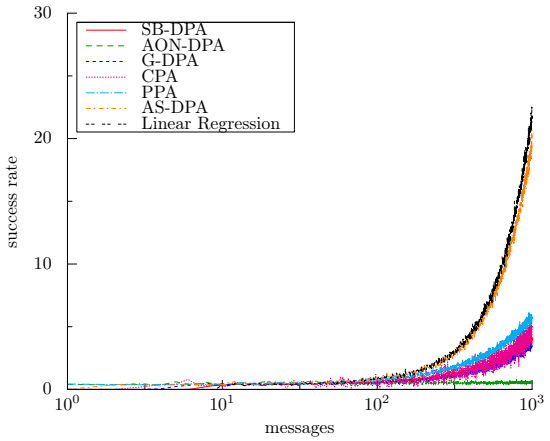
Among the attacks we simulated in the random model scenario, the linear regression attack and the AS-DPA are clearly the most efficient ones and they are the only ones that reach a success rate of 100%.



(a) No noise



(b) Mid noise (4.00)



(c) High noise (8.00)

Fig. 4: Evolution of the success rate (10,000 tries) for number of messages from 1 to 1,000 with some noise values.

5.3 Attacks Experiments in Real Life

In the previous sections, we have confronted our theoretical analyses with simulations in realistic scenarios. Two attacks

emerged, the CPA and the linear regression. In the following, we aim to confront our results against real measurements. Thus we only focus on CPA and linear regression attacks. Attack parameters are described below:

Attacks Target. The 8-bit output of the AES s-box, denoted by S, is targeted. Namely the variable V_k in (1) satisfies:

$$V_k = S(P \oplus k) \ , \tag{26}$$

where P corresponds to an 8-bit value known by the adversary.

Attack Types. We list below the attacks we have performed:

- CPA with m satisfying (22).
- Regression Attack with $\mathscr{B}_{\text{lin}} = (v_{\hat{k}}[i])_{0 \leqslant i \leqslant 7}$ as basis functions (Assumption 3 with $d = 1$).
- Regression Attack with $\mathscr{B}_{\text{quad}} = (v_{\hat{k}}[i] \cdot v_{\hat{k}}[j])_{0 \leqslant i \leqslant j \leqslant 7}$ as basis functions (Assumption 3 with $d = 2$).

Leakage Measurements. Power consumption leakages have been measured on a 8051 8-bit micro-controller. In each measurement curve, the part related to the manipulation of V_k is composed of 200 points. We suppose the curves to be synchronized (a glitch is used to be synchronized at the beginning of the manipulation processing). Before mounting the attacks, a pre-processing step has been performed on the curves to determine the most pertinent *point of interest* for each attack. By definition, this point is the one among the 200 points per curve that optimizes the attack efficiency. As argued in Section 3.5, it corresponds for the CPA to the point when the error resulting from the approximation of the leakage by the attack model (*i.e.* the Hamming weight function) is minimum. For the regression attacks, the point of interest is the point on which the error resulting from the approximation of the leakage by a linear (resp. quadratic) combination of the coordinates of the manipulated variable is minimum. During the pre-processing, we have used the fact that we knew the values $v_{k,i}$ manipulated by the device. Even if this does not correspond to a real life adversary, pre-processing in this context allows us to determine the time/point when an attack performed by an adversary with no such a knowledge is the most efficient. In the following, we sum-up the pre-processing step for the three attacks.

- CPA: the coefficient $\mathsf{CPA}_{\text{HW}}(k)^2$ has been estimated for each of the 200 points of the curve – the estimation being performed for a sample of size $400,000$ – to determine the best attack time.
- Regression Attack: a model function m_{lin} (resp. m_{quad}) corresponding to the correct k has been computed for each of the 200 points of the curve, the estimation being performed for a sample of size $400,000$. Then, 200 determination coefficients R^2 have been performed (one for each model M_k and the corresponding leakage point) to determine the best attack time corresponding to the basis functions $\mathscr{B}_{\text{lin}}$ (resp. $\mathscr{B}_{\text{quad}}$)

Figure 5 illustrates the results of the pre-processing step for each attack and each of the 200 points.

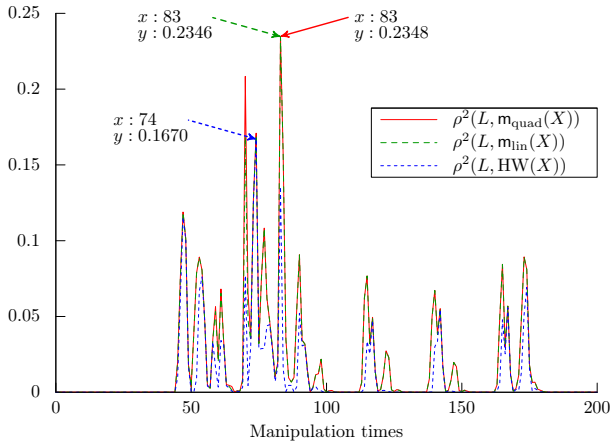


Fig. 5: Characterisation Timing Diagram. Max values are pinpointed by an arrow.

For the attack comparisons, only the point of interest resulting in the maximal distinguishing value has been considered for each attack.

Attack Comparaison. For each attack, the distinguishing coefficient (in y-axis) has been computed for each key candidate and for a given (increasing) number of power traces (x-axis). We recorded the minimal number of messages needed to have the real key ranked first (*i.e.* emerging from others). Results are drawn in Fig. 6,7 and 8. As expected linear regression with linear basis is clearly more efficient than CPA *i.e.* , a lower number of messages is required for the real key to emerge (68 messages is sufficient for the first one while 95 at least are needed for the CPA). As expected, the linear regression with quadratic basis needs more messages. In fact the information contained in the quadratic part of the leakage is not enough to compensate for the increase of noise resulting from the multiplication of leakage points (which is necessary to process the linear regression). Moreover the quadratic regression has to build a larger model (*i.e.* , from a larger basis) from data. We can remark that even with quadratic basis, the minimum number of messages needed to discriminate the real key is still very close to the one for CPA (≈ 95).

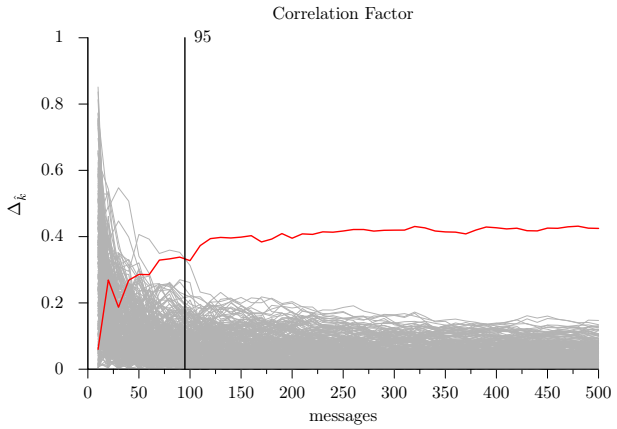


Fig. 6: Evolution of the distinguishing value (y-axis) with the number of messages (x-axis) for all key candidates for CPA. The curve of the real key used in the device is plotted in red.

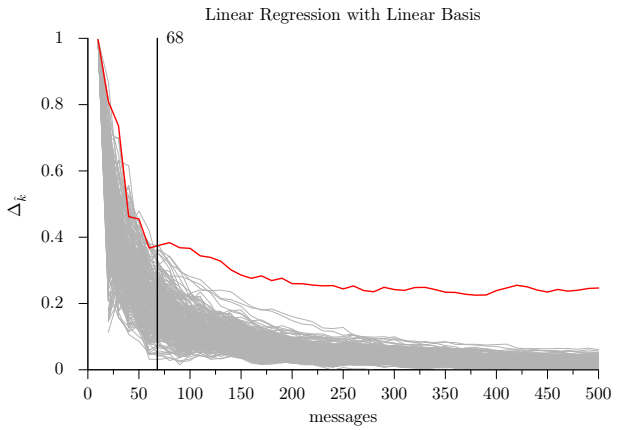


Fig. 7: Evolution of the distinguishing value (y-axis) with the number of messages (x-axis) for all key candidates for linear regression with linear basis. The curve of the real key used in the device is plotted in red.

5.4 Conclusion on the Attack Simulations and Experiments

When the chosen model exactly corresponds to the leakage function (perfect model case), each distinguisher reveals the key and the CPA and regression attacks are among the most efficient ones (actually except SB-DPA and AON-DPA all tested attacks have equivalent efficiency when the noise increases). Nevertheless in case of undersampling, CPA is ranked first. This can be explained by the fact that the linear regression attack has to rebuild the model from data while CPA is directly provided with the optimal model function and uses the observations only to corroborate a linear dependency.

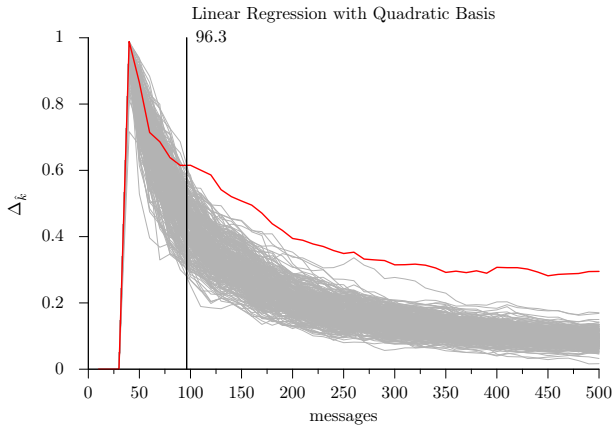


Fig. 8: Evolution of the distinguishing value (y-axis) with the number of messages (x-axis) for all key candidates for linear regression with quadratic basis. The curve of the real key used in the device is plotted in red.

When the model is unknown, the linear regression attack and the AS-DPA always succeed in revealing the key. they both are moreover more efficient than the model-based attacks. Nevertheless, collating both, the linear regression is always better than AS-DPA. That is, at a cost of a little computational overhead, linear regression attack shall be preferred to the other distinguishers.

Finally, if one has a good linear approximation of $\delta(\cdot)$ then CPA is an optimal way to perform an attack. In other cases, the linear regression attack will always perform better.

6 Conclusion and Future Works

In this paper, we have compared standard univariate side channel attacks and we have demonstrated that they all can be rewritten as a CPA. Our analyses show how important the model used for the attacks is. As a good model is not always known to the adversary, we have focused on another sound attack that is not parameterized by a model. This attack (introduced by Schindler *et al.* in [7]) is based on linear regression techniques. It is experimentally compared to CPA both in a favourable context for CPA (*i.e.* , the real leakage model is known) and in a more realistic context (*i.e.* , the real leakage model is linear but unknown and randomly generated). Eventually we have shown that in all cases the linear regression attack performs well independent of the leakage nature, provided that the key-dependent bits leak independently. We have moreover proposed an extension of the original attack in such a way that the latter assumption can be relaxed.

Based on our study, we think that the linear regression attacks are a relevant alternative to attacks based on an *a priori* model choice (as *e.g.* , the CPA). Our work moreover hilights the fact that any new attack should be compared at first

mathematically and experimentally if needed to the existing ones to reveal the core differences with the state-of-the-art. An interesting extension of our work will be to investigate the behavior of the linear regression attacks in multivariate contexts. Moreover, rewritting the side channel attack problematic in terms of a model estimation problematic opens the door to a large variety of stochastic tools that could be investigated for further research.

References

1. Kocher, P., Jaffe, J., Jun, B.: Differential Power Analysis. In Wiener, M., ed.: Advances in Cryptology – CRYPTO '99. Volume 1666 of Lecture Notes in Computer Science., Springer (1999) 388–397
2. Messerges, T.: Using Second-order Power Analysis to Attack DPA Resistant Software. In Koç, Ç., Paar, C., eds.: Cryptographic Hardware and Embedded Systems – CHES 2000. Volume 1965 of Lecture Notes in Computer Science., Springer (2000) 238–251
3. Brier, E., Clavier, C., Olivier, F.: Correlation Power Analysis with a Leakage Model. In Joye, M., Quisquater, J.J., eds.: Cryptographic Hardware and Embedded Systems – CHES 2004. Volume 3156 of Lecture Notes in Computer Science., Springer (2004) 16–29
4. Le, T.H., Clédière, J., Canovas, C., Robisson, B., Servière, C., Lacoume, J.L.: A Proposition for Correlation Power Analysis Enhancement. In Goubin, L., Matsui, M., eds.: Cryptographic Hardware and Embedded Systems – CHES 2006. Volume 4249 of Lecture Notes in Computer Science., Springer (2006) 174–186
5. Bévan, R., Knudsen, E.: Ways to Enhance Power Analysis. In Lee, P., Lim, C., eds.: Information Security and Cryptology – ICISC 2002. Volume 2587 of Lecture Notes in Computer Science., Springer (2002) 327–342
6. Mangard, S., Oswald, E., Standaert, F.X.: One for All - All for One: Unifying Standard DPA Attacks. Cryptology ePrint Archive, Report 2009/449 (2009) http://eprint.iacr.org/, to appear in IET Information Security.
7. Schindler, W., Lemke, K., Paar, C.: A Stochastic Model for Differential Side Channel Cryptanalysis. In Rao, J., Sunar, B., eds.: Cryptographic Hardware and Embedded Systems – CHES 2005. Volume 3659 of Lecture Notes in Computer Science., Springer (2005)
8. Messerges, T.: Power Analysis Attacks and Countermeasures for Cryptographic Algorithms. PhD thesis, University of Illinois (2000)
9. Brier, E., Clavier, C., Olivier, F.: Optimal Statistical Power Analysis. Cryptology ePrint Archive, Report 2003/152 (2003)
10. Coron, J.S., Giraud, C., Prouff, E., Rivain, M.: Attack and Improvement of a Secure S-Box Calculation Based on the Fourier Transform. [23] 1–14
11. Golić, J., Tymen, C.: Multiplicative Masking and Power Analysis of AES. In Kaliski Jr., B., Koç, Ç., Paar, C., eds.: Cryptographic Hardware and Embedded Systems – CHES 2002. Volume 2523 of Lecture Notes in Computer Science., Springer (2002) 198–212
12. Standaert, F.X., Gierlichs, B., Verbauwhede, I.: Partition vs. Comparison Side-Channel Distinguishers: An Empirical Evaluation of Statistical Tests for Univariate Side-Channel Attacks against Two Unprotected CMOS Devices. In Lee, P.J., Cheon, J.H., eds.: Information Security and Cryptology – ICISC 2008. Volume 5461 of Lecture Notes in Computer Science., Springer (2008) 253–267
13. Prouff, E., Rivain, M., Bévan, R.: Statistical Analysis of Second Order Differential Power Analysis. IEEE Trans. Comput. **58**(6) (2009) 799–811
14. Mangard, S., Oswald, E., Popp, T.: Power Analysis Attacks – Revealing the Secrets of Smartcards. Springer (2007)
15. Lemke-Rust, K.: Models and Algorithms for Physical Cryptanalysis. PhD thesis, Ruhr-Universität-Bochum, Germany (Jan 2007)
16. Renauld, M., Standaert, F.X., Veyrat-Charvillon, N., Kamel, D., Flandre, D.: A Formal Study of Power Variability Issues and Side-Channel Attacks for Nanoscale Devices. Advances in Cryptology – EUROCRYPT 2011 (2011)
17. Duan, C., Calle, V.H.C., Khatri, S.P.: Efficient On-Chip Crosstalk Avoidance CODEC Design. IEEE Trans. VLSI Syst. **17**(4) (2009) 551–560
18. Moll, F., Roca, M., Isern, E.: Analysis of dissipation energy of switching digital CMOS gates with coupled outputs. Microelectronics Journal **34**(9) (2003) 833–842
19. Agrawal, D., Rao, J., Rohatgi, P.: Multi-channel Attacks. In Walter, C., Koç, Ç., Paar, C., eds.: Cryptographic Hardware and Embedded Systems – CHES 2003. Volume 2779 of Lecture Notes in Computer Science., Springer (2003) 2–16
20. Bishop, C.M.: Pattern Recognition and Machine Learning (Information Science and Statistics). 1 edn. Springer (2007)
21. Mangard, S.: Hardware Countermeasures against DPA – A Statistical Analysis of Their Effectiveness. In Okamoto, T., ed.: Topics in Cryptology – CT-RSA 2004. Volume 2964 of Lecture Notes in Computer Science., Springer (2004) 222–235
22. Standaert, F.X., Archambeau, C.: Using Subspace-Based Template Attacks to Compare and Combine Power and Electromagnetic Information Leakages. [23] 411–425
23. Oswald, E., Rohatgi, P., eds.: Cryptographic Hardware and Embedded Systems - CHES 2008, 10th International Workshop, Washington, D.C., USA, August 10-13, 2008. Proceedings. In Oswald, E., Rohatgi, P., eds.: CHES. Volume 5154 of Lecture Notes in Computer Science., Springer (2008)

A Illustration of the Differences Between AS-DPA, M-DPA and CPA

Let us focus on an adversary targeting the manipulation of a 2-bit intermediate value V_k. For illustration purpose, we assume here that the attacked device leaks exactly the difference between the two bit-coordinates of V_k. Namely we assume that L satisfies $L = \delta(V_k)$, with $\delta(V_k) = V_k[0] - V_k[1]$. As explained in [22], such a situation is quite classical when leakage is measured by electro-magnetic analysis. If the adversary performs a single-bit DPA to exploit L, a natural choice for $M_{\hat{k}}$ is either $V_{\hat{k}}[0]$ or $V_{\hat{k}}[1]$ (namely in (2) the model function m is the projection related to one of the bit-coordinates of $V_{\hat{k}}$ and w_0 and w_1 equal 0 and 1 respectively). We denote by $\mathsf{DPA}(\hat{k})_0$ (respectively $\mathsf{DPA}(\hat{k})_1$) the distinguisher defined with respect to $M_{\hat{k}} = V_{\hat{k}}[0]$ (respectively $M_{\hat{k}} = V_{\hat{k}}[1]$). Under Assumption 2 which implies $\mathsf{var}(V_k[0]) = \mathsf{var}(V_k[1])$ and the independency between $V_k[0]$ and $V_k[1]$, we have

$$\mathsf{DPA}(\hat{k})_0 = \begin{cases} \mathbb{E}(V_k[0]) & \text{if } \hat{k} = k \\ 0 & \text{otherwise} \end{cases},$$

and

$$\mathsf{DPA}(\hat{k})_1 = \begin{cases} -\mathbb{E}(V_k[1]) & \text{if } \hat{k} = k \\ 0 & \text{otherwise} \end{cases}.$$

Since we have $\mathsf{DPA}(\hat{k})_0 = -\mathsf{DPA}(\hat{k})_1$ for every $\hat{k}$, the distinguisher $\mathsf{M\text{-}DPA}(\hat{k})$ always equals 0 whereas $\mathsf{AS\text{-}DPA}(\hat{k}) = 2 \cdot \mathbb{E}\left(V_{\hat{k}}[0]\right)$ if $\hat{k} = k$ and 0 otherwise.

Let us now focus on the case where the adversary performs a CPA with the Hamming weight as a model function. When computing the correlation between the leakage L and the model random variable $M_{\hat{k}} = \mathrm{HW}(V_{\hat{k}}) = V_{\hat{k}}[0] + V_{\hat{k}}[1]$, we have:

$$\mathrm{cov}\left(L, M_{\hat{k}}\right) = \mathrm{cov}\left(V_k[0] - V_k[1], V_{\hat{k}}[0] + V_{\hat{k}}[1]\right)$$

which can be rewritten:

$$\begin{aligned}\mathrm{cov}\left(L, M_{\hat{k}}\right) = \mathrm{cov}\left(V_k[0], V_{\hat{k}}[0]\right) + \mathrm{cov}\left(V_k[0], V_{\hat{k}}[1]\right) \\ - \mathrm{cov}\left(V_k[1], V_{\hat{k}}[0]\right) - \mathrm{cov}\left(V_k[1], V_{\hat{k}}[1]\right) \ ,\end{aligned}$$

from which we deduce $\mathsf{CPA}(\hat{k}) = 0$ independent of the relation between $\hat{k}$ and k.

To sum-up, this section gives an example of a leakage on a 2-bit variable for which the M-DPA and the CPA (with Hamming weight model function) fail, whereas the AS-DPA still succeeds.

B Why CPA can failed?

This section aims at explaining why the CPA fails in discriminating the correct key for some linear leakage models. Before starting our discussion, let us first have a look on the definition of the CPA distinguisher (6). Under Assumption 2, it involves standard deviations that tend to be independent of the key hypothesis when the sample size increases. As a consequence, the distinguisher in (6) discriminates key hypotheses in a similar way as the covariance $\mathrm{cov}\left(L, M_{\hat{k}}\right)$. Explaining the CPA failure hence amounts to explain the covariance failure when involved as a key-distinguisher.

Our analysis will be merely related to the following proposition.

Proposition 4 *Let f and g be two Boolean functions defined over $\mho_{2^n}$. If f and g are balanced, then we have:*

$$\mathrm{cov}\left(f, g\right) = \frac{1}{4} W(f \oplus g) \ , \tag{27}$$

where $W(f \oplus g)$ denotes the value $2^{-n} \sum_{x \in \mathbb{F}_2^n} (-1)^{f(x) \oplus g(x)}$.

Proof The result is a direct consequence of the following equality:

$$f + g = f \oplus g + 2fg \ . \tag{28}$$

Due to Assumption 1 and the fact that the leakage satisfies (1), we recall that $\mathrm{cov}\left(L, M_{\hat{k}}\right)$ equals $\mathrm{cov}\left(\delta\left(V_k\right), M_{\hat{k}}\right)$ independent of the targeted key k and the key hypothesis $\hat{k}$. If the model function m is the Hamming weight and if $\delta(\cdot)$ satisfies (18) with $d = 1$ (*i.e.* Assumption 3), then $\delta\left(V_k\right)$ and $M_{\hat{k}} = \mathsf{m}(V_{\hat{k}})$ respectively equal $\alpha_{-1} + \sum_i \alpha_i V_k[i]$ and $\sum_j V_{\hat{k}}[j]$. Under those two assumptions, we hence get:

$$\mathrm{cov}\left(L, M_{\hat{k}}\right) = \mathrm{cov}\left(\alpha_{-1} + \sum_i \alpha_i V_k[i], \sum_j V_{\hat{k}}[j]\right) \ , \tag{29}$$

i.e. ,

$$\mathrm{cov}\left(L, M_{\hat{k}}\right) = \sum_i \alpha_i \left(\sum_j \mathrm{cov}\left(V_k[i], V_{\hat{k}}[j]\right)\right) \ . \tag{30}$$

Since functions $V_k[i]$ and $V_{\hat{k}}[j]$ are both balanced under Assumption 2, Proposition 4 can be applied to develop the covariances in (30):

$$\mathrm{cov}\left(L, M_{\hat{k}}\right) = \frac{1}{4} \sum_i \alpha_i \sum_j W(V_k[i] \oplus V_{\hat{k}}[j]) \ , \tag{31}$$

That is we have $\mathrm{cov}\left(L, M_{\hat{k}}\right) = \frac{1}{4} \sum_i \alpha_i w_i(k, \hat{k})$ after denoting the term $\sum_j W(V_k[i] \oplus V_{\hat{k}}[j])$ by $w_i(k, \hat{k})$.

Let us study (31) when the correct key hypothesis is tested, *i.e.* , when $\hat{k}$ equals k. As V_k is balanced, the term $W(V_k[i] \oplus V_{\hat{k}}[j])$ is always zero except for $i = j$ where it equals 1. Equation (31) can thus be rewritten as:

$$\mathrm{cov}\left(L, M_{\hat{k}}\right) = \frac{1}{4} \sum_i \alpha_i \ . \tag{32}$$

In view of (32), $\mathrm{argmax}_{\hat{k}} |\mathrm{cov}\left(L, M_{\hat{k}}\right)|$ is not equal to the expected key (*i.e.* , the covariance distinguisher fails at discriminating the correct key), if there exists at least one key hypothesis $\hat{k} \neq k$ such that $V_{\hat{k}}$ satisfies:

$$\left|\sum_i \alpha_i\right| < \left|\sum_i \alpha_i w_i(k, \hat{k})\right| \ . \tag{33}$$

Actually, for the type of variables V_k involved in our attack simulations reported in Section 5.2, the condition (33) is often satisfied. In those simulations, V_k corresponds to the output of the AES s-box S parameterized by the key k. Namely, V_k takes the form $S(P \oplus k)$. In this context, $V_k[i] \oplus V_{\hat{k}}[j]$ corresponds to the function $P \mapsto S_i(P \oplus k) \oplus S_j(P \oplus \hat{k})$, where $S_1, \ldots, S_n$ denote the boolean coordinate functions of S. When P has a uniform distribution, the latter function shares the same distribution as the function $S^a_{i,j}$ defined by $S^a_{i,j}(P) = S_i(P \oplus a) \oplus S_j(P)$, with a denoting $k \oplus \hat{k}$. After denoting by $w_i(a)$ the sum $\sum_j W(S^a_{i,j})$, we therefore conclude on the equivalency between (33) and

$$\left|\sum_i \alpha_i\right| < \left|\sum_i \alpha_i w_i(a)\right| \ . \tag{34}$$

Since the coefficients $(\alpha_i)_i$ and $(w_i(a))_{i,a}$ have an amplitude upper bounded by 1 and the right hand of (34) is itself upper-bounded by $\min(\sum_i |\alpha_i|, \sum_i |w_i(a)|)$, we deduce two sufficient conditions for (34) to be never satisfied for $a \neq 0$:

- All the terms α_i's have the same sign.
- $\max_{a \neq 0} \sum_i |w_i(a)|$ is lower than or equal to $\sum_i |\alpha_i|$.

The first sufficient condition condition is device dependent and the second condition relies on the s-box properties. For the AES s-box for instance, it can be checked that $\max_a \sum_i |w_i(a)|$ equals 1.9375 (for $a = 53$). Thus, if $\sum_i |\alpha_i|$ is greater than 1.9375, then (34) cannot be satisfied for a value $a \neq 0$ and we deduce that the CPA is theoretically able to succeed in this case.

In the following, we give an example of such a case (*i.e.* , when CPA failed to discriminate the good key):

Example 2 Let $\{\alpha_i\}_{0 \leqslant i \leqslant 7} = \{0.5, 0.2, -0.5, 0.2, -0.5, 1, -0.8, 0.5\}$ be the coefficients of the leakage model, that is for every $x \in \mathbb{F}_{2^8}$:

$$\delta(x) = 0.5x_0 + 0.2x_1 - 0.5x_2 + 0.2x_3 - 0.5x_4 + x_5 - 0.8x_6 + 0.5x_7$$

where $(x_7, \ldots, x_0)$ is the binary decomposition of x. In this case we have $|\sum_i \alpha_i| = 0.6 < 1.9375$ and at least ten values (see Table 1) of a are such that $|\sum_i \alpha_i w_i(a)| > |\sum_i \alpha_i|$.

Table 2: Values of w_i and $\sum w_i$ for the AES s-box – values must be divided by 256.

a	w_0	w_1	w_2	w_3	w_4	w_5	w_6	w_7	$\sum w_i$
0	256	256	256	256	256	256	256	256	2048
1	-48	-28	0	-68	24	16	48	-72	304
2	-8	8	-16	-68	40	-16	40	4	200
3	0	40	32	8	0	84	48	4	216
4	64	-52	92	52	-44	-12	-12	40	368
5	4	24	-4	-28	-20	-20	-68	-88	256
6	64	-24	24	12	-48	20	40	16	248
7	36	48	-52	4	48	-40	84	-16	328
8	40	-60	20	24	76	84	-24	-96	424
9	-84	-40	-24	-52	12	40	-24	36	312

Table 1: Eleven highest values of $|\sum_i \alpha_i w_i(a)|$ obtained in Example 2.

a	$\lvert\sum_i \alpha_i w_i(a)\rvert$
101	0.8875
228	0.84375
109	0.775
25	0.7625
30	0.721875
176	0.66875
19	0.6578125
151	0.6515625
66	0.634375
158	0.6203125
0	0.6

10	60	0	56	4	-12	24	68	72	296
11	-48	-108	-60	-20	-60	-4	-4	-56	360
12	0	-36	32	-52	-108	32	12	40	312
13	40	-28	-64	60	80	-36	40	-52	400
14	48	-28	100	48	44	64	-20	-24	376
15	12	-12	-4	-56	32	56	-4	-40	216
16	-44	-16	-36	-52	-40	64	-48	-12	312
17	-36	28	-96	8	12	-44	20	-36	280
18	-36	24	-60	40	-32	-36	-12	-56	296
19	-44	-20	-76	-28	80	-64	36	-84	432
20	40	4	56	-68	-12	-24	24	76	304
21	-32	-8	-48	-48	-44	24	36	-32	272
22	20	44	60	140	12	0	16	68	360
23	-36	24	-52	-20	44	-48	-68	-4	296
24	-20	28	44	20	-12	52	-80	16	272
25	-40	12	8	-56	16	-72	68	-56	328
26	-16	8	-32	-16	-100	44	12	100	328
27	-20	36	-8	-104	32	-16	-20	-12	248
28	-56	48	-12	-28	-28	-20	-32	128	352
29	20	48	-104	-76	-68	-80	-40	-52	488
30	100	12	-44	-4	-56	24	-24	80	344
31	-104	-20	-32	4	8	0	-60	20	248
32	60	44	-16	88	56	-12	4	-104	384
33	-76	0	-88	-24	-44	-36	16	-60	344
34	8	24	12	-4	52	60	-60	76	296
35	-52	-48	-16	-28	-24	-64	-28	-68	328
36	20	24	16	44	-4	56	52	40	256
37	-48	76	-28	8	60	84	-12	-76	392
38	32	16	44	-84	52	60	20	44	352
39	-24	52	0	64	0	68	36	52	296
40	64	64	-68	-36	60	28	16	56	392
41	-32	-32	-52	-40	-32	-32	-8	-4	232
42	60	16	32	-36	16	-32	8	16	216
43	-20	-32	36	8	28	32	-24	-20	200
44	12	-60	44	-56	-64	20	-36	84	376
45	40	-28	-16	-12	32	4	4	-48	184
46	32	-20	-28	-84	-36	44	-48	-20	312
47	20	80	-92	32	64	8	-64	-8	368
48	68	20	16	36	28	-40	-16	104	328
49	-12	52	0	-60	-28	-32	-44	36	264
50	92	-28	48	-44	-8	0	76	40	336
51	24	-12	80	16	96	52	40	16	336
52	-4	24	24	40	40	-40	-16	68	256
53	-4	-60	-120	-84	68	-100	-12	-48	496
54	24	-36	44	-8	-16	-60	48	-44	280
55	4	68	8	-40	-60	16	-52	-24	272
56	24	-88	-8	24	-52	76	-28	12	312
57	24	12	-36	-32	-16	-32	-60	36	248
58	-32	-36	-32	40	-72	-4	-80	40	336
59	-72	60	16	-12	-84	-76	-32	-32	384
60	-28	-40	28	-112	-92	-8	-8	-44	360
61	-8	-28	84	24	-44	72	-20	-16	296
62	104	32	36	96	72	-4	36	12	392
63	-64	-40	-76	-124	-8	-20	16	-92	440
64	64	-12	60	-20	-16	-8	-104	28	312
65	12	-48	32	-60	-88	-8	-12	28	288
66	112	-16	-12	8	-32	48	0	76	304
67	-8	44	-24	68	-16	-16	-20	-60	256
68	20	64	-40	52	8	-28	-12	48	272
69	-28	28	-68	-36	-20	-28	44	44	296
70	-24	40	72	4	60	4	16	-36	256
71	20	20	-32	-32	-8	40	-96	-88	336
72	56	16	84	52	-20	20	12	-12	272
73	-20	-52	-32	0	76	-12	4	-12	208
74	-100	-40	-44	-8	-24	40	-36	-36	328
75	8	28	-12	36	-48	12	-28	-28	200
76	8	-12	-20	-16	-16	-96	0	-48	216
77	-32	-56	-32	-24	40	-12	-76	-80	352
78	-36	36	44	4	-32	-12	-12	-48	224
79	0	56	-60	-20	-60	48	24	-28	296
80	-4	-68	-36	40	-36	-16	24	-80	304
81	36	-40	-20	-24	32	-16	-44	-4	216
82	-12	-8	8	64	72	-68	48	80	360
83	8	-12	-40	-76	-12	-28	-20	-36	232
84	-8	-48	-16	8	16	-36	8	44	184
85	20	96	-52	-40	100	52	52	20	432
86	64	8	44	36	56	-16	-16	-24	264
87	-16	-76	-24	-44	-4	-24	-44	-16	248
88	68	20	-20	-20	40	-20	-56	12	256
89	-76	-20	84	-28	-36	28	4	-44	320
90	-24	40	-4	12	-32	-24	76	4	216
91	-12	-8	-24	-20	36	0	0	-4	104
92	-32	104	-4	40	48	24	36	56	344
93	44	40	-52	-20	40	12	24	-64	296
94	-60	-40	24	56	-76	36	32	12	336
95	16	92	-40	-20	0	8	24	72	272
96	32	-16	-60	-60	8	-4	0	-44	224
97	-32	24	-60	24	0	-8	-36	0	184
98	0	16	12	48	88	-40	-20	24	248
99	-56	-24	32	-52	20	48	12	-60	304
100	56	12	8	-24	48	12	76	4	240
101	96	-24	-64	28	-92	48	-68	-4	424
102	64	36	20	12	28	-80	76	-4	320
103	0	24	-24	-40	-64	20	16	-20	208
104	92	68	-16	-4	-16	-16	8	-36	256
105	-80	12	16	-56	-44	-20	-4	-80	312
106	-24	-68	-32	-96	36	12	12	-24	304
107	132	56	-40	0	-8	52	4	-44	336
108	12	44	92	-20	12	-60	4	92	336
109	-32	76	-36	20	4	88	-64	80	400
110	-32	-48	12	-24	-12	0	36	-68	232
111	-20	16	4	28	12	-44	-24	68	216
112	68	28	60	-8	76	0	-16	-8	264
113	-4	-20	-80	20	0	60	52	-20	256
114	-36	-40	-56	-12	-24	-8	104	32	312
115	36	32	44	-44	-48	12	-44	92	352
116	-76	-116	-20	-32	20	-40	12	28	344
117	16	120	64	28	36	88	-24	16	392
118	24	-32	-16	60	4	32	16	24	208
119	48	-28	-40	-60	76	-48	20	16	336
120	-76	-8	24	40	-32	20	72	-16	288
121	-32	8	20	60	-80	36	60	56	352
122	36	0	-72	32	80	-20	12	76	328
123	-20	32	-32	60	108	68	-40	40	400
124	-44	-76	-44	-56	-32	-24	32	44	352
125	-4	8	-60	68	-28	40	72	-24	304
126	-20	-48	-4	72	-12	-44	-48	8	256
127	-28	0	52	-36	-32	4	-16	-40	208
128	-24	-52	-36	-56	4	-16	-16	-100	304
129	52	16	-16	44	80	-32	-32	-40	312
130	-32	-32	-44	-12	-72	-36	-44	-88	360
131	-36	-4	-36	-92	-24	-52	-24	12	280
132	52	-44	12	28	28	-44	-40	24	272
133	-20	24	-12	-32	-92	-48	4	-40	272
134	12	36	20	8	20	-44	-4	-24	168
135	-28	0	64	56	0	24	-52	-16	240
136	8	24	8	-24	44	-44	-16	0	168
137	72	124	0	24	52	28	12	-8	320
138	4	36	-36	48	0	-60	16	-48	248
139	24	44	0	-100	-28	8	24	28	256
140	-8	-28	16	-24	36	-12	16	-4	144
141	4	-20	-24	0	-108	-52	-32	-24	264
142	-4	32	-16	64	-32	-36	56	-48	288
143	-12	24	12	8	-64	32	28	-44	224
144	24	24	0	4	112	-76	12	44	296
145	36	-12	8	92	-12	-32	-16	104	312
146	112	-4	84	32	-8	-32	-36	60	368
147	0	-80	72	12	-40	32	0	-28	264
148	-84	-8	72	-56	-68	-56	-20	-36	400
149	-4	-8	-12	52	-40	-20	-12	100	248
150	-24	8	-20	56	-20	24	80	8	240
151	0	0	-68	32	28	76	-48	52	304
152	8	-48	88	80	48	4	20	16	312
153	-120	-44	0	32	-60	0	-40	-96	392

154	12	-36	-32	0	-8	-28	-16	-44	176
155	-20	-48	-72	44	-40	-36	24	76	360
156	28	-8	60	32	104	0	96	32	360
157	-48	-96	-48	-44	-4	64	12	-4	320
158	-72	0	20	-8	52	-24	44	-52	272
159	-8	68	32	88	8	8	40	28	280
160	36	-36	80	20	-40	-8	-32	-92	344
161	60	0	12	32	28	72	-12	0	216
162	36	72	16	40	12	-44	36	64	320
163	0	-44	36	-56	16	4	-4	16	176
164	-12	20	36	-52	8	20	-40	-4	192
165	24	4	4	-4	24	-32	52	0	144
166	-56	-36	28	-4	12	44	-72	28	280
167	-84	-60	-12	-80	20	-56	-60	-52	424
168	-4	-56	16	68	-36	0	-4	24	208
169	-76	-60	-60	-24	-56	-40	36	-32	384
170	-68	-84	80	-12	-32	-52	-40	-24	392
171	20	-20	24	-12	44	40	-44	12	216
172	-32	-8	52	12	-4	-4	8	24	144
173	-12	-36	-32	-56	-52	24	64	-60	336
174	36	56	12	8	52	48	60	72	344
175	64	-24	-24	-12	40	40	-72	12	288
176	28	24	-28	-4	-12	68	-64	28	256
177	-72	8	-44	8	12	16	0	-56	216
178	-44	-28	4	32	-20	32	4	-68	232
179	32	-16	-76	-8	-52	-40	64	-24	312
180	-28	-4	32	0	4	-4	32	32	136
181	-8	-4	12	4	-88	12	-32	24	184
182	0	16	24	-8	-20	-20	40	0	128
183	0	-40	-28	20	48	28	24	-28	216
184	48	112	0	48	-8	72	36	52	376
185	-32	40	-8	-16	-4	0	16	4	120
186	4	20	100	-16	60	-48	-16	8	272
187	-24	-68	-60	-32	-72	16	44	-52	368
188	20	36	60	68	8	-4	4	-40	240
189	-132	28	-64	-16	-24	-28	-44	-40	376
190	-32	8	4	16	-4	-24	20	60	168
191	-12	52	4	20	-4	4	72	32	200
192	-64	0	-40	28	8	32	-20	40	232
193	-40	8	48	-28	-16	56	-20	-48	264
194	-72	-44	-56	24	56	4	-32	-32	320
195	-52	96	-32	20	52	4	-60	52	368
196	4	-28	-16	0	96	-20	0	-28	192
197	-32	-48	20	28	-16	92	44	-8	288
198	4	-44	40	60	-40	48	64	28	328
199	-68	0	52	-32	32	-4	-32	-44	264
200	-36	32	12	-4	-20	-28	-28	-56	216
201	68	48	4	20	32	60	4	76	312
202	48	64	8	40	64	4	68	72	368
203	-24	-32	8	-52	-64	32	-12	24	248
204	-44	12	44	24	92	-52	0	-4	272
205	72	16	-28	-24	-40	12	8	56	256
206	56	-32	-36	12	-12	-64	4	24	240
207	-12	-4	36	4	-8	36	-36	32	168
208	-72	52	156	20	-40	28	24	0	392
209	-20	-32	40	-40	-36	68	-68	-72	376
210	20	-36	-20	-32	16	-48	8	-20	200
211	-12	36	36	16	-80	64	80	92	416
212	92	60	-36	-12	76	16	-4	56	352
213	12	-8	24	20	8	56	36	28	192
214	-72	4	-16	-96	-44	12	-16	-44	304
215	0	44	28	40	-92	8	44	-8	264
216	56	16	-32	24	4	-68	8	-80	288
217	-12	-48	-8	24	-68	-44	64	-4	272
218	0	-36	-4	76	28	-36	36	-8	224
219	4	-20	0	-8	-32	-52	-72	-92	280
220	28	4	-32	-12	-8	0	-28	-40	152
221	-72	-4	88	8	-84	-44	-48	-20	368
222	32	-76	-12	-48	56	16	32	8	280
223	-36	-36	-24	-8	-44	36	-52	-52	288
224	-64	-52	24	48	-12	8	4	36	248
225	12	20	-16	0	-44	-28	88	40	248
226	64	-28	0	28	-12	0	8	12	152
227	-32	-20	84	-76	24	4	48	16	304
228	-28	28	-28	56	56	120	-84	80	480
229	-48	-36	-16	-44	56	-16	32	-32	280
230	68	-44	24	4	36	-64	4	12	256
231	32	-28	44	28	60	68	0	28	288
232	-20	-48	32	64	-76	-20	-84	16	360
233	48	-28	-20	68	8	-52	-28	-44	296
234	8	64	-4	136	-12	4	44	32	304
235	12	16	28	4	8	-52	32	32	184
236	116	-8	-12	12	76	-52	28	0	304
237	-24	-32	32	20	12	68	100	16	304
238	-16	4	-12	96	-44	-16	8	-4	200
239	44	-28	32	0	-80	-52	-12	-120	368
240	-16	60	-28	8	-24	-64	-32	48	280
241	-60	-4	-40	-20	-4	40	-8	0	176
242	-104	-8	60	16	8	56	-40	12	304
243	-4	-8	-28	-68	36	-60	-40	28	272
244	40	44	0	20	24	0	-52	-4	184
245	-104	-92	-56	-16	-72	4	-44	-92	480
246	-56	-40	16	-44	-28	0	8	-40	232
247	48	4	28	40	4	24	8	4	160
248	48	16	88	-64	52	-60	12	12	352
249	44	-28	32	-84	-28	-52	-24	-44	336
250	68	40	-24	-16	60	-32	-48	56	344
251	-76	32	-56	-112	-56	28	20	-4	384
252	44	-24	20	56	32	-48	-96	0	320
253	-12	-24	-28	0	-68	-88	-32	60	312
254	100	28	16	28	-16	-24	-28	8	248
255	0	-28	-36	-36	40	-52	-76	-28	296

C Additional Material

In this article, several experimentations have been practiced, resulting in various sets. Some of the data sets have been adapted (*e.g.* , fitted, truncated) to become more readable. For informational purpose we plot the whole data set in this section. Figure 9 shows us the raw data use for fitting in Fig. 1.

Figure 10 shows us the evolution of the success rate according to the number of messages and the noise deviation in the random leakage scenario. Figures 3 and 4 are extracted from Fig. 10.

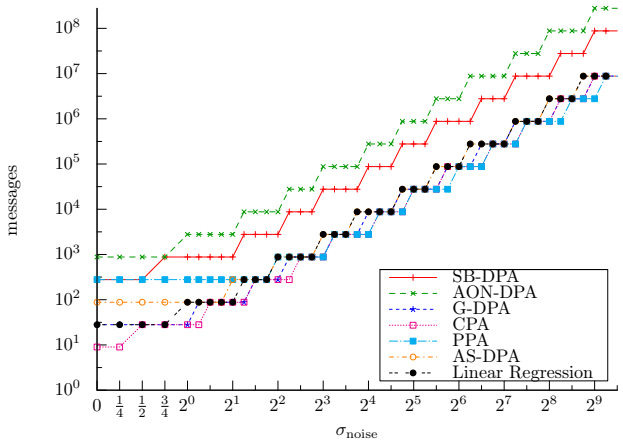


Fig. 9: Evolution of the number of messages needed to achieve a success rate of 90% for different noise values.

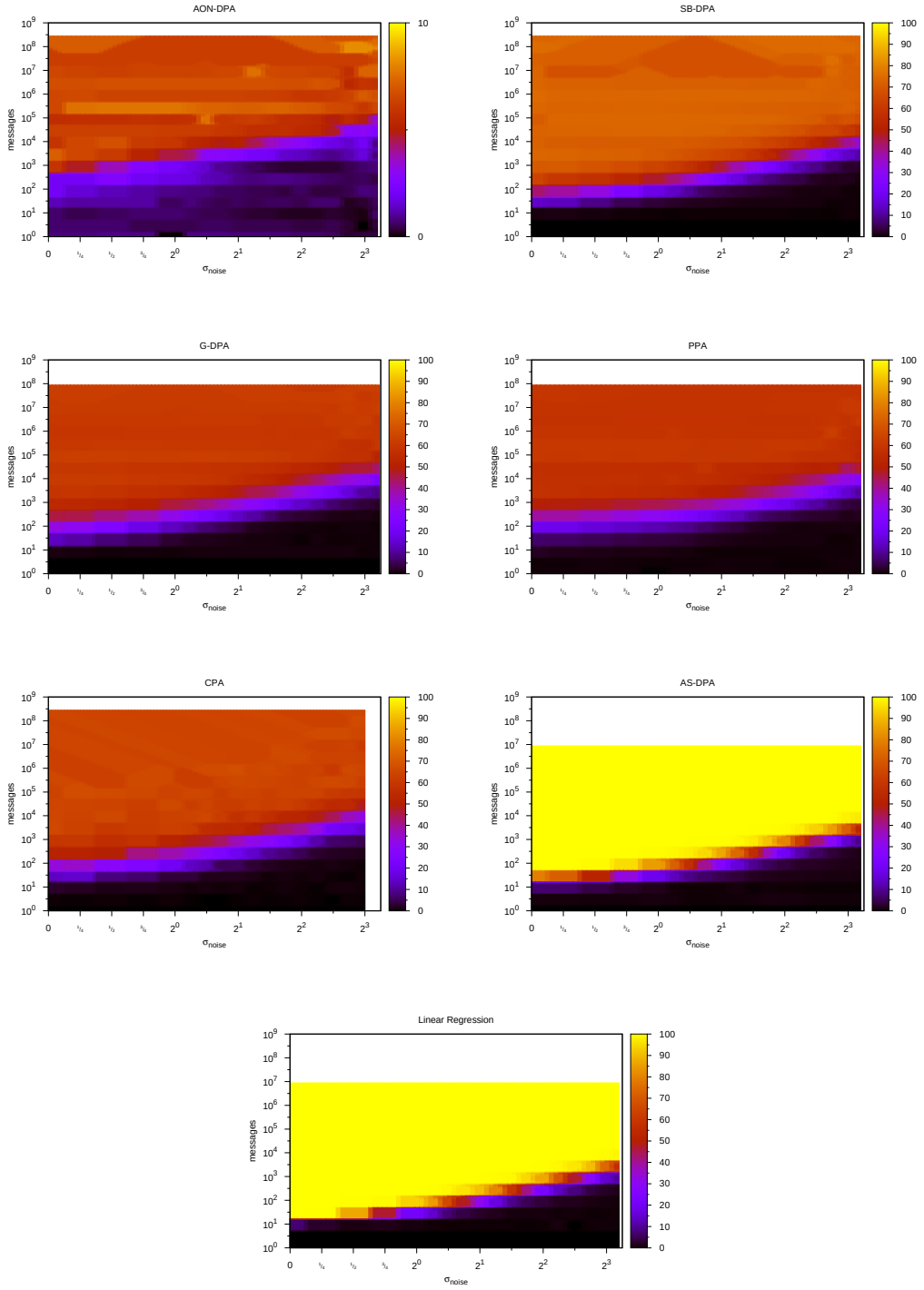


Fig. 10: evolution of the success rate according to the number of messages and the noise deviation.