Univariate Side Channel Attacks and Leakage Modeling

Julien Doget[1,2,3], Emmanuel Prouff[1], Matthieu Rivain[4], and François-Xavier Standaert[2] *

[1] Oberthur Technologies,
71-73 rue des Hautes Pâtures, F-92 726 Nanterre, France
{j.doget, e.prouff}@oberthur.com
[2] Université Catholique de Louvain-la-Neuve, UCL Crypto Group,
B-1348 Louvain-la-Neuve, Belgium
fstandae@uclouvain.be
[3] Université Paris 8, Département de Mathématiques,
2, rue de la Liberté, F-93 526 Saint-Denis, France
[4] CryptoExperts,
Paris, France
matthieu.rivain@cryptoexperts.com

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.

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, thoses 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

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

(*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 wether 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 is the impact of the model choice on the attack efficiency. Since a good leakage model is not always available to the attacker, we study in a 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** 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 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 an Subsitution-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

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

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

Assumption 1 (Independent Noise Assumption). *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

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

$v_{k,i}$ 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 system 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.

1.1 Notations

Let X be a random variable and let x and Ω be respectively an element and a subset of the domain $\mathcal{X}$ 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 $\widehat{\mathrm{P_r}}(X = x)$ respectively. For instance, if N denotes the size of the sample $(\ell_{k,i})_i$, notations $\widehat{\mathbb{E}}(L)$ and $\widehat{\mathrm{P_r}}(L = \ell)$ 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 $\mathcal{X}$ a same number of times. Clearly, the size N of such a sample is a multiple of the cardinality of $\mathcal{X}$.

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 $M_{\hat{k}} = \mathsf{m}(V_{\hat{k}})$.

2 Reduction Between Various Side Channel Attacks

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

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

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 used 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* $\widehat{\mathrm{cov}}\left(L\,,\, M_{\hat{k}}\right)$ *denotes their covariance which equals* $\widehat{\mathbb{E}}\left(LM_{\hat{k}}\right) - \widehat{\mathbb{E}}\left(L\right)\widehat{\mathbb{E}}\left(M_{\hat{k}}\right)$*.*

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* $\hat{k}$*.*

In what follows, we state the SCA-reductions between DPA, PPA and CPA (Sections 2.2 and 2.3). We show that all 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. These results emphasize the importance of making a good choice for the model according to the attack context specificities, which is eventually discussed (Section 2.4).

2.2 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 2.1 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 same reasoning holds for an all-or-nothing DPA 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 it can be rewritten as a PPA distinguisher $\mathsf{PPA}_{(\alpha_i)_i}(\hat{k})$ by stating:

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

Under Assumption 2, coefficients α_i are constant (namely independent of the sample size and of the key hypothesis). ◇

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.

2.3 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 family of coefficients* $(\alpha_i)_i$ *and a model function* m. *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:

$$\mathsf{f}(\omega_i) = \frac{\alpha_i}{\widehat{\mathrm{P_r}}\left(M_{\hat{k}} = \omega_i\right)} \ . \tag{8}$$

Under Assumption 2, probabilities $\widehat{\mathrm{P_r}}\left(M_{\hat{k}} = \omega_i\right)$, 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:

$$\mathsf{PPA}_{(\alpha_i)_i,\mathsf{m}}(\hat{k}) = \sum_{\omega_i \in \mathrm{Im}(\mathsf{m})} \mathsf{f}(\omega_i) \cdot \widehat{\mathrm{P_r}}\left(M_{\hat{k}} = \omega_i\right) \cdot \widehat{\mathbb{E}}\left(L \mid M_{\hat{k}} = \omega_i\right) \ .$$

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:

$$\mathsf{PPA}_{(\alpha_i)_i,\mathsf{m}}(\hat{k}) = \widehat{\mathbb{E}}\left(LM'_{\hat{k}}\right) \ . \tag{9}$$

On the other hand under Assumption 2, $\widehat{\mathbb{E}}(L)$, $\widehat{\sigma}(L)$, $\widehat{\mathbb{E}}\left(M_{\hat{k}}\right)$ 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:

$$\widehat{\mathbb{E}}\left(LM'_{\hat{k}}\right) = a \cdot \mathsf{CPA}_{\mathsf{f}\circ\mathsf{m}}(\hat{k}) + b \ , \tag{10}$$

$\diamond$

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.

2.4 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 which 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 $\delta()$ of the device leakage. Actually, they only require some general assumptions on the algebraic properties of $\delta()$ (*e.g.* the output value of the function is any linear combination of the bits of the input value).

3 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 $\delta()$ is greatly correlated to the attack model m, by the weaker requirement that $\delta()$ 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 $\delta()$ 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:

$$\delta(x) = \alpha_{-1} + \sum_{i=0}^{n-1} \alpha_i x_i + \sum_{i_1,i_2=0}^{n-1} \alpha_{i_1,i_2} x_{i_1} x_{i_2} + \cdots + \sum_{i_1,\ldots,i_d=0}^{n-1} \alpha_{i_1,\ldots,i_d} x_{i_1} x_{i_2} \cdots x_{i_d} \ . \quad (11)$$

In view of (11), a side channel adversary could use his knowledge of the device technology to make an assumption on the degree d of $\delta()$ 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 $\delta()$ of the leakage is upper bound 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 (11), 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 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 encompass 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, 17]) that the coefficients of the quadratic terms in (11) 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 $d = 2$ 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 sections we present a side channel attack that is 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, its efficiency does not rest on the adversary ability to find a model m which is a good affine approximation of $\delta()$ as it was the case for CPA-like attacks. The attack is 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 attack outlines. Eventually, in Section 3.2 we briefly explain how it 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).

3.1 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 [18]. 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 $\left(\ell_{k,1}, \ell_{k,2}, \ldots, \ell_{k,N}\right)$ 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 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 $\boldsymbol{\alpha}_{\hat{k}}$ defined such that:

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

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

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

Under Assumption 3, the distinguishing value $\Delta_{\hat{k}}$ is expected to be minimal for the good hypothesis $\hat{k} = k$.

Remark 1. In the literature, the common way to describe how well a model fits a set of observations is called *goodness of fit*. 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 coefficient of determination:

$$R^2 = \frac{\|\mathbf{L} - \mathbf{M} \cdot \boldsymbol{\alpha}_{\hat{k}}\|^2}{\mathsf{var}\,(L)} \ .$$

It first permits to have a value in the range $[0, 1]$. 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.

3.2 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 a consequence of 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.

4 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 one pertinent way of attacking where some constraints on the model can be relaxed. 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:

$$V_k = S(P \oplus k) \ , \tag{12}$$

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

Attack Types. We hereafter list 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. 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 and regression attack do not. In our attacks simulation, we have assumed that the adversary did not know the definition of the function $\delta(\cdot)$ in (1) 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:

$$\mathsf{m}(V_{\hat{k}}) = \mathrm{HW}(V_{\hat{k}}) = \sum_i V_{\hat{k}}[i] \ . \tag{13}$$

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) and (Ω_0, Ω_1) in AON-DPA and 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

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

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 (9) is satisfied for the model function m defined in (13) (*i.e.* $\mathsf{PPA}_{(\alpha_i)_i}(\hat{k}) = \widehat{\mathbb{E}}\left(L \cdot \mathrm{HW}(V_{\hat{k}})\right)$).

Leakage Simulations. Leakage measurements have been simulated according to (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 $\delta()$ in (1), leading to two different scenarios:

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

$$L = \mathrm{HW}(V_k) + B \ , \tag{14}$$

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 $\delta()$. It will be referred as the *perfect model* scenario.

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

$$L = \alpha_{-1} + \sum_{i=0}^{7} \alpha_i \cdot V_k[i] + B \ , \tag{15}$$

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 $\delta()$ 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.

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 4.1) and Scenario 2 (Section 4.2).

4.1 Attack Results in the Perfect Model Scenario

In this section we assume that L satisfies (14). In Fig. 1, the number of messages needed to achieve a success rate of 90% is recorded for each attack mentionned before*. Note that a success rate threshold has been fixed at 90% but in this configuration each attack can reach 100%.

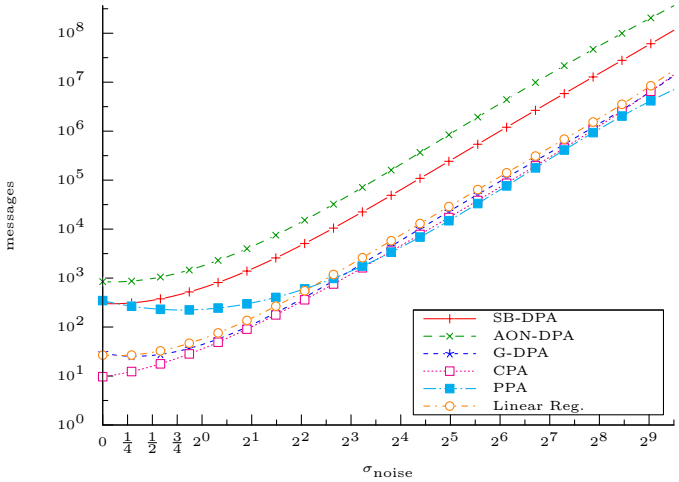


Fig. 1: Evolution of the number of messages (y-axis) to achieve a success rate of 90% according to the noise standard deviation (x-axis) – Fitted curves

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.

*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.

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 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 and regression attack are stacked. This 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 (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 $\delta()$ 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 and 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.

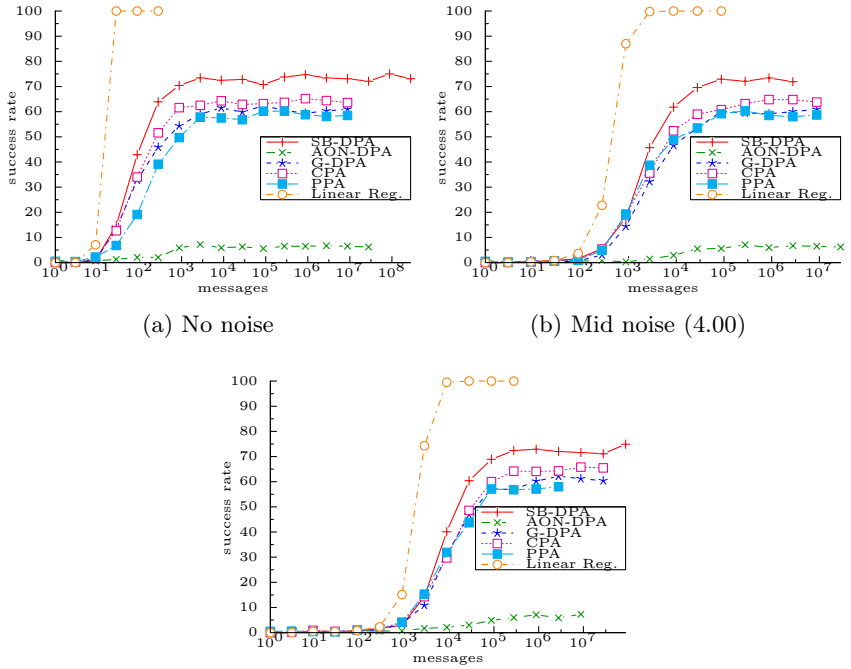
Undersampling. When the noise standard deviation is lower than 2^3, the number of messages needed to perform an attack is quite small (lower than 500). In this case, the statistical stability of the involved distinguisher plays a role. We detail our observations below:

- 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 regression attack. 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 whatever 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 $\delta()$ (G-DPA, CPA and PPA) or when a regression attack with $v_{\hat{k}}[0], \cdots, v_{\hat{k}}[7]$ as basis functions 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 the linear regression attack tends to be close to that of the CPA while the perfect model scenario is optimally suited for CPA.

4.2 Attack Results in the Random Linear Leakage Scenario

In this section we assume that L satisfies (15). In Fig. 2, we recorded the success rate for different numbers of messages and for different values of noise standard deviation.



(a) No noise

(b) Mid noise (4.00)

(c) High noise (8.00)

Fig. 2: Evolution of the success rate (1,000 tries) for different numbers of messages and according to some critic noise standard deviations

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 62% while G-DPA and PPA are still less efficient with a success rate limit of 58%) whatever the noise standard deviation. In other terms, for some linear functions $\delta()$, those attacks do not succeed in discriminating the good key candidate when the Hamming weight function is involved as model.
- At the opposite, the regression attack always succeeds in recovering the key and, actually, in a more efficient way than other attacks. Moreover, as it can be observed in Figures 2b–2c, this assessment is confirmed whatever 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 $\delta()$ 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 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. Our observations are detailled below:

- In this situation, all distinguishers have the same ranking as in oversampling.

– G-DPA, CPA and PPA are relatively less efficient than in the perfect model scenario. That is in the perfect model scenario they are more efficient than regression attack while not here.
– SB-DPA and AON-DPA still have a different behavior than others model based attacks due to the use of a suboptimal model (with respect to the attacker choice in (13)).

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 [19], 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 is clearly the most efficient one and it is the only one that reaches a success rate of 100%.

4.3 Conclusion on the Attack Simulations

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 the 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.

When the model is unknown, only the linear regression attack always succeeds in revealing the key. It is moreover more efficient than the model-based attacks. 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 $\delta()$ then CPA is an optimal way to perform an attack. In other cases, linear regression attack will always perform better.

5 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 is the model used for the attacks. 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 without care about 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. In Oswald, E., Rohatgi, P., eds.: CHES. Volume 5154 of Lecture Notes in Computer Science., Springer (2008) 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. 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
17. 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
18. Bishop, C.M.: Pattern Recognition and Machine Learning (Information Science and Statistics). 1 edn. Springer (2007)
19. 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