Partition *vs.* Comparison Side-Channel Distinguishers

An Empirical Evaluation of Statistical Tests for Univariate Side-Channel Attacks against Two Unprotected CMOS Devices*

François-Xavier Standaert[1]**, Benedikt Gierlichs[2], Ingrid Verbauwhede[2]

[1] UCL Crypto Group, Université catholique de Louvain, B-1348 Louvain-la-Neuve.
[2] K.U. Leuven, ESAT/SCD-COSIC and IBBT
e-mails: fstandae@uclouvain.be,{bgierlic,iverbauw}@esat.kuleuven.be

Abstract. Given a cryptographic device leaking side-channel information, different distinguishers can be considered to turn this information into a successful key recovery. Such proposals include *e.g.* Kocher's original DPA, correlation and template attacks. A natural question is therefore to determine the most efficient approach. In the last years, various experiments have confirmed the effectiveness of side-channel attacks. Unfortunately, these attacks were generally conducted against different devices and using different distinguishers. Additionally, the public literature contains more proofs of concept (*e.g.* single experiments exhibiting a key recovery) than sound statistical evaluations using unified criteria. As a consequence, this paper proposes a fair experimental comparison of different statistical tests for side-channel attacks. This analysis allows us to revisit a number of known intuitions and to put forward new ones. It also provides a methodological contribution to the analysis of physically observable cryptography. Additionally, we suggest an informal classification of side-channel distinguishers that underlines the similarities between different attacks. We finally describe a new (but highly inspired from previous ones) statistical test to exploit side-channel leakages.

1 Introduction

Showing the effectiveness of a side-channel attack usually starts with a proof of concept. An adversary selects a leaking device of his choice and exploits the available physical information with a distinguisher. Recovering a cryptographic key (*e.g.* from a block cipher) is then used to argue that the attack works. But as for any experimental observation, a proof of concept has to be followed by a sound statistical analysis. For example, one can compute the number of queries to a target cryptographic device required to recover a key with high confidence. Even better, one can compute the success rate or guessing entropy of a side-channel adversary in function of this number of queries. Various experimental and theoretical works describing different types of side-channel attacks can be found in

* Work supported in part by the IAP Programme P6/26 BCRYPT of the Belgian State, by FWO projects G.0475.05 and G.0300.07, by the European Commission under grant agreement 216646 ECRYPT NoE phase II, and by K.U. Leuven-BOF.
** Associate researcher of the Belgian Fund for Scientific Research (F.R.S.-FNRS).

the open literature. However, they are generally conducted independently and therefore are not straightforward to compare. Hence, we believe that a unifying analysis of actual distinguishers is important to enhance our understanding.

The contribution of this paper is threefold. First, we propose an informal classification of side-channel distinguishers into two categories, namely partition and comparison distinguishers. Second, we describe an alternative statistical test for partitioning attacks based on the sample variance. Most importantly and following the framework in [18], we provide a fair empirical comparison of statistical tests for univariate side-channel distinguishers against two unprotected software implementations of the AES Rijndael [5]. It includes Kocher's original Differential Power Analysis (DPA) [10], Pearson's correlation coefficient [2] and template attacks [3] as well as the recently proposed Mutual Information Analysis (MIA) [6]. Our results demonstrate the wide variety of flexibility *vs.* efficiency tradeoffs that can be obtained from different distinguishers and the effectiveness of template attacks when exploiting good leakage models. Additionally, they illustrate that claims on the efficiency of a given attack highly depend on an adversarial or implementation context. These results suggest that any new proposal of side-channel attack should come with a similar evaluation in order to show how these new proposals behave compared to former attacks. Note that we do not claim the novelty of our conclusions. As a matter of fact, several works already discussed similar comparison goals (see *e.g.* [4, 11]). However, we believe that the approach, metrics and number of experiments proposed in this paper allow improving the evaluation of side-channel attacks and pinpointing their limitations.

The rest of the paper is structured as follows. Sections 2 and 3 describe our target implementations and the side-channel adversaries that will be used in our comparisons. Section 4 proposes an informal classification for side-channel distinguishers and details the different statistical tests that we will consider in our comparisons. It additionally describes a variance-based statistical test for side-channel attacks. Section 5 defines our evaluation metrics. Section 6 discusses the limitations and features of our classification and methodology. The description of our experiments and results are in Section 7 and conclusions are in Section 8.

2 Target implementations

We target two implementations of the AES-128 in two 8-bit RISC-based microcontrollers. In the first setup, we used a PIC 16F877 running at a frequency around 4 MHz. In the second setup, we used an Atmel ATmega163 in a smart card body, clocked at 3.57 MHz. In both cases, our attacks aim to recover the first 8 bits of the block cipher master key k. We denote this part of the key as a key class s. The physical leakages were acquired with a digital oscilloscope, respectively a Tektronix 7140 with a 1 GHz bandwidth running at a 250 MS/s sampling rate for the PIC and an Agilent Infinium 54832D with a 1GHz bandwidth running at a 200 MS/s sampling rate for the Atmel. We note that although the PIC and Atmel devices seem to be very similar, their leakages are substantially different, as will be confirmed in the following sections.

3 Side-channel adversary

The present analysis aims to compare different statistical tests for side-channel attacks. But statistical tests are only a part of a side-channel adversary. A fair comparison between such tests consequently requires that the other parts of the adversary are identical. Following the descriptions and definitions that are introduced in [18], it means that all our attacks *against a given target device* exploit the same input generation algorithm, the same leakage function and the same leakage reduction mapping. In addition, in order to illustrate the wide variety of tradeoffs that can be considered for such adversaries, we used slightly different settings for our two devices (*i.e.* PIC and Atmel). Specifically:

- We fed both devices with uniformly distributed random (known) plaintexts.
- We provided the statistical tests with the same sets of leakages, monitored with two similar measurement setups (*i.e.* one setup by target device).
- Only univariate side-channel adversaries were considered in our comparison.
  - For the PIC device, we used a reduction mapping R that extracts the leakage samples corresponding to the computation of $\mathsf{S}(x_i \oplus s)$ in the traces (this clock cycle is illustrated in the left part of Figure 1), where S is the 8-bit substitution box of the AES and x_i the first 8 bits of the plaintext. Then, only the maximum value of this clock cycle was selected. Hence, to each input plaintext vector $\mathbf{x}_q = [x_1, x_2, \ldots, x_q]$ corresponds a q-sample leakage vector $\mathsf{R}(\mathbf{l}_q) = [\mathsf{R}(l_1), \mathsf{R}(l_2), \ldots, \mathsf{R}(l_q)]$.
  - For the Atmel implementation, all the samples corresponding to the computation of the first AES round were first tested independently in order to determine the sample giving rise to the best results, for each statistical test. Then, the actual analysis was only applied to this sample. Figure 2 illustrates this selection of time samples for two statistical tests to be defined in the next section, namely the Difference of Means (DoM) and Pearson's correlation coefficient. In other words, we used a different reduction mapping for each of the statistical tests in this case.

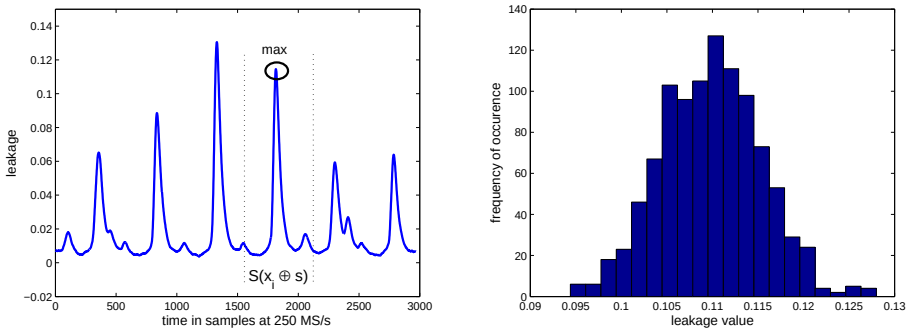

Fig. 1: Left: exemplary PIC power trace and selection of the meaningful samples. Right: Histogram for the statistical distribution of the electrical noise in the PIC leakages.

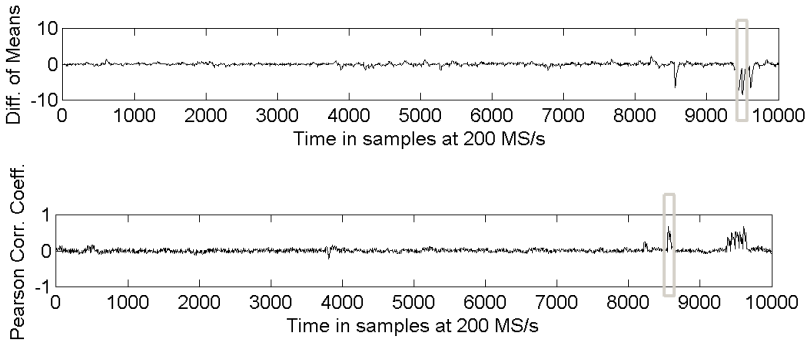

Fig. 2: Selection of the meaningful time samples for the Atmel.

We mention that these choices are arbitrary: the only goal is to provide each statistical test with comparable inputs, for each target device. They also correspond to different (more or less realistic) attack scenarios. For the PIC device, we assume that one knows which sample to target: it considerably reduces the attack's time and memory complexities. But the selected time sample may not be optimal for a given attack. For the Atmel, no such assumption is made.

4 Classification of distinguishers

In this section, we propose to informally classify the possible side-channel distinguishers as partition-based or comparison-based. More specifically:

- In a partition-based attack and for each key class candidate s^*, the adversary defines a partition of the leakages according to a function of the input plaintexts and key candidates. We denote such partitions as: $\mathsf{P}(s^*, \mathbf{v}^q_{s^*})$, where $\mathbf{v}^q_{s^*} = \mathsf{V}(s^*, \mathbf{x}_q)$ are some (key-dependent) values in the implementation that are targeted by the adversary. For example, the S-box output $\mathsf{S}(x_i \oplus s^*)$ is a usual target. Then, a statistical test is used to check which partition is the most meaningful with respect to the real physical leakages. We denote this test as $\mathsf{T}(\mathsf{P}(s^*, \mathbf{v}^q_{s^*}), \mathsf{R}(\mathbf{l}_q))$. For example, Kocher's original DPA [10] partitions the leakages according to one bit in the implementation.
- In comparison-based attacks, the adversary models a part/function of the actual leakage emitted by the target device, for each key class candidate s^*. Depending on the attacks, the model can be the approximated probability density function of a reduced set of leakage samples denoted: $\mathsf{M}(s^*, \mathsf{R}(\mathbf{l}_q)) = \hat{\Pr}[s^* | \mathsf{R}(\mathbf{l}_q)]$, as when using templates [3]. Or the model is a deterministic function (*e.g.* the Hamming weight) of some values in the implementation: $\mathsf{M}(s^*, \mathbf{v}^q_{s^*})$, as in correlation attacks [2]. Then, a statistical test is used to compare each model $\mathsf{M}(s^*, .)$ with the actual leakages. Similarly to partitioning attacks, we denote this test as $\mathsf{T}(\mathsf{M}(s^*, .), \mathsf{R}(\mathbf{l}_q))$.

We note that the previous classification is purely informal in the sense that it does not relate to the capabilities of an adversary but to similarities between

the way the different attacks are performed in practice. As the next sections will underline, it is only purposed to clarify the description of different statistical tests. As a matter of fact, one can partition according to (or model the leakage of) both single bits and multiple bits in an implementation. In both partition and comparison attacks, the expectation is that only the correct key class candidate will lead to a meaningful partition or good prediction of the actual leakages. Hence, for the two types of attacks, the knowledge of reasonable assumptions on the leakages generally improves the efficiency of the resulting key recovery. With this respect, the choice of the internal value used to build the partitions or models highly matters too. For example, one could use the AES S-box inputs $x_i \oplus s^*$ or outputs $\mathsf{S}(x_i \oplus s^*)$ for this purpose. But using the outputs generally gives rise to better attack results because the of the S-box non-linearity [14].

4.1 Statistical tests for partition distinguishers

In a partition attack, for each key class candidate s^* the adversary essentially divides the leakages in several sets and stores them in the vectors $\mathbf{p}^1_{s^*}$, $\mathbf{p}^2_{s^*}$, $\ldots$, $\mathbf{p}^n_{s^*}$. These sets are built according to a *hypothetical* function of the internal values targeted by the adversary that we denote as H. It directly yields a variable $\mathbf{h}^q_{s^*} = \mathsf{H}(\mathbf{v}^q_{s^*})$. In general, H can be any surjective function from the target values space $\mathcal{V}$ to a hypothetical leakage space $\mathcal{H}$. Examples of hypothetical leakages that can be used to partition a 16-element leakage vector $\mathsf{R}(\mathbf{l}_{16})$ include:

- a single bit of the target values (*i.e.* $n = 2$),
- two bits of the target values (*i.e.* $n = 4$),
- the Hamming weight of 4 bits of the target values (*i.e.* $n = 5$).

Such partitions are illustrated in Table 1 in which the indices of the $\mathsf{R}(l_i)$ values correspond to the input plaintexts $[x_1, \ldots, x_{16}]$. In a 1-bit partition, the 16 leakage values are stored in the vector $\mathbf{p}^1_{s^*}$ if the corresponding hypothetical leakage (*e.g.* one bit of $\mathsf{S}(x_i \oplus s^*)$) equals 0 and stored in $\mathbf{p}^2_{s^*}$ otherwise. As a result, we have one partition per key class candidate s^* and n vectors $\mathbf{p}^i_{s^*}$ per partition.

$\mathbf{p}^1_{s^*}$	$\mathbf{p}^2_{s^*}$
$\mathsf{R}(l_1)$	$\mathsf{R}(l_2)$
$\mathsf{R}(l_3)$	$\mathsf{R}(l_5)$
$\mathsf{R}(l_4)$	$\mathsf{R}(l_7)$
$\mathsf{R}(l_6)$	$\mathsf{R}(l_8)$
$\mathsf{R}(l_{10})$	$\mathsf{R}(l_9)$
$\mathsf{R}(l_{12})$	$\mathsf{R}(l_{11})$
$\mathsf{R}(l_{14})$	$\mathsf{R}(l_{13})$
$\mathsf{R}(l_{15})$	$\mathsf{R}(l_{16})$

$\mathbf{p}^1_{s^*}$	$\mathbf{p}^2_{s^*}$	$\mathbf{p}^3_{s^*}$	$\mathbf{p}^4_{s^*}$
$\mathsf{R}(l_3)$	$\mathsf{R}(l_1)$	$\mathsf{R}(l_5)$	$\mathsf{R}(l_6)$
$\mathsf{R}(l_4)$	$\mathsf{R}(l_2)$	$\mathsf{R}(l_9)$	$\mathsf{R}(l_7)$
$\mathsf{R}(l_{11})$	$\mathsf{R}(l_{10})$	$\mathsf{R}(l_8)$	$\mathsf{R}(l_{12})$
$\mathsf{R}(l_{15})$	$\mathsf{R}(l_{14})$	$\mathsf{R}(l_{16})$	$\mathsf{R}(l_{13})$

$\mathbf{p}^1_{s^*}$	$\mathbf{p}^2_{s^*}$	$\mathbf{p}^3_{s^*}$	$\mathbf{p}^4_{s^*}$	$\mathbf{p}^5_{s^*}$
$\mathsf{R}(l_5)$	$\mathsf{R}(l_2)$	$\mathsf{R}(l_1)$	$\mathsf{R}(l_3)$	$\mathsf{R}(l_{14})$
	$\mathsf{R}(l_7)$	$\mathsf{R}(l_4)$	$\mathsf{R}(l_6)$	
	$\mathsf{R}(l_9)$	$\mathsf{R}(l_8)$	$\mathsf{R}(l_{12})$	
	$\mathsf{R}(l_{16})$	$\mathsf{R}(l_{10})$	$\mathsf{R}(l_{13})$	
		$\mathsf{R}(l_{11})$		
		$\mathsf{R}(l_{15})$		

Table 1: Examples of 1-bit, 2-bit and Hamming weight partitions.

Kocher's DoM test. The first proposal for checking the relevance of a leakage partition is the difference of means test that was initially introduced in [10] and more carefully detailed in [12]. In this proposal and for each key class candidate s^*, the adversary only considers two vectors from each partition, respectively denoted as $\mathbf{p}_{s^*}^A$ and $\mathbf{p}_{s^*}^B$. Applying such a difference of means test simply means that the adversary computes the difference between the sample means[1]:

$$\Delta_{s^*} = \hat{\mathbf{E}}(\mathbf{p}_{s^*}^A) - \hat{\mathbf{E}}(\mathbf{p}_{s^*}^B) \quad (1)$$

In single-bit attacks as when using the 1-bit partition in Table 1, the vectors $\mathbf{p}_{s^*}^A$ and $\mathbf{p}_{s^*}^B$ correspond to the two only columns of the partition. In multiple-bit attacks as when using the 2-bit partition in Table 1, the vectors $\mathbf{p}_{s^*}^A$ and $\mathbf{p}_{s^*}^B$ either correspond to two columns (out of several ones) in the partition, as in "all-or-nothing" multiple-bit attacks or they correspond to two combinations of columns in the partition, as in "generalized" multiple-bit attacks (see [12]). Note that "all-or-nothing" attacks have the drawback that several leakage samples are not exploited (*i.e.* corresponding to the unexploited columns of the partition). Note also that the best selection of the (combination of) columns requires to make assumptions about the leakages. For example, "all-or-nothing" attacks implicitly assume that the behavior of several bits is the same so that "all-or-nothing" partitions yield the largest Δ_s. As a result of the attack, the adversary obtains a vector $\mathbf{g}_q$ with the key candidates rated according to the test result, the most likely key corresponding to the highest absolute value for Δ_{s^*}.

MIA. Another proposal for exploiting a leakage partition in a more generic way than using a difference of means test has been described in [6]. It notably aims to exploit all the samples in a multiple-bit partition, without making any assumption on the leakage model. For this purpose, the adversary attempts to approximate the mutual information between the hypothetical leakages $\mathbf{h}_{s^*}^q$ and the actual leakages $\mathsf{R}(\mathbf{l}_q)$. For each vector $\mathbf{p}_{s^*}^i$, he first builds histograms in order to evaluate the joint distribution $\hat{\Pr}[\mathsf{R}(\mathbf{L}_q), \mathbf{H}_{s^*}^q]$ and the marginal distributions $\hat{\Pr}[\mathsf{R}(\mathbf{L}_q)]$ and $\hat{\Pr}[\mathbf{H}_{s^*}^q]$, for each key class candidate. Then, he estimates:

$$\hat{\mathrm{I}}(\mathsf{R}(\mathbf{L}_q); \mathbf{H}_{s^*}^q) = \hat{\mathrm{H}}[\mathsf{R}(\mathbf{L}_q)] + \hat{\mathrm{H}}(\mathbf{H}_{s^*}^q) - \hat{\mathrm{H}}[\mathsf{R}(\mathbf{L}_q), \mathbf{H}_{s^*}^q]$$

As in a difference of means test, the adversary obtains a vector $\mathbf{g}_q$ containing the key candidates rated according to the test result, the most likely key corresponding to the largest value for the mutual information.

4.2 Statistical tests for comparison distinguishers

Pearson's correlation coefficient. In a correlation attack, the adversary essentially predicts a part/function of the leakage in the target device, for each key class candidate s^*. As a result, he obtains q-element vectors $\mathbf{m}_{s^*}^q = \mathsf{M}(s^*, \mathbf{v}_{s^*}^q)$.

[1] In statistical textbooks, Difference of Means tests usually refer to more complex hypothesis tests. We use this simple version for illustration because it was extensively used in the cryptographic hardware community.

For example, if a device is known to follow the Hamming weight leakage model, the vector typically contains the Hamming weights of the values $\mathsf{S}(x_i \oplus s^*)$. Since the reduced leakage vector $\mathsf{R}(\mathbf{l}_q)$ also contains q elements, the test can estimate the correlation between these two vectors, *e.g.* using Pearson's coefficient:

$$\rho_{s^*} = \frac{\sum_{i=1}^{q}(\mathsf{R}(l_i) - \hat{\mathbf{E}}(\mathsf{R}(\mathbf{l}_q))) \cdot (m_{s^*}^i - \hat{\mathbf{E}}(\mathbf{m}_{s^*}^q))}{\sqrt{\sum_{i=1}^{q}(\mathsf{R}(l_i) - \hat{\mathbf{E}}(\mathsf{R}(\mathbf{l}_q)))^2 \cdot \sum_{i=1}^{q}(m_{s^*}^i - \hat{\mathbf{E}}(\mathbf{m}_{s^*}^q))^2}} \quad (2)$$

Again, the adversary obtains a vector $\mathbf{g}_q$ with the key candidates rated according to the test result, the most likely key corresponding to the highest correlation.

Bayesian analysis. In template attacks, the adversary takes advantage of a probabilistic model for the leakages. He exploits an estimation of the conditional probabilities $\Pr[\mathsf{R}(\mathbf{l}_q)|s]$. From such an estimation, a straightforward strategy is to apply Bayes theorem and to select the keys according to their likelihood:

$$\lambda_{s^*} = \hat{\Pr}[s^*|\mathsf{R}(\mathbf{l}_q)] \quad (3)$$

It yields the same key candidate vector $\mathbf{g}_q$ as in the previous examples. Note that these attacks correspond to a stronger adversarial context than the other statistical tests in this section and require an estimation of the leakage probability distribution (*i.e.* to build templates). They should therefore be seen as a limit of what a side-channel adversary can achieve. We also note that in our simple context, the construction of templates was assumed to be unbounded[2]. But in more challenging scenarios, *i.e.* if the construction of templates is bounded, the use of stochastic models can be necessary for this purpose [15].

4.3 An alternative partition distinguisher using a variance test

The previous section described a number of statistical tests to evaluate the quality of a leakage model or partition. Of course, this list is not exhaustive: various other approaches have been and could be proposed. In this section, we suggest that under the common hypothesis of Gaussian noise in the physical leakages (confirmed in Figure 1 for the PIC), one can propose an alternative to the mutual information distinguisher [6]. Indeed, since in this context, the entropy of

[2] Following [3], we assumed the leakages to be drawn from a normal distribution:

$$\mathcal{N}(\mathsf{R}(l_i)|\mu_s^i, \sigma_s^i) = \frac{1}{\sigma_s^i\sqrt{2\pi}} \exp\frac{-(\mathsf{R}(l_i) - \mu_s^i)^2}{2{\sigma_s^i}^2}, \quad (4)$$

in which the means μ_s^i and standard deviations σ_s^i specify completely the noise associated to each key class s. In practice, these means and standard deviations were estimated during a preliminary profiling step in which the adversary characterizes the target device (we constructed one template for each value of $\mathsf{S}(s \oplus x_i)$). That is, the probabilities $\Pr[s^*|\mathsf{R}(l_i)]$ are approximated in our attacks using Bayes theorem and the estimated Gaussian distribution $\hat{\Pr}[\mathsf{R}(l_i)|s^*] = \mathcal{N}(\mathsf{R}(l_i)|\hat{\mu}_{s^*}^i, \hat{\sigma}_{s^*}^i)$, where $\hat{\mu}_{s^*}^i$ and $\hat{\sigma}_{s^*}^i$ respectively denote the sample mean and variance for a given leakage sample.

a good partition only depends on its variance, one can save the construction of histograms. Such a variance test can be described as follows. Let us denote the sample variance of the leakage and partition vectors as $\hat{\sigma}^2(\mathsf{R}(\mathbf{l}_q))$ and $\hat{\sigma}^2(\mathbf{p}^i_{s^*})$. From those variances, we compute the following statistic for each partition:

$$\sigma^2_{s^*} = \frac{\hat{\sigma}^2(\mathsf{R}(\mathbf{l}_q))}{\sum_{i=1}^{n} \frac{\#(\mathbf{p}^i_{s^*})}{q} \cdot \hat{\sigma}^2(\mathbf{p}^i_{s^*})} \tag{5}$$

where $\#(\mathbf{p}^i_{s^*})$ denotes the number of elements in a vector $\mathbf{p}^i_{s^*}$ of the partition. The most likely key is the one that gives rise to the highest variance ratio. Note that variance tests have been used in the context of timing attacks (*e.g.* in [9]). However, we could not find a reference using a similar test in the context of power analysis attacks. Any suggestion is welcome.

We finally mention that partition-based attacks generally require the partitions corresponding to different key candidates to be made of meaningful vectors $\mathbf{p}^i_{s^*}$. For example, an attack against the 8 key-bits corresponding to the first S-box of the AES using an 8-bit partition will give rise to vectors $\mathbf{p}^i_{s^*}$ containing only the leakages corresponding to one input x_i. Therefore, these partitions will not allow discriminating the key candidates. In other words, partition attacks cannot use bijective hypothetical leakage functions [6].

5 Evaluation metrics

We propose to quantify the effectiveness of our distinguishers with two security metrics, namely the success rates of order o and guessing entropy. Let $\mathbf{g}_q$ be the vector containing the key candidates sorted according to the test result after a side-channel attack has been performed: $\mathbf{g}_q := [g_1, g_2, \ldots, g_{|\mathcal{S}|}]$. A success rate of order 1 (*resp.* 2, ...) relates to the probability that the correct key class is sorted first (*resp.* among the two first ones, ...) by the adversary. More formally, we define the success function of order o against a key class s as: $\mathsf{S}^o_s(\mathbf{g}_q)$=1 if $s \in [g_1, \ldots, g_o]$, else $\mathsf{S}^o_s(\mathbf{g}_q)$=0. It leads to the o^{th}-order success rate:

$$\mathbf{Succ}^o_S = \mathop{\mathbf{E}}_{s} \mathop{\mathbf{E}}_{\mathbf{l}_q} \mathsf{S}^o_s(\mathbf{g}_q) \tag{6}$$

Similarly, the guessing entropy measures the average number of key candidates to test after a side-channel attack has been performed. Using the same notations as for the success rate, we define the index of a key class s in a side-channel attack as: $\mathsf{I}_s(\mathbf{g}_q) = i$ such that $g_i = s$. It corresponds to the position of the correct key class s in the candidates vector $\mathbf{g}_q$. The guessing entropy is simply the average position of s in this vector:

$$\mathbf{GE}_S = \mathop{\mathbf{E}}_{s} \mathop{\mathbf{E}}_{\mathbf{l}_q} \mathsf{I}_s(\mathbf{g}_q) \tag{7}$$

Intuitively, a success rate measures an adversarial strategy with fixed computational cost after the physical leakages have been exploited. The guessing entropy measures the average computational cost after this exploitation. For a theoretical discussion of these metrics, we refer to [18].

6 Limitations of our classification and methodology

Before moving to the description of our experimental results, let us emphasize again that the previous classification of attacks is informal. It is convenient to consider partition-based attacks since they all exploit a division of the leakages such as in Table 1. But as far as the adversarial capabilities are concerned, the most important classification relates to the need (or lack thereof) of a leakage model. For example, template attacks require the strongest assumptions, *i.e.* a precise characterization of the target device. Other attacks do not need but can be improved by such a characterization (*e.g.* correlation in [17]), with more or less resistance to a lack of knowledge on the target device. With this respect, the mutual information analysis is the most generic statistical test in the sense that it does not require any assumption on the leakage model.

Also, all our evaluations depend on the target implementations and attack scenarios, and hence are only valid within these fixed contexts. As will be shown in the next section, even two implementations of the same algorithm on similar platforms may yield contrasted results. Similarly, changing any part of the adversary in Section 3 (*e.g.* considering adaptively selected input plaintexts, another reduction mapping, ...) or modifying the measurement setups could affect our conclusions. Importantly, these facts should not be seen as theoretical limitations of the proposed framework but as practical limitations in its application, related to the complex device-dependent mechanisms in side-channel attacks. Hence, it motivates the repetition of similar experiments in various other contexts.

7 Experimental results

In this section, we present the different experiments that we carried out against our two target devices. We investigated partition distinguishers with DoM tests, MIA and variance tests. We also evaluated correlation attacks using Pearson's coefficient and template attacks. For this purpose and for each device, we generated 1000 leakage vectors, each of them corresponding to q=250 random input plaintexts. Then, for each of the previously described statistical tests, we computed different success rates and the guessing entropy for:

- various number of queries ($1 \leq q \leq 250$),
- various partitions and models (1-bit, 2-bit, ..., Hamming weight).

For each value of q, our metrics were consequently evaluated from 1000 samples. The results are represented in Figures 3, 4, 5, 6, 7, (the latter ones in Appendix) and lead to a number of observations that we now detail.

1. The two devices have significantly different leakage behaviors. While the PIC leakages closely follow Hamming weight predictions (*e.g.* Figure 3, correlation test, lower left part), the Atmel leakages give rise to less efficient attacks in this context (*e.g.* Figure 4, correlation test, lower left part).

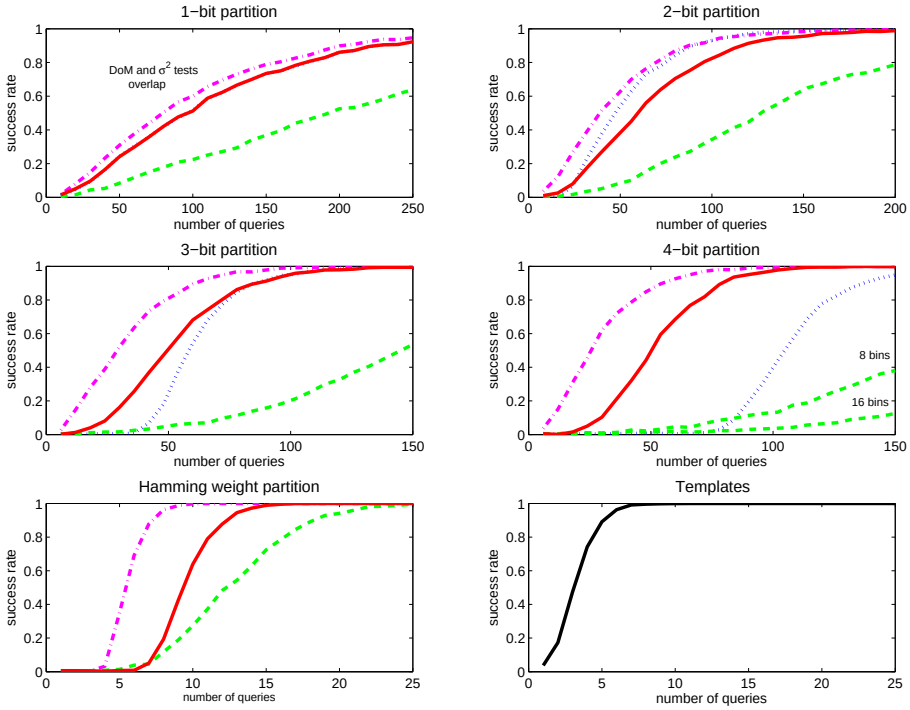

Fig. 3: PIC: 1st-order SR for different statistical tests, partitions, models (dotted: DoM test, dash-dotted: correlation test, dashed: MIA, solid: variance test).

2. By contrast 1-bit and 2-bit partitions give rise to more efficient attacks against the Atmel device than against the PIC (*e.g.* Figures 3 and 4 again).

The assumed reason for these observations is that different bits in the Atmel implementation contribute differently to the overall leakage. In particular, we observed experimentally that 1-bit attacks were the most efficient when targeting the S-box output LSB against the Atmel (the success rate was significantly lower with other bits). This assumption also explains why multiple-bit attacks lead to relatively small improvement of the attacks, compared to the PIC.

3. As far as the comparison of distinguishers is concerned, the main observation is that template attacks are the most efficient ones against both devices, confirming the expectations of [3, 18]. However, it is worth noting that while univariate templates directly lead to very powerful attacks against the PIC implementation, the exploitation of multiple samples significantly improves the success rate in the Atmel context (*e.g.* Figure 4, lower right part). Note again that the unbounded construction of our templates has a significant impact on this observation. The effect of a bounded construction of templates to the final attack effectiveness has been studied in [7].

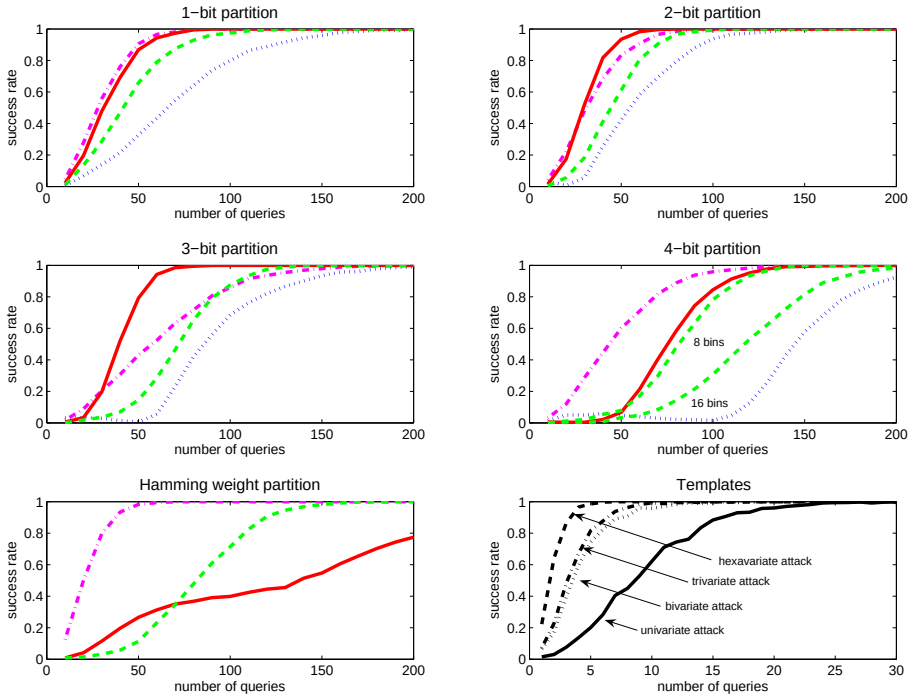

Fig. 4: Atmel: 1st-order SR for different statistical tests, partitions, models (dotted: DoM test, dash-dotted: correlation test, dashed: MIA, solid: variance test).

4. By contrast, no general conclusions can be drawn for the non-profiled distinguishers (DoM and variance tests, correlation attack, MIA). This confirms that different adversarial contexts (*e.g.* types of leakages, distributions of the noise, selections of the meaningful samples, ...) can lead to very different results for these attacks. A distinguisher can also be fast to reach low success rates but slow to reach high success rates. Or different distinguishers could be more or less immune to noise addition or other countermeasures against side-channel attacks. For example, in our experiments the DoM test shows an effectiveness similar the other distinguishers against the PIC with 1-bit partitions while it is the least efficient against the Atmel.

Next to these general observations, more specific comments can be made, *e.g.*:

– The results for the DoM test against the PIC (Figure 3) experimentally confirm the prediction of Messerges in [13]: in the context of "*all-or-nothing*" multiple-bit attacks using a DoM test, the best partition size is 3 bit out of 8 if the leakages have strong Hamming weight dependencies. It corresponds to the best tradeoff between the amplitude of the DoM peak (that increases with the size of the partition) and the number of traces that are not used by the test because not corresponding to an "*all zeroes/ones*" vector.

– The different metrics introduced, although correlated, bring different insights on the attacks efficiencies: Figures 5, 6 illustrate the guessing entropies of different attacks for our two devices; Figure 7 contains the 4^{th}-order success rates for the PIC. Interestingly, the variance test using 4-bit partitions against the Atmel allows a better 1^{st}-order success rate than guessing entropy, compared to other distinguishers (*e.g.* Figures 4, 6: middle right parts).
– The number of bins used to build the histograms in the MIA has a significant impact on the resulting attack efficiency. More bins generally allow a better estimation of the mutual information $\hat{\mathrm{I}}(\mathsf{R}(\mathbf{L}_q); \mathbf{H}^q_{s^*})$ but can lead to less discriminant attacks if the number of leakage samples is bounded. In general, we use as many bins as the number of vectors in our partitions. For the 4-bit partitions, we additionally considered 8-bins-based attacks to illustrate the impact of a change of this parameter. The optimal selection of these bins and their number is an interesting scope for further research.

This list of comments is of course not exhaustive and only points out exemplary facts that can be extracted from our experiments. We finally emphasize the importance of a sufficient statistical sampling in the approximation of the success rates or guessing entropy in order to provide meaningful conclusions. While an actual adversary only cares about recovering keys (*i.e.* one experiment may be enough for this purpose) the evaluation and understanding of side-channel attacks requires confidence in the analysis of different statistical tests. In practice, such evaluations are obviously limited by the amount of traces that one can acquire, store and process. With this respect, we computed our success rates and guessing entropies from sets of 1000 samples (*i.e.* 1000 leakage vectors of 250 encrypted plaintexts each). Both the smoothness of the curves in our figures and the confidence intervals that can be straightforwardly extracted for the success rates confirm that this sampling was enough to obtain sound observations.

8 Conclusions

This paper describes a fair empirical comparison of different side-channel distinguishers against two exemplary devices. Our results essentially highlight the implementation-dependent nature of such comparisons. It shows that any conclusion about the efficiency of a side-channel attack is only valid within a specific context. Therefore it emphasizes the importance of performing similar evaluations against various other implementations. In particular, countermeasures against side-channel attacks (*e.g.* masked [8] or dual-rail circuits [19]) are an interesting evaluation target. Other scopes for further research include the integration of more complex side-channel attacks in the comparisons, *e.g.* based on collisions [16] or the investigation of advanced statistical tools for key extraction, *e.g.* [1]. The methodology described in this work is expected to prevent wrong general claims on side-channel attacks and to allow a better understanding of both the target devices and the attacks used to exploit physical leakages.

References

1. L. Batina, B. Gierlichs, K. Lemke-Rust, *Comparative Evaluation of Rank Correlation based DPA on an AES Prototype Chip*, in the proceedings of ISC 2008, LNCS, vol 5222, pp 341-354, Taipei, Taiwan, September 2008.
2. E. Brier, C. Clavier, F. Olivier, *Correlation Power Analysis with a Leakage Model*, in the proceedings of CHES 2004, LNCS, vol 3156, pp 16-29, Boston, Massachussets, USA, August 2004.
3. S. Chari, J. Rao, P. Rohatgi, *Template Attacks*, in the proceedings of CHES 2002, LNCS, vol 2523, pp 13-28, Redwood Shores, CA, USA, August 2002.
4. J.S. Coron, D. Naccache, P. Kocher, *Statistics and Secret Leakage*, in the proceedings of FC 2000, LNCS, vol 1962, pp 157-173, Anguilla, British W. Indies, February 2000.
5. FIPS 197, *"Advanced Encryption Standard"*, Federal Information Processing Standard, NIST, U.S. Dept. of Commerce, November 26, 2001.
6. B. Gierlichs, L. Batina, P. Tuyls, B. Preneel *Mutual Information Analysis - A Generic Side-Channel Distinguisher*, in the proceedings of CHES 2008, LNCS, vol 5154, pp 426-442, Washington DC,USA, August 2008.
7. B. Gierlichs, K. Lemke, C. Paar, *Templates vs. Stochastic Methods*, in the proceedings of CHES 2006, LNCS, vol 4249, pp 15-29, Yokohama, Japan, October 2006.
8. L. Goubin, J. Patarin, *DES and Differential Power Analysis*, in the proceedings of CHES 1999, LNCS, vol 1717, pp 158-172, Worcester, MA, USA, August 1999.
9. P. Kocher, *Timing Attacks on Implementations of Diffie-Hellman, RSA, DSS and Other Systems*, in the proceedings of Crypto 1996, LNCS, vol 1109, pp 104-113, Santa-Barbara, CA, USA, August 1996.
10. P. Kocher, J. Jaffe, B. Jun, *Differential Power Analysis*, in the proceedings of Crypto 1999, LNCS, vol 1666, pp 398-412, Santa-Barbara, CA, USA, August 1999.
11. S. Mangard, E. Oswald, T. Popp, *Power Analysis Attacks*, Springer, 2007.
12. T.S. Messerges, E.A. Dabbish, R.H. Sloan, *Examining Smart-Card Security under the Threat of Power Analysis Attacks*, IEEE Transactions on Computers, vol 51, num 5, pp 541-552, May 2002.
13. T.S. Messerges, *Power Analysis Attacks and Countermeasures for Cryptographic Algorithms*, PhD Thesis, University of Illinois at Urbana Champaign, 2000.
14. E. Prouff, *DPA Attacks and S-Boxes* in the proceedings of FSE 2005, LNCS, vol 3557, pp 424-441, Paris, France, February 2005.
15. W. Schindler, K. Lemke, C. Paar, *A Stochastic Model for Differential Side-Channel Cryptanalysis*, in the proceedings of CHES 2005, LNCS, vol 3659, pp 30-46, Edinburgh, Scotland, September 2005.
16. K. Schramm, G. Leander, P. Felke, C. Paar, *A Collision-Attack on AES: Combining Side Channel and Differential Attack*, in the proceedings of CHES 2004, LNCS, vol 3156, pp 163-175, Cambridge, MA, USA, August 2004.
17. F.-X. Standaert, E. Peeters, F. Macé, J.-J. Quisquater, *Updates on the Security of FPGAs Against Power Analysis Attacks*, in the proceedings of ARC 2006, LNCS, vol 3985, pp 335-346, Delft, The Netherlands, March 2006.
18. F.-X. Standaert, T.G. Malkin, M. Yung, *A Unified Framework for the Analysis of Side-Channel Key Recovery Attacks*, Cryptology ePrint Archive, Report 2006/139.
19. K. Tiri, M. Akmal, I. Verbauwhede, *A Dynamic and Differential CMOS Logic with Signal Independent Power Consumption to Withstand DPA on Smart Cards*, in the proceedings of ESSCIRC 2003, Estoril, Portugal, September 2003.

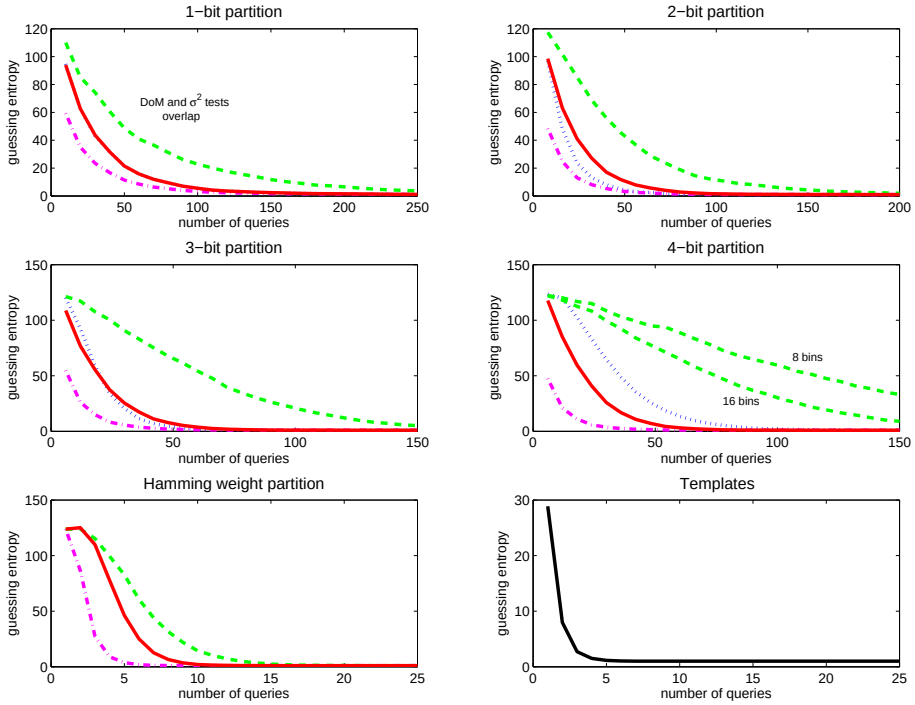

Fig. 5: PIC: Guessing entropy for different statistical tests, partitions, models (dotted: DoM test, dash-dotted: correlation test, dashed: MIA, solid: variance test).

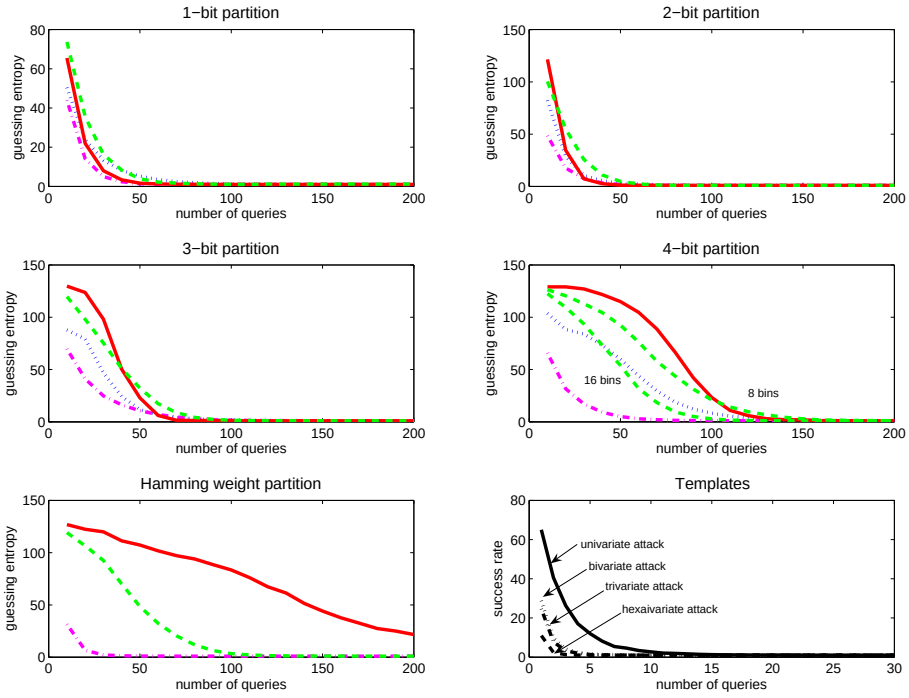

Fig. 6: Atmel: Guessing entropy for different statistical tests, partitions, models (dotted: DoM test, dash-dotted: correlation test, dashed: MIA, solid: variance test).

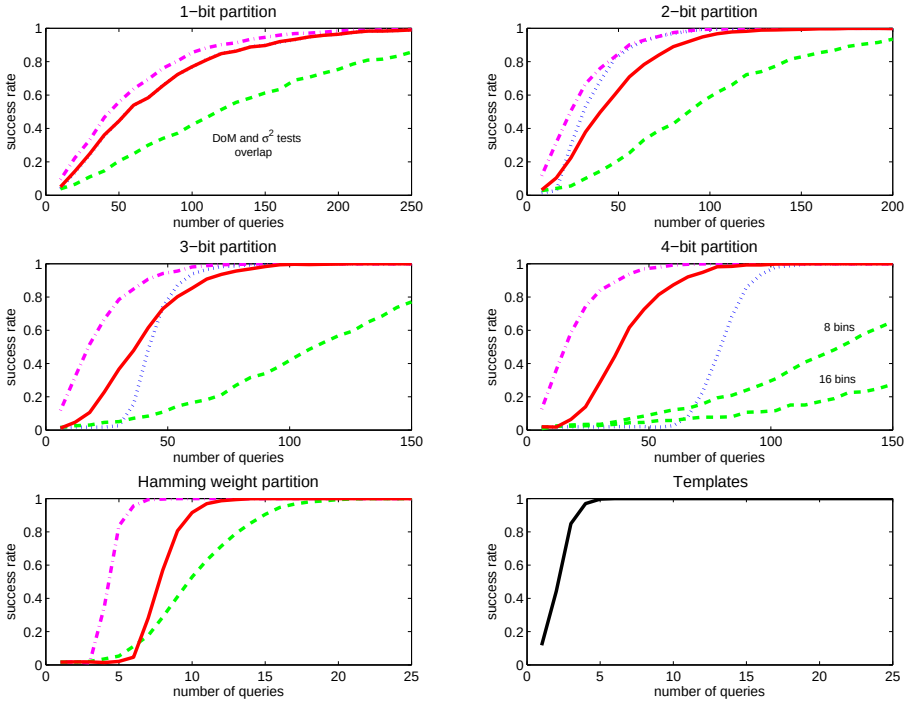

Fig. 7: PIC: 4^{th}-order SR for different statistical tests, partitions, models (dotted: DoM test, dash-dotted: correlation test, dashed: MIA, solid: variance test).