Combined Side-Channel Attacks

M. Abdelaziz Elaabid[1,2], Olivier Meynard[1,3],
Sylvain Guilley[1,4], and Jean-Luc Danger[1,4]

[1] Institut TELECOM / TELECOM ParisTech, CNRS LTCI (UMR 5141)
Département COMELEC, 46 rue Barrault, 75 634 Paris Cedex 13, France
[2] Université de Paris 8, Équipe MTII, Laga
2 rue de la liberté, 93 526 Saint-Denis Cedex, France
[3] DGA/MI (CELAR), La Roche Marguerite, 35 174 Bruz, France
[4] Secure-IC S.A.S., 37/39 rue Dareau, 75 014 Paris, France
{elaabid,meynard,guilley,danger}@TELECOM-ParisTech.fr

Abstract. The literature about side-channel attacks is very rich. Many side-channel distinguishers have been devised and studied; in the meantime, many different side-channels have been identified. Also, it has been underlined that the various samples garnered during the same acquisition can carry complementary information. In this context, there is an opportunity to study how to best combine many attacks with many leakages from different sources or using different samples from a single source. This problematic has been evoked as an open issue in recent articles. In this paper, we bring two concrete answers to the attacks combination problem. First of all, we experimentally show that two partitionings can be constructively combined. Then, we explore the richness of electromagnetic curves to combine several timing samples in such a way a sample-adaptative model attack yields better key recovery success rates than a mono-model attack using only a combination of samples (via a principal component analysis).

Keywords: Side-channel analysis; leakage models; attacks combination; multi-partitioning attacks; multi-modal leakage.

1 Introduction

Trusted computing platforms resort to secure components to conceal and manipulate sensitive data. Such components are in charge of implementing cryptographic protocols; for instance, the component is typically asked to encrypt the data with a cryptographic key. The secret key is protected against a direct readout from the circuit thanks to tamper-proof techniques. In general, the component is shielded by coatings to protect it from malevolent manipulations (active or passive micro-probing [1], modification, *etc.*). However, it has been noted that despite this protection, some externally measurable quantities can be exploited without touching the component. Typically, without special care, internal data are somehow modulating the computation timing, the instant current drawn from the power supply, and the radiated fields. Thus, those unintentional

Y. Chung and M. Yung (Eds.): WISA 2010, LNCS 6513, pp. 175–190, 2011.

physical emanations can be analyzed in a view to derive from them some sensitive information. Such analyses are referred to as side-channel attacks. The way the observed measurements are affected by the internal data is *a priori* unknown by the attacker, although in some cases an hypothetical, hence imperfect, physical model can be assumed. The link between the data and the side-channel is called the leakage model.

Most side-channel attacks start by a tentative partitioning of the measurements, indexed by key hypotheses [2]. Then, the adversary assesses the quality of each partitioning. This information is typically summarized by a figure of merit. This figure of merit can be a difference of means (in case there are only two partitions [3]), a correlation (case of the CPA [4]), a likelihood (case of template attacks [5]) or a mutual information (case of the MIA [6]), to cite only the few most widespread. Such figures of merit are often referred to as distinguishers, as they are able to successfully distinguish between the key candidates to select the correct one. The comparison of these distinguishers on the same acquisitions has been already discussed in some papers [7–11]. It appears that for a given partitioning, some distinguishers are better than the others to rank the correct key first, some other distinguishers are better than the others to optimize the average rank of the correct key [10]. Moreover, the conclusions depend on the target, since the leakage structure is inherent to each device. The definition of new distinguishers is an active research area; indeed, every new distinguisher contributes to feed a battery of attacks suitable to be launched in parallel on a device under test.

Another research direction is to attempt to make the most of the existing distinguishers. One interesting option is to constructively combine the wealth of cited attacks on common side-leakage traces. Another option consists in combining different samples of traces or even different traces acquired concomitantly.

The rest of the paper is structured as follows. The section 2 tackles the question of the multiple-partitioning attacks. The section 3 reports an original multi-sample electromagnetic (EM) trace, where the leakage model depends on the sample within the trace. We investigate attacks that could take advantage of this originally rich leakage and show that a combined attack indeed outperforms classical ones. The conclusions and the perspectives are in Sec. 4.

2 Combined Attacks and Metrics Based on Multiple Partitions

We explore in this section the combination of multiple partitionings on template attacks. Indeed, some "comparison" attacks that require a physical model of the leakage fail if the leakage function does not match enough the leaking modality of the device.

In [12], a framework is presented in order to evaluate the security of a cryptographic device. This approach relies on two different views: on the one hand the robustness of a circuit against a leakage function, and on the other the strength of an adversary. The information theory and specially the conditional entropy is chosen to quantify the information leaked during encryption. This very concept is thus promoted in order to measure the robustness. Indeed, the more the circuit is leaking the more it is vulnerable. The strength of the adversary is determined for example by its success rate to retrieve the encryption key.

2.1 Information Theoretic Metric

We adopt the idea that the quality of a circuit is assessed by the amount of information given by a leakage function. Thus, if S_K is the random variable representing the secret (ideally the key values), and $\mathbf{L}$ is the random variable representing the values of the leakage function.

The residual uncertainty on S_K knowing $\mathbf{L}$ is given by $\mathbf{H}(S_K \mid L)$. $\mathbf{H}$ is the conditional entropy introduced by Claude E. Shannon [12, 13]. Note that this value will depend on sensitive variables chosen, and thus the quality of the leakage function. The more the sensitive variable leaks, the smaller is the entropy and more vulnerable is the circuit.

2.2 Template Attacks

Template attacks are among the most powerful forms of side channel attacks. They are able to break implementations and countermeasures which assumes that the attacker cannot get more than a very small number of samples extracted from the attacked device. To this end, the adversary needs a hardware identical to the target, which allows him to obtain some information under the form of leakage realizations. The main step is to perform a modeling process; its goal is to build classes for side-channel traces that will help identify the secret values during the on-line phase of the attack. Said differently, the information provided by profiling are used to classify some part of encryption key. Actually, the full round key has obviously too many bits to be guessed in one go by exhaustive search. In general, the key bits at entering substitution boxes (sboxes) are targeted. In fact, they all contribute to activate the same logic, which explains why it is beneficial to guess them together. An adversary can also select other key bits if they are more vulnerable. In other words, the attacker itself selects the bits of the key best for his attack. Guessing the correct key is a problem of decision theory. To solve it, we introduce a statistical model that is directly applicable in principle to the problem of classification. This application is mainly based on Bayes' rule, which allows to evaluate an *a posteriori* probability (that is after the effective observation), knowing the conditional probability distributions *a priori* (*i.e.* independent of any constraint on observed variables). The maximum likelihood approach helps provide the most appropriate model.

2.2.1 Profiling Process

For this step, we need a set of traces $\mathcal{S}_o, o \in [0, N'[$ corresponding to each N' operation that are also values of the sensitive variable. Traces, denoted by t, are vectors of N dimensions related to random values of plaintext and keys needed to algorithm encryption. These observations are then classified according to functions of leakage $\mathcal{L}$. These leakage functions must depend on the configuration of the circuit, and of the implemented algorithm. This provides a framework for the estimation of the leakage during encryption. For each set $\mathcal{S}_o, o \in [0, N'[$ the attacker computes the average $\mu_o = \frac{1}{|\mathcal{S}_o|} \sum_{t \in \mathcal{S}_o} t$ and the covariance matrix $\Sigma_o = \frac{1}{|\mathcal{S}_o|-1} \sum_{t \in \mathcal{S}_o} (t - \mu_o)(t - \mu_o)^\mathsf{T}$. The ordered pair (μ_o, Σ_o) associated with value o of the leakage function outputs, is called *template* and will be used in the attack to retrieve subkeys. It allows to build the ideal probability density function (PDF) of a multivariate Gaussian distribution.

2.2.2 Principal Component(s) Analysis

One of the main contributions of the template attack is that an adversary may use all the information given by any trace. However, he is confronted with enormous data he has on hand, especially the covariance matrices. This poses some difficulties for calculations, since, because of algorithmic noise, large covariance matrices are poorly conditioned. For this purpose, the principal component analysis (PCA) is used to get round those drawbacks. It allows to analyze the structure of the covariance matrix (variability, dispersion of data). The aim of PCA is to reduce the data to $q \ll N$ new descriptors, that summarize a large part of (if not all) the variability. Also, it allows to better visualize the data in 2 or 3 dimensions (if $q = 2$ or 3).

These new descriptors are given by the data projection on the most significant eigenvectors given by PCA. Let EV be the matrix containing the eigenvectors classified according to the decreasing eigenvalues. The mean traces and covariance matrices are then expressed in this basis by: $p\mu_o = (EV)^\mathsf{T} \mu_o$ and $P\Sigma_o = (EV)^\mathsf{T} \Sigma_o (EV)$.

2.2.3 Online Attack and Success Rate

The *online attack* consists in first capturing one trace t of the target device during an encryption using the secret key κ. Knowing that each trace corresponds to one leakage value, the secret key will be retrieved from this trace by using maximum likelihood: $\kappa = \operatorname{argmax}_{s_{Kc}} Pr(s_{Kc} \mid t)$, where s_{Kc} is the candidate key. Indeed, for each key candidate, we estimate the value of leakage by using the message or the ciphertext that are *a priori* known. The success rate is given by the average number of times where the adversary succeeds to retrieve the key $s_{Kc} = \kappa$. For each attempt the adversary can use one trace corresponding to one query, or a set of traces corresponding to different queries.

2.3 Sensitive Variables

In the paper [13] a study is made on the choice of the best suited sensitive variable for an adversary attacking publicly available traces [14]. From a comparison between five different models, it is shown that the most appropriate model for the targeted circuit is the Hamming distance between two registers. However, "partitioning attacks" (in the sense of [2]) on various sensitive values (such as the linear and nonlinear functions inputs) also allows an adversary to recover the key, but with many more traces. The knowledge of circuit architecture provides definitely much more information about the main leakage function. In this article we elaborate by combining these models to retrieve the key with fewer traces, and watch the behavior of entropy as a function of the number of eigenvectors retained in the attack.

2.3.1 Combined Models

The goal is to combine two partitionings. The security of the resulting compound model is evaluated by template attacks; identically, the robustness of the circuit is measured under this new model. Can an adversary that combines models be considered as "higher order" [15]? Is he able to recover the secret key faster? The experiment described in this section attempts to address these issues. Let

1. **Model M1** be the value of the first round corresponding to the fanout of the first sbox. It is a 4-bit model, and
2. **Model M2** be the first bit transition of model M1. It is a mono-bit model, belonging to the general class of "Hamming distance" models.

From those two models, we derive a third one referred to as **Model M3**. M3 combines the 4-bit model M1 and the 1-bit model M2. In other words, M3 is considered as a "bit-field structure" where the value of the most significant bit (MSB) is the model M2. The others 4 bits correspond to the model M1. M3 is the concatenation of MA and M2, and we note M3 $\doteq$ (M1, M2). Hence M3 is a $4 + 1 = 5$ bit model, which means that M3 is based on 32 partitions. Said differently, the partitioning for M3 is equal to the Cartesian product of that of M1 and M2.

The fair comparison between the models is not a trivial operation. Typically, the number of templates for models M1, M2 and M3 differs. Basically, regarding the training (*i.e.* templates building) phase:

1. either the adversary has an equal number of traces by classes,
2. or the adversary has an equal number of traces for all the set of classes.

The choice will influence the success rate as we will see in the forthcoming experiment. The first case is the most realistic: it consists in saying that the precharacterization time is almost unbounded; the valuable asset being the traces taken on-line from the attacked device. We model this situation by taking the same number of traces for each partition. Therefore, in total, much less training traces are used for mono-partition models; but this really represents the case where models are evaluated with as identical conditions as possible. The second one reflects the case where the precharacterization cost is non-negligible. Under this assumption, the advantage of combined attacks is less clear, since the number of available traces to estimate each template gets lower. Thus, in a single-model attack, the greater accuracy of the templates will certainly compensate the loss of benefit conveyed by the combination.

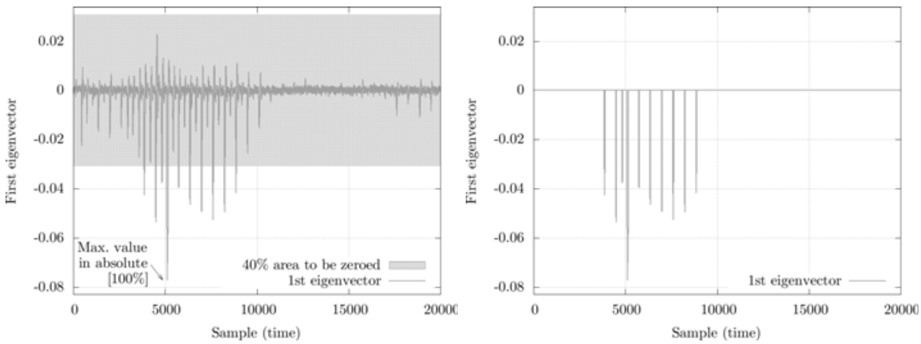


Fig. 1. Main eigenvector without thresholding (*left*), and the same with a 40% thresholding level (*right*)

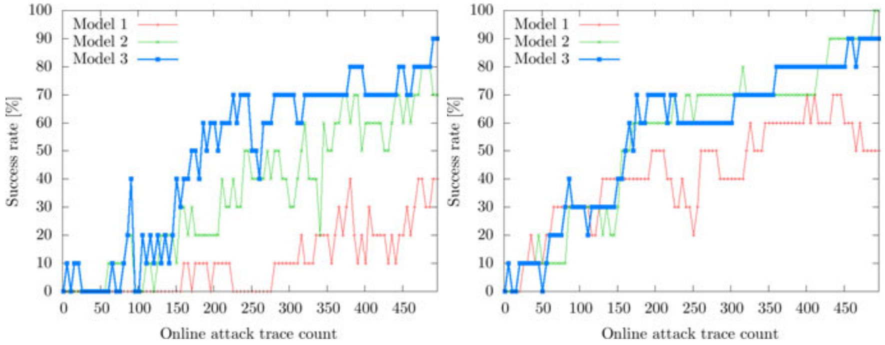
2.3.2 First Choice: Matching-Limited Evaluation

We use an equal number of traces per class. In our experiment we take 1,000 traces per class for models **M1**, **M2**, and **M3**. The comparison is made with and without the use of the thresholding method as presented in [13]. This method consists in accelerating the estimation of the principal directions in a PCA by forcing to zero the samples that are too small in the eigenvectors. The Fig. 1 illustrates the method. The idea is that most samples with low amplitude would actually be equal to zero with more traces in the estimation of the PCA. The thresholding allows to filter those samples out, so that they do not bring noise to the protection. In the same time, the thresholding keeps the samples with the greatest variance, which makes it a good tool to separate POIs from others. There is of course a trade-off in the choice for the best threshold. A too small threshold keeps too many irrelevant samples, whereas a too large threshold filters out even some weak POIs. For the implementation studied in this section, we found that a value of 40 % is a fair compromise. The figure 2 shows the success rate of the template attacks with the three models. We recall that the higher the success rate, the better the attack. We see in Fig. 2 that in the case of non-thresholding, the template attack based on the combined model is better than that on other models. It is much better than model **M1**, and slightly better than model **M2**.

Incidentally, when we resort to thresholding, the model M2 and M3 are equivalent and obviously always better than M1, that models in a less appropriate way the leakage function. The fact only the first PCA eigenvector is used in the comparison accounts for the equivalence between M2 and M3. Indeed, the other eigenvectors among the 31 possible in the case of combined model M3 also contain information, while the model M2 has only one significant direction.

2.3.3 Second Choice: Training-Limited Evaluation

If we follow the first option, we take 32,000 traces in general. Thus, for a constant number of traces per class, we have 32,000/16 = 2,000 traces by class for model **M1** and 32,000/2 = 16,000 traces by class for **M2**. The combined model **M3**



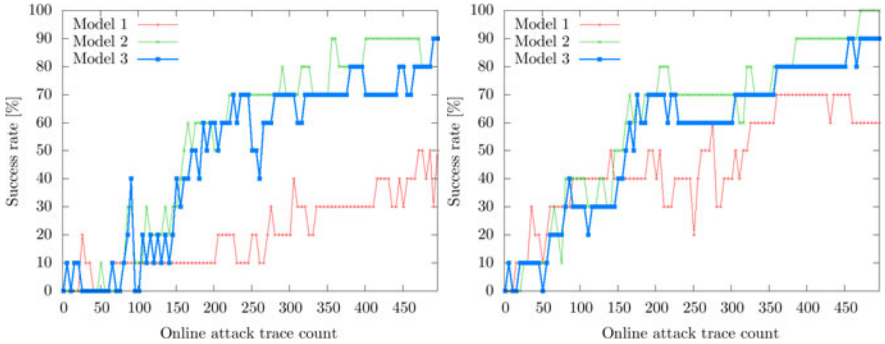
(a) Without threshold.

(b) With threshold at 40 %.

Fig. 2. Success rate comparison between mono-partitioning models M1, M2 and combined model M3 for two different thresholds and 1,000 traces per class

corresponds therefore to an amount of 32,000/32 = 1,000 traces by class. In this second case, we use systematically 32,000 for the training of all models M1, M2 and M3. As a consequence, model M2, that has the fewer number of partitions, will have its template evaluated more accurately than M1 and M3.

The two plots in Fig. 3 show that the models combination does not so much gain on the attack. Indeed, the success rate of model M3 is very close to the success rate of the model M1.



(a) Without threshold.

(b) With threshold at 40 %.

Fig. 3. Success rate comparison between mono-partitioning models M1, M2 and combined model M3 for two different thresholds and 32,000 traces in total for the training (to be divided between respectively 16, 2 and 32 classes)

2.4 Conditional Entropy

As explained above in Sec. 2.1, the conditional entropy gives an idea about the robustness of the circuit, irrespective of any attack. The value of the conditional entropy tends to a limit value in function to the number of traces used for profiling [13]. For our experiment, we took a large number of traces during the profiling phase to have an approximation of this limit value. This will help us compare the circuit robustness against attacks using models M1, M2 or M3. Is our circuit very vulnerable against an attacker who combines model? The figure 4 attempts to answer this question.

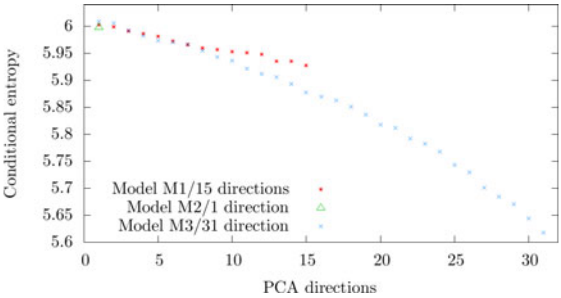


Fig. 4. Conditional entropy comparison between different models

The use of PCA provides new directions corresponding to different eigenvectors. The number of these directions depends on the cardinality of the sensitive variable. For example, in this study, we have 15 directions for the model M1, 1 direction for the model M2, and 31 directions for model M3. The first direction summarizes a large percentage of variance of data. Making a comparison of robustness using only this first direction may seem satisfactory, but this study shows that the more directions, the greatest the estimated leakage (*i.e.* the smallest the conditional entropy). Combined models are thus an opportunity to discover new leakage modes, as already noted for multi side-channel (power+EM) combination in [16]. This noting is actually a warning to the security evaluators: the robustness of an implementation can be underestimated if the models are either inappropriate (since incomplete, and thus should be completed with another or some other models) or contain too few partitions.

3 Combined Correlation Attacks

One difficulty for improving the side channel analysis or the template attack in presence of large noise is to identify the leaking samples, also called *Points Of Interest* (POIs). They correspond to the dates when the sensitive data is indeed processed and leaking the most. As already mentioned in the previous section when discussing the thresholding method, there is an obvious trade-off in the selection process for POIs. The more of them are selected, the more information is collected, but the more noise is kept. The difficult task consists in separating the signal from the noise.

Several techniques have been proposed to identify the POIs. The *Sum Of Squared pairwise (T-)Differences* (or sosd [17] and sost in [18]), the mutual information (MI [19]) and the *Principal Component Analysis* (PCA [20]) are four widespread examples. In this section, we study these methods and compare their efficiency, by applying them on two sets of measurements, one at short distance from the chip and another, one more noisy, at 25 cm from the chip. For these experiments we used a SASEBO-G board [21] embedding an AES hardware implementation. For these two sets of electromagnetic measurements $\mathbf{O}(t)$ we notice that a CPA can be successfully performed, by using the Hamming distance model between the penultimate and the last round state of the AES.

3.1 Techniques for Revealing the POIs

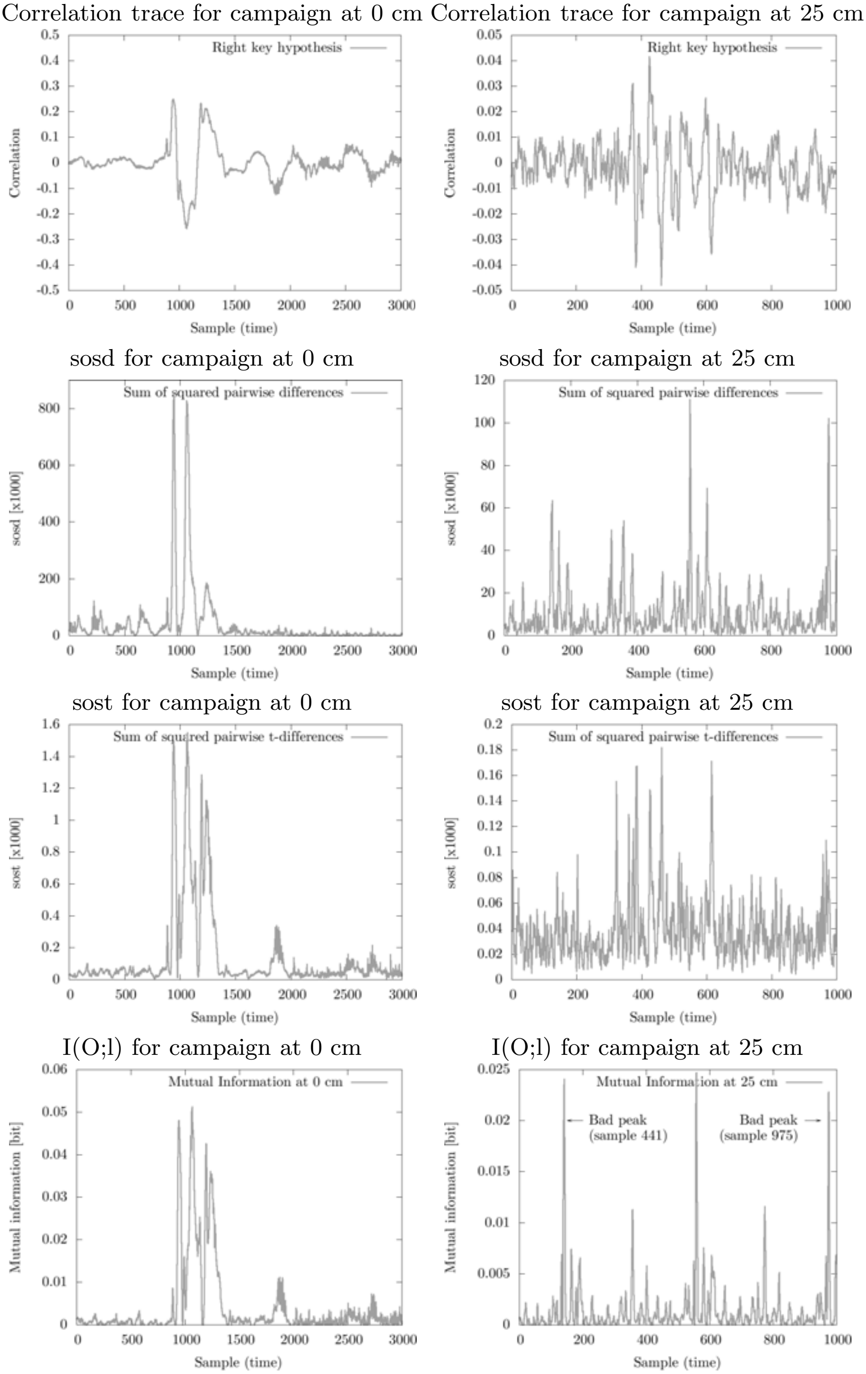
3.1.1 The sosd *versus* sost *versus* MI

The computation of the sosd leakage indicator metric requires to average the traces in a given partitioning. In the original proposal [17], the partitioning concerns all the 256 values of an AES state byte. The SASEBO-G implementation is known to leak the Hamming distance between the penultimate and the last round. Indeed, we succeed CPA for the both sets of measurements in this model. Therefore, we decide to restrict the values of the leakages to the interval $[0, 8]$, according to $\mathcal{L} = HW(\text{state}_9[sbox] \oplus \text{ciphertext}[sbox])$, where $sbox \in [0, 16[$ is the substitution box index. If we denote $o_i(t)$ all the samples (t) of the i^{th} realization of observation $\mathbf{O}(t)$, then the averages $\mu_j(t)$ in each class $j \in [0, 8]$ is given by the mean of set $\{o_i(t) \mid l_i = j\}$. Then their squared pairwise difference is summed up to yield the sosd.

The sost is based on the T-Test, which is a standard statistical tool to meet the challenge of distinguishing noisy signals. This method has the advantage to consider not only the difference between their means μ_j,$\mu_{j'}$ but as well their variability $(\sigma_j^2, \sigma_{j'}^2)$ in relation to the number of samples $(n_j, n_{j'})$. The definition of the sosd and sost is given below:

$$\text{sosd} \doteq \sum_{j,j'=0}^{8} (\mu_j - \mu_{j'})^2 \qquad \text{and} \qquad \text{sost} \doteq \sum_{j,j'=0}^{8} \left(\frac{\mu_j - \mu_{j'}}{\sqrt{\frac{\sigma_j^2}{n_j} + \frac{\sigma_{j'}^2}{n_{j'}}}} \right)^2 .$$

The sosd and the sost for the two EM observation campaigns are plotted in Fig. 5. We notice that the correlation trace, the sosd and sost curves are matching for the measurement at 0 cm. But, although we use for the partitioning the same leakage function $\mathcal{L}$ and although we find the right key with a CPA on the measurement at 25 cm, the sosd curve does not highlight the right time sample, *i.e.* that where the key can be retrieved by CPA. This figure 5 shows that the sosd metric is not always an efficient metric for revealing the points of interest. Indeed, we have tried to execute CPAs on the samples highlighted, but they all fail. Regarding the sost on the measurement at 25 cm, several POIs are revealed among samples that are not related to the secret data. Thus sost is neither a trustworthy tool to identify POIs.


Correlation trace for campaign at 0 cm Correlation trace for campaign at 25 cm

sosd for campaign at 0 cm

sosd for campaign at 25 cm

sost for campaign at 0 cm

sost for campaign at 25 cm

I(O;l) for campaign at 0 cm

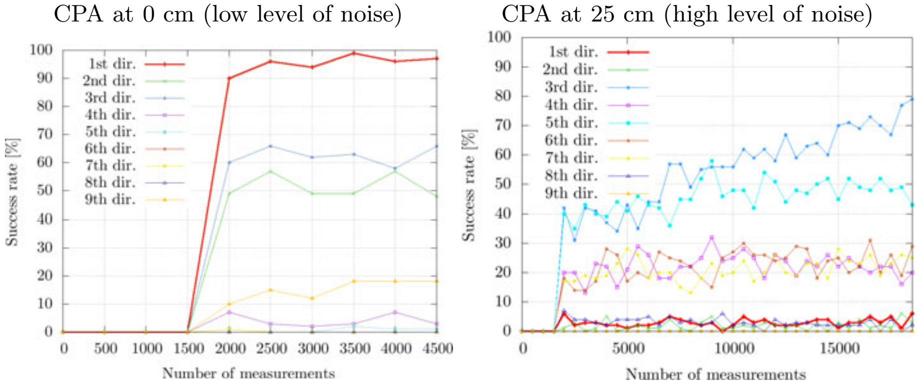
I(O;l) for campaign at 25 cm

Fig. 5. Correlation traces, sosd, sost and MI obtained for the right key hypothesis

Regarding the MI, also plotted in Fig. 5, it matches well the sost at short distances, but features peaks with no information (notably the samples 441 and 975). It is thus not a reliable tool. The principal reason is that the PDFs are poorly estimated in the presence of large amounts of noise.

3.1.2 The PCA

As previously explained in section 2.2.2, the PCA aims at providing a new description of the measurements by projection on the most significant eigenvector(s) of the empirical covariance matrix of (μ_j). If we compare the success rate of the CPA, applied after a PCA, we can notice, that in the case of the campaign at distance, featuring a high level of noise, the eigenvector corresponding to the greatest eigenvalue is not necessarily suitable. The success rate of the CPA after a projection onto each of the nine eigenvectors is given in Fig. 6. At 25 cm, we notice that the projection onto the first eigenvector is not necessarily the most suitable, since it does not yield the best attack success rate. The projection onto the third eigenvector turns out, quite surprisingly, to be more efficient. At the opposite, when the noise level is low and the electromagnetic probe set at short distance, the projection onto the first vector is indeed more efficient.


CPA at 0 cm (low level of noise)

CPA at 25 cm (high level of noise)

Fig. 6. Success rate of the CPA after PCA pre-processing

This phenomena can be explained by the fact that the number of curves in the sub-set corresponding to the Hamming distances 0 and 8 are in same proportion, nevertheless the level of noise is higher, since they contain the fewest number of traces. Indeed, the proportion of traces available for the training is equal to $\frac{1}{2^8} \cdot \binom{8}{l}$, which is lowest for $l = 0$ or 8. The estimation of those classes is thus less accurate.

In order to improve the PCA, we have reduced the number of partitions from 9 to 7 sub-sets depending on the Hamming distance $HD \in [1, 7] = [0, 8]\backslash\{0, 8\}$. We observe that, under this restriction, the best success rate is obtained for the projection on the first eigenvector. In the meantime, the condition number of the empirical covariance matrix decreases, which confirms that the weakly populated classes $l \in \{0, 8\}$ added more noise than signal to the PCA. Amazingly enough, this approach is antinomic with the multi-bit DPA of Messerges [22]. If we transpose from DES to AES, Messerges suggests at the opposite to get rid of the classes $l = [1, 7]$ and to retain only $l = \{0, 8\}$. Those extremal samples have two ambivalent properties. They convey the most information, as shown in Tab. 1, but also are the rarest samples, and thus are the most noisy coefficient in the covariance matrix. As Messerges does not make use of extra-diagonal coefficients, his attack is not concerned by this fact.

Table 1. Information and probability of the Hamming weight of an 8-bit uniformly distributed random variable

Class index l	0	1	2	3	4	5	6	7	8
Information [bit]	8.00	5.00	3.19	2.19	1.87	2.19	3.19	5.00	8.00
Probability [%]	0.4	3.1	10.9	21.9	27.3	21.9	10.9	3.1	0.4

3.2 Combining Time Samples

3.2.1 Observations

The correlation trace obtained for the right key with measurements at distance is given in Fig. 7. We observe that the correlation traces are extremely noisy. Moreover for some time samples, identified in as Sample{1,2,3,4} in Fig. 7, the magnitude of the correlation trace obtained for the right key is clearly higher than the magnitude of the correlation traces for bad key hypotheses. These samples are all located within the same clock period that corresponds to the last round of the AES. At the four identified dates, the sample are undoubtedly carrying secret information.

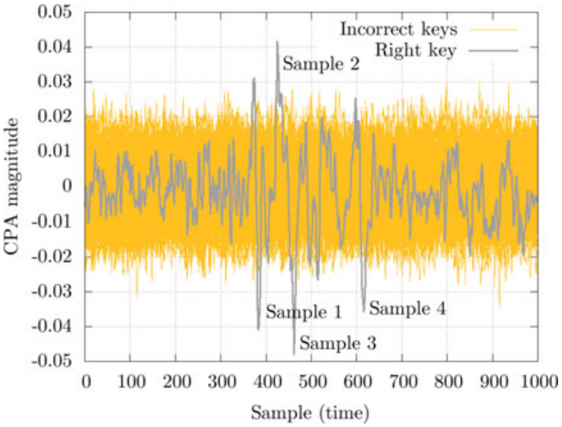


Fig. 7. Correlation traces obtained for the right key hypotheses and for incorrect key hypotheses at 25 cm

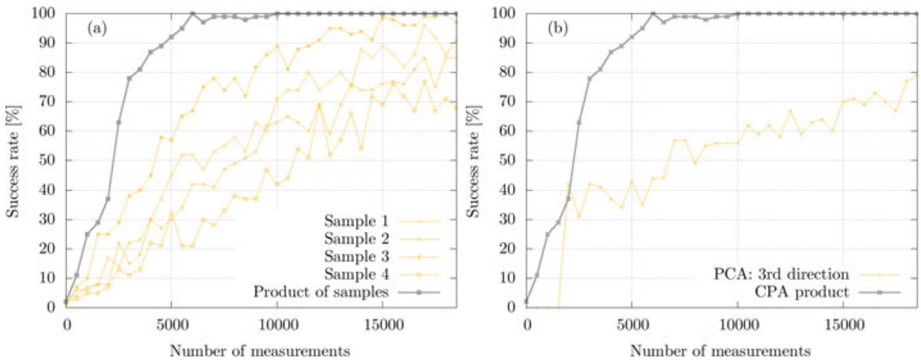


Fig. 8. (a)–*left*: Success rate of the mono-sample attack, and product of correlations attack; (b)–*right*: Comparison between a CPA using the pre-treatment by PCA and our product of correlation, introduced in Eqn. (1)

3.2.2 Sample Combination Principle and Results

We aim at showing that there is a gain in combining the leaks from the four identified dates. First of all, we confirm that the four samples of peak CPA are actually POIs. To do so, we perform successful CPAs at these time samples. The result is shown in Fig. 8: all four attacks pass over a success rate of 50 % after 12,000 traces. Second, we devise a method to attack that exploits at once all those samples. Similar methods have already be introduced in the context of combining samples in order defeat masking countermeasures [23]. In [24], Chari *et al.* suggest to use the product of two leakage models. In [25], Joye *et al.* recommend to combine two samples with the absolute value of the difference. As in our case we intend to combine more than two samples, we resort to the product for the combination function. We apply it to Pearson empirical correlation coefficients $\hat{\rho}_t$, where t are the four identified dates. The new distinguisher we promote is thus:

$$\hat{\rho}_{\text{combined}} \doteq \prod_{t \in \text{Sample}\{1,2,3,4\}} \hat{\rho}_t. \qquad (1)$$

This technique applies well to the Pearson correlation coefficients, that are already centered by design. Thus it indeed puts forward the simultaneous coincidences of high correlation, while it demotes incorrect hypotheses for which at least one $\hat{\rho}_t$ is close to zero. As shown in Fig. 8(a), the success rate of this new attack is greater than that for mono-samples attacks. Additionally, we confirm in Fig. 8(b) that our combination defined in Eqn. (1), although simple in its setup, clearly outperforms a PCA after performing PCA.

However, we have only shown that when knowing some POIs in the curve, a powerful combining multi-sample attack can be devised. Now, for the time being, the only method to exhibit those POIs has been to apply a successful attack (a

CPA in our case). Therefore, an open question is to locate those POIs without knowing the key beforehand or without conducting another less powerful attack. We suggest two solutions to spot the POIs: either online or by precharacterization on an open sample assuming the position of the POIs do not depend on the secret key.

4 Conclusion and Perspectives

In this paper, we have studied two examples of side-channel attacks combinations. The first contribution is the demonstration of a constructive multi-partitioning attack. We show that two partitioning can enhance the convergence of the success rate to one hundred percent; such attacks benefit from an exhaustive pre-characterization, since the number of templates increases, and that the training phase length is the product of the training phase for each partitioning. The second contribution is to highlight the existence of the leakage model in far field EM signals. We show how the leakage of each sample can be combined better than usual leakage reduction methods (*e.g.* the sosd, the sost or the PCA). This improvement comes from the fact each sample features a leakage of different nature that can be exploited individually, which is out of the reach of global techniques that consist in identifying points with large variation. Our improved combining distinguisher consists in multiplying the Pearson correlation coefficients for several POIs. Although this attack leads to better success rates than other attacks using different state-of-the-art pre-processing, we do think it can still be enhanced by another method to identify the points of interest accurately even when the side-channel observations are extremely noisy. As a perspective, we intend to apply those ideas to an online only attack, typically the MIA.

References

1. Gammel, B.M., Mangard, S.: On the duality of probing and fault attacks. Cryptology ePrint Archive, Report 2009/352 (2009), `http://eprint.iacr.org/`
2. 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.) ICISC 2008. LNCS, vol. 5461, pp. 253–267. Springer, Heidelberg (2009)
3. Kocher, P.C., Jaffe, J., Jun, B.: Differential Power Analysis. In: Wiener, M. (ed.) CRYPTO 1999. LNCS, vol. 1666, pp. 388–397. Springer, Heidelberg (1999)
4. Brier, É., Clavier, C., Olivier, F.: Correlation Power Analysis with a Leakage Model. In: Joye, M., Quisquater, J.-J. (eds.) CHES 2004. LNCS, vol. 3156, pp. 16–29. Springer, Heidelberg (2004)
5. Chari, S., Rao, J.R., Rohatgi, P.: Template Attacks. In: Kaliski Jr., B.S., Koç, Ç.K., Paar, C. (eds.) CHES 2002. LNCS, vol. 2523, pp. 13–28. Springer, Heidelberg (2003)

6. Gierlichs, B., Batina, L., Tuyls, P., Preneel, B.: Mutual information analysis. In: Oswald, E., Rohatgi, P. (eds.) CHES 2008. LNCS, vol. 5154, pp. 426–442. Springer, Heidelberg (2008)
7. Coron, J.S., Kocher, P.C., Naccache, D.: Statistics and Secret Leakage. In: Frankel, Y. (ed.) FC 2000. LNCS, vol. 1962, pp. 157–173. Springer, Heidelberg (2001)
8. Mangard, S., Oswald, E., Popp, T.: Power Analysis Attacks: Revealing the Secrets of Smart Cards, p. 338. Springer, Heidelberg (2006), `http://www.springer.com/`
9. Le, T.H., Canovas, C., Clédière, J.: An overview of side channel analysis attacks. In: ASIACCS, ASIAN ACM Symposium on Information, Computer and Communications Security, Tokyo, Japan, pp. 33–43 (2008), doi:10.1145/1368310.1368319
10. Gierlichs, B., De Mulder, E., Preneel, B., Verbauwhede, I.: Empirical comparison of side channel analysis distinguishers on DES in hardware. In: IEEE (ed.) ECCTD. European Conference on Circuit Theory and Design, Antalya, Turkey, pp. 391–394 (2009)
11. Veyrat-Charvillon, N., Standaert, F.X.: Mutual Information Analysis: How, When and Why? In: Clavier, C., Gaj, K. (eds.) CHES 2009. LNCS, vol. 5747, pp. 429–443. Springer, Heidelberg (2009)
12. Standaert, F.X., Malkin, T., Yung, M.: A Unified Framework for the Analysis of Side-Channel Key Recovery Attacks. In: Joux, A. (ed.) EUROCRYPT 2009. LNCS, vol. 5479, pp. 443–461. Springer, Heidelberg (2009)
13. Elaabid, M.A., Guilley, S.: Practical improvements of profiled side-channel attacks on a hardware crypto-accelerator. In: Bernstein, D.J., Lange, T. (eds.) AFRICACRYPT 2010. LNCS, vol. 6055, pp. 243–260. Springer, Heidelberg (2010), doi:10.1007/978-3-642-12678-9_15
14. TELECOM ParisTech SEN research group: DPA Contest, 1st edn. (2008-2009), `http://www.DPAcontest.org/`
15. Messerges, T.S.: Using Second-Order Power Analysis to Attack DPA Resistant Software. In: Paar, C., Koç, Ç.K. (eds.) CHES 2000. LNCS, vol. 1965, pp. 238–251. Springer, Heidelberg (2000)
16. Standaert, F.X., Archambeau, C.: Using Subspace-Based Template Attacks to Compare and Combine Power and Electromagnetic Information Leakages. In: Oswald, E., Rohatgi, P. (eds.) CHES 2008. LNCS, vol. 5154, pp. 411–425. Springer, Heidelberg (2008)
17. Gierlichs, B., Batina, L., Tuyls, P., Preneel, B.: Mutual information analysis. In: Oswald, E., Rohatgi, P. (eds.) CHES 2008. LNCS, vol. 5154, pp. 426–442. Springer, Heidelberg (2008)
18. Gierlichs, B., Lemke-Rust, K., Paar, C.: Templates vs. stochastic methods. In: Goubin, L., Matsui, M. (eds.) CHES 2006. LNCS, vol. 4249, pp. 15–29. Springer, Heidelberg (2006)
19. Macé, F., Standaert, F.X., Quisquater, J.J.: Information theoretic evaluation of side-channel resistant logic styles. In: Paillier, P., Verbauwhede, I. (eds.) CHES 2007. LNCS, vol. 4727, pp. 427–442. Springer, Heidelberg (2007)
20. Archambeau, C., Peeters, É., Standaert, F.X., Quisquater, J.J.: Template Attacks in Principal Subspaces. In: Goubin, L., Matsui, M. (eds.) CHES 2006. LNCS, vol. 4249, pp. 1–14. Springer, Heidelberg (2006)
21. Satoh, A.: (Side-channel Attack Standard Evaluation Board, SASEBO) Project of the AIST – RCIS (Research Center for Information Security), `http://www.rcis.aist.go.jp/special/SASEBO/`

22. Messerges, T.S., Dabbish, E.A., Sloan, R.H.: Investigations of Power Analysis Attacks on Smartcards. In: USENIX — Smartcard 1999, Chicago, Illinois, USA, pp. 151–162 (1999)
`http://www.usenix.org/publications/library/proceedings/smartcard99/messerges.html`
23. Prouff, E., Rivain, M., Bevan, R.: Statistical Analysis of Second Order Differential Power Analysis. IEEE Trans. Computers 58, 799–811 (2009)
24. Chari, S., Jutla, C.S., Rao, J.R., Rohatgi, P.: Towards Sound Approaches to Counteract Power-Analysis Attacks. In: Wiener, M. (ed.) CRYPTO 1999. LNCS, vol. 1666, p. 398. Springer, Heidelberg (1999)
25. Joye, M., Paillier, P., Schoenmakers, B.: On Second-Order Differential Power Analysis. In: Rao, J.R., Sunar, B. (eds.) CHES 2005. LNCS, vol. 3659, pp. 293–308. Springer, Heidelberg (2005)