Using Subspace-Based Template Attacks to Compare and Combine Power and Electromagnetic Information Leakages

F.-X. Standaert[1⋆] C. Archambeau[2]

[1] UCL Crypto Group, Université catholique de Louvain, [2] Centre for Computational Statistics and Machine Learning, University College London.

fstandae@uclouvain.be, c.archambeau@cs.ucl.ac.uk

Abstract. The power consumption and electromagnetic radiation are among the most extensively used side-channels for analyzing physically observable cryptographic devices. This paper tackles three important questions in this respect. First, we compare the effectiveness of these two side-channels. We investigate the common belief that electromagnetic leakages lead to more powerful attacks than their power consumption counterpart. Second we study the best combination of the power and electromagnetic leakages. A quantified analysis based on sound information theoretic and security metrics is provided for these purposes. Third, we evaluate the effectiveness of two data dimensionality reduction techniques for constructing subspace-based template attacks. Selecting automatically the meaningful time samples in side-channel leakage traces is an important problem in the application of template attacks and it usually relies on heuristics. We show how classical statistical tools such as Principal Component Analysis and Fisher Linear Discriminant Analysis can be used for efficiently preprocessing the leakage traces.

1 Introduction

Power Analysis Attacks have been introduced in the late nineties as a powerful cryptanalytic technique to extract secret data from cryptographic hardware devices [12, 13]. Shortly after, the ElectroMagnetic (EM) radiation of a chip appeared as an alternative source of physical leakages [8, 16]. Since the publication of these seminal papers, different lines of research have been followed, mainly ranging between attempts to prevent and counteract side-channel attacks and attempts to develop their understanding and discuss their optimality.

For example, Template Attacks (TAs) were introduced in [7] as the most powerful type of side-channel attack from an information theoretic point of view. TAs assume a probabilistic noise model for the leakages and use maximum likelihood as a similarity measure between actual leakage traces and their key-dependent predictions. Because of computational restrictions, TAs usually rely on heuristics in order to determine the leakage samples for which a model will be estimated. A more systematic approach is to use dimensionality reduction techniques for

⋆ Postdoctoral researcher of the Belgian Fund for Scientific Research (FNRS).

this purpose. Assuming that the information content of a leakage trace resides mainly at the time instants of maximum variability, Principal Subspace Template Attacks (PSTAs) have been introduced in [4]. Principal Component Analysis (PCA) [11] is then used to find the linear transformation that maximizes the inter-class distance when projecting the data into a low-dimensional subspace. Finally, Multi-Channel Attacks (MCAs) [2] exploit similar statistical techniques as TAs, but utilize simultaneously several side-channels like power and EM.

In parallel to these algorithmic advances, a number of practical experiments have been conducted and underlined the advantages of EM attacks compared to power analysis. For example, the ability to carefully position a small probe in the near field of a chip and to defeat certain countermeasures against power analysis attacks has been exhibited [1]. More recently, the exploitation of better leakage models than the usual Hamming weight/distance ones by monitoring the EM field sign has been detailed in [14]. Such improved models involve the improvement of non-profiled side-channel attacks, *e.g.* using the correlation coefficient [6]. They suggest that the EM leakage of a chip generally provides the adversary with more information than the power consumption of the same chip.

In this paper, we first intend to take both the comparison of the power and EM leakages and their combination one step further. For this purpose, we use different types of template attacks to evaluate the information theoretic and security metrics introduced in [18]. We apply these tools and metrics to an exemplary leaking implementation and bring a quantified confirmation of the previous intuitions: when accessible, near-field EM measurements provide more information than power leakages. Our results also demonstrate that the real power and EM channels are significantly more informative than their idealized models based on Hamming weights/distances (or even signed distances [14]). Similarly, we confirm the advantages of a multi-channel approach in which one channel is used to correct/improve the weaknesses of the other one.

In addition, we take advantage of the available power and EM measurements to compare different data dimensionality reduction techniques. PSTAs are very powerful in practice since a "small sample size" PCA can be applied in cases where there are much more time samples in the traces than key classes in the attack. But PSTAs only maximize the variance between average traces (each trace corresponding to a key class candidate) without considering the intra-class scatter. Hence, the resulting subspace might be suboptimal. Fisher's Linear Discriminant Analysis (LDA) [10] is more appropriate with this respect. LDA seeks the subspace that maximizes the ratio between inter-class and intra-class scatter. It finds the subpace in which the average traces are maximally separated, while minimizing the spread of the individual traces within their respective classes.

The rest of this paper is structured as follows. Section 2 describes our target implementation. Section 3 defines the evaluation metrics for the analysis of our experiments. Section 4 describes the PCA- and LDA-based template attacks. Experimental results are presented in Section 5 and conclusions are in Section 6.

2 Target implementation

Our target device for the following experiments is a PIC 16F877 8-bit RISC-based microprocessor. Our measurements exploit the setup described in [14]. The microchip was clocked at a frequency around 4 MHz. We monitored the power consumption by inserting a small resistor at the ground pin of the device. The value of the resistor was chosen so that it disrupts the voltage supply by at most 5% of its reference. In order to get accurate near field EM measurements, the chip was depackaged following the guidelines given in [3]. We monitored the EM leakages with a small hand-made loop probe that was soldered on a coax mounted on a SMA connector. The signal was then amplified with a large band, low noise pre-amplifier and sampled with a 1 GHz bandwidth oscilloscope.

Since the primary goal of this paper is to evaluate the effectiveness of the power consumption and EM side-channels and to discuss their relation with different leakage models, we did not directly target a cryptographic algorithm. Rather, we programmed the microchip so that it processed all the possible 4-bit transitions in order to determine the extent to which these transitions could be recovered through physical observations. That is, we targeted the 256 possible transitions between two 4-bit values x_1 and x_2 on the bus. Considering 4-bit (rather than 8-bit) transitions was justified by the need to keep a reasonable number of transitions (hence, templates) under investigation. In practice, recovering a transition through side-channel measurements can be straightforwardly exploited *e.g.* to recover the secret key of a block cipher. This is similar to recovering, *e.g.* the Hamming weight of a key dependent intermediate value after the application of a substitution box. Consequently and for simplicity, we will denote each 4-bit transition as a a key class s in the rest of the paper.

3 Evaluation metrics

Following the framework introduced in [18], we will evaluate our different experiments with a combination of information theoretic and security metrics.

Information theoretic metric. Let S be a discrete random variable indicating the target key class associated to a side-channel attack and s be a realization of this variable corresponding to one particular transition $x_1 \rightarrow x_2$ on the bus. Let $\mathbf{L}_q$ be a random vector denoting the side-channel observations generated with q queries to the target physical computer and $\mathbf{l}_q = [l_1, l_2, \ldots, l_q]$ be a realization of this random vector. Let finally $\Pr[s|\mathbf{l}_q]$ be the conditional probability of a key class s given a leakage $\mathbf{l}_q$. We first define a conditional entropy matrix as:

$$\mathbf{H}^q_{s,s^*} = -\sum_{\mathbf{l}_q} \Pr[\mathbf{l}_q|s] \cdot \log_2 \Pr[s^*|\mathbf{l}_q], \tag{1}$$

where s^* denotes a possible key class candidate in the attack. Second, we derive Shannon's conditional entropy as follows:

$$\mathrm{H}[S|\mathbf{L}_q] = -\sum_{s} \Pr[s] \sum_{\mathbf{l}_q} \Pr[\mathbf{l}_q|s] \cdot \log_2 \Pr[s|\mathbf{l}_q] = \mathop{\mathbf{E}}_{s}\ \mathbf{H}^q_{s,s},$$

where $\mathbf{E}$ denotes the mathematical expectation and $\Pr[s|\mathbf{l}_q]$ is derived from the Bayes law. We note that this definition is equivalent to the classical one since:

$$\begin{aligned}\mathrm{H}[S|\mathbf{L}_q] &= -\sum_{\mathbf{l}_q} \Pr[\mathbf{l}_q] \sum_s \Pr[s|\mathbf{l}_q] \cdot \log_2 \Pr[s|\mathbf{l}_q] \\ &= -\sum_s \Pr[s] \sum_{\mathbf{l}_q} \Pr[\mathbf{l}_q|s] \cdot \log_2 \Pr[s|\mathbf{l}_q]\end{aligned}$$

Then, we define an entropy reduction matrix: $\widetilde{\mathbf{H}}^q_{s,s^*} = \mathrm{H}[S] - \mathbf{H}^q_{s,s^*}$, where $\mathrm{H}[S]$ is the entropy of the key class variable S before any side-channel attack has been performed: $\mathrm{H}[S] = \mathbf{E}_s - \log_2 \Pr[s]$. It directly yields the mutual information:

$$\mathrm{I}(S; \mathbf{L}_q) = \mathrm{H}[S] - \mathrm{H}[S|\mathbf{L}_q] = \mathop{\mathbf{E}}_s \widetilde{\mathbf{H}}^q_{s,s} \tag{2}$$

Security metric. We consider a side-channel key recovery adversary of which the aim is to guess a key class s with non negligible probability. For this purpose and for each candidate s^*, he compares the actual observation of a leaking device $\mathbf{l}_q$ with some key dependent model for these leakages $\mathsf{M}(s^*, .)$. The construction of these models (otherwise said templates) will be detailed in the next section. Let $\mathsf{T}(\mathbf{l}_q, \mathsf{M}(s^*, .))$ be the statistical test used in the comparison. We assume that the highest value of the statistic corresponds to the most likely key candidate. For each observation $\mathbf{l}_q$, we store the result of the statistical test T in a vector $\mathbf{g}_q = \mathsf{T}(\mathbf{l}_q, \mathsf{M}(s^*, .))$ containing the key candidates sorted according to their likelihood: $\mathbf{g}_q := [g_1, g_2, \ldots, g_{|\mathcal{S}|}]$ (*e.g.* in our present context $|\mathcal{S}|$=256). Then, for any side-channel attack exploiting a leakage vector $\mathbf{l}_q$ and giving rise to a result $\mathbf{g}_q$, 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 directly leads to the o^{th}-order success rate:

$$\mathbf{Succ}^o_S = \mathop{\mathbf{E}}_s \mathop{\mathbf{E}}_{\mathbf{l}_q} \mathsf{S}^o_s(\mathbf{g}_q) \tag{3}$$

Intuitively, a success rate of order 1 (*resp.* 2, ...) relates to the probability that the correct key is sorted first (*resp.* among the two first ones, ...) by the adversary. From a theoretical point of view, the information theoretic metric is purposed for the comparison of different implementations. It is therefore convenient to compare different side-channels and is central in the present analysis. A security metric is additionally provided for discussion (but we do not consider all the success rate orders and the guessing entropy defined in [18]).

4 Statistical tools

In this section, we present the different tools that will be used to extract the side-channel information from the actual observations of our target device. We first describe a classical template attack as it is the strongest form of side-channel attack from an information theoretic point of view. Hence, it provides the best way to evaluate the information theoretic metric of the previous section. Then we discuss how to solve the main practical problem that arises in the application of template attacks, namely the automatic selection of meaningful samples in a leakage trace. For this purpose, we propose to use PCA or LDA.

4.1 Template attacks

Templates construction. Suppose that an adversary is provided with N_t leakage vectors for a given operation, *e.g.* the transition between two values $x_1 \rightarrow x_2$ on the bus of a microchip represented by a key class s. In template attacks, a multivariate Gaussian noise model is generally considered, which means that these vectors $\{\mathbf{l}_q^{s,i}\}_{i=1}^{N_t}$ are assumed to be drawn from the multivariate distribution:

$$\mathcal{N}(\mathbf{l}_q^{s,i}|\boldsymbol{\mu}_s, \boldsymbol{\Sigma}_s) = \frac{1}{(2\pi)^{\frac{N}{2}}|\boldsymbol{\Sigma}_s|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(\mathbf{l}_q^{s,i} - \boldsymbol{\mu}_s)^\top \boldsymbol{\Sigma}_s^{-1}(\mathbf{l}_q^{s,i} - \boldsymbol{\mu}_s)\right\},$$

where the mean $\boldsymbol{\mu}_s$ and the covariance matrix $\boldsymbol{\Sigma}_s$ specify completely the noise distribution associated to each key class s. Constructing the templates consists then in estimating the sets of parameters $\{\boldsymbol{\mu}_s\}_{s=1}^{|\mathcal{S}|}$ and $\{\boldsymbol{\Sigma}_s\}_{s=1}^{|\mathcal{S}|}$. A standard approach is to use the empirical mean and covariance matrix associated to the observations $\{\mathbf{l}_q^{s,i}\}_{i=1}^{N_t}$: $\hat{\boldsymbol{\mu}}_s = \frac{1}{N_t}\sum_{i=1}^{N_t} \mathbf{l}_q^{s,i}$, $\widehat{\boldsymbol{\Sigma}}_s = \frac{1}{N_t}\sum_{i=1}^{N_t}(\mathbf{l}_q^{s,i} - \hat{\boldsymbol{\mu}}_s)(\mathbf{l}_q^{s,i} - \hat{\boldsymbol{\mu}}_s)^\top$.

Attack. Assume now that there are $|\mathcal{S}|$ possible secret key classes. In order to determine by which secret signal a new vector $\mathbf{l}_{\text{new}}$ was generated, we apply Bayes' rule. This leads to the following classification rule:

$$\tilde{s} = \underset{s^*}{\text{argmax}}\ \hat{\Pr}[s^*|\mathbf{l}_{\text{new}}] = \underset{s^*}{\text{argmax}}\ \hat{\Pr}[\mathbf{l}_{\text{new}}|s^*]\Pr[s^*], \tag{4}$$

where $\hat{\Pr}[\mathbf{l}_{\text{new}}|s^*] = \mathcal{N}(\mathbf{l}_{\text{new}}|\hat{\boldsymbol{\mu}}_{s^*}, \widehat{\boldsymbol{\Sigma}}_{s^*})$ and $\Pr[s^*]$ is the prior probability of the class candidate s^*. The classification rule assigns $\mathbf{l}_{\text{new}}$ to the candidate s^* with the highest posterior probability. In general, we have $\Pr[s^*] = \frac{1}{|\mathcal{S}|}$.

Limitations. Although template attacks are theoretically the strongest ones, computational issues arise in their application. Mainly, the number of samples N per leakage trace can be very large which prevents the direct computation of a leakage covariance matrix. As a consequence, a number of heuristics are usually deployed in order to reduce the dimensions of the traces before the construction of the templates. In summary, the first template attacks (*e.g.* in [7]) selected the time samples showing the largest difference between the mean traces $\{\hat{\boldsymbol{\mu}}_s\}_{s=1}^{|\mathcal{S}|}$ associated to the classes $[1, 2, \ldots, s]$. The PCA-based attacks [4] that are described below were suggested as a way to improve and automatize this process.

4.2 PCA-based template attacks

Consider the 256 mean (power and EM) traces associated to the 256 possible transitions $x_1 \rightarrow x_2$ on the microchip bus that are represented in Figure 1. PCA can be applied in order to find a single linear transform that maximizes the inter-class variance between these different empirical mean traces $\{\hat{\boldsymbol{\mu}}_s\}_{s=1}^{|\mathcal{S}|}$ associated to the classes $[1, 2, \ldots, s]$. PCA looks for the first principal directions $\{\mathbf{w}_m\}_{m=1}^{N_c}$ such that $N \geq N_c$ and which form an orthonormal basis of the N_c-dimensional subspace capturing maximal variance of $\{\hat{\boldsymbol{\mu}}_s\}_{s=1}^{|\mathcal{S}|}$. It can be shown [11] that the principal directions are the eigenvectors $\mathbf{U}$ of the empirical covariance matrix:

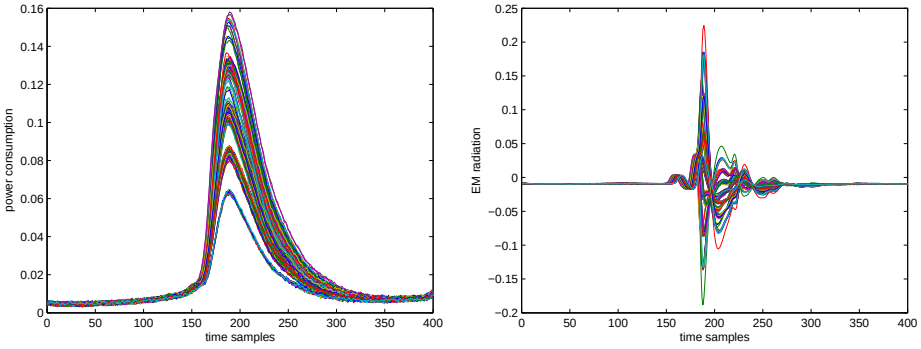

Fig. 1: Average power and EM traces.

$$\bar{\mathbf{S}} = \frac{1}{|\mathcal{S}|}\sum_{s=1}^{|\mathcal{S}|}(\hat{\boldsymbol{\mu}}_s - \bar{\boldsymbol{\mu}})(\hat{\boldsymbol{\mu}}_s - \bar{\boldsymbol{\mu}})^\top, \qquad \bar{\mathbf{S}} = \mathbf{U}\boldsymbol{\Delta}\mathbf{U}^\top.$$

The quantity $\bar{\boldsymbol{\mu}} = \frac{1}{|\mathcal{S}|}\sum_{s=1}^{|\mathcal{S}|}\hat{\boldsymbol{\mu}}_s$ is the average of the mean traces. The principal directions $\{\mathbf{w}_m\}_{m=1}^{N_c}$ are the columns of $\mathbf{U}$ corresponding to the N_c largest eigenvalues of $\boldsymbol{\Delta}$. We denote these eigenvalues by the diagonal matrix $\boldsymbol{\Lambda} \in \mathbb{R}^{N_c \times N_c}$ and the corresponding matrix of principal directions by $\mathbf{W} \in \mathbb{R}^{N \times N_c}$.

In order to built principal subspace templates, we simply estimate the projected means $\{\boldsymbol{\nu}_s\}_{s=1}^{|\mathcal{S}|}$ and the covariance matrices of the projected traces along the (retained) principal directions $\{\boldsymbol{\Lambda}_s\}_{s=1}^{|\mathcal{S}|}$. These parameters are given by:

$$\boldsymbol{\nu}_s = \mathbf{W}^\top\hat{\boldsymbol{\mu}}_s, \qquad \boldsymbol{\Lambda}_s = \mathbf{W}^\top\widehat{\boldsymbol{\Sigma}}_s\mathbf{W}.$$

As in standard template attacks, the noise model is a multivariate Gaussian distribution. However, the number of principal directions N_c is much smaller than N. In practice, a direction can be considered as not being principal if the associated eigenvalue is small compared to the largest one.

Then, in order to classify a new trace $\mathbf{l}_{\text{new}}$, we apply Bayes theorem in the principal subspace of the empirical means which leads to the following rule:

$$\tilde{s} = \underset{s^*}{\text{argmax}}\ \hat{\Pr}[\mathbf{W}^\top\mathbf{l}_{\text{new}}|s^*]\Pr[s^*], \tag{5}$$

with (as in classical template attacks) $\hat{\Pr}(\mathbf{W}^\top\mathbf{l}_{\text{new}}|s^*) = \mathcal{N}(\mathbf{W}^\top\mathbf{l}_{\text{new}}|\boldsymbol{\nu}_{s^*}, \boldsymbol{\Lambda}_{s^*})$.

Properties and Limitations. PSTAs improve simple heuristics selecting time samples according to the variance between the mean traces $\{\hat{\boldsymbol{\mu}}_s\}_{s=1}^{|\mathcal{S}|}$ because they first project the traces in a subspace where these variances are maximized. Additionally, they have the significant advantage of having a "small sample size" variant (see [4] for details), which is particularly useful in contexts where there

are much more time samples in the traces than key classes in the attack. However, PSTAs do not consider the intra-classes variance which may theoretically result in poor classification performances. A natural extension (described in the next section) is to exploit LDA [5, 10], which allows projecting the traces in a subspace that maximizes the ratio between the inter- and intra-class variance.

4.3 LDA-based template attacks

Instead of seeking the directions maximizing the variance of the mean traces, it is intuitively more appropriate to seek for the directions $\{\tilde{\mathbf{w}}_m\}_{m=1}^{N_c}$ that maximize the ratio between the inter-class scatter $\mathbf{S}_B$ and the total intra-class scatter $\mathbf{S}_W$ after projection. That is, to maximize the objective: $\frac{\tilde{\mathbf{w}}^\top \mathbf{S}_B \tilde{\mathbf{w}}}{\tilde{\mathbf{w}}^\top \mathbf{S}_W \tilde{\mathbf{w}}}$, where:

$$\mathbf{S}_B = \sum_{s=1}^{|\mathcal{S}|} N_t (\hat{\boldsymbol{\mu}}_s - \bar{\boldsymbol{\mu}})(\hat{\boldsymbol{\mu}}_s - \bar{\boldsymbol{\mu}})^\top,$$
$$\mathbf{S}_W = \sum_{s=1}^{|\mathcal{S}|} \sum_{i=1}^{N_t} (\mathbf{l}_q^{s,i} - \hat{\boldsymbol{\mu}}_s)(\mathbf{l}_q^{s,i} - \hat{\boldsymbol{\mu}}_s)^\top.$$

Note that these quantities are equal to covariances up to multiplicative constants that do not play a role in the subsequent derivation. Since $\mathbf{S}_B$ is positive definite and symmetric and since $\tilde{\mathbf{w}}$ is scale invariant, the maximization problem can be replaced by the following eigendecomposition:

$$\mathbf{S}_B^{1/2} \mathbf{S}_W^{-1} \mathbf{S}_B^{1/2} = \widetilde{\mathbf{U}} \widetilde{\boldsymbol{\Delta}} \widetilde{\mathbf{U}}^\top,$$

where $\mathbf{S}_B = \mathbf{U}_B \boldsymbol{\Delta}_B \mathbf{U}_B^\top \rightarrow \mathbf{S}_B^{1/2} = \mathbf{U}_B \boldsymbol{\Delta}_B^{1/2} \mathbf{U}_B^\top$. The projection directions are subsequently given by $\widetilde{\mathbf{V}} = \mathbf{S}_B^{-1/2} \widetilde{\mathbf{U}}$. We denote the directions corresponding to the N_c largest eigenvalues of $\widetilde{\boldsymbol{\Delta}}$ as $\{\tilde{\mathbf{w}}_m\}_{m=1}^{N_c}$ and stack them in a projection matrix $\widetilde{\mathbf{W}} \in \mathbb{R}^{N \times N_c}$. The templates are then constructed as in PSTAs, but in the subspace obtained by LDA. Therefore and as previously, the parameters of the multivariate Gaussian noise model are given by:

$$\tilde{\boldsymbol{\nu}}_s = \widetilde{\mathbf{W}}^\top \hat{\boldsymbol{\mu}}_s, \qquad \widetilde{\boldsymbol{\Lambda}}_s = \widetilde{\mathbf{W}}^\top \widehat{\boldsymbol{\Sigma}}_s \widetilde{\mathbf{W}}.$$

Finally, the attack is also performed in the same way as in PSTAs, *i.e.* by applying (5), but using the projection matrix $\widetilde{\mathbf{W}}$ found by LDA, as well as the projected means $\{\tilde{\boldsymbol{\nu}}_s\}_{s=1}^{|\mathcal{S}|}$ and the projected covariances $\{\widetilde{\boldsymbol{\Lambda}}_s\}_{s=1}^{|\mathcal{S}|}$.

Properties and Limitations. While PSTAs can perform well in practice, LDA-based TAs (LDTAs for short) optimize an objective function which is more meaningful. Its limitation arises from the fact that we have to compute the (total) intra-class scatter matrix, which becomes singular when the number of traces N_t is smaller than the number of time samples in the traces N. Hence, for very long traces, one needs to take a lot of measures which may be a practical issue. Also,

the resulting matrices have to be computed and stored which may be another issue. In other words, LDA is not suitable in the "small sample size" case in contrast to PCA. But this is not always a problem since side-channel traces can be reasonably short (as in the next section). And when they are not, they can always be divided into several pieces or preliminarily reduced by heuristics or PCA. In summary, LDTAs bring another tradeoff to the side-channel toolbox: they optimize a more meaningful criteria at the cost of more constraints on the size and amount of measurements performed by the adversary.

5 Experimental results

In this section, we present our experimental results for which we used the following parameters. For each of the 256 possible transitions on the bus, we generated simultaneously 1000 traces (such as those illustrated in Figure 1). Among those traces, N_t=500 were used for the construction of the PCA-based and LDA-based templates. The remaining 500 traces were used for testing the templates and evaluate the information and security metrics of Section 3. From these experiments, we detail both the comparison of the PCA and LDA dimensionality reductions and the comparison/combination of the power and EM side-channels. Note that for the combination of the power and EM leakages, we simply use straightforwardly concatenated traces containing an EM leakage followed by a power one, as initially suggested in [2]. This leads to 800-sample traces that are illustrated in Figure 2. In order to keep $N < N_t$, we simply rejected one every two samples.

5.1 PCA versus LDA

Selection of the time samples. Before detailing the information theoretic and security metrics, an intuitive way to analyze the behavior of the PCA and LDA is to observe how they select the meaningful time samples in the traces. For this purpose, it is convenient to plot the eigenvectors of the transforms, as in Figures 3 and 4 for the power + EM combination.

It yields the following observations:

1. In Figure 3, the eigenvectors corresponding to the first three components of the PCA and LDA are pictured. They clearly show that the EM leakage is dominating in the first component while the power one comes as a backup in the second component. The same figure shows that PCA and LDA select time samples in a similar (and intuitive) way. Namely, they select the points where most of the variability occurs in the curves.
2. In Figure 4, the same eigenvectors are pictured for the last three components of the PCA and LDA. They confirm the expectation that these last components mainly (only) contain noise and therefore can be discarded. Interestingly, these figures show a clear difference between the PCA and LDA. While the noise is uniformly distributed in the PCA eigenvectors, there is a significant absence of noise in the (intuitive) regions of interest for the LDA. This suggest a possibly better information extraction for LDTAs.

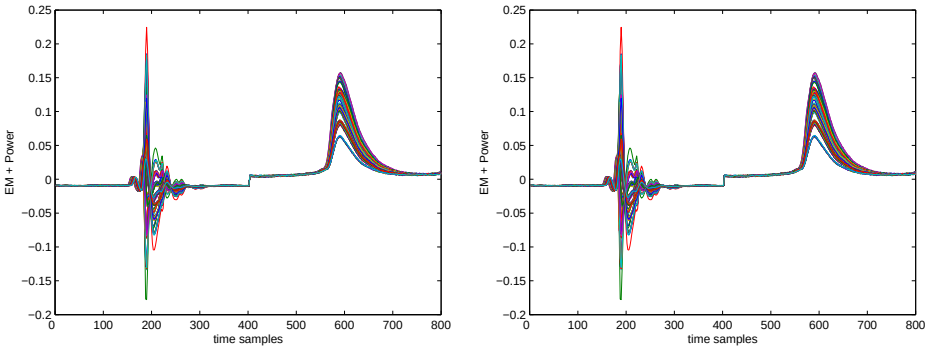

Fig. 2: Average combined power and EM traces.

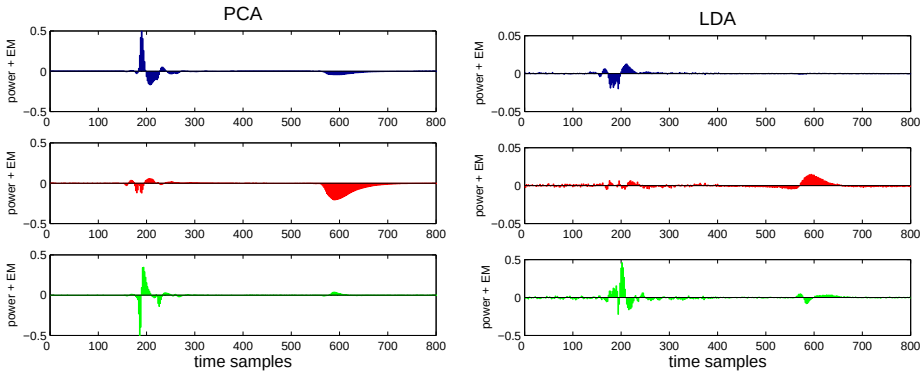

Fig. 3: Eigenvectors for the combined power and EM leakage, first three components.

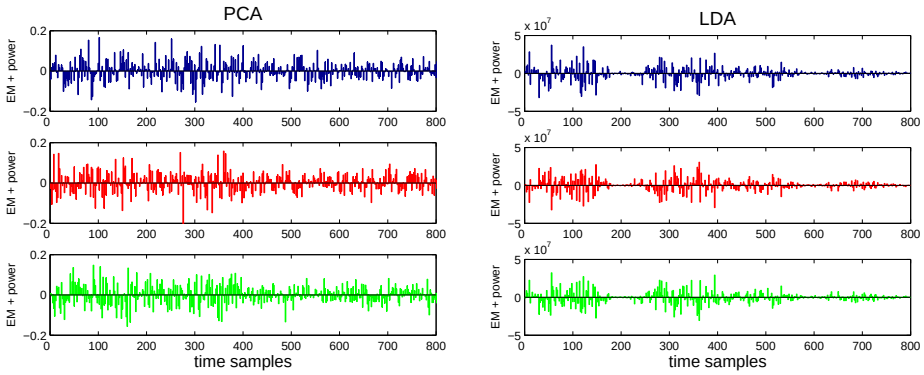

Fig. 4: Eigenvectors for the combined power and EM leakage, last three components.

Figures 6 and 7 in appendix represent the same eigenvectors for the power and EM channels taken independently. They highlight similar intuitions. Note that the dominance of the EM channel in the first component of the combined power and EM eigenvectors already suggest that EM leakages are the most informative.

Entropy scores. Let $\mathbf{H}^1_{s,s^*}$ be the entropy matrix defined in Section 3, Equation (1), where the superscript 1 comes from the fact that we classify the different transitions on the bus based on single traces (or queries). From the estimated probability distributions $\hat{\Pr}[\mathbf{l}_1|s]$ that are provided by the PCA- or LDA-based templates and the set of 500 traces to test these templates, one can derive an estimation $\hat{\mathbf{H}}^1_{s,s^*}$ of this matrix. We say that a leakage model is (*i.e.* that our templates are) sound if for each line of the estimated entropy matrix (corresponding to a key class or transition s), the minimum value occurs for $s^* = s$. The entropy score is simply the fraction of key classes for which this condition is respected. As demonstrated in [18], it corresponds to the faction of key classes for which a Bayesian side-channel attack will be be asymptotically successful.

The entropy scores of the power, EM and power + EM template attacks exploiting both the PCA and LDA are represented in Figure 5. A first observation from these pictures is that none of these channels leads to a 100% entropy score. This is natural since we do not aim to perform a real attack but to evaluate the effectiveness of different side-channels. Similarly, *e.g.* in the Hamming weight leakage model, some key classes will remain undistinguishable (*i.e.* those corresponding to the same Hamming weight values). But since in a real attack, each actual key class (that are not transitions as in this paper but real parts of *e.g.* a block cipher key) can be identified thanks to all the transitions possibly generated by different input plaintexts, a practical attack will be successful.

More importantly, these pictures exhibit (as expected) that the best entropy score occurs for the power + EM channel, followed by the EM and the power channels. They also highlight that LDTAs lead to (slightly) better results than PSTAs, in particular for the combination of the power and EM channels.

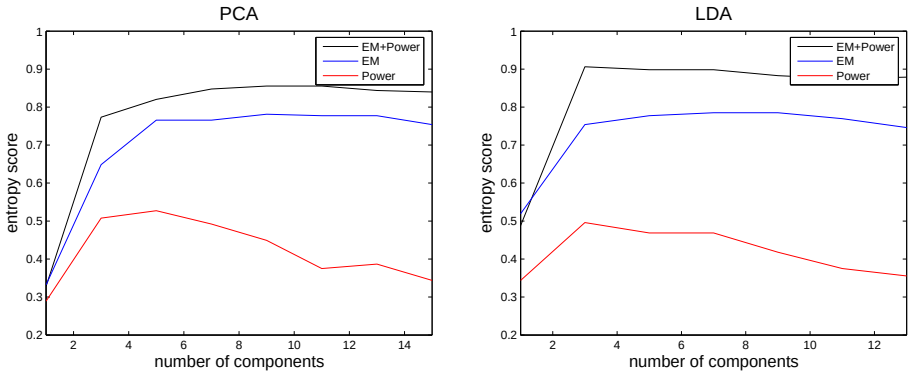

Fig. 5: Entropy scores for the PCA and LDA.

Figure 8 in appendix shows a similar evaluation of the success rates. Interestingly, they do not reach as high values as the entropy scores. This follows the theoretical expectation that success rates do not measure the quality of an implementation (nor the information leakage of a side-channel) but the effectiveness of an adversary. Again, the combination of several leakages would lead to higher success rates. Under the conditions discussed in [18], the more information leaked (measured with the conditional entropy defined in Section 3), the faster a Bayesian adversary exploiting a sound leakage model will converge towards a 100% success rate. Note that even if the success rates are lower than the entropy scores, they underline that the EM side-channel leads to extremely powerful attacks. Indeed, even the leakage of a single clock cycle leads to a non negligible success rate in the recovery of the transitions on the bus.

Entropy values. Since as previously mentioned, less entropy in the traces involves (under certain conditions) a more efficient Bayesian side-channel attack, this section finally provides the conditional entropy values extracted from the leakage traces for both the PCA- and LDA-based templates and different number of components. They are summarized in Table 1.

It yields the following observations:

1. This table confirms all the previous conclusions. Namely, LDA leads to a better information extraction than PCA; the EM channel is significantly more informative than the power one; and the combination of both channels (*i.e.* power + EM) leads to the most powerful type of attack.
2. The lowest conditional entropy values and the maximum entropy scores do not occur for the same number of components (although they are strongly correlated). This highlights that there exist situations where more key classes can be asymptotically recovered, but reaching a 100% success rate will be slower than in a context where less classes can be asymptotically recovered.
3. Compared to previous works on multi-channel or template attacks, this quantified comparison is justified by theoretical statements on the evaluation metrics. Therefore, it is expected that Table 1 provides the best possible comparison of the information leakages for the different channels.

Table 1: Conditional entropy of the power, EM and power + EM traces.

Number of components	3	5	7
power (PCA)	4.62	4.49	4.57
power (LDA)	**4.41**	4.48	4.62
EM (PCA)	3.92	3.65	3.55
EM (LDA)	3.21	**3.15**	3.24
power + EM (PCA)	3.57	3.36	3.20
power + EM (LDA)	2.92	**2.80**	2.87

5.2 Power versus EM, Power + EM

While comparing LDA and PCA in the previous section, most important quantitative analyzes and conclusions on the respective effectiveness of the power and EM channels have already been drawn. However, one question that has not yet been tackled is: "how far as these real leakages from the idealized (*e.g.* Hamming weight) models that are frequently considered in the side-channel literature?".

In order to answer this question, we can fortunately use the same framework again. Let us start with the power channel for which a usual assumption is to correlate it with the Hamming distances of the transitions on the bus. In our present example, 5 possible Hamming distances can be observed (*i.e.* $h_d \in [0 \dots 4]$) which leads to the following conditional entropy:

$$\mathrm{H}[S|\mathbf{L}_1] = -\sum_{h_d=0}^{4} \frac{2^4 \cdot \binom{4}{h_d}}{2^8} \cdot \log_2 \left(\frac{1}{2^4 \cdot \binom{4}{h_d}} \right) = 5.9694$$

This indicates that such a model is significantly less informative than a real power consumption channel which would reduce the conditional entropy down to 4.41 in exactly the same context. Looking back at Figure 1, this simply means that traces corresponding to the same Hamming distance $H_W(x_1 \oplus x_2)$ can actually be distinguished by a carefully profiled template adversary. Such traces can be seen as included in one of the packets of curves in the figure.

A very similar analysis can be performed for the EM channel. Let us for example take the signed distance model proposed in [14]. Such a model is purposed to better incorporate the specificities of the EM channel since it allows to distinguish between $x_1 \rightarrow x_2$ and $x_2 \rightarrow x_1$ transitions. In practice, it means that 9 possible signed distances can be observed by the adversary (*i.e.* $s_d \in [-4, \dots, 4]$) which leads to the following conditional entropy:

$$\mathrm{H}[S|\mathbf{L}_1] = -\sum_{s_d=-4}^{4} \frac{\binom{8}{s_d+4}}{2^8} \cdot \log_2 \frac{1}{\binom{8}{s_d+4}} = 5.4558$$

While such a model is slightly more informative than the standard Hamming distance model, it is again by far less informative than the real EM channel that would reduce the entropy down to 3.15 in exactly the same context.

6 Conclusions

Following recent developments in physically observable cryptography, this paper provides theoretical and practical insights in the analysis of the power and EM side-channels and their efficient exploitation with powerful statistical tools.

First, we use fair information theoretic and security metrics to evaluate and compare these two side-channels. The resulting analysis demonstrates the significantly higher information leakages of the EM channel when near field measurements of a cryptographic chip are available.

Second, we propose the Linear Discriminant Analysis as an alternative to the Principal Component Analysis for the best selection of the meaningful leakage samples in template attacks. We apply these tools to both the power and EM channels as well as to their comparison in a multi-channel attack context. The results show that Linear Discriminant Analysis is an interesting alternative, bringing a better information extraction at the cost of more constraints in the size and amount of measurements performed by a side-channel adversary. It is therefore a very interesting tool to combine with Principal Component Analysis in any practical application of the template attacks.

Finally, we compare the information leakages of the power and EM channels with some idealized (*e.g.* Hamming weight) models used to predict these leakages. Our results confirm that not only the distinguishers that usually exploit these models (*e.g.* the correlation coefficient) are suboptimal, but also that the models themselves are far less informative than the actual power and EM observations. This highlights the importance of template attacks when the provable (or arguable) security of cryptographic implementations is discussed and the relevance of strong models such as the noisy identity leakages introduced in [15] in this context. In summary, if you are an adversary trying to recover the key of a cryptographic device, Hamming weight (or similar) models can be useful. But if you are a designer trying to convince that your cryptographic implementation is secure against side-channel attacks, they are definitely not sufficient.

We note that these results only considered the application of PCA and LDA to the original template attacks of [7]. However, these techniques could be similarly applied to the stochastic models of [9, 17]. The direct use of templates in our experiments was reasonable since we were not limited in the number of samples to build the templates (due to our evaluation goal). But stochastic models could be very efficient in more constrained contexts. Combining data dimensionality reduction techniques with stochastic models is a scope for further research.

References

1. D. Agrawal, B. Archambeault, J. Rao, P. Rohatgi, *The EM Side-Channel(s)*, in the proceedings of CHES 2002, Lecture Notes in Computer Science, vol 2523, pp 29-45, Redwood City, CA, USA, August 2002.
2. D. Agrawal, J.R. Rao, P. Rohatgi, *Multi-Channel Attacks*, in the proceedings of CHES 2003, LNCS, vol 2779, pp 2-16, Cologne, Germany, September 2003.
3. R. Anderson, M. Kuhn, *Tamper Resistance - a Cautionary Note*, in the proceedings of USENIX Electronic Commerce, pp 1-11, Oakland, CA, USA, November 1996.
4. C. Archambeau, E. Peeters, F.-X. Standaert, J.-J. Quisquater, *Template Attacks in Principal Subspaces*, CHES 2006, Lecture Notes in Computer Science, vol 4249, pp. 1-14, Yokohama, Japan, October 2006.

5. C.M. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.
6. E. Brier, C. Clavier, F. Olivier, *Correlation Power Analysis with a Leakage Model*, in the proceedings of CHES 2004, Lecture Notes in Computer Science, vol 3156, pp 16-29, Boston, Massachusetts, USA, August 2004.
7. S. Chari, J.R. Rao, P. Rohatgi, *Template Attacks*, in the proceedings of CHES 2002, Lecture Notes in Comuter Science, vol 2523, pp 13-28, CA, USA, August 2002.
8. K. Gandolfi, C. Mourtel, F. Olivier, *Electromagnetic Analysis: Concrete Results*, in the proceedings of CHES 2001, Lecture Notes in Computer Science, vol 2162, pp 251-261, Paris, France, May 2001.
9. B. Gierlichs, K. Lemke, C. Paar, *Templates vs. Stochastic Methods*, CHES 2006, LNCS, vol 4249, pp 15-29, Yokohama, Japan, October 2006.
10. T. Hastie, R. Tibshirani, J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference and Prediction*, Springer, 2001.
11. I.T. Jolliffe, *Principal Component Analysis*, Springer, New York, 1986.
12. 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.
13. 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.
14. E. Peeters, F.-X. Standaert, J.-J. Quisquater, *Power and Electromagnetic Analysis: Improved Models, Consequences and Comparisons*, in Integration, the VLSI Journal, vol 40, pp 52-60, Elsevier, Spring 2007.
15. C. Petit, F.-X. Standaert, O. Pereira, T.G. Malkin, M. Yung, *A Block Cipher based PRNG Secure Against Side-Channel Key Recovery*, to appear in the proceedings of ASIACCS 2008, available from: http://eprint.iacr.org/2007/356.
16. J.-J. Quisquater, D. Samyde, *ElectroMagnetic Analysis (EMA): Measures and Counter-Measures for Smart Cards*, in the proceedings of E-smart 2001, Lecture Notes in Computer Science, vol 2140, pp 200-210, Cannes, France, September 2001.
17. W. Schindler, K. Lemke, C. Paar, *A Stochastic Model for Differential Side-Channel Cryptanalysis*, CHES 2005, Lecture Notes in Computer Science, vol 3659, pp 30-46, Edinburgh, Scotland, September 2005.
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.

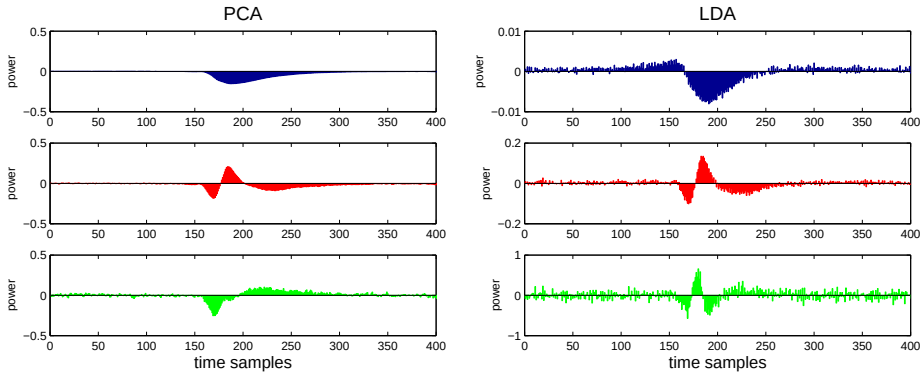

Fig. 6: Eigenvectors for the power leakage, first three components.

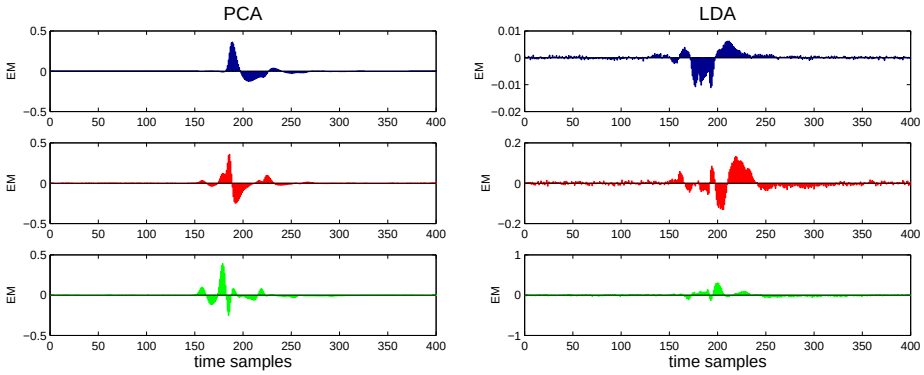

Fig. 7: Eigenvectors for the EM leakage, first three components.

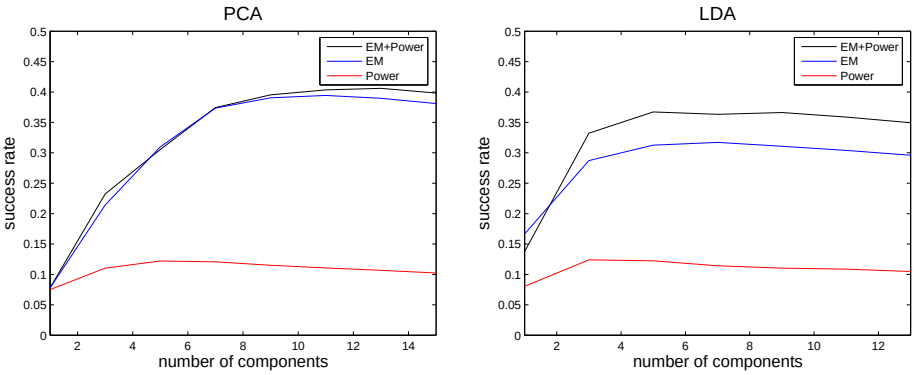

Fig. 8: Success rates for the PCA and LDA.