Chapter 7

Conclusions and Perspectives

7.1 Conclusions

In this thesis, we focused over issues related to the side-channel profiling attacks, which play a fundamental role in the context of the evaluation of cryptographic secure devices. The opportunity of performing a characterisation of the device leakages opens the way to an optimal approach, allowing the estimation of the conditional probabilities needed to identify the target key through maximum likelihood. Nevertheless, the attempt to estimate the probability distributions of highly multi-dimensional data is hindered by the curse of dimensionality. Our first efforts were thus focused over the development of dimensionality reduction techniques, and we proposed two works on this topic.

First, in Chapter 4, we presented an analysis of linear dimensionality reduction techniques, that had already been introduced in side-channel context before 2014, the PCA and the LDA. These techniques extract interesting features from data by means of linear combinations of time samples. Despite the fact that the LDA is mainly a technique that allows to build a linear classifier, it is its dimensionality reduction version, known as Fisher's Linear Discriminant that raised attention in side-channel context. We followed this trail, and exploited both PCA and LDA as preliminary phases for a Gaussian template attack. In this context, we tackled some open issues, in particular the problem of the component selections, proposing an automatic criterion to perform the choice, namely the ELV. The obtained results were published at CARDIS 2015 [CDP16a].

In a second work we enlarged the considered models from linear to non-linear ones, in order to treat the dimensionality reduction issue in presence of masking countermeasure. We focused on the rarely considered case in which the profiling phase does not enable the access to the randomly drawn masks. In this context we proposed a non-linear generalisation of the LDA method, namely the KDA equipped with a polynomial kernel function. This KDA extracts features from signals through

products of time samples (up to a fixed polynomial degree) and linear combinations. Even in this case, despite the KDA may naturally provide a non-linear classifier, the KDA application in our study were intended as preliminary phase of a Gaussian template attacks. The obtained results of this contribution were published at CARDIS 2016 [CDP16b].

The third contribution of this thesis, presented in Chapter 6, explores the Neural Networks models. Such models are a further generalisations of techniques like LDA and KDA: they extract features from data by means of several layers of linear combinations and non-linear functions. Neural Networks are widely used to build non-linear classifiers. Differently from the LDA classifier, NN ones may be easily constructed in a multi-class manner, and in such a way that classification scores have a probabilistic meaning. In this way they are directly suitable for advanced side-channel attacks. Choosing this kind of construction, we could substitute the typical side-channel profiling routine divided into dimensionality reduction and Gaussian profiles estimation, with an integrated approach that directly extracts significant feature and estimates *a posteriori* probabilities. In this case, such an estimation dispensed of the Gaussian hypothesis about data distribution, not justifiable in general. The estimation is guided by a single optimization criterion, aiming at reducing the classification error. The optimization algorithm is not in a closed form as for the LDA and KDA technique, and there is no guaranties about the existence/uniqueness of a solution and about the fact that the learning algorithm is eventually able to find the solution. Anyway, many ML techniques are funded over the acceptance of this intrinsic non-optimality, and face in this way the curse of dimensionality that prohibits perfect estimations. Anyway, ML techniques demonstrate their validity in many real applications, in side-channel analysis as well. In our contribution, we took advantage of the Convolutional Neural Network models, and we proposed some Data Augmentation techniques, to tackle the hiding countermeasures inducing misalignment in side-channel acquisitions. The obtained results were published at CHES 2017 [CDP17].

7.2 Tracks for Future Works

The common thread of this thesis is the constantly growing awareness of the fact that practical problems we were facing in side-channel domain, were almost identical to those faced in many other domains. In particular, today an immense and still expansing number of applicative fields are based on the sense and the analysis of a huge quantify of highly multi-dimensional data, and all of them have to tackle the

curse of dimensionality. The preliminary purpose of these researches was to determine a way to select features among the sensed ones, *i.e.* select leaking time samples. This approach implies a critical waste of potentially useful information, and we turned soon to methodologies aiming to construct new interesting features by means of increasingly complex models. This required a conversion of side-channel problems from a classical statistical asset into an ML one, and we believe that this conversion process should be pursued in future works.

A first issue we left open is explicitly related to such a conversion: it is the definition of a DPA-specific ML task. Indeed, by now we exploited classifiers to perform advanced side-channel attacks. Nevertheless we observed that the classification task perfectly match with the simple attack scenario. Specialized metrics and optimization criteria (*e.g.* loss functions, evaluation metrics) should be proposed to tackle advanced attacks, instead: the final goal of an advanced attack is indeed the identification of a secret value by means of several observations, and it does not coincide in general with the classification of the observations by the sensitive variable labels. To explicit this final goal to the ML tool may lead to great advantages. Moreover, a Bayesian statistical approach should even be explored in the attempt of defining a DPA-specific ML strategy. Indeed, a secret key chunk may be viewed as a discrete parameter for a sort of regression model that describes the side-channel traces. Before starting an attack, such key chunk parameter has in general a uniform distribution over its definition set, *i.e.* to the attacker any value is equally probable. Applying a Bayesian approach means considering every model parameter with the probability distribution modelling the attacker uncertainty over it, and building a system that updates such distributions as long as the attacker observes new traces and gains new information. This process should stop once the key chunk parameter distribution has a sufficiently low entropy, showing high probability concentrated over few values. Interestingly, recently a new field is arising, known as Bayesian Deep Learning (BDL) [Gal16], which provides a deep learning framework, able to achieve state-of-the-art results, at least in imaging domain, while also modelling uncertainty.

As a second track for future works, we remarked that the classical ML verification task perfectly matches with the current collision attacks in side-channel domain. This topic is not developed in this thesis, but we already focused on the possibility of exploiting some so-called *Siamese Neural Networks*, specialised for the verification task, to perform collision attacks. We obtained some promising preliminary results.

In general, we are convinced of the importance of further exploring DL techniques in side-channel context. At the same time we are aware of the lack of clear theoretical foundations that give guides about the choice of the hyper-parameters that influence so much the performances of the DL architectures. For this reason we believe that researchers should share their efforts in developing and analysing *ad hoc* methodologies to tune side-channel-oriented neural networks. To this aim, a publication appeared in the *Cryptology ePrint archive* on January 2018 proposing a fully-reported set of benchmarks performed over some electromagnetic emanation acquisitions. The whole acquisitions database were published as well, including all the sources of the target implementation [Pro+18b]. We wish this open platform may serve as a common basis for researchers willing to compare their new architectures or their improvements of existing models. This kind of public databases have been central tools in the development of deep learning solutions in many other domains, for example in image recognition context.

Finally, in the optic of enhancing cryptanalysis in order to make cryptography stronger, there is a missing key-stone in this work. We adopted methods to extract new features from data, by means of complex models, most of all neural networks, instead of selecting leaking points of interests. Once an evaluator obtains a model allowing a successful attack, his role should be to point out the vulnerabilities of the attacked device, eventually explaining their origin. If the model the attack bases on exploits abstract features impossible to interpret, such a role is impossible to play. A methodology to unroll the construction of the abstract feature and understand which part of the cryptographic algorithm execution most contributes to the success of the attack is indispensable in the optic of strengthen the embedded security against the powerful increasing deep learning attackers.