Improving the Rules of the DPA Contest

François-Xavier Standaert*, Philippe Bulens,
Giacomo de Meulenaer, Nicolas Veyrat-Charvillon

UCL Crypto Group, Université catholique de Louvain, B-1348 Louvain-la-Neuve.
e-mails: fstandae,philippe.bulens,giacomo.demeulenaer,nicolas.veyrat@uclouvain.be

Abstract. A DPA contest has been launched at CHES 2008. The goal of this initiative is to make it possible for researchers to compare different side-channel attacks in an objective manner. For this purpose, a set of 80 000 traces corresponding to the encryption of 80 000 different plaintexts with the Data Encryption Standard and a fixed key has been made available. In this short note, we discuss the rules that the contest uses to rate the effectiveness of different distinguishers. We first describe practical examples of attacks in which these rules can be misleading. Then, we suggest an improved set of rules that can be implemented easily in order to obtain a better interpretation of the comparisons performed.

1 Introduction

Comparing side-channel attacks in a fair manner is a challenging problem due to the large amount of parameters that enter into account in their implementation.

On the one hand, one can hardly compare two attacks against the same algorithm implemented on two different platforms. This is because the amount of information leakage provided by different devices can be significantly different. As a consequence, the fair comparison of attacks implies the need of standardized measurement platforms such as the DPA workstation of CRI [1] or the SASEBO boards [2]. Alternatively, the use of public databases containing a large set of traces can be used for comparing different attacks. This is the approach chosen by the DPA contest [7] that was launched at CHES 2008 [3].

On the other hand, having comparable measurements is not sufficient to obtain fair comparisons. Comparable methodologies in the exploitation of the leakages also have to be considered. This is typically the goal of the framework introduced in [5] in which the comparison of different physically observable implementations and side-channel attacks is discussed in details.

In this note, we consequently aim to show how such a framework can be used to improve the relevance of the rules in a DPA Contest. For this purpose, Section 2 recalls the original rules described in [7] and highlights their limitations. Then, Section 3 suggests a set of improved rules. Finally, Section 4 shows an application of these new rules to an exemplary set of attacks.

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

2 Former rules of the DPA contest and limitations

The DPA contest presented in [7] relies on the three main rules that follow:

(1) Comparison criteria. The only metric that is considered is the amount of traces needed to guess the key. A key is supposed to be guessed correctly if a stability criteria is fulfilled, *i.e.* the side-channel distinguisher has to continuously output the correct key candidate when accumulating the traces. A threshold of 100 iterations with the correct key guess is arbitrarily chosen.
(2) Computational complexity. The reference attack of [7] is a mix of side-channel attack and brute-force search. Namely, $6 \times 8 = 48$ bits of the key are recovered by a side-channel analysis of the subkey in the first DES round. The remaining 8 bits are searched exhaustively. It is a rule of the contest that brute force search is not used for more than those 8 bits.
(3) Template attacks are not (trivially) realizable, because the 80 000 traces correspond to the encryption of 80 000 plaintexts with the same secret key.

We pinpoint the following limitations.

With respect to (1), a first concern is the need to define a threshold in the stability criteria. It implies that the comparisons between different attacks could be different for different thresholds, which is typically something to avoid. More importantly, the use of stability as a criteria is not directly related to a security notion (*e.g.* key recovery that is the most frequently considered in the context of side-channel attacks). As a consequence, it may lead to controversial statements. For example, imagine an attack in which the stability criteria is never fulfilled but the side-channel distinguisher outputs the correct key candidate 99% of the time. Can we claim this is a secure implementation?

As far as (2) is concerned, the problem is that only a part of the computational capabilities in a side-channel attack is considered by this rule. But in practice, computation can be exploited in different parts of an attack. As a typical example, the reference attack in [7] is made of 8 sub-attacks targeting 6-bit parts of the block cipher key. This means that the adversary has to build 8×2^6 partitions or predictions of the leakages. But one could also design an attack made of 4 sub-attacks targeting 12-bit parts of the block cipher key. This new attack would require to build 4×2^{12} partitions or predictions of the leakages. Both attacks fall into the contest rules, but the second one has a higher computational cost and is therefore not directly comparable with the first one.

Eventually, (3) is more an informal rule than a strict one. But similarly to the previous cases, it may result in some fuzzyness in the interpretation of the results. First, templates could be built using the single key given in the contest, *e.g.* by exploiting an EIS (Equal Images under different Subkeys) property as defined in [4]. Second, performing a side-channel attack can always be seen as a profiling step for a subsequent side-channel attack. For example, performing single-bit DPA attacks against the 4 output bits of a DES S-box can be used to improve the model in another attack (*e.g.* by assigning weights to these bits).

3 New proposal of rules for the DPA Contest

The previous limitations can be mitigated by using the following guidelines.

(1) Use sound comparison criteria. Following the framework in [5], we can use the *success rates of different orders* and the *guessing entropy* to compare different side-channel attacks. Both metrics measure the extent to which a given adversary is efficient in turning the side-channel leakage into a key recovery. Informally, 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 a side-channel adversary. It measures an adversarial strategy with a fixed computational cost after the physical leakages have been exploited. Similarly, the guessing entropy measures the average number of key candidates to test after a side-channel attack has been performed. For more details and justifications about these metrics, we refer to [5].

(2) Provide a clear description of the adversary. Again following [5], it is interesting to detail the different steps of the side-channel attack. In particular, the assumptions about the leakage model, the statistical test and the reduction mapping (*i.e.* the mapping that possibly reduces the leakages dimensions) that are used in the attacks are worth being clearly specified.

(3) Detail the attack data, time and memory complexities. The data complexity (*i.e.* the number of measured traces) is usually the most important parameter in a side-channel attack. But the time and memory complexities cannot be neglected. With this respect, it is important to specify at least:

   – The size of the keyguesses in the attack since a model or partition of the leakage usually has to be built for each of those guesses.
   – The number of leakage samples for which the side-channel attack will be applied (this typically depends on the acquisition device sampling rate and the possible use of a leakage reduction mapping).
   – The remaining workload after the side-channel attack has been performed. This can be brute force search of a part of the key as in the reference attack of [7]. But it can also correspond to the test of key candidates if high-order success rates are considered. Eventually, it can correspond to any other type of cryptanalytic computation.

(4) Evaluate the profiling and exploitation of the leakages separately. If profiling is used in the attacks (there is no reason for not doing it anytime it is possible), the complexities of the profiling and exploitation phases (as described in [5]) should be explicitly evaluated. For example, if template attacks are considered, the same traces cannot be used to build and evaluate the effectiveness of the templates. Note again that most attacks can be viewed as profiling for a subsequent attack. Note also that if the set of traces is limited (*e.g.* 80 000 in the DPA contest of [7]), spending traces for profiling also involves that less traces will be available to evaluate the metrics afterwards. Hence, the confidence in the statistical sampling will be reduced.

4 Exemplary application of the new rules

In practice, the previous evaluation metrics have to be evaluated from a sufficient number of samples. Since 80 000 traces are provided in the DPA contest of [7] and assuming that an attack does not require more than 1000 traces to be successful, one can perform 80 independent experiments exploiting 1000 traces and estimate the success rates and guessing entropy from those experiments.

As an illustration, Figures 1 and 3 represent the success rate of order 1 and the guessing entropy in function of the number of traces in the four attacks presented in [7] on November 1st, 2008. They include a reference attack (single-bit DPA), a multiple-bit DPA, a multiple-bit correlation power analysis and a multiple-bit correlation power analysis with pre-processed leakage traces.

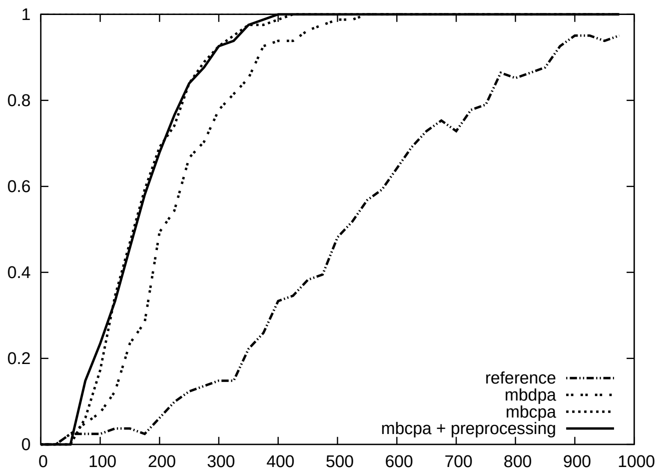

Fig. 1: X: number of queries, Y: success rates of the attacks in [7].

Following the previous guidelines, we first mention that those four attacks rely on similar assumptions. Namely, they target the 6 key bits entering the first S-box in the first DES round. The same metrics could be computed for all the 48 bits of this first-round subkey. We limited our investigations to these 6 bits for convenience. No leakage reduction mapping was used in the experiments: the side-channel distinguishers are applied to all the leakage samples independently (which results in relatively long computation times). Multiple-bit attacks exploit a Hamming weight leakage model. Only the attack with pre-processing has a slightly higher computational cost which can however be neglected in first approximation since this pre-computation is applied once to the leakage traces.

Interestingly, while the metric in [7] suggests that pre-processing the traces slightly improves the multiple-bit correlation attack (see the hall of fame in Table 1 for illustration), our metrics rather show that this precomputation is quite inefficient in the investigated context. Also, a significant advantage of the proposed success rates and guessing entropy is that they can be computed in function of the number of traces in the attacks. Hence, it is possible to see how the remaining security of the implementation (*i.e.* the amount of exhaustive search required to perform a successful key recovery) evolves with this quantity.

Rate	Author	Attack	Criteria
1	S. Guilley	mbcpa+prep.	569 traces
2	S. Guilley	mbcpa	584 traces
3	S. Guilley	mbdpa	866 traces
4	F. Flament	reference	2766 traces
...	...	...	...

Table 1: Hall of fame of the DPA contest (November 1^{st} 2008).

4.1 Improving the figures with re-sampling techniques

Before concluding this note, we mention that the smoothness of the curves in our figures can possibly be improved by using re-sampling techniques. For example, one could repeatedly (*e.g.* N times) select 1000 traces randomly among the 80 000 ones in the DPA contest in order to compute the metrics from $N > 80$ experiments. More advanced bootstrapping techniques could also be considered. As an illustration, Figure 2 illustrates the success rate of the multiple-bit DPA, with re-sampling ($N = 1000$) and without it.

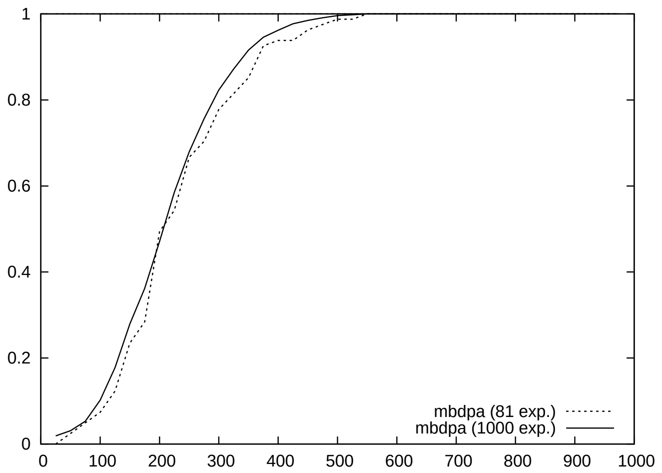

Fig. 2: X: number of queries, Y: success rate of the mbdpa attack in [7].

5 Conclusion

We proposed an improved set of rules to evaluate a DPA contest. By experimentally applying these rules, we showed how they can enhance the understanding of different side-channel attacks. We note that the proposed rules are not claimed to be perfect and can still hide certain interesting phenomenons. For example, success rates could be very different for different keys which is not detectable with average evaluation criteria. If the number of traces provided to evaluate the metrics is limited, agreeing on a re-sampling strategy may also be necessary. Eventually, rating attacks in a hall of fame still requires to fix some thresholds (*e.g.* the exact success rate for which the attacks will be compared). But even if a threshold success rate is fixed, we suggest to always provide the metrics in function of the number of measurements in order to better illustrate the context-dependent nature of side-channel attacks. We believe that the proposed rules provide a reasonably fair picture of the effectiveness of a side-channel attack. More experiments using such metrics can be found in [6]. As far as the DPA contest is concerned, we mention that attacks could be enhanced with profiling, but it would require to clearly separate the traces used for preparation from those used in the online phase. Eventually, we mention that the DPA contest is about comparing side-channel adversaries. But comparing leaking devices is an equally interesting (and in fact more challenging) problem, as detailed in [5].

References

1. CRI (Cryptographic Research Incorporated), *Hardware Testing - DPA Workstation*, http://www.cryptography.com/technology/dpa/workstation.html
2. RCIS (Research Center for Information Security), *SASEBO (Side-Channel Attack Standard Evaluation Boards)*, http://www.rcis.aist.go.jp/special/SASEBO/
3. E. Oswald, P. Rohatgi, *Cryptographic Hardware and Embedded Systems - CHES 2008*, LNCS, vol. 5154, Washington, D.C., USA, August 2008.
4. 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.
5. 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.
6. F.-X. Standaert, B. Gierlichs, I. Verbauwhede, *Partitions vs. Comparison Side-Channel Attacks: An Empirical Evaluation of Statistical Tests for Univariate Side-Channel Attacks against Two Unprotected CMOS Devices*, in the proceedings of ICISC 2008, LNCS, vol xxxx, pp yy-zz, Seoul, Korea, December 2008.
7. VLSI Research Group, TELECOM ParisTech, *The DPA Contest 2008/2009*, http://www.dpacontest.org/

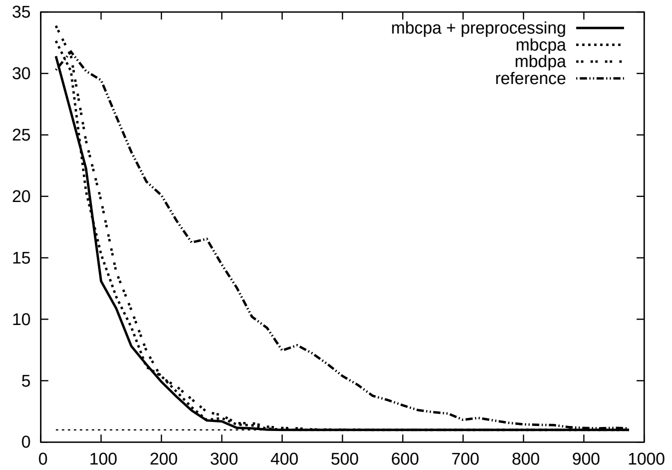

Fig. 3: X: number of queries, Y: guessing entropies of the attacks in [7].