Revisiting Higher-Order DPA Attacks: Multivariate Mutual Information Analysis

Benedikt Gierlichs¹, Lejla Batina^{1,2}, Bart Preneel¹, and Ingrid Verbauwhede¹

¹ K.U. Leuven, ESAT/SCD-COSIC and IBBT

Kasteelpark Arenberg 10, B-3001 Leuven-Heverlee, Belgium

`firstname.lastname@esat.kuleuven.be`

² Radboud University Nijmegen, CS Dept./Digital Security group

Heyendaalseweg 135, 6525 AJ Nijmegen, The Netherlands

`lejla@cs.ru.nl`

Abstract. Security devices are vulnerable to side-channel attacks that perform statistical analysis on data leaked from cryptographic computations. Higher-order (HO) attacks are a powerful approach to break protected implementations. They inherently demand multivariate statistics because multiple aspects of signals have to be analyzed jointly. However, most works on HO attacks follow the approach to first apply a pre-processing function to map the multivariate problem to a univariate problem and then to apply established 1st order techniques. We propose a novel and different approach to HO attacks, Multivariate Mutual Information Analysis (MMIA), that allows to directly evaluate joint statistics without pre-processing. While this approach can benefit from a good power model, it also works without an assumption. We present the first experimental results for 2nd and 3rd order MMIA as well as state-of-the-art HO attacks based on real measurements. A thorough empirical evaluation confirms the advantage of the new approach: 3rd order MMIA attacks require about 800 measurements to achieve 100% success while state-of-the-art HODPA requires 1000 measurements to achieve about 40% success. As a consequence, the security provided by the masking countermeasure needs to be reconsidered as 3rd and possibly higher order attacks become more practical.

1 Introduction

Embedded devices such as smart cards, mobile phones, and RFID tags are becoming increasingly pervasive. In order to secure the applications, these devices execute cryptographic algorithms and protocols to authenticate data and entities and to protect the confidentiality of sensitive data. An embedded device is often physically accessible and it is very likely that such a device falls into the hands of a malicious user. The physical accessibility has led to a number of very powerful attacks that include physical tampering and side-channel attacks. A typical example is Differential Power Analysis (DPA) [11]. The technique explores weaknesses of implementations rather than algorithms, allowing an attacker to extract the secret of a device by monitoring its power dissipation, if

no special countermeasures are taken. A successful DPA attack is subject to two conditions: i) there exists an intermediate variable in the implementation that is correlated with the power consumption and ii) this variable can be predicted with knowledge of the plaintext (or ciphertext) and by guessing a small part of the key and possibly other constants.

In order to protect devices against DPA, one can get rid of the second condition by data randomization or masking [4,8]. The idea is to conceal intermediate values through addition or multiplication with random values, which makes it impossible to correctly predict the intermediate variable.

However, this so-called 1st order masking succumbs to higher-order DPA attacks (HODPA) as originally proposed by Messerges [13] and Chari et al. [2]. These HODPA attacks are based on the joint statistical properties of multiple aspects of the signal, typically joint analysis of the power consumption at two (or more) points in time. In this case one would think of multivariate analysis but all established techniques [2,10,13,15,17,21] rely on a pre-processing step to map the multivariate problem to a univariate problem before attacking the result with a standard DPA attack. The sole exception is a recent and independent work by Prouff and Rivain [16], which we will address in Sect. 4.1.

HODPA attacks imply higher costs in terms of number of samples and computational complexity. In addition, the identification of points in time at which to take the signals is a hard problem. Another common issue is the pre-processing step: while this problem has been studied by many authors, finding the optimal transformation is still an open problem. Furthermore, it is evident that none of the solutions is generic as each pre-processing is tightly linked to a leakage model, that is not always met in practice [18]. Eventually, it is unclear how the pre-processing functions can be generalized for attacks of order higher than two without accepting enormous drawbacks.

Our contribution solves all but one of the aforementioned problems. At CHES 2008 we introduced a side-channel distinguisher called Mutual Information Analysis (MIA) [7]. This 1st order attack is effective without any knowledge or restrictive assumption about the particular dependencies between processed data and observable power consumption. Our proposal, Multivariate MIA (MMIA), inherits this important property. Also the issue of pre-processing unravels because MMIA is explicitly multivariate, which furthermore allows to easily adapt it to attacks of order higher than two. What remains is the problem of identifying the points in time at which to take the signals. Our theory is confirmed by an extensive empirical evaluation of MMIA, the approach of [16], and state-of-the-art HODPA attacks against 1st and 2nd order masked software implementations of a DES like mini-cipher.

This paper is organized as follows. In Sect. 2 we introduce our notation and summarize related work on MIA, the masking countermeasure, and in particular HODPA attacks. In Sect. 3 we discuss our motivation and formulate a generic 2nd order attack problem, for which we present a sound solution in Sect. 4 and compare it to the approach of [16]. Section 5 deals with the empirical evaluation of our proposal and state-of-the-art HODPA attacks. We conclude our work in Sect. 6.

2 Preliminaries

A device performs a cryptographic computation $E_k(x)$ under some key k from a keyspace $\mathcal{K} = \{0, 1\}^m$. Since the key is unknown it is modeled as a random variable (RV) $\mathbf{K}$ on $\mathcal{K}$ with a priori uniform probability mass function (PMF). The information leakage of the device, due to its physical properties, is called side-channel leakage. Side-channel key-recovery attacks exploit this side-channel leakage to recover the key k . Typically, a differential attack consists of statistically analyzing the side-channel leakage related to a well chosen, key and input dependent intermediate value of the computation $E_k(x)$. We model this intermediate value as a RV $\mathbf{W}_k$ on a space $\mathcal{W} = \{0, 1\}^n$ where n is the word length of the device. The corresponding side-channel leakage is denoted $\mathbf{L}_k$. Non-profiled side-channel attacks usually ignore the presence of noise and measurement inaccuracy, and assume that the measurement of a physical observable (here power consumption) provides direct access to the leakage $\mathbf{L}_k$. In this idealized case $\mathbf{L}_k$ exists on a space with at most 2^n elements.

The central problem in differential attacks is to decide whether two RVs are statistically (in-)dependent. The RV $\mathbf{L}_k$ is the measured side-channel leakage. The other RV $\mathbf{P}_{k'}$ is the predicted side-channel leakage of the same computation that depends on a key guess k' , the input, and a leakage model. DPA attacks apply a statistical test $T(\mathbf{L}_k, \mathbf{P}_{k'})$ for all key guesses k' that measures whether the RVs are statistically dependent or not. The value of k' that leads to the highest dependence is an adversary's best guess. In this context, statistical tests are frequently called distinguishers in the literature [19].

MIA's core is the mutual information based distinguisher that implements the test T by computing the mutual information values

$$\mathbf{I}(\mathbf{L}_k; \mathbf{P}_{k'}) = \mathbf{H}(\mathbf{L}_k) - \mathbf{H}(\mathbf{L}_k | \mathbf{P}_{k'}), \quad (1)$$

between $\mathbf{L}_k$ and $\mathbf{P}_{k'}$ for all key hypotheses k' . The value of k' that leads to the highest value of mutual information is an adversary's best guess. In the above formula $\mathbf{H}(\cdot)$ denotes Shannon entropy. Throughout the paper we assume that the sampling of probability distributions is sufficient such that entropy estimations $\mathbf{H}(\cdot)$ are good. In our experiments we apply our histogram method from [7] to estimate said distributions.

2.1 Masking

A typical way to protect an implementation of a cryptographic algorithm against such 1st order attacks is to mask the occurring intermediate values. Masking is usually implemented by combining the intermediate values with random data and by adapting the algorithm accordingly. The effect of masking is that each intermediate value that is predictable by an attacker is pairwise uncorrelated to the masked intermediate values that are actually processed.

In the following we formalize these notions. Let $\mathbf{M}$ denote a RV on $\mathcal{W}$ with uniform PMF. Masking is implemented by replacing the intermediate value $\mathbf{W}_k$

by $\mathbf{W}_{k,M} = \mathbf{W}_k \circ \mathbf{M}$ where $\circ$ is a suitable operation. In masked block cipher implementations, for example, one often uses the exclusive-or operation and replaces $\mathbf{W}_k$ by $\mathbf{W}_{k,M} = \mathbf{W}_k \oplus \mathbf{M}$. Non-linear operations that are typically implemented as S-box tables $\text{S-box}(\mathbf{W}_k^{in})$ can be implemented with a recomputed S-box such that $\mathbf{W}_{k,M}^{out} = \text{S-box}'(\mathbf{W}_{k,M}^{in}) = \text{S-box}(\mathbf{W}_k^{in}) \oplus \mathbf{M}$, see Fig. 1 (left).

If $\mathbf{M}$ is a RV with uniform PMF, intermediate results $\mathbf{W}_k$ predictable by an adversary (i.e. $k' = k$) are not correlated to computed intermediate results $\mathbf{W}_k \oplus \mathbf{M}$. It follows that the actual leakage $\mathbf{L}_{k,M}$ of a masked value is independent of any predictable leakage $\mathbf{P}_k$ and therefore $\mathbf{I}(\mathbf{L}_{k,M}; \mathbf{P}_k) = 0$.

2.2 Higher-Order Attacks

The mounting point for 2^{nd} order attacks is the fact that the side-channel leakage $\mathbf{L}_{k,M}$ of a masked value depends on a predictable value $\mathbf{W}_k$ and an unpredictable value $\mathbf{M}$. The core idea is to jointly analyze the leakage of the masked value and the mask (or a second value masked with the same mask) to establish a relation to the predictable $\mathbf{W}_k$ or its predictable leakage $\mathbf{P}_k$. In the example in Fig. 1 (left) one could for example combine the leakage $\mathbf{L}_M$ of the mask at time τ_1 with the leakage $\mathbf{L}_{k,M}$ of the masked S-box output at time τ_4 .

Fig. 1. Left: Masked S-box lookup with recomputed S-box. Right: Schematic of 2^{nd} order DPA with pre-processing functions g and h that output correlated values.

2^{nd} order DPA requires a suitable pre-processing that combines the leakage of two masked RVs to construct a signal that is correlated to the unmasked intermediate result (or a function thereof) and can be attacked with 1^{st} order DPA. More formally that is: one looks for functions $g(\mathbf{L}_M, \mathbf{L}_{k,M})$ and $h(\mathbf{W}_k)$ that yield highly correlated values [14], see Fig. 1 (right). These values can be attacked with 1^{st} order DPA.

Most proposals for the pre-processing do not consider the function h but focus on a function g whose outputs are correlated with $\mathbf{P}_k$. Such proposals typically assume that the leakage at all relevant time instants approximately follows the Hamming weight model and use Pearson's correlation coefficient (ρ) as distinguisher. More precisely and sticking to the above example, the adversary evaluates

$$\rho(g(\mathbf{L}_M, \mathbf{L}_{k,M}), \text{HW}(\mathbf{W}_{k'})) \quad (2)$$

for all hypotheses k' . We provide background information on HODPA attacks in Appendix 6. The state-of-the-art combining function is the normalized product [17]

$$g(\mathbf{L}_M, \mathbf{L}_{k,M}) = (\mathbf{L}_M - E(\mathbf{L}_M)) \cdot (\mathbf{L}_{k,M} - E(\mathbf{L}_{k,M})) \quad (3)$$

where $E(\cdot)$ denotes expectation or the empirical mean. This combining function was first hinted by Chari et al. [2] and more thoroughly analyzed by Prouff, Rivain, and Bevan [17]. They showed that this combining function maximizes the Pearson coefficient in Eq. (2) for the correct key guess k in the Hamming weight model. We note, however, that this combining function is not necessarily optimal, since it was not shown that wrong key guesses can not lead to similarly high Pearson coefficients.

3 Motivation

An adversary faces three essential problems when mounting a HODPA:

1. How to identify the points of interest τ_j when the interesting intermediate values leak?
2. How to model the power consumption at these points in time? This question is particularly interesting as the power consumption model need not be the same at several instants, e.g. while a random number generator is active vs. during a table lookup.
3. How to choose the functions g and h for the pre-processing? This question is particularly interesting because the answer is tightly linked to the previous question.

In this paper we will not deal with the first problem but assume that the instants τ_j are known.¹ The same assumption is made in all related literature except for [15].

In our view most previous work discusses 2^{nd} and HODPA in specific contexts and under restrictive assumptions, that are not always met in practice [15,18] but the drawbacks have been accepted. In particular we point out that most contributions assume i) the linear Hamming weight or distance model and implicitly require them for the attack to work, due to the choice of g (and h where applicable); and ii) that the leakage functions associated to the computation of different intermediate results, where different parts of the device may be active, are the same or very similar.

It is surprising that, while DPA and its higher-order variants have been published more than 10 years ago, the problem of finding an optimal pre-processing (except for very specific contexts) remains unsolved.

A weakness of earlier work is that tight link between leakage model(s) and pre-processing. It is evident that a pre-processing tailored to specific leakage functions loses all meaning if the leakage models are not met.

---

¹ Some advice can be found in Appendix 6.

Two works that relax the above mentioned assumptions or that could be accordingly adapted [1,14] deal with variants of template attacks [3], which consider an adversary who is able to characterize the leakage function(s) of the target device and the implementation as well as the electrical properties of the measurement setup. The adversarial context of a profiled attack is, however, beyond the scope of this paper. The third work relaxing above assumptions is a recent and independent paper by Prouff and Rivain [16], which we will address in Sect. 4.1.

3.1 Problem Statement

We formulate a generic 2^{nd} order DPA problem that relaxes all (but one) of the above mentioned assumptions and requirements. Informally speaking, we ask “what is possible” if i) the power models at the two (known) instants are unknown and possibly substantially different, which implies ii) that the best choice for g and h in the pre-processing is *a priori* unknown. Further, the sought method should naturally extend to attacks of order greater than two. A sound solution would be a powerful tool that allows successful HO attacks in a range of scenarios otherwise resistant or inaccessible to standard attacks due to intrinsic errors introduced in the pre-processing [2,10,13,15,18,21].

Formally, let $\mathbf{L}_{\tau_1}$ denote the leakage of a masked value at time τ_1 and $\mathbf{L}_{\tau_2}$ denote the leakage of the corresponding mask (or a value masked with the same mask) at instant τ_2 . Given leakage $(\mathbf{L}_{\tau_1}, \mathbf{L}_{\tau_2})$ determine k with non-negligible advantage over a random guess. Note that solving the problem does not necessarily require a transformation step (i.e. a pre-processing).

4 Extending MIA to Multivariate Analysis

It is natural to look for methods that can solve the above stated problems without transforming them to other problems while relying on possibly wrong assumptions. In [7] we showed that MIA is a 1^{st} order attack that works without restrictive assumptions about the leakage function. We can thus use MIA to solve the problem of unknown leakage functions at instants τ_1 and τ_2 . Since MIA is well suited to exploit dependencies between RVs without making a restrictive assumption about *how* the RVs are related, it appears natural to also use this technique to solve the second problem, i.e. combining the information contained in $\mathbf{L}_{\tau_1}$ and $\mathbf{L}_{\tau_2}$.

Our extension of MIA to a multivariate scenario is straight forward: one merely computes the multivariate mutual information of three RVs

$$\mathbf{I}(\mathbf{P}_{k'}; \mathbf{L}_{k,M}; \mathbf{L}_M). \quad (4)$$

In [5,6,12] it is shown that

$$\mathbf{I}(\mathbf{X}; \mathbf{Y}; \mathbf{Z}) = \mathbf{I}(\mathbf{X}; \mathbf{Y}) - \mathbf{I}(\mathbf{X}; \mathbf{Y}|\mathbf{Z}) \quad (5)$$

where both terms on the right hand side of the equation can be computed using Eq. (1).

Depending on the source, Eq. (5) is either called multivariate mutual information or mutual interaction. It is clear that Eq. (5) can have positive and negative values depending on the relation between the arguments. For example, if $\mathbf{X}$ and $\mathbf{Y}$ are independent but possibly related through $\mathbf{Z}$ as in our context, then

$$\mathbf{I}(\mathbf{X}; \mathbf{Y}; \mathbf{Z}) = \mathbf{I}(\mathbf{X}; \mathbf{Y}) - \mathbf{I}(\mathbf{X}; \mathbf{Y}|\mathbf{Z}) = 0 - \mathbf{I}(\mathbf{X}; \mathbf{Y}|\mathbf{Z}) \leq 0$$

and one says that $\mathbf{Z}$ explains the (in-)dependence of $\mathbf{X}$ and $\mathbf{Y}$. Note that the choice of how to substitute the arguments is arbitrary, any combination works. The MMIA key recovery attack decides for the key hypothesis k' that minimizes expression (4). For the more general case of n^{th} order MMIA attacks one computes

$$\mathbf{I}(\mathbf{X}_1; \dots; \mathbf{X}_{N+1}) = \mathbf{I}(\mathbf{X}_1; \dots; \mathbf{X}_N) - \mathbf{I}(\mathbf{X}_1; \dots; \mathbf{X}_N | \mathbf{X}_{N+1}).$$

We want to emphasize that our proposal has one clear advantage: there is no need to assume leakage functions neither to choose the functions g and h for the pre-processing. This makes it generic and applicable in virtually any scenario.

4.1 Generalized MIA

Simultaneous but independent of this work, Prouff and Rivain proposed a different extension of MIA to multivariate analysis [16]. In the case of a 2^{nd} order attack they suggest to evaluate the classical formula for mutual information

$$\mathbf{I}(\mathbf{P}_{k'}; (\mathbf{L}_{k,\mathbf{M}}, \mathbf{L}_{\mathbf{M}})) \quad (6)$$

between the predicted leakage and the pair of leakage observations. For n^{th} order attacks they suggest to compute $\mathbf{I}(\mathbf{X}_1; (\mathbf{X}_2, \dots, \mathbf{X}_{N+1}))$. We note that this approach shares all the advantages over HODPA we pointed out for our approach. In the next section we investigate how both approaches perform in practice and how they compare to state-of-the-art HODPA.

5 Reality Check

We study our approach and confront its performance with HODPA using the normalized product preprocessing [2,17] and the generalized MIA variant of [16] in two scenarios.

We consider a Boolean masking scheme as for example described in [8] for DES or triple-DES. For simplicity we focus on a representative mini-cipher consisting of a masked S-box lookup of the first DES S-box (S_1). In practice, we pre-compute the values $\mathbf{M} = m_i$ and $\mathbf{W}_k \circ \mathbf{M} = S_1(x_i \oplus k) \oplus m_i$ for each encryption i on a desktop computer and send them to the card, which successively moves the values over its data-bus generating leakage $\mathbf{L}_M$ and $\mathbf{L}_{k,\mathbf{M}}$. The data bus is reset to 0x00 before and after each memory access. All attacks are provided with the physical measurements of $\mathbf{L}_M$ and $\mathbf{L}_{k,\mathbf{M}}$. Note that unmasked values are never processed by the card and that $\mathbf{M}$ as well as $\mathbf{W}_{k,\mathbf{M}}$ exist on $\{0, 1\}^4$.

For our experiments we use an 8-bit RISC microcontroller based smartcard. The power measurements represent the voltage drop over a 10Ω resistor inserted in the smart-card's GND. We implemented two scenarios: A) 1^{st} order masking exactly as described above; B) 2^{nd} order masking, i.e. the S-Box output is concealed by two masks $\mathbf{M}_1$ and $\mathbf{M}_2$ and the card generates leakage $\mathbf{L}_{M_1}$, $\mathbf{L}_{M_2}$, and $\mathbf{L}_{k, \mathbf{M}_1, \mathbf{M}_2}$. In both scenarios the leakage of the device is very close to the linear Hamming weight model and affected by little noise. Given the mask values, we obtain Pearson coefficients $\rho > 0.99$.

We apply the MIA variants and HODPA assuming the Hamming weight leakage model. In this configuration the scenarios are very well suited to study the impact of the pre-processing in HODPA, i.e. a potential loss of information, versus the multivariate approaches without pre-processing. We also apply the attacks without assuming the Hamming weight leakage model but instead assume a generic leakage model, namely the identity function. In this configuration our experiments allow to study how much an attack depends on a good leakage model.

More precisely, all attacks target the same unmasked intermediate result $\mathbf{W}_k = S_1(\mathbf{X} \oplus k)$ which does not give rise to 1^{st} order leakage. When assuming the Hamming weight leakage model we make leakage predictions of the form $\mathbf{P}_{k'} = \text{HW}(\mathbf{W}_{k'})$. Without this assumption predictions are of the form $\mathbf{P}_{k'} = \mathbf{W}_{k'}$.

For the MIA variants the assumed leakage model further affects the number of bins for the histograms which we use to estimate densities [7]. The densities are required to compute entropy and mutual information values. When making the Hamming weight assumption we use five bins, since the Hamming weight of a 4-bit variable can take five distinct values, and otherwise we use sixteen bins, since $2^4 = 16$.

For HODPA we use the normalized product combining function (see Eq. (3)) and Pearson's correlation coefficient.

Following the framework for the comparison of side-channel distinguishers [20] we use the first-order success rate to assess the performance of the attacks. The first-order success rate expresses the probability that, given n measurements, the attack's best guess is the correct key. For each scenario, we acquired a set of 100 000 power curves using random masks and plaintexts. To evaluate the attacks under the Hamming weight assumption in the 2^{nd} order attack scenario, for example, we split the set into 1000 subsets v_i ($i = 1, \dots, 1000$) of 100 curves and do the following:

```

for  $n := 10$  to 100
  counter  $\leftarrow 0$ 
  for  $i := 1$  to 1000
    i. select the first  $n$  curves from set  $v_i$ 
    ii. run the attack for  $k' \in \{0, 1\}^6$ 
    iii. increase counter if attack successful
  compute success rate for  $n$  curves as counter/1000.

```


In other configurations the attacks require more measurements and we use less subsets each containing more curves instead. In both scenarios a HODPA attack is considered successful if the correct key guess leads to the highest correlation coefficient in absolute terms, a MMIA attack (this paper) is considered successful if the correct key guess minimizes the multivariate mutual information, and a generalized MIA attack [16] is considered successful if the correct key guess maximizes the mutual information.

Figure 2 shows the results of the 2^{nd} order attacks in scenario A.

Fig. 2. First-order success rates of 2^{nd} order attacks: MMIA (solid), generalized MIA (dash-dotted), HODPA (dotted), left: assuming a HW leakage model, right: assuming an identity leakage model; the vertical line shall remind the reader that the X-axis' scale is not the same

On the left hand side of the figure we show how the attacks perform under the Hamming weight assumption. We can see that all three attacks perform very similar. They require about 40 measurements to achieve a success rate of 50% and about 100 measurements to reach a success rate close to 100%. The results for both MIA variants using the (correct) Hamming weight assumption are particularly interesting as these attacks take uncertainty about the leakage functions out of the equation and show the impact of sound joint statistics. The fact that the HODPA performs similarly well in this experiment supports the optimality claim for normalized product 2^{nd} order DPA in the Hamming weight model of [17].

On the right hand side of the figure we show how the attacks perform under the assumption of an identity leakage model. We can see that all attacks are affected by this change, but that the performance of both MIA variants decreases more than the performance of HODPA. HODPA requires about 200 measurements to reach a success rate close to 100% while both MIA variants require about 500 measurements. This observation can be explained by two facts. First, the identity function and the Hamming weight of a 4-bit variable have a strong linear correlation [16]. This explains why the performance of the HODPA decreases not that much. Second, the MIA variants now use sixteen instead of five bins for the histograms. Therefore more samples are required to obtain good density estimations. This explains why the performance of both MIA variants decreases similarly and quite drastically.

We conclude that each 2^{nd} order attack that uses the correct power model and sound joint statistics works very efficient. However, we remind that scenario A is a somewhat easy target because device's leakage almost perfectly follows the Hamming weight model and because the noise level in the measurements is low. In [16] it was shown that generalized MIA is less affected by an increase of noise than HODPA. It is reasonable to assume that MMIA behaves similarly.

In scenario B we want to study how the MIA variants and HODPA scale with respect to attacks of order greater than two, and extended scenario A to 2^{nd} order masking. The S-box output values are now concealed by two independent random masks. The combining function of HODPA computes the normalized product

$$(\mathbf{L}_{M_1} - E(\mathbf{L}_{M_1})) \cdot (\mathbf{L}_{M_2} - E(\mathbf{L}_{M_2})) \cdot (\mathbf{L}_{k,M_1,M_2} - E(\mathbf{L}_{k,M_1,M_2})).$$

The MIA variants compute

$$I(\mathbf{L}_{M_1}; \mathbf{L}_{M_2}; \mathbf{L}_{k,M_1,M_2}; \mathbf{P}_{k'}) \text{ and } I((\mathbf{L}_{M_1}, \mathbf{L}_{M_2}, \mathbf{L}_{k,M_1,M_2}); \mathbf{P}_{k'})$$

respectively. Note that the leakage behavior of the card is still close to the Hamming weight model and that the noise level is low.

Figure 3 shows our experimental results for the 3^{rd} order attacks in scenario B. On the left hand side of the figure we show how the attacks perform under

Fig. 3. First-order success rates of 3^{rd} order attacks: MMIA (solid), generalized MIA (dash-dotted), HODPA (dotted), left: assuming a HW leakage model, right: assuming an identity leakage model

the Hamming weight assumption. We can see that both MIA variants perform again very similar. They require about 800 measurements to achieve a success rate close to 100%. HODPA performs significantly worse. Even using 1000 measurements the attack merely achieves a success rate of about 40%. As before the results for both MIA variants using the (correct) Hamming weight assumption are particularly interesting because they allow us to focus on the impact of sound joint statistics. The fact that HODPA performs significantly worse can not be assigned to a wrong assumption in the leakage model and can therefore be best explained by information loss during the pre-processing. Note that the optimality claim of [17] only holds for 2^{nd} order attacks.

On the right hand side of the figure we show how the attacks perform under the assumption of an identity leakage model. We can see that the performance of HODPA is only slightly decreased by this change. Using 1000 measurements the attack achieves a success rate of about 30%. This observation supports our theory that the efficiency of HODPA in a 3^{rd} order attack scenario is mostly limited by sub-optimal pre-processing. The impact of the change in the assumed leakage model on the MIA variants is more visible but much weaker than in scenario A. Both attacks require about 900 measurements to achieve close to 100% success. This rather minor increase (recall that both attacks now generate histograms using sixteen instead of five bins) indicates that the large number of samples is not required to estimate each separate density sufficiently well. Instead, it is required to extract the complex interrelation from all of them at once.

We conclude that an information theoretic approach using multivariate statistics is clearly preferable over HODPA in attacks of order greater than two. While the prevailing opinion is that the measurement cost of HO attacks grows exponentially with the order of the attack [2,15,18], we demonstrate that attacks of order up to three are realistic and practical.

6 Conclusion

Confronted with a new problem, one typically first tries to transform it into another problem for which one knows the solution. HODPA attacks seem to be such a problem. They inherently demand multivariate statistics because multiple aspects of signals have to be analyzed jointly. However, most publications on HO-attacks follow the approach to first apply a pre-processing function to map the multivariate problem to a univariate problem and then to apply established 1^{st} order techniques. All proposed pre-processing functions have drawbacks that are accepted at the price of an exponential growth of the measurement cost with the attack order.

We propose a novel and different approach for HO attacks that does not suffer from intrinsic errors but solves the initial problem directly. Further, we present the first experimental results for the considered 2^{nd} and 3^{rd} order attacks based on real measurements. The empirical evidence confirms the advantage of MMIA over established HODPA in particular in 3^{rd} order attack scenarios. Our results also show that MMIA and generalized MIA [16] perform similar in practice. As a consequence, the security provided by the masking countermeasure needs to be reconsidered as 3^{rd} and possibly higher order attacks become more practical. The typically implemented combination of masking and temporal randomization [9] should render attacks using either MIA variant more difficult.

Directions for further research are: a detailed comparison of MMIA and generalized MIA, combining functions for HODPA of order greater than two, and methods for the identification of the interesting points in time without device profiling step.

Acknowledgments and Disclaimer

This work was supported in part by the IAP Programme P6/26 BCRYPT of the Belgian State, by FWO project G.0300.07, by the European Commission under contract number ICT-2007-216676 ECRYPT NoE phase II, and by the K.U. Leuven-BOF (OT/06/40).

The information in this document reflects only the author's views, is provided as is and no guarantee or warranty is given that the information is fit for any particular purpose. The user therefore uses the information at its sole risk and liability.

References

1. Agrawal, D., Rao, J.R., Rohatgi, P., Schramm, K.: Templates as Master Keys. In: Rao, J.R., Sunar, B. (eds.) CHES 2005. LNCS, vol. 3659, pp. 15–29. Springer, Heidelberg (2005)
2. 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, pp. 398–412. Springer, Heidelberg (1999)
3. 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. 172–186. Springer, Heidelberg (2003)
4. Coron, J.-S., Goubin, L.: On Boolean and Arithmetic Masking against Differential Power Analysis. In: Paar, C., Koç, Ç.K. (eds.) CHES 2000. LNCS, vol. 1965, pp. 231–237. Springer, Heidelberg (2000)
5. Cover, T.M., Thomas, J.A.: Elements of Information Theory. John Wiley & Sons, Chichester (2006)
6. Fano, R.M.: Transmission of Information: A Statistical Theory of Communications. MIT Press, Cambridge (1961)
7. Gierlichs, B., Batina, L., Tuyls, P., Preneel, B.: Mutual Information Analysis - A Generic Side-Channel Distinguisher. In: Oswald, E., Rohatgi, P. (eds.) CHES 2008. LNCS, vol. 5154, pp. 426–442. Springer, Heidelberg (2008)
8. Goubin, L., Patarin, J.: DES and Differential Power Analysis (The “Duplication” Method). In: Koç, Ç.K., Paar, C. (eds.) CHES 1999. LNCS, vol. 1717, pp. 158–172. Springer, Heidelberg (1999)
9. Herbst, C., Oswald, E., Mangard, S.: An AES Smart Card Implementation Resistant to Power Analysis Attacks. In: Zhou, J., Yung, M., Bao, F. (eds.) ACNS 2006. LNCS, vol. 3989, pp. 239–252. Springer, Heidelberg (2006)
10. 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)
11. Kocher, P., Jaffe, J., Jun, B.: Differential Power Analysis. In: Wiener, M. (ed.) CRYPTO 1999. LNCS, vol. 1666, pp. 388–397. Springer, Heidelberg (1999)
12. McGill, W.J.: Multivariate Information Transmission. *Psychometrika* (19), 97–116 (1954)
13. 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)

14. Oswald, E., Mangard, S.: Template Attacks on Masking - Resistance Is Futile. In: Abe, M. (ed.) CT-RSA 2007. LNCS, vol. 4377, pp. 243–256. Springer, Heidelberg (2006)
15. Oswald, E., Mangard, S., Herbst, C., Tillich, S.: Practical Second-Order DPA Attacks for Masked Smart Card Implementations of Block Ciphers. In: Pointcheval, D. (ed.) CT-RSA 2006. LNCS, vol. 3860, pp. 192–207. Springer, Heidelberg (2006)
16. Prouff, E., Rivain, M.: Theoretical and Practical Aspects of Mutual Information Based Side Channel Analysis. In: Abdalla, M., Pointcheval, D., Fouque, P.-A., Vergnaud, D. (eds.) ACNS 2009. LNCS, vol. 5536, pp. 499–518. Springer, Heidelberg (2009)
17. Prouff, E., Rivain, M., Bevan, R.: Statistical Analysis of Second Order Differential Power Analysis. IEEE Transactions on Computers (58-6), 799–811 (2009)
18. Schramm, K., Paar, C.: Higher Order Masking of the AES. In: Pointcheval, D. (ed.) CT-RSA 2006. LNCS, vol. 3860, pp. 208–225. Springer, Heidelberg (2006)
19. Standaert, F.-X., Gierlichs, B., Verbauwhede, I.: Partition vs. Comparison Side-Channel Distinguishers. In: Lee, P.J., Cheon, J.H. (eds.) ICISC 2008. LNCS, vol. 5461, pp. 253–267. Springer, Heidelberg (2009)
20. 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)
21. Waddle, J., Wagner, D.: Towards Efficient Second-Order Power Analysis. In: Joye, M., Quisquater, J.-J. (eds.) CHES 2004. LNCS, vol. 3156, pp. 1–15. Springer, Heidelberg (2004)

Appendix

A More Background on HODPA

Early works mentioned two essential options for the pre-processing: the product of the two leaked values and the absolute value of their difference. The first work showing a practical higher-order attack to defeat the masking countermeasure came from Messerges [13]. He assumed that the device leaks the Hamming weight of intermediate values (i.e. $\mathbf{L}_k = \text{HW}(\mathbf{W}_k)$) and proposed to compute the absolute difference $|\mathbf{L}_{k,\mathbf{M}} - \mathbf{L}_{\mathbf{M}}|$ in the pre-processing (abs-diff-DPA). Messerges showed that, when focusing on a single bit,

$$|\text{HW}(\mathbf{W}_k \oplus \mathbf{M}) - \text{HW}(\mathbf{M})| = \text{HW}(\mathbf{W}_k \oplus \mathbf{M} \oplus \mathbf{M}) = \text{HW}(\mathbf{W}_k). \quad (7)$$

Thus, the pre-processing reveals the unmasked $\mathbf{L}_k = \text{HW}(\mathbf{W}_k)$ which can be attacked with 1st order DPA. If one wants to attack more than a single bit simultaneously Eq. (7) changes to

$$\text{HW}(\mathbf{W}_k) = \text{HW}(\mathbf{W}_k \oplus \mathbf{M}) + \text{HW}(\mathbf{M}) - 2 \cdot \text{HW}((\mathbf{W}_k \oplus \mathbf{M}) \wedge \mathbf{M}) \quad (8)$$

where $\wedge$ denotes bitwise AND. However, an adversary cannot evaluate this function because $\mathbf{W}_k \oplus \mathbf{M}$ and $\mathbf{M}$ are unknown.

The Hamming weight assumption was also used by Oswald et al. in [15]. They showed that the idea of Eq. (7) can still be used to attack multiple bits

although the equality no longer holds. For 8-bit variables, the Pearson correlation coefficient (ρ) between the predictable $\text{HW}(\mathbf{W}_k)$ and the output of the abs-diff pre-processing decreases to 0.24.

Chari et al. [2] suggested to use the product $\text{HW}(\mathbf{W}_k \oplus \mathbf{M}) \cdot \text{HW}(\mathbf{M})$ in the pre-processing (product-DPA). Their technique does not require the ideal Hamming weight model but still makes some restrictive assumptions about the leakage and power consumption behavior and is in practice more vulnerable to deviations from the model. Waddle and Wagner [21] were the first to clearly split higher-order attacks into pre-processing and attack step as we present them here. They proposed a few variants of product-DPA that differ in complexity. The work of Joye et al. [10] introduces a more theoretical approach to 2^{nd} order DPA. The authors analyzed single bit 2^{nd} order abs-diff-DPA in the Hamming weight model, as introduced by Messerges, and in the Hamming distance model. They suggest to use a power of the absolute difference in the pre-processing, which yields a slightly higher coefficient ρ [15].

B Identifying the Points of Interest

One approach towards identifying these instants may be to examine the empirical variance of several power traces obtained during processing of constant input data. In this case, the variance in the power traces is mostly caused by the masking and thus reveals the points in time when masked values are processed. Another approach is to select a small time window based on an educated guess and to perform an exhaustive search over all pairs of time instants [15].