Making Masking Security Proofs Concrete
Or How to Evaluate the Security of Any Leaking Device

Alexandre Duc[1], Sebastian Faust[1,2], and François-Xavier Standaert[3](✉)

[1] EPFL, Lausanne, Switzerland
{alexandre.duc,sebastian.faust}@epfl.ch
[2] Horst Görtz Institute, Ruhr-University Bochum, Bochum, Germany
[3] ICTEAM/ELEN/Crypto Group, Université catholique de Louvain,
Louvain-la-neuve, Belgium
fstandae@uclouvain.be

Abstract. We investigate the relationships between theoretical studies of leaking cryptographic devices and concrete security evaluations with standard side-channel attacks. Our contributions are in four parts. First, we connect the formal analysis of the masking countermeasure proposed by Duc et al. (Eurocrypt 2014) with the Eurocrypt 2009 evaluation framework for side-channel key recovery attacks. In particular, we re-state their main proof for the masking countermeasure based on a mutual information metric, which is frequently used in concrete physical security evaluations. Second, we discuss the tightness of the Eurocrypt 2014 bounds based on experimental case studies. This allows us to conjecture a simplified link between the mutual information metric and the success rate of a side-channel adversary, ignoring technical parameters and proof artifacts. Third, we introduce heuristic (yet well-motivated) tools for the evaluation of the masking countermeasure when its independent leakage assumption is not perfectly fulfilled, as it is frequently encountered in practice. Thanks to these tools, we argue that masking with non-independent leakages may provide improved security levels in certain scenarios. Eventually, we consider the tradeoff between measurement complexity and key enumeration in divide-and-conquer side-channel attacks, and show that it can be predicted based on the mutual information metric, by solving a non-linear integer programming problem for which efficient solutions exist. The combination of these observations enables significant reductions of the evaluation costs for certification bodies.

1 Introduction

Side-channel attacks are an important concern for the security of cryptographic hardware, and masking is one of the most investigated solutions to counteract them. Its underlying principle is to randomize any sensitive data manipulated by a leaking implementation by splitting it into d shares, and to perform all the

E. Oswald and M. Fischlin (Eds.): EUROCRYPT 2015, Part I, LNCS 9056, pp. 401–429, 2015.
DOI: 10.1007/978-3-662-46800-5_16

computations on these shared values only. Intuitively, such a process is expected to force the adversary to combine several leakages corresponding to the different shares in order to recover secret information. As a result, it has first been shown by Chari et al. that the measurement complexity of a specialized attack – namely a single-bit Differential Power Analysis (DPA) [35] – against a carefully implemented masked computation (i.e. where the leakages of all the shares are independent and sufficiently noisy) increases exponentially with d [14]. Following this seminal work, a number of progresses have been made in order to state the security guarantee of masking in both general and rigorous terms. For example, Ishai, Sahai and Wagner introduced a compiler (next referred to as the ISW compiler), able to encode any circuit into an equivalent (secret-shared) one, and proved its security against so-called probing adversaries, able to read a bounded number of wires in the implementation [33]. A practical counterpart to these results was published at Asiacrypt 2010, where Standaert et al. analyzed the security of several masked implementations [61], using the information theoretic framework introduced in [60]. While this analysis was specialized to a few concrete case studies, it allowed confirming the exponential security increase provided by masking against actual leakages, typically made of a noisy but arbitrary function of the target device's state. Following, Faust et al. attempted to analyze the ISW compiler against more realistic leakage functions, and succeeded to prove its security against computationally bounded (yet still unrealistic) ones, e.g. in the AC^0 complexity class [25]. Prouff and Rivain then made a complementary step towards bridging the gap between the theory and practice of masking schemes, by providing a formal information theoretic analysis of a wide (and realistic) class of so-called noisy leakage functions [49]. Eventually, Duc et al. turned this analysis into a simulation-based security proof, under standard conditions (i.e. chosen-message rather than random-message attacks, without leak-free components, and with reduced noise requirements) [22]. The central and fundamental ingredient of this last work was a reduction from the noisy leakage model of Prouff and Rivain to the probing model of Ishai et al.

Our Contribution. In view of this state-of-the-art, one of the main remaining questions regarding the security of the masking countermeasure is whether its proofs can be helpful in the security evaluation of concrete devices. That is, can we state theorems for masking so that the hypotheses can be easily fulfilled by hardware designers, and the resulting guarantee is reflective of the actual security level of the target implementation. For this purpose, we first observe that the proofs in [22,49] express their hypothesis for the amount of noise in the shares' leakages based on a statistical distance. This is in contrast with the large body of published work where the mutual information metric introduced in [60] is estimated for various implementations (e.g. [4,12,27,30,32,42,50,51,57,63,66]). Since the latter metric generally carries more intuition (see, e.g. [3] in the context of linear cryptanalysis), and benefits from recent advances in leakage certification, allowing to make sure that its estimation is accurate and based on sound assumptions [23], we first provide a useful link between the statistical distance and mutual information, and also connect them with easy-to-interpret (but more specialized) tools such as the Signal-to-Noise Ratio (SNR). We then re-state the theorems of Duc et al. based on the mutual information metric in two relevant scenarios. Namely, we consider both the security of an idealized implementation with a "leak-free refreshing" of the shares, and the one of a standard ISW-like encoding (i.e. capturing any type of leaking computation).

Interestingly, the implementation with leak-free refreshing corresponds to the frequently investigated (practical) context where a side-channel attack aims at key recovery, and only targets the d shares' leakage of a so-called sensitive intermediate variable (i.e. that depends on the plaintext and key) [17]. So despite being less interesting from a theoretical point of view, this scenario allows us to compare the theorem bounds with concrete attacks. Taking advantage of this comparison, we discuss the bounds' tightness and separate parameters that are physically motivated from more "technical ones"(that most likely result of proof artifacts). As a result, we conjecture a simplified link between the mutual information metric and the success rate of a side-channel adversary, which allows accurate approximations of the attacks' measurement complexity at minimum (evaluation) cost. We further illustrate that the noise condition for masking has a simple and intuitive interpretation when stated in terms of SNR.

Next, we note that the published results about masking (including the previously mentioned theorems and conjecture) assume independence between the leakages corresponding to different shares in an implementation. Yet, concrete experiments have shown that small (or even large) deviations from this assumption frequently occur in practice (see, e.g. [5,16,41,54]). Hence, we complete our discussion by providing sound heuristics to analyze the impact of "non-independent leakages" which allow, for the first time, to evaluate and predict the security level of a masked implementation in such imperfect conditions.

Eventually, we consider the tradeoff between measurement complexity and time complexity in the important context of divide-and-conquer attacks. Previously known approaches for this purpose were based on launching key enumeration and/or rank estimation algorithms for multiple attacks, and to average results to obtain a success rate [64,65]. We provide an alternative solution, where success rates (possibly obtained from estimations of the mutual information metric) are estimated/bounded for all the target key bytes of the divide-and-conquer attack first, and the impact of enumeration is evaluated only once afterwards. We also connect the problem of approximating the enumeration cost for a given number of measurements with a non-linear integer programming problem, and provide simple heuristics to estimate bounds on this enumeration cost.

Summarizing, the combination of these observations highlights that the security evaluation of a masked implementation boils down to the estimation of the mutual information between its shares and their corresponding leakages. Incidentally, the tools introduced in this paper apply identically to unprotected implementations, or implementations protected with other countermeasures, as long as one can estimate the same mutual information metric for the target intermediate values. Therefore, our results clarify the long standing open question whether the (informal) link between information theoretic and security metrics in the Eurocrypt 2009 evaluation framework [60] can be proved formally. They also have important consequences for certification bodies, since they translate the (worst-case) side-channel evaluation problem into the well-defined challenge of estimating a single metric, leading to significantly reduced evaluation costs.

Notations. We next use capital letters for random variables, small caps for their realizations and hats for estimations. Vectors will be denoted with bold notations, functions with sans serif fonts, and sets with calligraphic ones.

2 Background

2.1 Leakage Traces and Assumptions

Let y be a n-bit sensitive value manipulated by a leaking device. Typically, it could be the output of an S-box computation such that $y = \mathsf{S}(x \oplus k)$ with n-bit plaintext/key words x and k. Let $y_1, y_2, \ldots, y_d$ be the d shares representing y in a Boolean masking scheme (i.e. $y = y_1 \oplus y_2 \oplus \ldots \oplus y_d$). In a side-channel attack, the adversary is provided with some information (aka leakage) on each share. Typically, this leakage takes the form of a random variable $\boldsymbol{L}_{y_i}$ that is the output of a leakage function L with y_i and a noise variable $\boldsymbol{R}_i$ as arguments:

$$\boldsymbol{L}_{y_i} = \mathsf{L}(y_i, \boldsymbol{R}_i) \ . \tag{1}$$

The top of Fig. 1 represents a leakage trace corresponding to the manipulation of d shares. Concretely, each subtrace $\boldsymbol{L}_{y_i}$ is a vector of which the elements represent time samples. Whenever accessing a single time sample t, we use the notation $\boldsymbol{L}_{y_i}^t = \mathsf{L}^t(y_i, \boldsymbol{R}_i^t)$. From this general setup, a number of assumptions are frequently used in the literature. We will consider the following three.

a. Selection of Points-of-Interest / Dimensionality Reduction. For convenience, a number of attacks start with a pre-processing in order to reduce each leakage subtrace $\boldsymbol{L}_{y_i}$ to a scalar random variable L_{y_i}. Such a pre-processing is motivated both by popular side-channel distinguishers such as Correlation Power Analysis (CPA) [11], which can only deal with univariate data, and by the easier representation of small dimensional data spaces. In this respect, even distinguishers that naturally extend towards multivariate data (such as Template attacks (TA) [15], Linear Regression (LR) [58] or Mutual Information Analysis (MIA) [28]) generally benefit from some dimensionality reduction. This step can be achieved heuristically, by looking for leakage samples where one distinguisher works best, or more systematically using tools such as Principal Component Analysis (PCA) [2] or Linear Discriminant Analysis (LDA) [59]. An example of reduced leakage trace is represented at the bottom of Fig. 1.

b. Additive Noise. A standard assumption in the literature is to consider leakage functions made of a deterministic part $\mathsf{G}(y_i)$ and additive noise $\boldsymbol{N}_i$ [40]:

$$\boldsymbol{L}_{y_i} = \boldsymbol{L}(y_i, \boldsymbol{R}_i) \approx \mathsf{G}(y_i) + \boldsymbol{N}_i \ . \tag{2}$$

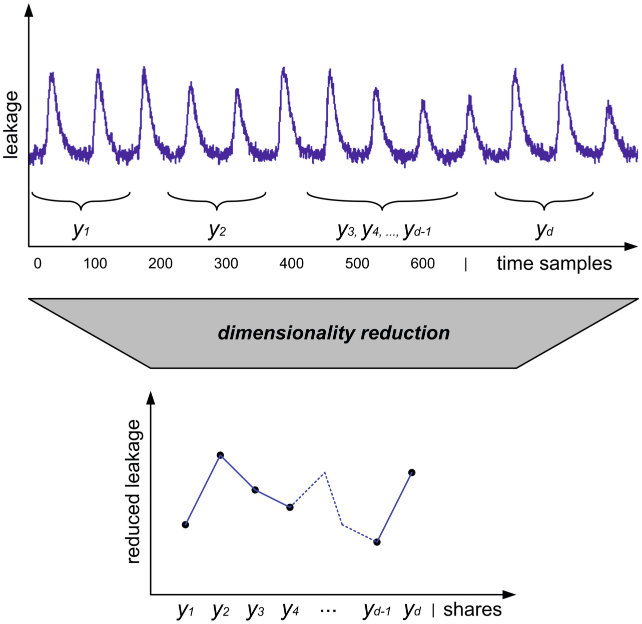

Fig. 1. Leakage trace & reduced leakage trace of a d-shared secret

For example, a typical setting is to assume reduced leakages to be approximately generated as the combination of a Hamming weight function (or some other simple function of the shares' bits [58]) with additive Gaussian noise.

c. Independence Condition. A secure implementation of the masking countermeasure requires that the leakage vectors $\boldsymbol{L}_{y_i}$ are independent random variables. If respected, it implies that a d-share masking will lead to a $(d-1)^{\text{th}}$-order secure implementation as defined in [17]. That is, it guarantees that every d-tuple of leakage vectors is independent of any sensitive variable. This means that any adversary targeting the implementation will have to "combine"the information of at least d shares, and that extracting information from these d shares will require to estimate a d^{th}-order moment of the leakage PDF (conditioned on a sensitive variable) – a task that becomes exponentially hard in d if the noise is sufficient. As witnessed by several prior works, this condition may be hard to fulfill in practice. In software implementations, it typically requires avoiding transition-based leakages (i.e. leakages that depend on the distance between shares rather than directly on the shares) [5,16]. In hardware implementations, physical defaults such as glitches are another usual issue that can invalidate the independence assumption [41], which motivates various research efforts to mitigate this risk, both at the hardware level (e.g. [43]) and at the algorithmic level (e.g. [46]).

Note that only this last (independence) assumption is strictly needed for the following proofs of Sect. 3 to hold. By contrast, the previous assumptions (a) and (b) will be useful to provide practical intuition in Sect. 4. Furthermore, it is worth noting that slight deviations from this independence assumption (i.e. slight dependencies between the shares' leakages) may still lead to concrete security improvements, despite falling outside the proofs' formal guarantees. Such (practically meaningful) contexts will be further analyzed in Sect. 4.2.

2.2 Evaluation Metrics

Following [60], one generally considers two types of evaluation metrics for leaking cryptographic devices. First, information theoretic metrics aim to capture the amount of information available in a side-channel, independent of the adversary exploiting it. Second, security metrics aim to quantify how this information can be exploited by some concrete adversary. As will be clear next, the two types of metrics are related. For example, in the context of standard DPA attacks [41], they both measure the prediction of the (true) leakage function with some model, the latter usually expressed as an estimation of the leakage Probability Density Function (PDF). Yet they differ since information theoretic metrics only depend on the leakage function and model, while security metrics also depend on the adversary's computational power. For example, the capacity to enumerate key candidates may improve security metrics, but has no impact on information theoretic ones [64,65]. Our goal in the following is to draw a formal connection between information theoretic and security metrics, i.e. between the amount of leakage provided by an implementation and its (worst-case) security level.

In the case of masking, proofs informally state that "*given that the leakage of each share is independent of each other and sufficiently noisy, the security of the implementation increases exponentially in the number of shares*". So we need the two types of metrics to quantify the noise condition and security level.

b. Metrics to Quantify the Noise Condition. In general (i.e. without assumptions on the leakage distribution), the noise condition on the shares can be expressed with an information theoretic metric. The Mutual Information (MI) advocated in [60] is the most frequently used candidate for this purpose:

$$\mathrm{MI}(Y_i;\boldsymbol{L}_{Y_i}) = \mathrm{H}[Y_i] + \sum_{y_i\in\mathcal{Y}} \Pr[y_i]\cdot \sum_{\boldsymbol{l}_{y_i}\in\mathcal{L}} \Pr[\boldsymbol{l}_{y_i}|y_i]\cdot\log_2 \Pr[y_i|\boldsymbol{l}_{y_i}]\ , \qquad (3)$$

where we use the notation $\Pr[Y_i = y_i] =: \Pr[y_i]$ when clear from the context. Note that whenever trying to compute this quantity from an actual implementation, evaluators face the problem that the leakage PDF is unknown and can only be sampled and estimated. As a result, one then computes the Perceived Information (PI), which is the evaluator's best estimate of the MI [54]:

$$\hat{\mathrm{PI}}(Y_i;\boldsymbol{L}_{Y_i}) = \mathrm{H}[Y_i] + \sum_{y_i\in\mathcal{Y}} \Pr[y_i]\cdot \sum_{\boldsymbol{l}_{y_i}\in\mathcal{L}} \Pr_{\text{chip}}[\boldsymbol{l}_{y_i}|y_i]\cdot\log_2 \hat{\Pr_{\text{model}}}[y_i|\boldsymbol{l}_{y_i}]\ , \qquad (4)$$

with $\Pr_{\text{chip}}$ the true chip distribution that can only be sampled and $\hat{\Pr}_{\text{model}}$ the adversary's estimated model. For simplicity, we will ignore this issue and use the MI in our discussions (conclusions would be identical with the PI).

Interestingly, when additionally considering reduced leakages with additive Gaussian noise, and restricting the evaluation to so-called "first-order information"(i.e. information lying in the first-order statistical moments of the leakage PDF, which is typically the case for the leakage of each share), simpler metrics can be considered [40]. For example, the SNR introduced by Mangard at CT-RSA 2004 in [38] is of particular interest for our following discussions:

$$\mathrm{SNR} = \frac{\hat{\mathsf{var}}_{Y_i}\left(\hat{\mathsf{E}}_{n_i}(L_{Y_i})\right)}{\hat{\mathsf{E}}_{Y_i}\left(\hat{\mathsf{var}}_{n_i}(L_{Y_i})\right)}\ , \tag{5}$$

where $\hat{\mathsf{E}}$ is the sample mean operator and $\hat{\mathsf{var}}$ is the sample variance. Summarizing, stating the noise condition based on the MI metric is more general (as it can capture any leakage PDF). By contrast, the SNR provides a simpler and more intuitive condition in a more specific but practically relevant context.

Eventually, the previous works of Prouff–Rivain and Duc et al. [22,49] consider the following Statistical Distance (SD) to state their noise condition:

$$\mathrm{SD}(Y_i; Y_i|\boldsymbol{L}_{Y_i}) = \sum_{\boldsymbol{l}_{y_i}\in\mathcal{L}} \Pr[\boldsymbol{l}_{y_i}] \cdot \mathsf{d}(Y_i; Y_i|\boldsymbol{l}_{y_i})\ , \tag{6}$$

with d the Euclidean norm in [49] and $\mathsf{d}(X_1, X_2){=}\frac{1}{2}\sum_{x\in\mathcal{X}} |\Pr[X_1 = x]{-}\Pr[X_2 = x]|$ in [22]. In their terminology, a leakage function L is then called "δ-noisy" if $\delta = \mathrm{SD}(Y_i; Y_i|\boldsymbol{L}_{Y_i})$, which was useful to connect different leakage models.

As previously mentioned, some of these metrics can be related under certain conditions. For example, in the context of univariate Gaussian random variables, the MI can be approximated from Pearson's correlation coefficient [40], which was also connected to the SNR by Mangard [38]. The combination of those links corresponds to the classical MI bound in Cover and Thomas [19]:

$$\mathrm{MI}(Y_i; \boldsymbol{L}_{Y_i}) \approx -\frac{1}{2}\log\left(1 - \left(\frac{1}{\sqrt{(1+\frac{1}{\mathrm{SNR}})}}\right)^2\right) \leq \frac{1}{2}\log\Big(1+\mathrm{SNR}\Big)\cdot \tag{7}$$

In Sect. 3.1, we show that the MI and SD metrics can be connected as well.

c. Metrics to Quantify the Security Result. Quantifying security requires defining the adversary's goal. Current side-channel attacks published in the literature mostly focus on key recovery. In this context, one can easily evaluate the exploitation of the leakages with the success rate defined in [60], i.e. the probability that an adversary recovers the key given the observation of some (typically known or chosen) plaintexts, ciphertexts and leakages. We will next denote it with $\mathrm{SR}^{\mathsf{kr}}$. Key recovery is a weak security notion from a cryptographic point of view. As a result, rigorous proofs for masking such as the one of Duc et al. in [22] rather define security using the standard real/ideal world paradigm, which consider two settings: the ideal world where the adversary attacks the algorithm of a cryptographic scheme in a black-box way, and the real world where he additionally obtains leakages. A scheme is said to be secure in the real world, if for any adversary in the real world there exists an adversary in the ideal world. In other words: any attack that can be carried out given the leakages can also be carried out in a black-box manner. A proof of security usually involves constructing an efficient simulator that is able to simulate the leakages just giving black-box access to the attacked cryptographic scheme. Whenever considering this (standard) indistinguishability-based security notion, we will denote the adversary's success probability of distinguishing the two worlds with $\mathrm{SR}^{\mathsf{dist}}$.

3 Making Proofs Concrete: Theory

In this section, we discuss theoretical tweaks allowing to improve the concreteness of masking proofs. For this purpose, we recall three important leakage models that are relevant for our work. First, the t-probing and ϵ-probing (aka random probing) models were introduced in [33]. In the former one, the adversary obtains t intermediate values of the computation (e.g. can probe t wires if we compute in binary fields). In the latter one, he obtains each of these intermediate values with probability ϵ, and gets $\perp$ with probability $1-\epsilon$ (where $\perp$ means no knowledge). Using a Chernoff-bound it is easy to show that security in the t-probing model reduces to security in the ϵ-probing model for certain values of ϵ. Second, the noisy leakage model describes many realistic side-channel attacks and allows an adversary to obtain each intermediate value perturbed with a δ-noisy leakage function L [49]. As mentioned in the previous section, a leakage function L is called δ-noisy if for a uniformly random variable Y (over the field $\mathbb{F}$) we have $\mathrm{SD}(Y;Y|\boldsymbol{L}_Y) \leq \delta$. In contrast with the conceptually simpler ϵ-probing model, the adversary obtains noisy leakages on each intermediate variable. For example, in the context of masking, he obtains $\mathsf{L}(Y_i, \boldsymbol{R})$ for all the shares Y_i, which is more reflective of actual implementations where the adversary can potentially observe the leakage of all these shares, since they are all present in leakage traces such as in Fig. 1. Recently, Duc et al. showed that security against probing attacks implies security against noisy leakages (up to a factor $|\mathbb{F}|$, where $\mathbb{F}$ is the underlying field in which the operations are carried out) [22]. In the rest of this section, we first connect the statistical distance SD with the mutual information metric MI, which shows that both can be used to quantify the noise condition required for masking. Next, we provide alternative forms for the theorems of Duc et al. and show (i) the security of the encoding used in (e.g. Boolean) masking and (ii) the security of a complete circuit based on the ISW compiler.

3.1 From Statistical Distance to MI

The results from Duc et al. require to have a bound on the SD between the shares and the shares given the leakage. For different reasons, expressing this distance based on the MI metric may be more convenient in practice (as witnessed by the numerous works where this metric has been computed, for various types of devices, countermeasures and technologies – see the list in introduction). For example, the MI metric is useful to determine whether the leakage model used in a standard DPA is sound (see the discussion in Sect. 4.1) and for analyzing the impact of key enumeration in divide-and-conquer attacks (see the discussion in Sect. 4.3). Very concretely, Equations (3) and (4) are also expressed in a way that requires summing over the intermediate values first and on the leakages afterwards, which corresponds to the way security evaluations are performed (i.e. fix the target device's state, and then perform measurements). Thus, we now show how to express the SD in function of the MI. We use a previous result from Dodis [21], which proofs follows [9] that we rephrase with our notations.

Lemma 1 ([21], Lemma 6). *Let Y_i and $\boldsymbol{L}_{Y_i}$ be two random variables. Then:*

$$\frac{1}{2}\left(\sum_{(y\in\mathcal{Y},\ell\in\mathcal{L})} \left|\Pr[Y_i=y,\boldsymbol{L}_{Y_i}=\ell]-\Pr[Y_i=y]\Pr[\boldsymbol{L}_{Y_i}=\ell]\right|\right)^2 \leq \mathrm{MI}(Y_i;\boldsymbol{L}_{Y_i})\;.$$

Using this lemma, we can now express the SD in function of the MI as follows.

Theorem 1. *Let Y_i and $\boldsymbol{L}_{Y_i}$ be two random variables. Then:*

$$2\cdot\mathrm{SD}(Y_i;Y_i\mid\boldsymbol{L}_{Y_i})^2\leq\mathrm{MI}(Y_i;\boldsymbol{L}_{Y_i})\;.$$

Proof. The proof follows the proof of [8], Lemma 4.4. We have:

$$\begin{aligned}&\sum_{(y\in\mathcal{Y},\ell\in\mathcal{L})}\left|\Pr[Y_i=y,\boldsymbol{L}_{Y_i}=\ell]-\Pr[Y_i=y]\Pr[\boldsymbol{L}_{Y_i}=\ell]\right|,\\&=\sum_{\ell\in\mathcal{L}}\Pr[\boldsymbol{L}_{Y_i}=\ell]\sum_{y\in\mathcal{Y}}\left|\Pr[Y_i=y\mid\boldsymbol{L}_{Y_i}=\ell]-\Pr[Y_i=y]\right|,\\&=2\cdot\mathrm{SD}(Y_i;Y_i\mid\boldsymbol{L}_{Y_i})\;.\end{aligned}$$

The final result directly derives from Lemma 1. □

3.2 Security of the Encoding

In this section, we analyze the security of an encoding when m measurements are performed and the encoding is refreshed between each measurements using a leak-free gate. More precisely, we assume that a secret y is secret-shared into d shares $y_1,\ldots y_d$, using an additive masking scheme over a finite field $\mathbb{F}$. Between each measurement, we assume that we take fresh $y_1,\ldots,y_d$ values such that $y=y_1+\cdots+y_d$ (e.g. it could be the Boolean encoding of Sect. 2.1). We also assume that this refreshing process does not leak and first recall a previous result from [22] that relates the random probing model to the noisy model. For conciseness, we call an adversary in the random-probing model a "random-probing adversary", an adversary in the δ-noisy model a "δ-noisy adversary", and an adversary having access to leakages such that $\mathrm{MI}(Y;Y|\boldsymbol{L}_Y)\leq\delta$ a "δ-MI-adversary". However, note that the physical noise (and its quantification with the MI) is a property of the implementation rather than of the adversary.

Lemma 2 ([22], Lemma 3). *Let $\mathcal{A}$ be a δ-noisy adversary on $\mathbb{F}^d$. Then, there exists a $\delta \cdot |\mathbb{F}|$-random-probing adversary $\mathcal{S}$ on $\mathbb{F}^d$ such that for every $(y_1, \ldots, y_d)$, $\mathcal{A}$ and $\mathcal{S}$ produce the same view when applied on $(y_1, \ldots, y_d)$.*

This result enables us to work directly in the random-probing model instead of the noisy leakage model. Next, we study the security of the encoding. As mentioned in introduction, the adversary's goal in this case is to recover the encoded value, which is equivalent to key recovery if this value is a key. In order to make it completely comparable with actual attacks, we also add the number of measurements m used by the adversary as a parameter in our bounds.

Theorem 2. *Let d be the number of shares used for a key encoding, m be the number of measurements, and $\mathrm{MI}(Y_i, \boldsymbol{L}_{Y_i}) \leq t$ for some $t \leq 2/|\mathbb{F}|^2$. Then, if we refresh the encoding in a leak-free manner between each measurement, the probability of success of a key recovery adversary under independent leakage is:*

$$\mathrm{SR}^{\mathsf{kr}} \leq 1 - \left(1 - \left(|\mathbb{F}|\sqrt{t/2}\right)^d\right)^m . \tag{8}$$

Proof. In the random probing model with parameter ϵ, an adversary learns nothing about the secret if there is at least one share that did not leak. Since all the measurements are independent and we use leak-free refreshing gates, we have:

$$\mathrm{SR}^{\mathsf{kr}} \leq 1 - \left(1 - \epsilon^d\right)^m . \tag{9}$$

Let A be a t-MI-adversary on $\mathbb{F}^d$. From Theorem 1, we know that A implies a $\sqrt{t/2}$-noisy-adversary on $\mathbb{F}^d$ and, by Lemma 2, we obtain a $|\mathbb{F}|\sqrt{t/2}$-random-probing adversary on $\mathbb{F}^d$. Letting $\epsilon := |\mathbb{F}|\sqrt{t/2}$ in (9) gives us the result. □

Note that Equation (9) focuses on the impact of the adversary's measurement complexity m on the success rate, which is usually the dominating factor in concrete side-channel analyses. Yet, the impact of time complexity when considering key enumeration will be discussed in Sect. 4.3. Besides and for readability, this equation only includes the terms corresponding to attacks taking advantage of the leakages. We ignore the additional terms corresponding to mathematical cryptanalysis (e.g. exhaustive search) that should be added for completeness. In order to allow us comparing this result with the case where we study the security of a complete circuit encoded with the ISW compiler, we also write our result according to the following corollary (which is less general than Theorem 2).

Corollary 1. *Let d be the number of shares used for a key encoding and m the number of measurements. Then, if we refresh the encoding in leak-free manner between each measurement and for any $\alpha > 0$, the probability of success of a key recovery adversary under independent leakage is:*

$$\mathrm{SR}^{\mathsf{kr}} \leq m \cdot \exp\left(-\alpha d\right), \tag{10}$$

if we have:

$$\mathrm{MI}(Y_i; \boldsymbol{L}_{Y_i}) \leq 2\left(\frac{1}{e^{\alpha}|F|}\right)^2 . \tag{11}$$

Proof. We have:

$$1 - \left(1 - \epsilon^d\right)^m \leq m e^{\log(\epsilon) d} .$$

We want $\log(\epsilon) = -\alpha$. Hence, from Theorem 2, we get our result. □

3.3 Security of the Whole Circuit

In this section, we restate the theorems from Duc et al. when securing a whole circuit with the seminal ISW compiler. The main theorem from [22] bounds the probability of success of a distinguishing adversary in the noisy leakage model. We provide an alternative version of their theorem and, as in the previous section, we relate it to the mutual information instead of the statistical distance.

Theorem 3. *Suppose that we have a circuit of size $|\Gamma|$ protected with the ISW compiler with d shares. Then, the probability of success of a distinguishing adversary under independent leakage is:*

$$\mathrm{SR}^{\mathsf{dist}} \leq |\Gamma| \cdot \exp\left(-\frac{d}{12}\right) = |\Gamma| \cdot 2^{\left(-\frac{d \cdot \log_2(e)}{12}\right)} \leq |\Gamma| \cdot 2^{-d/9}, \tag{12}$$

if we have:

$$\mathrm{MI}(Y_i; \boldsymbol{L}_{Y_i}) \leq 2 \cdot \left(\frac{1}{|F| \cdot (28d + 16)}\right)^2 . \tag{13}$$

Similarly to what we did in the previous section, we also write this corollary.

Corollary 2. *Suppose that we have a circuit of size $|\Gamma|$ protected with the ISW compiler with d shares. Then, if $\mathrm{MI}(Y_i, \boldsymbol{L}_{Y_i}) \leq t$, a distinguisher adversary under independent leakage needs:*

$$d \geq \frac{1 - 16|F|\sqrt{\frac{1}{2}t}}{28|F|\sqrt{\frac{1}{2}t}} \tag{14}$$

shares in order to obtain:

$$\mathrm{SR}^{\mathsf{dist}} \leq |\Gamma| \cdot \exp\left(-\frac{d}{12}\right) \leq |\Gamma| \cdot \exp\left(-\frac{1 - 16|F|\sqrt{\frac{1}{2}t}}{336|F|\sqrt{\frac{1}{2}t}}\right) . \tag{15}$$

Note that the ISW compiler can actually be used to efficiently compute any circuit. For example, the work of Rivain and Prouff at CHES 2010 showed how to adapt the compiler to $|F| = 256$ which leads to efficient masked implementations of the AES [56] (see also various following works such as [13,18,31,57]).

4 Making Proofs Concrete: Practice

In this section, we complement the previous theoretical results with an experimental analysis. Our contributions are threefold. First, we provide and empirical evaluation of the encoding scheme in Sect. 3.2, which allows us to discuss the noise condition and tightness of the bounds in our proofs. We use this discussion to conjecture a simple connection between the mutual information metric and the success rate of a (worst-case) side-channel adversary, and argue that it can lead to quite accurate approximations of the attacks' measurement complexity. Next, we discuss possible deviations from the independent leakage assumption and provide tools allowing one to approximate the security level of concrete devices in such cases. Eventually, we consider the tradeoff between measurement complexity and time complexity in the context of divide-and-conquer side-channel attacks. We show how one can build a side-channel security graph (i.e. a plot of the adversary's success probability bounds in function of both parameters [65]), based only on the estimation of the MI metric for each share of a masking scheme. Along these lines, we eventually provide a formal justification for the physical security evaluation framework proposed at Eurocrypt 2009 [60].

4.1 Experimental Validation

In order to discuss the relevance of the proofs in the previous section, we take the (usual) context of standard DPA attacks defined in [40]. More precisely, we consider the simple case where an adversary targets a single S-box from a block cipher (e.g. the AES) as specified in Sect. 2.1, and obtains leakage variables $\boldsymbol{L}_{y_i} = \mathsf{L}(y_i, \boldsymbol{R}_i)$ for $1 \leq i \leq d$ (the case of multiple S-boxes will be studied in Sect. 4.3). For convenience, we mainly consider the context of mathematically-generated Gaussian Hamming weight leakages, where $\boldsymbol{L}_{y_i} = \mathsf{HW}(y_i) + N_i$, with HW the Hamming weight function and N_i a Gaussian-distributed noise, with variance σ^2. In this respect, we note that we did not mount concrete attacks since we would have had to measure hundreds of different implementations to observe useful trends in practice. Our experiments indeed correspond to hundreds of different noise levels. Yet, we note that devices that exhibit close to Hamming weight leakages are frequently encountered in practice [39]. Furthermore, such a simulated setting is a well established tool to analyze masking schemes (see, e.g. [18] for polynomial masking, [4] for inner product masking and [12] for leakage squeezing). Besides, we also consider random Gaussian leakage functions, of which the deterministic part corresponds to random functions over $\mathcal{Y}$, to confirm that all the trends we put forward are also observed with leakage functions that radically differ from the usual Hamming weight one.

a. Computing the MI Metric. In this DPA setting, we aim to compute the MI between the key and the plaintext and leakages. For conciseness, we use the notations $\overline{Y} = [Y_1, \ldots, Y_d]$ and $\overline{\boldsymbol{L}} = [\boldsymbol{L}_{Y_1}, \ldots, \boldsymbol{L}_{Y_d}]$ for vectors containing the d shares and their corresponding leakages. Then we compute:

$$\begin{aligned}\mathrm{MI}(K;X,\overline{\boldsymbol{L}}_Y) = \mathrm{H}[K] + \sum_{k\in\mathcal{K}} \Pr[k]\cdot \\ \sum_{x\in\mathcal{X},\overline{y}\in\mathcal{Y}^d} \Pr[x,\overline{y}] \cdot \sum_{\overline{\boldsymbol{l}}_y\in\mathcal{L}^d} \Pr[\overline{\boldsymbol{l}}_y|k,x,\overline{y}]\cdot\log_2\Pr[k|x,\overline{\boldsymbol{l}}_y]\,.\end{aligned} \tag{16}$$

While this expression may look quite involved, we note that it is actually simple to estimate in practice, by sampling the target implementation. Evaluators just have to set keys k in their device and generate leakage traces corresponding to (known) plaintexts x and (unknown) shares $\overline{y}$. Say there are $|\mathcal{K}| = n_k$ key candidates and we generate n_t leakage traces $\overline{\boldsymbol{l}}_i$, then, one just assigns probabilities $\hat{p}_i^j$ to each key candidate k_j^*, for each measured trace, as in Table 1. This is typically done using TA or LR. Following, if the correct key candidate is k, the second line of (16) can be computed as $\hat{\mathsf{E}}_i \log_2(\hat{p}_i^k)$. Note that whenever considering the standard DPA setting where the target operations follow a key addition, it is not even necessary to sum over the keys since $\mathrm{MI}(K=k;X,\overline{\boldsymbol{L}}_Y)$ is identical for all k's, thanks to the key equivalence property put forward in [40].

Table 1. Computing key candidate probabilities for MI metric estimation

	Key candidates			
State & leakage	k_1^*	k_2^*	...	$k_{N_k}^*$
$(k,x_1) \rightsquigarrow \overline{\boldsymbol{l}}_1$	$\hat{p}_1^1$	$\hat{p}_1^2$	...	$\hat{p}_1^{n_k}$
$(k,x_2) \rightsquigarrow \overline{\boldsymbol{l}}_2$	$\hat{p}_2^1$	$\hat{p}_2^2$	...	$\hat{p}_2^{n_k}$
...	...	...	...	...
$(k,x_{n_t}) \rightsquigarrow \overline{\boldsymbol{l}}_{n_t}$	$\hat{p}_{n_t}^1$	$\hat{p}_{n_t}^2$	...	$\hat{p}_{n_t}^{n_k}$

Intuitively, $\mathrm{MI}(K;X,\overline{\boldsymbol{L}}_Y)$ measures the amount of information leaked on the key variable K. The framework in [60] additionally defines a Mutual Information Matrix (MIM) that captures the correlation between any key k and key candidates k^*. Using our sampling notations, it can be simply defined as $\mathrm{MIM}_{k,k^*} = \mathrm{H}[K] + \sum_i \log_2(\hat{p}_i^{k^*})$, which directly leads to $\mathrm{MI}(K;X,\overline{\boldsymbol{L}}_Y) = \mathsf{E}_k(\mathrm{MIM}_{k,k})$.

b. Intuition Behind the Noise Condition. Theorems 2 and 3 both require that the MI between the shares and their corresponding leakage is sufficiently small. In other words, they require the noise to be sufficiently large. In this section, we compute the MI metric for both an unprotected implementation (i.e. $d = 1$) and a masked one (i.e. $d = 2$) in function of different parameters.[1] In order to illustrate the computation of this metric, we provide a simple open source code that evaluates the MI between a sensitive variable Y and its Hamming weights, for different noise levels, both via numerical integration (that is only possible for mathematically-generated leakages) and sampling (that is more reflective of the evaluation of an actual device) [1]. In the latter case, an evaluator additionally

[1] For the masked case, we consider univariate leakages corresponding to the parallel setting in [7], for which computing the MI is slightly faster than in the serial one.

has to make sure that his estimations are accurate enough. Tools for ensuring this condition are discussed in [23]. In the following, this sufficient sampling is informally confirmed by the smooth shape of our experimental curves.

We start with the simplest possible plot, where the MI metric is computed in function of the noise variance σ^2. Figure 2 shows these quantities, both for Hamming weight leakage functions and for random ones with output range N_l (in the latter context, the functions for different N_l's were randomly picked up prior to the experiments, and stable across experiments). We also considered different bit sizes ($n = 2, 4, 6, 8$). Positively, we see that in all cases, the curves reach a linear behavior, where the slope corresponds to the number of shares d. Since the independent leakage condition is fulfilled in these experiments, this d corresponds to the smallest key-dependent moment in the leakage distribution. And since the measurement (aka sampling) cost for estimating such moments is proportional to $(\sigma^2)^d$, we observe that the MI decreases exponentially in d for large enough noises. Note that this behavior is plotted for $d = 1, 2$, but was experimented for d's up to 4 in [61], and in fact holds for any d, since it exactly corresponds to Theorem 2 in a context where its assumptions are fulfilled.

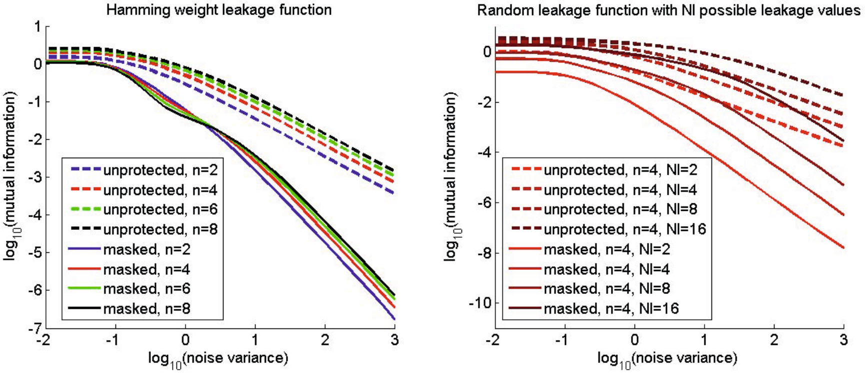

Fig. 2. MI metric in function of σ^2. HW (left) and random (right) leakages.

Negatively, we also see that the noise level that can be considered as high enough depends on the leakage functions. For example, the random leakage functions in the right part of the figure have signals that vary from approximately $\frac{2}{4}$ for $N_l = 2$ to $\frac{16}{4}$ for $N_l = 16$. It implies that the linearly decreasing part of the curves is reached for larger noises in the latter case. Yet, this observation in fact nicely captures the intuition behind the noise condition. That is, the noise should be high enough for hiding the signal. Therefore, a very convenient way to express it is to plot the MI metric in function of shares' SNR, as in Fig. 3. Here, we clearly see that as soon as the SNR is below a certain constant (10^{-1}, typically), the shape of the MI curves gets close to linear. This corroborates the condition in Theorem 2 that masking requires $\mathrm{MI}(K_i; X, \boldsymbol{L}_{Y_i})$ to be smaller than a given constant. Our experiments with different bit sizes also suggest that the $|\mathbb{F}|$

factor in this noise condition is a proof artifact. This is now formally proven by Dziembowski, Faust and Skorski in [24]. Of course, and as discussed in Sect. 2.2, the SNR metric is only applicable under certain conditions (univariate Gaussian leakages). So concretely, an evaluator may choose between computing it after dimensionality reduction (leading to a heuristic but intuitive condition), or to directly state the condition in function of the MI. For completeness, we also plot the MI metric for an unprotected and masked implementation in function of the share's MI in Appendix, Fig. 10. It clearly exhibits that as the share's MI decreases, this reduction is amplified by masking (exponentially in d).

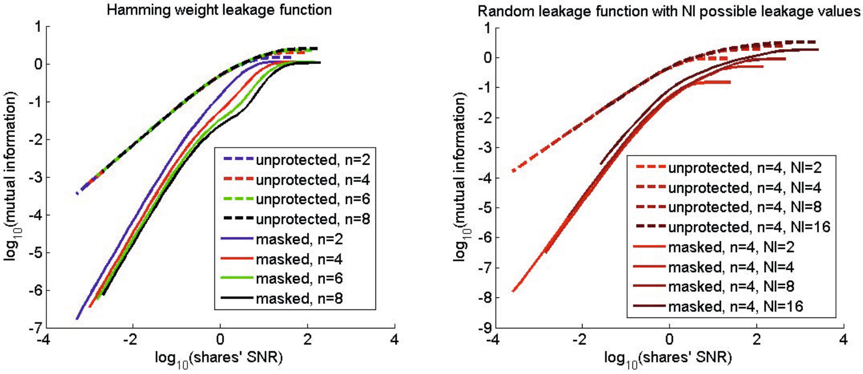

Fig. 3. MI metric in fct. of the shares' SNR. HW (left) and random (right) leakages.

c. Tightness of the Bounds. Given that the noise is high enough (as just discussed), Theorems 2 and 3 guarantee that the success rate of a side-channel adversary can be bounded based on the value of the share's leakage, measured with $\mathrm{MI}(K_i; X, \boldsymbol{L}_{Y_i})$. This directly leads to useful bounds on the measurement complexity to reach a given success rate, e.g. from (8) we can compute:

$$m \geq \frac{\log(1 - \mathrm{SR}^{\mathsf{kr}})}{\log\left(1 - \left(|\mathbb{F}|\sqrt{\frac{\mathrm{MI}(K_i;X,\boldsymbol{L}_{Y_i})}{2}}\right)^{d}\right)} \cdot \tag{17}$$

We now want to investigate how tight this bound is. For this purpose, we compared it with the measurement complexity of concrete key recovery TA (using a perfect leakage model).[2] As previously mentioned, the $|\mathbb{F}|$ factor in this equation can be seen as a proof artifact related to the reduction in our theorems – so we tested a bound excluding this factor. For similar reasons, we also tested a bound additionally excluding the square root loss in the reductions (coming

[2] Our attacks exploit the leakages of an S-box output, as specified in Sect. 2.1. We took the PRESENT S-box for $n = 4$, the AES one for $n = 8$, and picked up two random S-boxes for $n = 2, 6$, as we did for the random leakage functions.

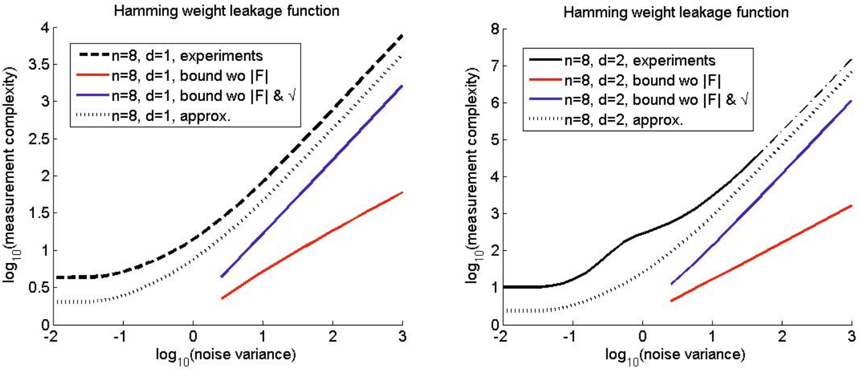

Fig. 4. Measurement complexity and bounds/approximations for concrete TA

from Theorem 1). As illustrated in Fig. 4, the measurement complexity of the attacks is indeed bounded by Equation (17), and removing the square root loss allows the experimental and theoretical curves to have similar slopes. The latter observation fits with the upper bound $\mathrm{MI}(Y_i; \boldsymbol{L}_{Y_i}) \leq \frac{|\mathbb{F}|}{\ln(2)} \cdot \mathrm{SD}(Y_i; Y_i \mid \boldsymbol{L}_{Y_i})$ given in [49] that becomes tight as the noise increases.[3] As expected, the bounds become meaningless for too low noise levels (or too large SNRs, see Appendix, Fig. 11). Intuitively, this is because we reach success rates that are stuck to one when we deviate from this condition. For completeness, we added approximations obtained by normalizing the shares' MI by $\mathrm{H}[K]$ to the figure, which provide hints about the behavior of a leaking device when the noise is too low.

Interestingly, these results also allow us to reach a comprehensive view of the parameters in Theorem 3, where the security of a complete circuit encoded according to the ISW compiler is proven. That is, in this case as well we expect the $|\mathbb{F}|$ and 1/9 factors in Equation (12) to be due to proof technicalities. By contrast, the $|\Gamma|$ factor is physically motivated, since it corresponds to the size of the circuit and fits the intuition that more computations inevitably means more exploitable leakage. The d factor appearing in the noise condition of Equation (13) can also be explained, since it directly relates to the fact that in the ISW compiler, any multiplication will require to manipulate each share d times. It typically reflects the distance between standard (divide-and-conquer) side-channel attacks (such as analyzed in this section) and more powerful (multivariate) adversaries trying the exploit the leakage of all the intermediate computations in a block cipher, e.g. based on algebraic cryptanalysis (see [52,53] and follow up works). Taking all these observations into account, we summarize the concrete security of any masking scheme with the following informal conjecture.

Informal Conjecture. *Suppose that we have a circuit of size $|\Gamma|$ masked with d shares such that the information leakage on each of these shares (using all available time samples) is bounded by $\mathrm{MI}(Y_i; \boldsymbol{L}_{Y_i})$. Then, the probability of success of*

[3] Since their inequality comes from a $\log(1+x) < \log(x)$ inequality that gets close to an equality when x gets close to 0, which happens for large noise levels.

a distinguishing adversary using m measurements and targeting a single element (e.g. gate) of the circuit under independent and sufficiently noisy leakage is:

$$\mathrm{SR}_1^{\mathsf{dist}} \leq 1 - \left(1 - \mathrm{MI}(Y_i; \boldsymbol{L}_{Y_i})^d\right)^m \ , \tag{18}$$

and the probability of success targeting all $|\Gamma|$ elements independently equals:

$$\mathrm{SR}_{|\Gamma|}^{\mathsf{dist}} \leq 1 - (1 - \mathrm{SR}_1^{\mathsf{dist}})^{|\Gamma|} \ . \tag{19}$$

Interestingly, Equation (19) (like Theorem 3) assumes that the leakages of the $|\Gamma|$ gates (or target intermediate values) are exploited independently. This perfectly corresponds to the probing model in which the adversary gains either full knowledge or no knowledge of such computing elements. Thanks to [22], it also implies a similar result against noisy leakages if the noise condition is fulfilled. However, as the noise level decreases, some advanced (e.g. algebraic) side-channel attacks can sometimes take advantage of different computations jointly in a more efficient manner. Note that this informal conjecture is backed up by the results in [3] (Theorem 6) where a similar bound is given in the context of statistical cryptanalysis. By using the approximation $\log(1-x) \approx -x$ that holds for x's close to 0, Equation (18) directly leads to the following simple approximation of a standard DPA's measurement complexity for large noise levels:

$$m \geq \frac{\log(1 - \mathrm{SR}_1^{\mathsf{dist}})}{\log(1 - \mathrm{MI}(Y_i; \boldsymbol{L}_{Y_i})^d)} \approx \frac{c}{\mathrm{MI}(Y_i; \boldsymbol{L}_{Y_i})^d} \ , \tag{20}$$

where c is a small constant that depends on the target success rate. A similar approximation can be obtained from Equation (19) for multi-target attacks.

d. Relation with the Eurocrypt 2009 Evaluation Framework. The evaluation of leaking cryptographic implementations with a combination of information and security metrics was put forward by Standaert et al. at Eurocrypt 2009. In this reference, the authors showed a qualitative connection between both metrics. Namely, they proved that the model (i.e. the approximation of the leakage PDF) used by a side-channel adversary is sound (i.e. allows key recoveries) if and only if the mutual information matrix (defined in paragraph (a) of this section) is such that its diagonal values are maximum for each line. By contrast, they left the quantitative connection between these metrics as an open problem (i.e. does more MI imply less security?). Our results provide a formal foundation for this quantitative connection. They prove that for any implementation, decreasing the MI of the target intermediate values is beneficial to security. This can be achieved by ad hoc countermeasures, in which case it is the goal of an evaluation laboratory to quantify the MI metric, or by masking, in which case we can bound security based only on the value of this metric for each share taken separately.

4.2 Beyond Independent Leakage

The previous section evaluated an experimental setting where the leakage of each share is independent of each other, i.e. $\boldsymbol{L}_{y_i} = \mathsf{G}(y_i) + N_i$. But as discussed in introduction, this condition frequently turns out to be hard to fulfill and so far, there are only limited (in)formal tools allowing to analyze the deviations from independent leakages that may be observed in practice. In order to contribute to this topic, we first launched another set of experiments (for 2-share masking), where the leakage of each share can be written as:

$$\begin{aligned}\boldsymbol{L}_{y_1} &= \mathsf{G}_1(y_1) + f \cdot \mathsf{G}_{1,2}(y_1, y_2) + N_1 \ , \\ \boldsymbol{L}_{y_2} &= \mathsf{G}_2(y_2) + f \cdot \mathsf{G}_{2,1}(y_1, y_2) + N_2 \ .\end{aligned}$$

Here the G_i functions manipulate the shares independently, while the $G_{i,j}$ functions depend on both shares. We additionally used the f (for flaw) parameter in order to specify how strongly we deviate from the independent leakage assumption. As in the previous section, we considered Hamming weight and random functions for all G's (and we used $\mathsf{G}_{i,j}(y_i, y_j) = \mathsf{G}(y_i \oplus y_j)$ for illustration). Exemplary results of an information theoretic analysis in this context are given in Fig. 5 for the $n =$ 4-, and 8-bit cases (and in Appendix, Fig. 12 for the $n =$ 2- and 6-bit S-box cases). We mainly observe that as the noise increases, even small flaws are exploitable by an adversary. Indeed, breaking the independence condition makes smaller-order moments of the leakage distribution key-dependent. Consequently, for large enough noise, it is always this smaller-order moment that will be the most informative. This is empirically confirmed by the slopes of the IT curves in the figures, that gradually reach one rather than two.

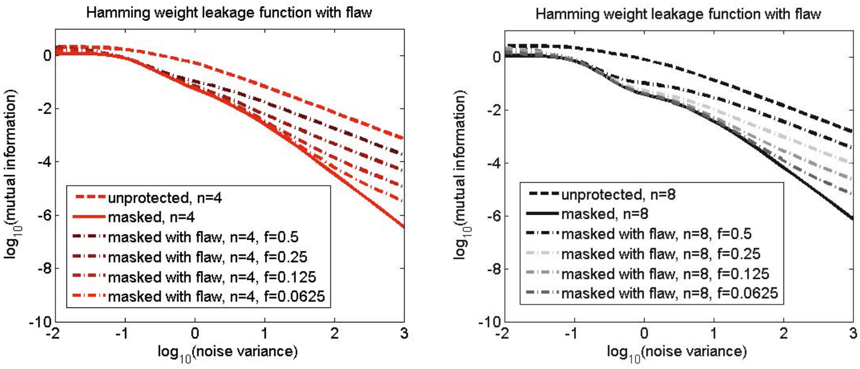

Fig. 5. MI metric for masked implementation with flaw ($n = 4, 8$)

Following these experiments, let us consider a chip that concretely exhibits such a flaw for a given noise level σ^2_{exp} (corresponding to its actual measurements). Despite falling outside the masking proofs' guarantees, an important question is whether we can still (approximatively) predict its security level based on sound statistical tools. In this respect, a useful observation is that the MI metric cannot directly answer the question since it captures the information lying in all the statistical moments of the leakage PDF. So we need another ingredient in order to reveal the informativeness of each moment of the leakage PDF, separately. The Moments-Correlating DPA (MC-DPA) recently introduced in [44] is a natural candidate for this purpose. We now describe how it can be used to (informally) analyze the security of a flawed masked implementation.

In this context, we first need to launch MC-DPA for different statistical moments, e.g. the first- and second-order ones in our 2-share example. They are illustrated by the circle and square markers in the left part of Fig. 6. For concreteness, we take the (most revealing) case where the second-order moment is more informative than the first-order one. Assuming that the noise condition in our theorems is fulfilled, the impact of increasing the noise on the value of the MC-DPA distinguisher can be predicted as indicated by the curves of the figure. That is, with a slope of 1/2 for the first-order moment and a slope of 1 for the second-order one.[4] Hence, we can directly predict the noise level $\sigma^2_{\text{exp}} + \varDelta$ such that the first-order moment becomes more informative. Eventually, we just observe that concrete side-channel attacks always exploit the smallest key-dependent moment in priority (which motivates the definition of the security-order for masking schemes [17]). So starting from the value of the MI at σ^2_{exp} (represented by a circle in the right part of the figure), we can extrapolate that this MI will decrease following a curve with slope 2 until $\sigma^2_{\text{exp}} + \varDelta$ and a curve with slope 1 afterwards. Taking advantage of the theorems in the previous sections, this directly leads to approximations of the best attacks' measurement complexity. Furthermore, extending this reasoning to more shares and higher-order statistical moments is straightforward: it just requires to add MC-DPA curves in the left part of Fig. 6, and to always consider the one leading to the highest MC-DPA value to set the slope of the MI curves, in the right part of the figure. To the best of our knowledge, such figures (despite informal) provide the first concrete tools to approximate the security level in such contexts.

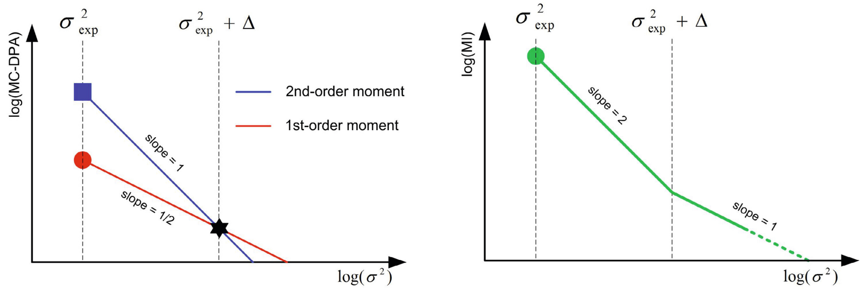

Fig. 6. Evaluating non-independent leakages with MC-DPA (left) and MI (right)

[4] Slopes are divided by 2 when considering Pearson's correlation rather than the MI since this correlation is essentially proportional to the square root of the SNR. This is also reflected by the measurement complexity of CPA, that is proportional to the inverse of the squared correlation vs. the inverse of the MI for TA [62].

Note finally that the shape of the non-independent leakages (i.e. the $G_{i,j}$ functions) observed in practice highly depends on the implementations. For example in hardware, multiple shares can leak jointly in a hardly predictable manner [41,54]. By contrast in software, the most usual issue (due to transition-based leakages) is easier to analyse [5]. It typically divides the order of the smallest key-dependent moment in the leakage distribution by two, which corresponds to the additional square root loss in the security bounds of Duc et al. when considering leakages that depend on two wires simultaneously (see [22], Sect. 5.5).

4.3 Exploiting Computational Power

In this section, we finally tackle the problem of divide-and-conquer DPA attacks, where the adversary aims to combine side-channel information gathered from a number of measurements, and computational power. That is, how to deal with the practically critical situation where the number of measurements available is not sufficient to exactly recover the key? As discussed in [64,65], optimal enumeration and key ranking algorithms provide a concrete answer to this question. They allow building security graphs, where the success rate is plotted in function of a number of measurements and computing power, by repeating attacks multiple times. We next discuss more efficient and analytical strategies.

a. Why MI Is Not Enough? Whenever trying to exploit both side-channel leakage and brute-force computation (e.g. key enumeration) the most challenging aspect of the problem is to capture how measurements and computation actually combine. This is easily illustrated with the following example. Imagine two hypothetical side-channel attacks that both succeed with probability 1/100. In the first case, the adversary gains nothing with probability 99/100 and the full key with probability 1/100. In the second case, he always gains a set of 100 equally likely keys. Clearly, enumeration will be pretty useless in the first case, while extremely powerful in the second one. More generally, such examples essentially suggest that the computational cost of an enumeration does not only depend on the informativeness of the leakage function (e.g. measured with the MI) but also on its shape. For illustration, a line of the mutual information matrix computed from Hamming weight leakages for two noise levels is given in Fig. 7, where we can clearly identify the patterns due to this leakage model. While $\mathrm{MIM}_{k,k}$ only corresponds to a single value of the matrix line (here $k = 111$), which bounds the measurement complexity to recover this key without additional computation (as previously discussed), how helpful is enumeration will additionally depend on the relative distance between the $\mathrm{MIM}_{k,k}$ and MIM_{k,k^*} values [68]. Incidentally, this example also puts forward some limitations of the probing leakage model when measuring computational cost, since it describes an all-or-nothing strategy – as already mentioned in Sect. 4.1, paragraph (c) – which is not the case for the noisy leakage setting. Hence, whereas the probing model is easier to manipulate in proofs, and therefore useful to obtain asymptotic results, noisy leakages are a more accurate tool to quantify concrete security levels as in this section.

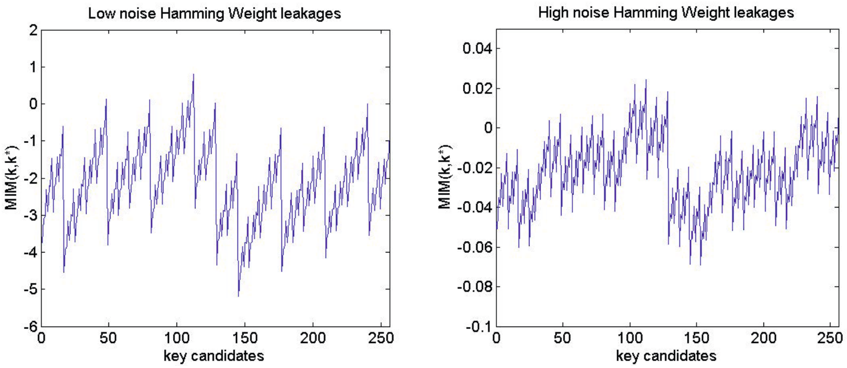

Fig. 7. Exemplary line of the mutual information matrix (for $k = 111$)

b. Measurement and Computational Bounds Per S-Box. Interestingly, one can easily derive bounds for attacks combining side-channel measurements and enumeration power against a single S-box, by re-using exactly the same material as we anyway need to estimate $\mathrm{MI}(K; X, \boldsymbol{L}_{Y_i})$ for a single secret share. In the following, we will assume that the key equivalence property mentioned in Sect. 4.1, paragraph (a) holds, and focus on a single line of the mutual information matrix (if it does not, evaluators simply have to compute all its lines), next denoted as $\mathrm{MIM}_{k,-}$. In order to characterize the distance between a key k and its close candidates k^*, we first sort this line and produce $s = \mathsf{sort}(\mathrm{MIM}_{k,-})$. As a result, the key candidate $k^*_{s(1)}$ is the best rated one (i.e. the correct k if the leakage model is sound), $k^*_{s(2)}$ is the second best, ... From there, we compute a "computational version"of the mutual information matrix as:

$$\mathrm{MIM}^c_{k,k} = \mathrm{H}[K_i] + \mathsf{E}_j \left(\log \left(\sum_{l=1}^{c} \hat{p}_j^{s(l)} \right) \right) . \tag{21}$$

It essentially corresponds to the amount of information an adversary obtains about a random variable that aggregates the c most likely key candidates. Assuming that these c key candidates are equally likely (which can only be pessimistic), it directly provides simple bounds on the success rate of an attack combining m measurements with the enumeration of c keys:

$$\mathrm{SR}^{\mathsf{kr}}(m, c) \leq 1 - \left(1 - (\mathrm{MIM}^c_{k,k})^d\right)^m \ , \tag{22}$$

For illustration, a couple of such bounds are given in Fig. 8, where we see the impact of increasing the number of shares d and number of measurements m. Note that despite requiring similar characterization efforts, these bounds are conceptually different from the previous approaches to approximate the success rate of side-channel attacks. In particular, works like [20,26,37,55] are specific to popular distinguishers (and usually require specialized assumptions about the distribution of these distinguishers), while our results directly connect to security

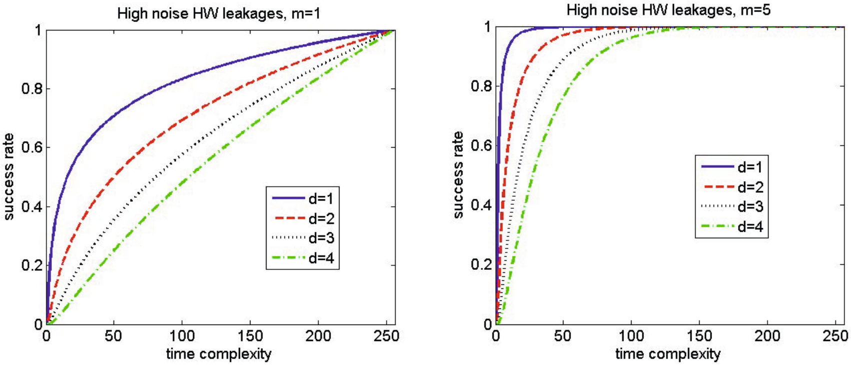

Fig. 8. Key recovery success rate against a single S-box, in function of the time complexity for the leakage function of Fig. 7 (right), after m measurements

proofs that are independent of the adversarial strategy and hold for any leakage distribution. Nevertheless, the only requirement to analyze the combination of multiple S-boxes in the next paragraph (c) is to have success rates curves for each S-box. So while this paragraph (b) describes an efficient way to build such curves, the following contribution is in fact general, and could be used as a complement to any security evaluation obtained for separate S-boxes.

c. Combining Multiple S-Boxes. We finally generalize our analysis of the previous paragraph to the case where we target n_s S-boxes (e.g. $n_s = 16$ for the AES), gained information about their respective input key bytes, and want to recover the full master key. We assume that we perform the same amount of measurements m on each S-box. This can be easily justified in practice, since a leakage trace usually contains samples corresponding to all S-boxes. By contrast, we make no assumption about how informative the leakages of each S-box are. For example, it could completely happen that one S-box is very leaky, and another one perfectly protected (so that enumeration is the only option to recover its corresponding key byte). As just explained, we then characterize the measurement vs. complexity tradeoff with n_s success rate curves $\mathrm{SR}_i^{\mathsf{kr}}(m, c_i)$ with $1 \le i \le n_s$. Typically, we will then set a limit β to the adversary's computational power and try to solve the following optimization problem:

$$\begin{aligned} \max_{c_1,\ldots,c_{n_s}} \qquad & \prod_{i=1}^{n_s} \mathrm{SR}_i^{\mathsf{kr}}(m, c_i) \ , \\ \text{subject to} \qquad & \prod_{i=1}^{n_s} c_i \le \beta \ . \end{aligned} \tag{23}$$

Taking the logarithm of both products, we get:

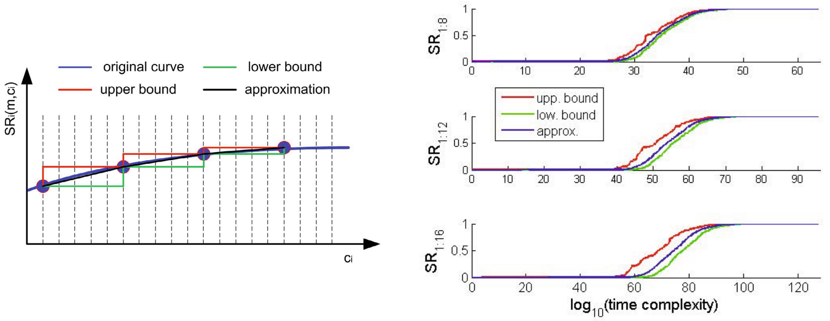

Fig. 9. Downsampling SR curves (left) and enumeration complexity bounds (right)

$$\begin{aligned} \max_{c_1,\dots,c_{n_s}} \quad & \sum_{i=1}^{n_s} \log\left(\mathrm{SR}_i^{\mathsf{kr}}(m,c_i)\right), \\ \text{subject to} \quad & \sum_{i=1}^{n_s} \log(c_i) \leq \log(\beta)\ . \end{aligned} \tag{24}$$

For general functions $\mathrm{SR}_i^{\mathsf{kr}}$, this problem is known as a "separable, non-linear integer programming problem". Surveys about non-linear integer programming problems are various (e.g. [10,36]). There exist many well-studied heuristics to solve them, including branch-and-bounds and convex envelop techniques. Note that the problem generally becomes easier when dealing with convex functions.

We conclude this section with a simple and cheap heuristic algorithm which approximates well the optimal solution for the problem sizes and leakage functions we considered. The approach we propose is inspired by [29], and based on a tradeoff between the computational cost and accuracy of the solutions found, that is controlled by downsampling the success rate curves and keeping track of quantization errors. Intuitively, enumerating the combination of the possible success rates for two n-bit S-boxes requires the computation of 2^{2n} product complexities $c_i \cdot c_j$. Since combining more S-boxes exhaustively will increase the complexity exponentially (i.e. $2^{n_s \cdot n}$ for n_s n-bit S-boxes), the idea of our heuristic is simply to ignore some samples. Namely, we will fix a bound $N_{\max}$ which will designate the maximum number of samples we save per success rate curve (or combination of them). Such a well-known downsampling process is informally illustrated in the left part of Fig. 9, where we can see that the original curve can be easily upperbounded and lowerbounded, since it is increasing.

This solution is described more formally in Algorithm 1 and works as follows. First, we downsample each success rate curve $\mathrm{SR}_i(m,c)$ to $N_{\max}$ linearly spaced points that we can write as $N_{\max}$ pairs $(s_{i,1}, c_{i,1}), \dots, (s_{i,N_{\max}}, c_{i,N_{\max}})$. Next, we take the first S-Box and combine it with the second one, obtaining ${N_{\max}}^2$ values. These values are then downsampled again to $N_{\max}$ linearly spaced points, so that we can iteratively combine them with the next S-boxes. We denote the aggregation of the i first success rate curves with $\mathrm{SR}_{1:i}$. We also add an additional

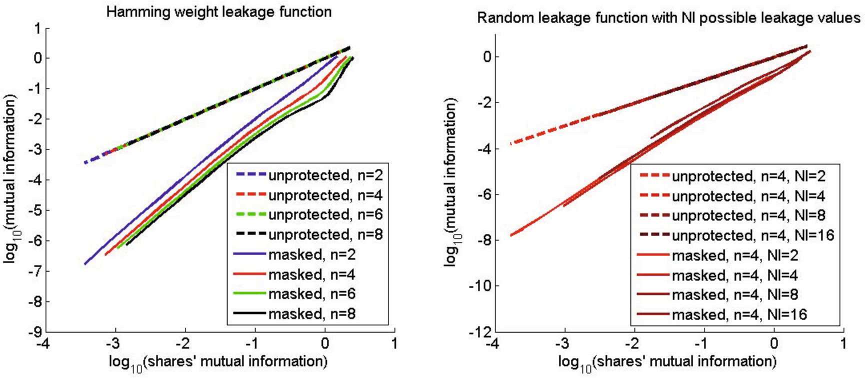

Fig. 10. MI metric in fct. of the shares' MI. HW (left) and random (right) leakages.

output to our algorithm, namely a list of complexities ℓ_i, describing how the effort is distributed among the S-boxes. Indeed, suppose for example that we combine the success rate pair $(0.1, 2^4)$ of S-box 1 with the success rate pair $(0.2, 2^5)$ of S-box 2. We obtain a success rate of 0.02 for a complexity of 2^9, but nothing tells us how the effort is distributed between S-box 1 and S-box 2. Hence, we write the result as $(0.02, 2^9, \{2^4, 2^5\})$ which shows how the complexities are shared.

For illustration, the right part of Fig. 9 provides such bounds for the combination of 8, 12 and 16 AES S-boxes, for a noise level and number of measurement

Algorithm 1. Heuristic to combine the SR curves of n_s S-boxes

Require: Pairs $[(s_{i,1}, c_{i,1}), \ldots, (s_{i,N_{\max}}, c_{i,N_{\max}})] =: \mathrm{SR}_i$ for each S-box i.
Require: A bound $N_{\max}$ on the number of samples and a bound β on the complexity.
Ensure: Triplets $(s_1, c_1, \ell_1), \ldots, (s_{N_{\max}}, c_{N_{\max}}, \ell_{N_{\max}})$ approximating the success rate curve of the combination of the n_s S-Boxes, where the ℓ_i-s are ordered lists of complexities showing how they should be distributed among the S-boxes.

```
$\mathrm{SR}_{1:1} \leftarrow [(s_{1,1}, c_{1,1}, \{c_{1,1}\}), \ldots, (s_{1,N_{\max}}, c_{i,N_{\max}}, \{c_{i,N_{\max}}\})]$;
for $i = 2$ to $n_s$ do
    $\mathrm{SR}_{1:i} \leftarrow \varnothing$;
    ▷ Combination of aggregated S-Boxes $1 : i\text{-}1$ with S-Box $i$.
    for $(s_j, c_j, \ell_j) \in \mathrm{SR}_{1:i-1}$ do        ▷ For all $N_{\max}$ values in aggregated curve.
        for $(s_{i,k}, c_{i,k}) \in \mathrm{SR}_i$ do        ▷ For all $N_{\max}$ values of the new S-Box.
            if $c_j \cdot c_{i,k} < \beta$ then        ▷ If we did not reach the bound $\beta$.
                $\mathrm{SR}_{1:i} \leftarrow \mathrm{SR}_{1:i} \cup (s_j \cdot s_{i,k}, c_j \cdot c_{i,k}, \ell_j \cup c_{i,k})$;        ▷ Merge.
            end if
        end for
    end for
    ▷ Downsampling.
    Sort $\mathrm{SR}_{1:i}$ and keep only $N_{\max}$ linearly spaced pairs.
end for
return $\mathrm{SR}_{1:n_s}$
```

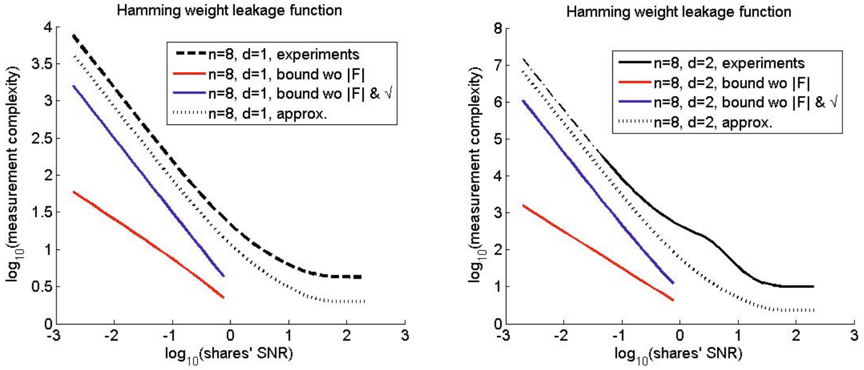

Fig. 11. Measurement complexity and bounds/approximations for concrete TA

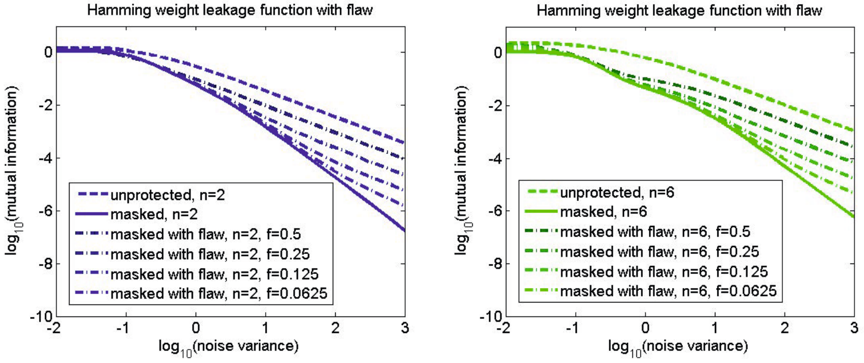

Fig. 12. MI metric for masked implementation with flaw ($n = 2, 6$)

such that the rank estimation problem is challenging (i.e. with full key rank for the 16-byte master key around 2^{80}). The complexity of this heuristic is proportional to $n_s \cdot (N_{\max}^2 + \log(N_{\max}))$ and the results in the figure were obtained within seconds of computation on a desktop computer, using a simple Matlab prototype code. We leave the investigation of better solutions to obtain accurate time complexity bounds with minimum efforts as a scope for further research.

Summarizing, our results show that the (complex) task of evaluating the worst-case security level of a masked implementation against (divide-and-conquer) DPA can be simplified to the evaluation of a couple of MI values, even in contexts where the independence assumption is not fulfilled. This provides a solid foundation for the Eurocrypt 2009 evaluation framework. It also makes it easier to implement, since success rate curves for full keys can now be derived from the MI values, rather than sampled experimentally by repeating (many) subkey recovery experiments and key rank estimations, which is an expensive task. Taking advantage of the tools in this paper therefore allow reducing both the number of measurements and the time needed to evaluate leaking devices.

Acknowledgments. Alexandre Duc is supported by the Swiss National Science Foundation, grant 200021 143899/1. Sebastian Faust received funding from the Marie Curie IEF/FP7 project GAPS (grant 626467). François-Xavier Standaert is a research associate of the Belgian Fund for Scientific Research. This work has been funded in parts by the ERC project 280141 (CRASH).

References

1. http://perso.uclouvain.be/fstandae/PUBLIS/154.zip
2. Archambeau, C., Peeters, E., Standaert, F.-X., Quisquater, J.-J.: Template attacks in principal subspaces. In: Goubin, L., Matsui, M. (eds.) CHES 2006. LNCS, vol. 4249, pp. 1–14. Springer, Heidelberg (2006)
3. Baignères, T., Junod, P., Vaudenay, S.: How far can we go beyond linear cryptanalysis? In: Lee, P.J. (ed.) ASIACRYPT 2004. LNCS, vol. 3329, pp. 432–450. Springer, Heidelberg (2004)
4. Balasch, J., Faust, S., Gierlichs, B., Verbauwhede, I.: Theory and practice of a leakage resilient masking scheme. In: Wang and Sako [67], pp. 758–775
5. Balasch, J., Gierlichs, B., Grosso, V., Reparaz, O., Standaert, F.-X.: On the Cost of Lazy Engineering for Masked Software Implementations. IACR Cryptology ePrint Archive 2014:413 (2014)
6. Batina, L., Robshaw, M. (eds.) CHES 2014. LNCS, vol. 8731. Springer, Heidelberg (2014)
7. Belaïd, S., Grosso, V., Standaert, F.-X.: Masking and Leakage-Resilient Primitives: One, the Other(s) or Both? Cryptography and Communications **7**(1), 163–184 (2015)
8. Bellare, M., Tessaro, S., Vardy, A.: A Cryptographic Treatment of the Wiretap Channel. IACR Cryptology ePrint Archive 2012:15 (2012)
9. Bellare, M., Tessaro, S., Vardy, A.: Semantic security for the wiretap channel. In: Safavi-Naini, R., Canetti, R. (eds.) CRYPTO 2012. LNCS, vol. 7417, pp. 294–311. Springer, Heidelberg (2012)
10. Bertsekas, D.P.: Nonlinear Programming. Athena Scientific (1999)
11. Brier, E., Clavier, C., Olivier, F.: Correlation power analysis with a leakage model. In: Joye, M., Quisquater, J.-J. (eds.) CHES 2004. LNCS, vol. 3156, pp. 16–29. Springer, Heidelberg (2004)
12. Carlet, C., Danger, J.-L., Guilley, S., Maghrebi, H.: Leakage Squeezing: Optimal Implementation and Security Evaluation. J. Mathematical Cryptology **8**(3), 249–295 (2014)
13. Carlet, C., Goubin, L., Prouff, E., Quisquater, M., Rivain, M.: Higher-order masking schemes for S-boxes. In: Canteaut, A. (ed.) FSE 2012. LNCS, vol. 7549, pp. 366–384. Springer, Heidelberg (2012)
14. Chari, S., Jutla, C.S., Rao, J.R., Rohatgi, P.: Towards sound approaches to counteract power-analysis attacks. In: Wiener [69], pp. 398–412
15. Chari, S., Rao, J.R., Rohatgi, P.: Template attacks. In: Kaliski Jr., B.S., Koç, C.K., Paar, C. (eds.) CHES 2002. LNCS, vol. 2523, pp. 13–28. Springer, Heidelberg (2003)

16. Coron, J.-S., Giraud, C., Prouff, E., Renner, S., Rivain, M., Vadnala, P.K.: Conversion of security proofs from one leakage model to another: A new issue. In: Schindler, W., Huss, S.A. (eds.) COSADE 2012. LNCS, vol. 7275, pp. 69–81. Springer, Heidelberg (2012)
17. Coron, J.-S., Prouff, E., Rivain, M.: Side channel cryptanalysis of a higher order masking scheme. In: Paillier, P., Verbauwhede, I. (eds.) CHES 2007. LNCS, vol. 4727, pp. 28–44. Springer, Heidelberg (2007)
18. Coron, J.-S., Prouff, E., Rivain, M., Roche, T.: Higher-order side channel security and mask refreshing. In: Moriai, S. (ed.) FSE 2013. LNCS, vol. 8424, pp. 410–424. Springer, Heidelberg (2014)
19. Cover, T.M., Thomas, J.A.: Elements of Information Theory 2 edn. Wiley (2006)
20. Ding, A.A., Zhang, L., Fei, Y., Luo, P.: A statistical model for higher order DPA on masked devices. In: Batina and Robshaw [6], pp. 147–169
21. Dodis, Y.: Shannon impossibility, revisited. In: Smith, A. (ed.) ICITS 2012. LNCS, vol. 7412, pp. 100–110. Springer, Heidelberg (2012)
22. Duc, A., Dziembowski, S., Faust, S.: Unifying leakage models: From probing attacks to noisy leakage. In: Nguyen and Oswald [45], pp. 423–440
23. Durvaux, F., Standaert, F.-X., Veyrat-Charvillon, N.: How to Certify the Leakage of a Chip? In: Nguyen and Oswald [45], pp. 459–476
24. Dziembowski, S., Faust, S., Skorski, M.: Noisy leakage revisited. In: The Proceedings of EUROCRYPT (to appear 2015)
25. Faust, S., Rabin, T., Reyzin, L., Tromer, E., Vaikuntanathan, V.: Protecting circuits from leakage: The computationally-bounded and noisy cases. In: Gilbert, H. (ed.) EUROCRYPT 2010. LNCS, vol. 6110, pp. 135–156. Springer, Heidelberg (2010)
26. Fei, Y., Luo, Q., Ding, A.A.: A statistical model for DPA with novel algorithmic confusion analysis. In: Prouff, E., Schaumont, P. (eds.) CHES 2012. LNCS, vol. 7428, pp. 233–250. Springer, Heidelberg (2012)
27. Fumaroli, G., Martinelli, A., Prouff, E., Rivain, M.: Affine masking against higher-order side channel analysis. In: Biryukov, A., Gong, G., Stinson, D.R. (eds.) SAC 2010. LNCS, vol. 6544, pp. 262–280. Springer, Heidelberg (2011)
28. Gierlichs, B., Batina, L., Tuyls, P., Preneel, B.: Mutual information analysis. In: Oswald and Rohatgi [47], pp. 426–442
29. Glowacz, C., Grosso, V., Poussier, R., Schueth, J., Standaert, F.-X.: Simpler and more efficient rank estimation for side-channel security assessment. In: IACR Cryptology ePrint Archive 2014:920 (2014)
30. Goubin, L., Martinelli, A.: Protecting AES with Shamir's secret sharing scheme. In: Preneel and Takagi [48], pp. 79–94
31. Grosso, V., Prouff, E., Standaert, F.-X.: Efficient masked s-boxes processing – A step forward –. In: Pointcheval, D., Vergnaud, D. (eds.) AFRICACRYPT. LNCS, vol. 8469, pp. 251–266. Springer, Heidelberg (2014)
32. Grosso, V., Standaert, F.-X., Prouff, E.: Low entropy masking schemes, revisited. In: Francillon, A., Rohatgi, P. (eds.) CARDIS 2013. LNCS, vol. 8419, pp. 33–43. Springer, Heidelberg (2014)
33. Ishai, Y., Sahai, A., Wagner, D.: Private circuits: Securing hardware against probing attacks. In: Boneh, D. (ed.) CRYPTO 2003. LNCS, vol. 2729, pp. 463–481. Springer, Heidelberg (2003)
34. Johansson, T., Nguyen, P.Q. (eds.): EUROCRYPT 2013. LNCS, vol. 7881. Springer, Heidelberg (2013)
35. Kocher, P.C., Jaffe, J., Jun, B.: Differential power analysis. In: Wiener [69], pp. 388–397

36. Li, D., Sun, X.: Nonlinear knapsack problems. In: Nonlinear Integer Programming. International Series in Operations Research & Management Science, vol. 84, pp. 149–207. Springer, US (2006)
37. Lomné, V., Prouff, E., Rivain, M., Roche, T., Thillard, A.: How to estimate the success rate of higher-order side-channel attacks. In: Batina and Robshaw [6], pp. 35–54
38. Mangard, S.: Hardware countermeasures against DPA – A statistical analysis of their effectiveness. In: Okamoto, T. (ed.) CT-RSA 2004. LNCS, vol. 2964, pp. 222–235. Springer, Heidelberg (2004)
39. Mangard, S., Oswald, E., Popp, T.: Power Analysis Attacks - Revealing the Secrets of Smart Cards. Springer (2007)
40. Mangard, S., Oswald, E., Standaert, F.-X.: One for All - All for One: Unifying Standard Differential Power Analysis Attacks. IET Information Security **5**(2), 100–110 (2011)
41. Mangard, S., Popp, T., Gammel, B.M.: Side-channel leakage of masked CMOS gates. In: Menezes, A. (ed.) CT-RSA 2005. LNCS, vol. 3376, pp. 351–365. Springer, Heidelberg (2005)
42. Medwed, M., Standaert, F.-X.: Extractors against Side-Channel Attacks: Weak or Strong? J. Cryptographic Engineering **1**(3), 231–241 (2011)
43. Moradi, A., Mischke, O.: Glitch-Free implementation of masking in modern FPGAs. In: HOST, pp. 89–95. IEEE (2012)
44. Moradi, A., Standaert, F.-X.: Moments-correlating DPA. In: IACR Cryptology ePrint Archive, 2014:409 (2014)
45. Nguyen, P.Q., Oswald, E. (eds.): EUROCRYPT 2014. LNCS, vol. 8441. Springer, Heidelberg (2014)
46. Nikova, S., Rijmen, V., Schläffer, M.: Secure Hardware Implementation of Nonlinear Functions in the Presence of Glitches. J. Cryptology **24**(2), 292–321 (2011)
47. Oswald, E., Rohatgi, P. (eds.): CHES 2008. LNCS, vol. 5154. Springer, Heidelberg (2008)
48. Preneel, B., Takagi, T. (eds.): CHES 2011. LNCS, vol. 6917. Springer, Heidelberg (2011)
49. Prouff, E., Rivain, M.: Masking against side-channel attacks: A formal security proof. In: Johansson and Nguyen [34], pp. 142–159
50. Prouff, E., Roche, T.: Attack on a higher-order masking of the AES based on homographic functions. In: Gong, G., Gupta, K.C. (eds.) INDOCRYPT 2010. LNCS, vol. 6498, pp. 262–281. Springer, Heidelberg (2010)
51. Renauld, M., Kamel, D., Standaert, F.-X., Flandre, D.: Information theoretic and security analysis of a 65-Nanometer DDSLL AES S-box. In: Preneel and Takagi [48], pp. 223–239
52. Renauld, M., Standaert, F.-X.: Algebraic side-channel attacks. In: Bao, F., Yung, M., Lin, D., Jing, J. (eds.) Inscrypt 2009. LNCS, vol. 6151, pp. 393–410. Springer, Heidelberg (2010)
53. Renauld, M., Standaert, F.-X., Veyrat-Charvillon, N.: Algebraic side-channel attacks on the AES: Why time also matters in DPA. In: Clavier, C., Gaj, K. (eds.) CHES 2009. LNCS, vol. 5747, pp. 97–111. Springer, Heidelberg (2009)
54. Renauld, M., Standaert, F.-X., Veyrat-Charvillon, N., Kamel, D., Flandre, D.: A formal study of power variability issues and side-channel attacks for nanoscale devices. In: Paterson, K.G. (ed.) EUROCRYPT 2011. LNCS, vol. 6632, pp. 109–128. Springer, Heidelberg (2011)

55. Rivain, M.: On the exact success rate of side channel analysis in the Gaussian model. In: Avanzi, R.M., Keliher, L., Sica, F. (eds.) SAC 2008. LNCS, vol. 5381, pp. 165–183. Springer, Heidelberg (2009)
56. Rivain, M., Prouff, E.: Provably secure higher-order masking of AES. In: Mangard, S., Standaert, F.-X. (eds.) CHES 2010. LNCS, vol. 6225, pp. 413–427. Springer, Heidelberg (2010)
57. Roche, T., Prouff, E.: Higher-order Glitch Free Implementation of the AES using Secure Multi-Party Computation Protocols - Extended Version. J. Cryptographic Engineering **2**(2), 111–127 (2012)
58. Schindler, W., Lemke, K., Paar, C.: A stochastic model for differential side channel cryptanalysis. In: Rao, J.R., Sunar, B. (eds.) CHES 2005. LNCS, vol. 3659, pp. 30–46. Springer, Heidelberg (2005)
59. Standaert, F.-X., Archambeau, C.: Using subspace-based template attacks to compare and combine power and electromagnetic information leakages. In: Oswald and Rohatgi [47], pp. 411–425
60. Standaert, F.-X., Malkin, T.G., 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)
61. Standaert, F.-X., Veyrat-Charvillon, N., Oswald, E., Gierlichs, B., Medwed, M., Kasper, M., Mangard, S.: The world is not enough: Another look on second-order DPA. In: Abe, M. (ed.) ASIACRYPT 2010. LNCS, vol. 6477, pp. 112–129. Springer, Heidelberg (2010)
62. Standaert, F.-X., Peeters, E., Rouvroy, G., Quisquater, J.-J.: An Overview of Power Analysis Attacks against Field Programmable Gate Arrays. Proceedings of the IEEE **94**(2), 383–394 (2006)
63. Standaert, F.-X., Petit, C., Veyrat-Charvillon, N.: Masking with randomized look up tables - Towards preventing side-channel attacks of all orders. In: Naccache, D. (ed.) Cryphtography and Security: From Theory to Applications. LNCS, vol. 6805, pp. 283–299. Springer, Heidelberg (2012)
64. Veyrat-Charvillon, N., Gérard, B., Renauld, M., Standaert, F.-X.: An optimal key enumeration algorithm and its application to side-channel attacks. In: Knudsen, L.R., Wu, H. (eds.) SAC 2012. LNCS, vol. 7707, pp. 390–406. Springer, Heidelberg (2013)
65. Veyrat-Charvillon, N., Gérard, B., Standaert, F.-X.: Security evaluations beyond computing power. In: Johansson and Nguyen [34], pp. 126–141
66. Veyrat-Charvillon, N., Medwed, M., Kerckhof, S., Standaert, F.-X.: Shuffling against side-channel attacks: A comprehensive study with cautionary note. In: Wang and Sako [67], pp. 740–757
67. Wang, X., Sako, K. (eds.): ASIACRYPT 2012. LNCS, vol. 7658. Springer, Heidelberg (2012)
68. Whitnall, C., Oswald, E.: A comprehensive evaluation of mutual information analysis using a fair evaluation framework. In: Rogaway, P. (ed.) CRYPTO 2011. LNCS, vol. 6841, pp. 316–334. Springer, Heidelberg (2011)
69. Wiener, M. (ed.): CRYPTO 1999. LNCS, vol. 1666. Springer, Heidelberg (1999)