A Nearly Tight Proof of Duc et al.'s Conjectured Security Bound for Masked Implementations

Loïc Masure[1], Olivier Rioul[2], François-Xavier Standaert[1]

[1] ICTEAM Institute, Université catholique de Louvain, Louvain-la-Neuve, Belgium.
[2] LTCI, Télécom Paris, Institut Polytechnique de Paris, Palaiseau, France.

Abstract. We prove a bound that approaches Duc et al.'s conjecture from Eurocrypt 2015 for the side-channel security of masked implementations. Let Y be a sensitive intermediate variable of a cryptographic primitive taking its values in a set $\mathcal{Y}$. If Y is protected by masking (a.k.a. secret sharing) at order d (i.e., with $d+1$ shares), then the complexity of any non-adaptive side-channel analysis — measured by the number of queries to the target implementation required to guess the secret key with sufficient confidence — is lower bounded by a quantity inversely proportional to the product of mutual informations between each share of Y and their respective leakage. Our new bound is nearly tight in the sense that each factor in the product has an exponent of -1 as conjectured, and its multiplicative constant is $\mathcal{O}\big(\log|\mathcal{Y}|\cdot|\mathcal{Y}|^{-1}\cdot C^{-d}\big)$, where $C \leq 2\log(2) \approx 1.38$. It drastically improves upon previous proven bounds, where the exponent was $-1/2$, and the multiplicative constant was $\mathcal{O}\Big(|\mathcal{Y}|^{-d}\Big)$. As a consequence for side-channel security evaluators, it is possible to provably and efficiently infer the security level of a masked implementation by simply analyzing each individual share, under the necessary condition that the leakage of these shares are independent.

1 Introduction

Evaluating the side-channel security of a cryptographic implementation is a sensitive task, in part due to the challenge of defining the adversary's capabilities [ABB+20]. One approach to deal with this problem is to consider the worst-case security of the implementation, which is characterized by the mutual information between its sensitive intermediate variables and the leakage [SMY09]. Worst-case analysis can be viewed as a natural extension of Kerckhoffs' laws to side-channel security, where all the implementation details are given to the evaluator who can even profile (i.e., estimate the statistical distribution of the leakage) in an offline phase where he controls the implementation (including its keys and the random coins used in countermeasures). This approach has the advantage of leading to a simple definition of security matching the standard practice of modern cryptography. It has been known for a while that the link between the mutual information metric and the security of an unprotected implementation is nearly tight [MOS11]. It is also known that if an evaluator can estimate this metric for an implementation with countermeasures, the link

with its security level is nearly tight as well [dCGRP19]. This state of the art essentially leaves evaluators with the problem of estimating the mutual information between sensitive intermediate variables and the leakage of an implementation protected with countermeasures, which can be much more challenging since countermeasures typically make leakage distributions more complex.

In this paper, we are concerned with the important case of the masking countermeasures [CJRR99]. Its high-level idea is to split any sensitive variable of an implementation into $d+1$ shares and to compute on those shares only. As a result, evaluating the worst-case security of a masked implementation requires the characterization of high-order and multivariate distributions, which rapidly turns out to be expensive as the number of shares increases [SVO+10]. In order to mitigate this difficulty, a sequence of works has focused on the formal understanding of the masking countermeasure [PR13,DDF14], and its link with concrete evaluation practice [DFS15a]. In this last reference, a lower bound on the minimum number of queries $N_a^\star$ required to recover a target secret with a success rate at least β thanks to a side-channel attack was established as:

$$N_a^\star \geq \frac{\log(1-\beta)}{\log\left(1-\left(\frac{|\mathcal{Y}|}{\sqrt{2}}\right)^{d+1}\prod_{i=0}^{d}\mathsf{MI}(\mathrm{Y}_i;\mathbf{L}_i)^{1/2}\right)} \approx \frac{\log(1-\beta)}{\left(\frac{|\mathcal{Y}|}{\sqrt{2}}\right)^{d+1}}\prod_{i=0}^{d}\mathsf{MI}(\mathrm{Y}_i;\mathbf{L}_i)^{-1/2} \ ,$$

where $|\mathcal{Y}|$ stands for the size of the group over which masking is applied. Such bounds are interesting since they reduce the assessment of the success rate of an attack to the evaluation of the $d+1$ mutual information values between the shares and their corresponding leakage $\mathsf{MI}(\mathrm{Y}_i;\mathbf{L}_i)$. Each of these mutual information values is substantially simpler to estimate than $\mathsf{MI}(\mathrm{Y};\mathbf{L})$ since the distribution of the leakage random variable $\mathbf{L}_i$ is a first-order one. Unfortunately, is was also shown in the same paper that this proven bound is no tight. More precisely, empirical attacks suggest that the $-1/2$ exponent for each $\mathsf{MI}(\mathrm{Y}_i;\mathbf{L}_i)$ factor might be decreased to -1 and that the $|\mathcal{Y}|^d$ factor would actually be a proof artifact. Halving the exponent has a strong practical impact since it implies that the required number of shares needed to provably reach a given security level might be doubled compared to what is strictly necessary, and the implementation overheads caused by masking scale quadratically in the number of shares. As a result, Duc et al. conjectured that, provided that the shares' leakage is sufficiently small, the lower bound might be tightened as:

$$N_a^\star \geq f(\beta)\prod_{i=0}^{d}\mathsf{MI}(\mathrm{Y}_i;\mathrm{L}_i)^{-1} \ , \tag{1}$$

where $f(\beta) = \mathsf{H}(\mathrm{Y}) - (1-\beta)\cdot\log_2(2^n-1) - \mathsf{H}_2(\beta)$ is a function of the attack success rate β, given by Fano's inequality [CT06], as shown in [dCGRP19].[1]

In this note, we prove a lower bound on $N_a^\star$ that fulfills almost all the conditions of Duc et al.'s conjecture. More precisely, we establish a lower bound

[1] H_2 stands for the *binary* entropy function [dCGRP19].

like the one in Equation 1 with the function $f(\beta)$ divided by a factor $|\mathcal{Y}| \cdot C^d$, with $C = 2\log(2) \approx 1.38$, regardless of the nature of the group $\mathcal{Y}$.

The proof is simple to establish. It mixes Chérisey et al.'s inequality and Dziembowski et al.'s *XOR lemma* [dCGRP19,DFS16], and holds for any group-based masking, such as Boolean or arithmetical masking. The former is expressed with the mutual information metric while the latter is expressed with the statistical distance. We bridge the gap between both by converting Dziembowski et al.'s XOR lemma into a variant that is based on the mutual information.

Related works. Prest et al. used the Rényi divergence in order to improve the tightness of masking security proofs but do not get rid of the square root loss (i.e., the $-1/2$ exponent) on which we focus [PGMP19]. Nevertheless, their bound has a logarithmic dependency on the field size $|\mathcal{Y}|$. Liu et al. used the α-information in order to improve Chérisey's bound [LCGR21]. It could be used to improve our results (at the cost of a slightly less readable bound). In a paper published on the IACR ePrint at the same time as ours (now accepted at CCS 2022), Ito et al. independently obtained a result very similar to ours [IUH22]. Although it is only valid for binary fields (whereas ours is valid for any finite field), their bound is slightly tighter, as the $|\mathcal{Y}|$ is replaced by a $|\mathcal{Y}| - 1$ factor. Interestingly, they also conjecture through the derivation of another bound that is leakage-dependent, and through experimental verifications similar to ours, that the obtained MI-dependent bound is far from being tight in practical cases.

2 Background

2.1 Problem Statement

Let $\mathcal{Y}$ be a finite set. Let $\mathrm{Y} \in \mathcal{Y}$ be a random variable denoting the sensitive intermediate variable targeted by a side-channel adversary. In the "standard DPA setting" we consider [MOS11], Y depends on both a uniformly distributed public plaintext and a secret chunk. We assume an implementation that is protected by a d-th order masking. This means that Y is encoded into $d+1$ shares $\mathrm{Y}_0, \ldots, \mathrm{Y}_d$ such that $\mathrm{Y}_1, \ldots, \mathrm{Y}_d$ are drawn uniformly at random from the group $(\mathcal{Y}, \star)$ and $\mathrm{Y}_0 = \mathrm{Y} \star (\mathrm{Y}_1 \star \ldots \star \mathrm{Y}_d)^{-1}$, with $\star$ the operation over which the masking is applied (e.g., $\oplus$ for Boolean masking, modular addition $+$ for arithmetic masking, ...). As required by masking security proofs, we further assume that the shares' leakage vectors $\mathbf{L}_i$ are the output of a memoryless side-channel and depend only on the realization Y_i, so that the random vectors $\mathbf{L}_0, \ldots, \mathbf{L}_d$ are mutually independent.[2] Intuitively, the goal of a worst-case side-channel security evaluation is to quantify the distance of the random variable Y to the uniform distribution over $\mathcal{Y}$ given the observation of $\mathbf{L} = (\mathbf{L}_0, \ldots, \mathbf{L}_d)$.

[2] This typically captures a software implementation manipulating the shares sequentially, but as discussed in [BDF+17], Lemma 1, it generalizes to parallel (e.g., hardware) implementations as long as the leakage due to the manipulation of shares in parallel can be written as a linear combination of the $\mathbf{L}_i$ vectors.

To simplify our computations, we use (in the proof of Theorem 3) Dziembowski et al.'s reduction to random walks [DFS16, Proof of Lemma 3]. Namely, it is equivalent to consider Y_0 to be uniformly distributed over $\mathcal{Y}$ and to quantify the distance of the variable $\mathrm{Y} = \mathrm{Y}_0 \star \ldots \star \mathrm{Y}_d$ to the uniform distribution given the observations of $\mathbf{L}$.

2.2 Quantifying the Distance to Uniform

To quantify the notion of distance to the uniform distribution over $\mathcal{Y}$, we will use two different metrics. The first one is widely known in information theory.

Definition 1 (Mutual Information). *Let p, m be two Probability Mass Functions (PMFs) over the finite set $\mathcal{Y}$.*[3] *We denote by $\mathsf{D}_{\mathsf{KL}}(\mathsf{p} \parallel \mathsf{m})$ the Kullback-Leibler (KL) divergence between p and m:*

$$\mathsf{D}_{\mathsf{KL}}(\mathsf{p} \parallel \mathsf{m}) = \sum_{y \in \mathcal{Y}} \mathsf{p}(y) \log_2 \left(\frac{\mathsf{p}(y)}{\mathsf{m}(y)} \right) \quad . \tag{2}$$

Then, we define the Mutual Information (MI) between a discrete random variable Y and a continuous random vector $\mathbf{L}$ as follows:

$$\mathsf{MI}(\mathrm{Y}; \mathbf{L}) = \mathop{\mathbb{E}}_{\mathbf{L}} \left[\mathsf{D}_{\mathsf{KL}} \left(\mathsf{p}_{\mathrm{Y} \mid \mathbf{L}} \,\middle\|\, \mathsf{p}_{\mathrm{Y}} \right) \right] \quad , \tag{3}$$

where p_{Y} and $\mathsf{p}_{\mathrm{Y} \mid \mathbf{L}}$ respectively denote the PMF of Y and the PMF of Y given a realization $\mathbf{l}$ of the random vector $\mathbf{L}$, with the expectation taken over $\mathbf{L}$.

The second metric is well-known in the cryptographic community.

Definition 2 (Statistical Distance). *Let p, m be two PMFs over the finite set $\mathcal{Y}$. We denote by $\mathsf{TV}(\mathsf{p}; \mathsf{m})$ the Total Variation (TV) between p and m:*

$$\mathsf{TV}(\mathsf{p}; \mathsf{m}) = \frac{1}{2} \sum_{y \in \mathcal{Y}} |\mathsf{p}(y) - \mathsf{m}(y)| \quad . \tag{4}$$

Then, we define the Statistical Distance (SD) as follows:

$$\mathsf{SD}(\mathrm{Y}; \mathbf{L}) = \mathop{\mathbb{E}}_{\mathbf{L}} \left[\mathsf{TV} \left(\mathsf{p}_{\mathrm{Y} \mid \mathbf{L}}; \mathsf{p}_{\mathrm{Y}} \right) \right] \quad . \tag{5}$$

Note that both the MI and the SD are so-called *global* metrics and are similarly constructed as the expectation over the marginal leakage distribution of so-called *local* quantities, namely the KL divergence and the TV.

Remark 1. Equation 5 is not a distance between the distributions of $\mathbf{L}$ and Y *per se*, as both random variables are not even defined on the same support. Actually, the TV (local) metric is denoted as SD in the work of Dziembowski et al. [DFS16], whereas our definition of SD coincides with their definition of bias. Nevertheless, we keep this notation for the SD metric in order to keep consistency with the notations for the MI, as previously done by Prest et al. [PGMP19].

[3] We assume without loss of generality that both p and m have full support over $\mathcal{Y}$.

In previous works such as the ones of Prouff et al. [PR13], Duc et al. [DFS15a] or Prest et al. [PGMP19], all the inequalities used are stated in terms of global metrics. The idea of our proof is to rely on some similar inequalities between the local quantities, since they are arguably stronger. We introduce such inequalities hereafter. The first one is the well-known Pinsker's inequality.

Proposition 1 (Pinsker's inequality [CT06, Lemma 11.6.1]). *Let* p, m *be two distributions over the — not necessarily finite — set* $\mathcal{Y}$*. Then:*

$$\mathsf{TV}(\mathsf{p};\mathsf{m})^2 \cdot 2\log_2(e) \leq \mathsf{D}_{\mathsf{KL}}(\mathsf{p} \parallel \mathsf{m}) \ . \tag{6}$$

Pinsker's local inequality implies the global inequality

$$2\,\mathsf{SD}(\mathrm{Y};\mathbf{L})^2 \leq \mathsf{MI}(\mathrm{Y};\mathbf{L})$$

used by Duc et al. [DFS15a, Thm. 1], and Prest et al. [PGMP19, Prop. 2].

Remark 2. It is possible to find tighter distribution-dependent constants for Equation 6 [OW05]. Nevertheless, the universal constant denoted in the inequality remains the tightest possible if m is the uniform distribution — which is our case of interest.

The second inequality we need is a reversed version of Pinsker's inequality.

Theorem 1 (Reversed Pinsker's Inequality [SV15, Thm. 1]). *Let* p *be a PMF over the finite set* $\mathcal{Y}$*, and let* u *denote the uniform PMF over* $\mathcal{Y}$*. Then:*

$$\mathsf{D}_{\mathsf{KL}}(\mathsf{p} \parallel \mathsf{u}) \leq \log_2\Big(1 + 2\,|\mathcal{Y}|\,\mathsf{TV}(\mathsf{p};\mathsf{u})^2\Big) \leq 2\log_2(e)\,|\mathcal{Y}| \cdot \mathsf{TV}(\mathsf{p};\mathsf{u})^2 \ . \tag{7}$$

Again, the reversed Pinsker's inequality is stronger than some previous results from Prest et al. as it implies the following global inequality established by the authors of [PGMP19]: $\mathsf{MI}(\mathrm{Y};\mathbf{L}) \leq 2\,\mathsf{RE}(\mathrm{Y};\mathbf{L}) \cdot \mathsf{SD}(\mathrm{Y};\mathbf{L})$, where $\mathsf{RE}(\mathrm{Y};\mathbf{L})$ stands for the Relative Error (RE) between Y and $\mathbf{L}$, another global distance metric introduced in this reference. The reversed Pinsker's (local) inequality is even strictly stronger than Prest et al.'s global inequality, as the former one enables to show that the KL divergence and the squared TV are equivalent metrics (up to a multiplicative constant) whereas, to the best of our knowledge, it is not possible to state that the MI and the squared SD are equivalent global metrics, as the latter one is always bounded by 1, whereas the former one is only bounded by $\log_2|\mathcal{Y}|$, which can be arbitrarily high. The third inequality we need is the so-called XOR lemma stated by Dziembowski et al. at TCC 2016 [DFS16, Thm. 2]. We next provide a slightly looser version of this result with a simpler proof.

Theorem 2 (XOR Lemma). *Let* $\mathrm{Y}_1, \mathrm{Y}_2$ *be independent random variables on a group* $\mathcal{Y}$*, and let* u *denote the uniform PMF over* $\mathcal{Y}$*. Then:*

$$\mathsf{TV}\big(\mathsf{p}_{\mathrm{Y}_1 \star \mathrm{Y}_2};\mathsf{u}\big) \leq 2 \cdot \mathsf{TV}\big(\mathsf{p}_{\mathrm{Y}_1};\mathsf{u}\big) \cdot \mathsf{TV}\big(\mathsf{p}_{\mathrm{Y}_2};\mathsf{u}\big) \ . \tag{8}$$

Proof. Equation 8 is actually a corollary of Young's convolution inequality [Gra14, Thm. 1.2.10]. Denoting by $*$ the convolution product between two PMFs over $\mathcal{Y}$, we have:

$$\mathsf{TV}\big(\mathsf{p}_{\mathrm{Y}_1 \star \mathrm{Y}_2}; \mathsf{u}\big) = \frac{1}{2} \left\| \mathsf{p}_{\mathrm{Y}_1 \star \mathrm{Y}_2} - \mathsf{u} \right\|_1 = \frac{1}{2} \left\| \mathsf{p}_{\mathrm{Y}_1} * \mathsf{p}_{\mathrm{Y}_2} - \mathsf{u} \right\|_1 \tag{9}$$

$$= \frac{1}{2} \left\| \left(\mathsf{p}_{\mathrm{Y}_1} - \mathsf{u}\right) * \left(\mathsf{p}_{\mathrm{Y}_2} - \mathsf{u}\right) \right\|_1 \tag{10}$$

$$\leq \frac{1}{2} \left\| \left(\mathsf{p}_{\mathrm{Y}_1} - \mathsf{u}\right) \right\|_1 \cdot \left\| \left(\mathsf{p}_{\mathrm{Y}_2} - \mathsf{u}\right) \right\|_1 \tag{11}$$

$$= 2\, \mathsf{TV}\big(\mathsf{p}_{\mathrm{Y}_1}; \mathsf{u}\big) \cdot \mathsf{TV}\big(\mathsf{p}_{\mathrm{Y}_2}; \mathsf{u}\big) \quad , \tag{12}$$

where Equation 10 comes from the fact that the uniform PMF is absorbing for the convolution, and Equation 11 comes from Young's convolution inequality. □

Theorem 2 is the core of the noise amplification result of Dziembowski et al. and will also be used to argue about it in our proofs. More precisely, we will use the following corollary that is straightforwardly implied by Theorem 2.

Corollary 1. *Let $\mathrm{Y}_0, \ldots, \mathrm{Y}_d$ be independent random variables on a group $\mathcal{Y}$. Denote $\mathrm{Y}_0 \star \ldots \star \mathrm{Y}_d$ by Y. Then, we have:*

$$\mathsf{TV}(\mathsf{p}_{\mathrm{Y}}; \mathsf{u}) \leq 2^d \cdot \prod_{i=0}^{d} \mathsf{TV}\big(\mathsf{p}_{\mathrm{Y}_i}; \mathsf{u}\big) \quad . \tag{13}$$

3 Nearly Tight Bounds

We now provide new provable bounds for the worst-case side-channel security of masked cryptographic implementations. We start with an upper bound on the mutual information and follow with a lower bound on the security level.

3.1 Upper Bounding the Mutual Information

We first establish noise amplification in terms of KL divergence.

Proposition 2. *Let $\mathrm{Y}_0, \ldots, \mathrm{Y}_d$ be independent but not necessarily identically distributed random variables over $\mathcal{Y}$, with PMFs respectively worth $\mathsf{p}_{\mathrm{Y}_i}$. Let $C = 2\log(2) \approx 1.3862$. Denote the PMF of $\mathrm{Y}_0 \star \ldots \star \mathrm{Y}_d$ as p_{Y}. Then:*

$$\mathsf{D}_{\mathsf{KL}}(\mathsf{p}_{\mathrm{Y}} \parallel \mathsf{u}) \leq \log_2\left(1 + |\mathcal{Y}| \cdot \prod_{i=0}^{d} \left(C \cdot \mathsf{D}_{\mathsf{KL}}\big(\mathsf{p}_{\mathrm{Y}_i} \parallel \mathsf{u}\big)\right)\right) \quad . \tag{14}$$

Proof. Using the inequalities introduced in subsection 2.2, we get:

$$\mathsf{D}_{\mathsf{KL}}(\mathsf{p}_{\mathrm{Y}} \parallel \mathsf{u}) \underset{(7)}{\leq} \log_2\Big(1 + 2\,|\mathcal{Y}|\,\mathsf{TV}(\mathsf{p}_{\mathrm{Y}};\mathsf{u})^2\Big) \tag{15}$$

$$\underset{(13)}{\leq} \log_2\left(1 + |\mathcal{Y}| \cdot \prod_{i=0}^{d} \left(2\,\mathsf{TV}\big(\mathsf{p}_{\mathrm{Y}_i};\mathsf{u}\big)\right)^2\right) \tag{16}$$

$$\underset{(6)}{\leq} \log_2\left(1 + |\mathcal{Y}| \cdot \prod_{i=0}^{d} \left(\frac{2}{\log_2(e)}\mathsf{D}_{\mathsf{KL}}\big(\mathsf{p}_{\mathrm{Y}_i} \parallel \mathsf{u}\big)\right)\right) \quad . \tag{17}$$

□

Having established an amplification result at a local scale, we can now extend it towards the (global) MI metric by taking the expectation over the marginal distribution of the leakage, as stated by the following theorem.

Theorem 3 (MI upper bound (main result)). *Let* $\mathrm{Y}_0, \ldots, \mathrm{Y}_d$ *be* $d+1$ *Independent and Identically Distributed (IID) shares uniformly distributed over* $\mathcal{Y}$*. Let* $\mathbf{L}_0, \ldots, \mathbf{L}_d$ *be the leakages occurred by each share. Denote* $\mathrm{Y} = \mathrm{Y}_0 \star \ldots \star \mathrm{Y}_d$*, and* $\mathbf{L} = (\mathbf{L}_0, \ldots, \mathbf{L}_d)$*. Then:*

$$\mathsf{MI}(\mathrm{Y};\mathbf{L}) \leq \log_2\left(1 + |\mathcal{Y}| \cdot \prod_{i=0}^{d} C \cdot \mathsf{MI}(\mathrm{Y}_i;\mathbf{L}_i)\right) \quad . \tag{18}$$

Proof. We apply Proposition 2 to the random variables

$$\mathrm{Y}'_0 = (\mathrm{Y}_0 \mid \mathbf{L}_0), \ldots, \mathrm{Y}'_d = (\mathrm{Y}_d \mid \mathbf{L}_d).$$

Therefore, we introduce $\mathrm{Y}' = \mathrm{Y}'_0 \star \ldots \star \mathrm{Y}'_d$. As a consequence, each term $\mathsf{D}_{\mathsf{KL}}\left(\mathsf{p}_{\mathrm{Y}'_i} \,\middle\|\, \mathsf{u}\right)$ becomes a random variable depending only on the realization of $\mathbf{L}_i$. Furthermore, as stated in subsection 2.1, thanks to Dziembowski et al.'s reduction to random walks, the random variables $\mathbf{L}_0, \ldots, \mathbf{L}_d$ are mutually independent. As a consequence:

$$\mathsf{MI}(\mathrm{Y};\mathbf{L}) = \mathop{\mathbb{E}}_{\mathbf{L}}\left[\mathsf{D}_{\mathsf{KL}}\left(\mathsf{p}_{\mathrm{Y}\mid\mathbf{L}} \,\middle\|\, \mathsf{p}_{\mathrm{Y}}\right)\right] = \mathop{\mathbb{E}}_{\mathbf{L}}\left[\mathsf{D}_{\mathsf{KL}}(\mathsf{p}_{\mathrm{Y}'} \parallel \mathsf{u})\right] \tag{19}$$

$$\underset{(14)}{\leq} \mathop{\mathbb{E}}_{\mathbf{L}}\left[\log_2\left(1 + |\mathcal{Y}| \cdot \prod_{i=0}^{d}\left(C \cdot \mathsf{D}_{\mathsf{KL}}\left(\mathsf{p}_{\mathrm{Y}'_i} \,\middle\|\, \mathsf{u}\right)\right)\right)\right] \tag{20}$$

$$\leq \log_2\left(1 + |\mathcal{Y}| \cdot \mathop{\mathbb{E}}_{\mathbf{L}}\left[\prod_{i=0}^{d}\left(C \cdot \mathsf{D}_{\mathsf{KL}}\left(\mathsf{p}_{\mathrm{Y}'_i} \,\middle\|\, \mathsf{u}\right)\right)\right]\right) \tag{21}$$

$$\leq \log_2\left(1 + |\mathcal{Y}| \cdot \prod_{i=0}^{d}\left(C \cdot \mathop{\mathbb{E}}_{\mathbf{L}_i}\left[\mathsf{D}_{\mathsf{KL}}\left(\mathsf{p}_{\mathrm{Y}'_i} \,\middle\|\, \mathsf{u}\right)\right]\right)\right) \tag{22}$$

$$\leq \log_2\left(1 + |\mathcal{Y}| \cdot \prod_{i=0}^{d} C \cdot \mathsf{MI}(\mathrm{Y}_i;\mathbf{L}_i)\right) \quad , \tag{23}$$

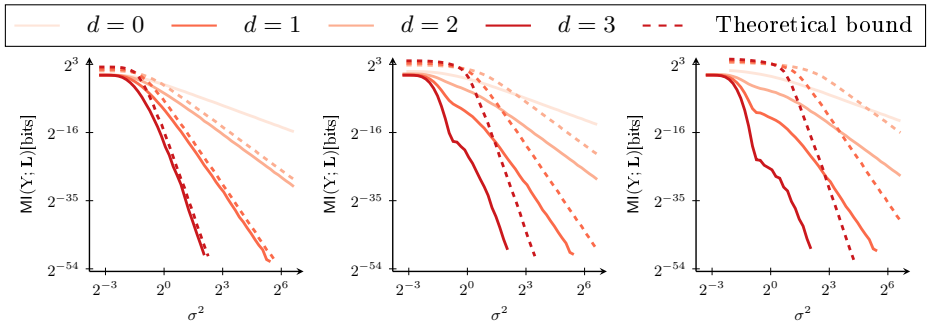
where Equation 21 comes from Jensen's inequality applied to the logarithm, as it is concave, Equation 22 comes from the independence of the leakages, and Equation 19, Equation 23 come from the definition of MI in Equation 3. □

Corollary 2. *Using the same notations as in Theorem 3, we have:*

$$\mathsf{MI}(\mathrm{Y};\mathbf{L}) \leq 2 \cdot |\mathcal{Y}|\, C^d \prod_{i=0}^{d} \mathsf{MI}(\mathrm{Y}_i;\mathbf{L}_i) \quad . \tag{24}$$

Proof. Direct by applying the inequality $\log_2(1+x) \leq x \log_2(e)$ to Equation 18. □

Verification on Simulated Measurements. To verify the soundness of the previous bound, we consider a standard simulated setting where the leakage of each share corresponds to its Hamming weight with additive Gaussian noise with variance σ^2. We first estimate the exact $\mathsf{MI}(\mathrm{Y};\mathbf{L})$ with Monte-Carlo simulations for one, two, four and eight shares [MDP20]. Then, we use the estimated $\mathsf{MI}(\mathrm{Y}_i;\mathbf{L}_i)$ of one share (assuming it is equal for all the shares) to derive an upper bound for two, four and eight shares. Figure 1 shows the resulting information theoretic curves in function of the variance of the additive Gaussian noise. It confirms that the bound is nearly tight for binary targets. By contrast, as the the size of the masking field increases, the factor $|\mathcal{Y}|$ of Equation 24 makes it less tight. Based on the results of [DFS15b], we expect it to be a proof artifact, of which the removal in our bound is an interesting open problem.



(a) $|\mathcal{Y}| = 2$ (bit-slice). (b) $|\mathcal{Y}| = 16$. (c) $|\mathcal{Y}| = 256$.

Fig. 1: MI (plain) and new MI upper bound (dashed) for different field sizes.

3.2 From a MI Upper Bound to a Security Lower Bound

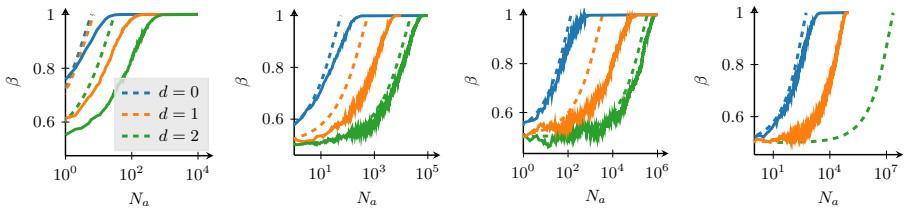
Combining Chérisey's bound in [dCGRP19] with Equation 18 leads to following corollary that bounds the number of measurement queries needed to guess a target secret with sufficient confidence thanks to side-channel leakage.

Corollary 3. *Let $\mathcal{A}$ be any random-plaintext side-channel adversary against an implementation masked at the order d, with each share leaking respectively an amount of information $\mathsf{MI}(\mathrm{Y}_i; \mathbf{L}_i)$. Let $\frac{1}{|\mathcal{Y}|} \leq \beta \leq 1$. Then, for $\mathcal{A}$ to succeed in guessing the secret Y with probability higher than β, at least*

$$N_a^\star \geq \frac{f(\beta)}{\log_2\left(1 + |\mathcal{Y}| \cdot \prod_{i=0}^{d} C \cdot \mathsf{MI}(\mathrm{Y}_i; \mathbf{L}_i)\right)} \geq \frac{f(\beta)}{2\,|\mathcal{Y}|\,C^d} \prod_{i=0}^{d} \mathsf{MI}(\mathrm{Y}_i; \mathbf{L}_i)^{-1} \tag{25}$$

measurement queries to the target leaking implementation are needed.

Verification on Simulated Measurements. To verify the soundness of the proposed bound, we simulate a bit recovery using the same leakage model as in our previous simulations, for one, two and three shares, and different levels of noise — captured by the Gaussian noise variance σ^2.[4] The results are depicted in plain curves on Figure 2. Based on the Monte-Carlo simulation of the MI for one share — still assuming that the shares verify the same leakage model — we also compute the right hand-side of Equation 25, for $\beta \in [0.5, 1]$. This gives the dotted upper bounds of β depicted on Figure 2. Figure 2a shows that the bound



(a) $\sigma^2 = 1$. (b) $\sigma^2 = 10$. (c) $\sigma^2 = 25$. (d) $\sigma^2 = 100$.

Fig. 2: Success rate of concrete bit recoveries and MI-based upper bounds.

derived in Equation 25 may not be tight when the MI is high. Nevertheless, we can see from Figure 2b, Figure 2c and Figure 2d that the higher the noise variance (and the lower the MI) the tighter the expected upper bounds.

4 Conclusions

We prove a new bound approaching by Duc et al.'s conjecture for the security of masked implementations. Our result is tight in $\mathbb{F}_2$, which makes it practically-relevant since bitslice masking is currently the most efficient way to implement

[4] The success rate is estimated with bootstrapping, which gives good estimations with a negligible bias provided that the number of simulated traces is far higher than the value of $N_a^\star$ such that $\beta = 1$. Due to memory constraints, not enough samples could be drawn to get a consistent simulation in the case where $\sigma^2 = 100, d = 2$.

masking for binary ciphers (especially lightweight ones) [GR17]. For larger field sizes, a factor corresponding to the field size $|\mathcal{Y}|$ makes it less tight. Getting rid of this last source of non-tightness therefore remains as an interesting direction for further improvements. We finally note that we can improve another bound from TCC 2016 — which is stated in terms of statistical distance — as a side-effect of our investigations. We detail this last result in section A.

Acknowledgments. François-Xavier Standaert is a Senior Research Associate of the Belgian Fund for Scientific Research (FNRS-F.R.S.). This work has been funded in part by the ERC project number 724725 (acronym SWORD).

References

ABB+20. Melissa Azouaoui, Davide Bellizia, Ileana Buhan, Nicolas Debande, Sébastien Duval, Christophe Giraud, Éliane Jaulmes, François Koeune, Elisabeth Oswald, François-Xavier Standaert, and Carolyn Whitnall. A systematic appraisal of side channel evaluation strategies. In *SSR*, volume 12529 of *Lecture Notes in Computer Science*, pages 46–66. Springer, 2020. 1

BDF+17. Gilles Barthe, François Dupressoir, Sebastian Faust, Benjamin Grégoire, François-Xavier Standaert, and Pierre-Yves Strub. Parallel implementations of masking schemes and the bounded moment leakage model. In *EUROCRYPT (1)*, volume 10210 of *Lecture Notes in Computer Science*, pages 535–566, 2017. 3

CJRR99. Suresh Chari, Charanjit S. Jutla, Josyula R. Rao, and Pankaj Rohatgi. Towards sound approaches to counteract power-analysis attacks. In *CRYPTO*, volume 1666 of *Lecture Notes in Computer Science*, pages 398–412. Springer, 1999. 2

CT06. Thomas M. Cover and Joy A. Thomas. *Elements of information theory (2. ed.)*. Wiley, 2006. 2, 5

dCGRP19. Eloi de Chérisey, Sylvain Guilley, Olivier Rioul, and Pablo Piantanida. Best information is most successful mutual information and success rate in side-channel analysis. *IACR Trans. Cryptogr. Hardw. Embed. Syst.*, 2019(2):49–79, 2019. 2, 3, 8

DDF14. Alexandre Duc, Stefan Dziembowski, and Sebastian Faust. Unifying leakage models: from probing attacks to noisy leakage. *IACR Cryptol. ePrint Arch.*, page 79, 2014. 2

DFS15a. Alexandre Duc, Sebastian Faust, and François-Xavier Standaert. Making masking security proofs concrete - or how to evaluate the security of any leaking device. In *EUROCRYPT (1)*, volume 9056 of *Lecture Notes in Computer Science*, pages 401–429. Springer, 2015. 2, 5

DFS15b. Stefan Dziembowski, Sebastian Faust, and Maciej Skorski. Noisy leakage revisited. In *EUROCRYPT (2)*, volume 9057 of *Lecture Notes in Computer Science*, pages 159–188. Springer, 2015. 8

DFS16. Stefan Dziembowski, Sebastian Faust, and Maciej Skórski. Optimal amplification of noisy leakages. In *TCC (A2)*, volume 9563 of *Lecture Notes in Computer Science*, pages 291–318. Springer, 2016. 3, 4, 5, 11, 12

GR17. Dahmun Goudarzi and Matthieu Rivain. How fast can higher-order masking be in software? In *EUROCRYPT (1)*, volume 10210 of *Lecture Notes in Computer Science*, pages 567–597, 2017. 10

Gra14. Loukas Grafakos. Classical fourier analysis, 2014. 6

IUH22. Akira Ito, Rei Ueno, and Naofumi Homma. On the success rate of side-channel attacks on masked implementations: Information-theoretical bounds and their practical usage. *IACR Cryptol. ePrint Arch.*, page 576, 2022. 3

LCGR21. Yi Liu, Wei Cheng, Sylvain Guilley, and Olivier Rioul. On conditional alpha-information and its application to side-channel analysis. In *ITW*, pages 1–6. IEEE, 2021. 3

MDP20. Loïc Masure, Cécile Dumas, and Emmanuel Prouff. A comprehensive study of deep learning for side-channel analysis. *IACR Trans. Cryptogr. Hardw. Embed. Syst.*, 2020(1):348–375, 2020. 8

MOS11. Stefan Mangard, Elisabeth Oswald, and François-Xavier Standaert. One for all - all for one: unifying standard differential power analysis attacks. *IET Inf. Secur.*, 5(2):100–110, 2011. 1, 3

OW05. Erik Ordentlich and Marcelo J. Weinberger. A distribution dependent refinement of pinsker's inequality. *IEEE Trans. Inf. Theory*, 51(5):1836–1840, 2005. 5

PGMP19. Thomas Prest, Dahmun Goudarzi, Ange Martinelli, and Alain Passelègue. Unifying leakage models on a rényi day. In *CRYPTO (1)*, volume 11692 of *Lecture Notes in Computer Science*, pages 683–712. Springer, 2019. 3, 4, 5

PR13. Emmanuel Prouff and Matthieu Rivain. Masking against side-channel attacks: A formal security proof. In *EUROCRYPT*, volume 7881 of *Lecture Notes in Computer Science*, pages 142–159. Springer, 2013. 2, 5

SMY09. François-Xavier Standaert, Tal Malkin, and Moti Yung. A unified framework for the analysis of side-channel key recovery attacks. In *EUROCRYPT*, volume 5479 of *Lecture Notes in Computer Science*, pages 443–461. Springer, 2009. 1

SV15. Igal Sason and Sergio Verdú. Upper bounds on the relative entropy and rényi divergence as a function of total variation distance for finite alphabets. In *ITW Fall*, pages 214–218. IEEE, 2015. 5

SVO+10. François-Xavier Standaert, Nicolas Veyrat-Charvillon, Elisabeth Oswald, Benedikt Gierlichs, Marcel Medwed, Markus Kasper, and Stefan Mangard. The world is not enough: Another look on second-order DPA. In *ASIACRYPT*, volume 6477 of *Lecture Notes in Computer Science*, pages 112–129. Springer, 2010. 2

A Side-Effect: Improving TCC 2016's Bounds

In subsection 3.1, we have proven that the MI between Y and the whole leakage vector is bounded (up to a multiplicative constant) by the product of the shares' MIs. Since the squared TV and the KL divergence are consistent for finite sets like $\mathcal{Y}$, we may wonder whether an upper bound implying the SD as global metric can be derived from the XOR lemma as well. Dziembowski et al. actually provided such an upper bound as recalled hereafter.[5]

[5] The original version of the theorem is stated for non-uniform secrets. In order to avoid an unfair comparison with respect to Dziembowski et al.'s work, we present an intermediate result of their proof [DFS16, Sec. 3.1.5].

Proposition 3 ([DFS16, Thm. 1(i)], restated). *Let* Y *be a uniform random variable on a group* $\mathcal{Y}$*, encoded by the* $(d+1)$*-sharing* $\mathrm{Y}_0, \ldots, \mathrm{Y}_d$*. Suppose that all the leakages are* δ*-noisy for* $i = 0, \ldots, d$*, i.e., for* $0 \leq \delta < 1/2$*,*

$$\mathsf{SD}(\mathrm{Y}_i; \mathbf{L}_i) \leq \delta \ .$$

Define the noise *parameter* $\theta = 1/2 - \delta$*. Then, for all* $\epsilon > 0$*, in order to get* $\mathsf{SD}(\mathrm{Y}; \mathbf{L}) \leq \epsilon$*, it is sufficient that the masking order verifies:*

$$d \geq 8\theta^{-2} \log\left(\frac{3}{2}\epsilon^{-1}\right) \ . \tag{26}$$

Informally, Equation 26 gives the sufficient masking order d in order to achieve a desired security level ϵ (expressed in terms of Statistical Distance (SD)), depending on the noise level θ that the developer may leverage — the higher θ, the noisier the leakage model. Unfortunately, the bound (26) is not tight, as the authors also derive the following necessary condition [DFS16, Eq. (12)]:

$$d \geq \frac{\log\left((2\epsilon)^{-1}\right)}{\log((1-2\theta)^{-1})} \ , \tag{27}$$

which is asymptotically linear in θ^{-1} when $\theta \to 0$, whereas it is quadratic in Equation 26. Actually, this is mostly due to the overhead term that occurs from the authors' so-called "reduction to unconditional random walks". We shall show that this reduction is not necessary, by leveraging the independence of the leakages to compute the expectation, as in our proof of Theorem 3. As a result, we end up with a tight upper bound, no longer involving the overhead term.

Proposition 4 (Improved bound). *With the same notations as in Prop. 3, it is sufficient that the masking order verifies Equation 27:*

Proof. Starting from the definition of SD (Definition 2), we have:

$$\begin{aligned}
\mathsf{SD}(\mathrm{Y}; \mathbf{L}) &= \mathop{\mathbb{E}}_{\mathbf{L}} \left[\mathsf{TV}\left(\mathsf{p}_{\mathrm{Y}}; \mathsf{p}_{\mathrm{Y} \mid \mathbf{L}} \right) \right] && (28) \\
&\underset{(13)}{\leq} 2^d \mathop{\mathbb{E}}_{\mathbf{L}} \left[\prod_{i=0}^{d} \mathsf{TV}\left(\mathsf{p}_{\mathrm{Y}_i}; \mathsf{p}_{\mathrm{Y}_i \mid \mathbf{L}_i} \right) \right] && (29) \\
&= 2^d \prod_{i=0}^{d} \mathop{\mathbb{E}}_{\mathbf{L}_i} \left[\mathsf{TV}\left(\mathsf{p}_{\mathrm{Y}_i}; \mathsf{p}_{\mathrm{Y}_i \mid \mathbf{L}_i} \right) \right] && (30) \\
&= 2^d \prod_{i=0}^{d} \mathsf{SD}(\mathrm{Y}_i; \mathbf{L}_i) \ , && (31)
\end{aligned}$$

where Equation 30 comes from the mutual independence of the leakages $\mathbf{L}_i$. Now, assuming that for all $i \in [\![0, d]\!]$ we have $\mathsf{SD}(\mathrm{Y}_i; \mathbf{L}_i) \leq \delta < \frac{1}{2}$ and defining $\theta = \frac{1}{2} - \delta$, we have $\mathsf{SD}(\mathrm{Y}; \mathbf{L}) \leq \epsilon$ if $(1-2\theta)^d \leq 2\epsilon$. Hence, the inequality $\mathsf{SD}(\mathrm{Y}; \mathbf{L}) \leq \epsilon$ holds if Equation 27 holds. □