Prouff and Rivain's Formal Security Proof of Masking, Revisited

Loïc Masure, François-Xavier Standaert

Prouff & Rivain's Formal Security Proof of Masking, Revisited

Tight Bounds in the Noisy Leakage Model

Loïc Masure and François-Xavier Standaert

ICTEAM Institute, Université catholique de Louvain, Louvain-la-Neuve, Belgium
loic.masure@uclouvain.be

Abstract. Masking is a counter-measure that can be incorporated to software and hardware implementations of block ciphers to provably secure them against side-channel attacks. The security of masking can be proven in different types of threat models. In this paper, we are interested in directly proving the security in the most realistic threat model, the so-called *noisy leakage* adversary, that captures well how real-world side-channel adversaries operate. Direct proofs in this leakage model have been established by Prouff & Rivain at Eurocrypt 2013, Dziembowski *et al.* at Eurocrypt 2015, and Prest *et al.* at Crypto 2019. These proofs are complementary to each other, in the sense that the weaknesses of one proof are fixed in at least one of the others, and conversely. These weaknesses concerned in particular the strong requirements on the noise level and the security parameter to get meaningful security bounds, and some requirements on the type of adversary covered by the proof — *i.e.*, chosen or random plaintexts. This suggested that the drawbacks of each security bound could actually be proof artifacts. In this paper, we solve these issues, by revisiting Prouff & Rivain's approach.

1 Introduction

1.1 Context

Side-Chanel Analysis (SCA) represents an important threat for cryptographic implementations on embedded devices such as smart-cards, Micro-Controller Units (MCUs), *etc.* [35,36]. In such attacks, the adversary has a physical access to the target device. More precisely, the adversary is assumed to measure some physical metrics of the device called *leakages* — *e.g.* the power consumption of the device or the Electro-Magnetic (EM) emanations around the target — during one or several encryptions. It is then possible to use this side information — beside leveraging plaintexts and ciphertexts — to guess the values of *sensitive* variables, *i.e.* the values of intermediate calculations depending on some chunks of secret. This way, an SCA adversary may independently recover the secret in a divide-and-conquer approach, making the typical complexity of such attacks often negligible compared to a regular cryptanalysis. That is why the SCA threat should carefully be taken into account in the design of cryptographic implementations.

Thankfully, this does not prevent the deployment and the use of embedded cryptography, as this threat can be mitigated by incorporating counter-measures in the implementation. At a very high level, most of the counter-measures such as *masking* or *shuffling* turn a deterministic cryptographic primitive into a non-deterministic implementation by injecting some randomness during the execution of the primitive, either at a physical level or at an algorithmic level. In this paper, we focus on the main counter-measure considered so far in SCA, namely masking [27,16], a.k.a. "Multi-Party Computation (MPC) on silicon" [32]. In a nutshell, any sensitive variable is submitted to a $(d+1)$-linear secret-sharing, where d is the security parameter that the designer may control in order to achieve the desired security level. The implementation is then modified in a way such that all the subsequent calculations involving a sensitive variable are now replaced by some *gadgets* operating on the shares separately, as in multi-party computation. As a result, any SCA adversary must have access to the *noisy* observation of every share of secret to be able to recover any piece of information about a sensitive variable. If any noisy observation induces some uncertainty on the actual value of the corresponding share, it results in an *amplified* uncertainty on the actual value of the target sensitive variable — an intuition that dates back to the seminal works of Chari *et al.* at Crypto 99 [16]. As a consequence, the complexity of any SCA attack increases exponentially fast with the security parameter d, at the price of quadratic (or super-linear) runtime and memory overheads in the implementation only [32].

1.2 Provable Security of Masking

The latter intuition has been formalized over the past few years by masking security proofs. Generally speaking, a masking security proof takes as inputs an abstract representation of the implementation, the number of shares $d+1$ (where d act as he security parameter) and a measure of the noisiness of the leakage, usually characterized from the device embedding the implementation. The masking security proof then returns an upper bound on a metric depicting the security level of the implementation.

There exists different strategies to establish a masking security proof. In this paper, we focus on masking security bounds directly stated in the most realistic threat model. This approach has been first considered by Chari *et al.* [16], before being formalized by Prouff and Rivain [45]. Concretely, a *noisy* observation of an intermediate calculation is a Probability Mass Function (p.m.f.) over all the hypothetical values that the operands may take: the closer the p.m.f. to the uniform distribution, the noisier the leakage.

The idea of security proofs in the noisy leakage model is to assume that any noisy leakage accessed by the adversary is δ-close to the uniform distribution, for some real-valued parameter δ stated in a metric that can be measured by the practitioner.[1] Then, the goal is to prove that the p.m.f. of the secret key, given an

[1] *e.g.*, the Statistical Distance (SD), the Euclidean Norm (EN), or the Mutual Information (MI). Notice that in our context, "noisier" means a *lower* δ.

access to the full leakage, is in turn ϵ-close to the p.m.f. that an adversary without access to side-channel would get, for some real-valued parameter ϵ depending on δ, the security parameter d, and some other specifications of the implementation.

This direct approach has gained the reputation of being "not convenient" [8,10] to work with, up to the point that most masking security proofs are now established in much simpler yet unrealistic threat models [32,6,8,9,15], relying on a non-tight reduction from the noisy leakage model to such simpler threat models [23]. As a result, only three previous works tackled masking security proofs through this direct way so far. These works, from Prouff and Rivain at Eurocrypt 2013 [45], Dziembowski *et al.* at Eurocrypt 2015 [25], and Prest *et al.* at Crypto 2019 [44], considered implementations of block ciphers protected with an Ishai-Sahai-Wagner (I.S.W.) masking scheme [32,47], assuming leak-free refreshings. The latter assumption is a drawback, as it is unrealistic — otherwise studying leaky computations would not be relevant — and some real-world refreshings could critically decrease the security level [19]. Interestingly, these three proofs are quite complementary to each other, in the sense that the weaknesses of one proof are fixed in at least one of the others, and conversely. We give hereafter a brief overview of these pros and cons — also synthesized in Table 1:

1. **Strong noise requirements [45].** Prouff and Rivain's bound required the baseline noise parameter δ to scale *polynomially* with the field size, which is prohibitive for concrete block ciphers, *e.g.*, the Advanced Encryption Standard (AES) whose field size is 256. On the opposite, Dziembowski *et al.*'s bound have a nearly tight noise requirement that does not depend on the field size.
2. **Lack of incentive for noisier leakage [25].** In Dziembowski *et al.*'s security bound assuming that the noise requirement is verified, the bound no longer depends on the actual baseline noise level δ. This suggests that to reach the desired security level ϵ, the designer would have no incentive in choosing a noisier device on which implementing the block cipher, which sounds unrealistic. In the extreme case where the device is so noisy enough that $\delta \leq \epsilon$, masking would not be necessary, whereas Dziembowski *et al.*'s bound would still require a prohibitive number of shares to be meaningful. On the opposite, the bounds of Prouff and Rivain and Prest *et al.* still carry some incentive towards noisier baseline leakage.
3. **Too conservative and hard to estimate metric [44].** Contrary to the other proofs, the baseline noise in Prest *et al.*'s security bound is assumed to be measured in a *worst-case* metric, the so-called Relative Error (RE). This contrasts with all the other works considering *average-case* metrics, such as the MI [45] or the SD [25], and does not fit either with SCA security metrics such as Guessing Entropy (GE) or Success Rate (SR) [50] that are averaged metrics as well. Using worst-case metrics has two main drawbacks. First, a baseline noise characterization made with a worst-case metric necessarily results in more conservative requirements than with average-case metrics. Second, worst-case metrics are by definition harder to estimate on concrete

devices by evaluators, and hereupon the RE may not be efficiently tractable — especially for high-dimensional leakage — nor even be formally defined in some cases. As an example, Prest *et al.* even needed to use tedious tail-cut arguments on the exemplary leakage distributions of their case study [44, Remark 2].

4. **Random message attacks [45].** Last but not least, Prouff and Rivain's security bounds are given for random message attacks, whereas Dziembowski *et al.* and Prest *et al.* state security bounds for chosen plaintext attacks. Even if most of state-of-the-art SCA adversaries consider random plaintext attacks, this contrasts with the common practice in cryptography, where the adversary is assumed to (adaptively) choose the message or the ciphertext.

Table 1: Comparison between all proofs in the Noisy Leakage model: Prouff & Rivain [45], Dziembowski *et al.* [25], Prest *et al.* [44].

Feature	[45]	[25]	[44]	Our work
Strong noise requirement	Yes	No	No	No
Leak-free refreshing	Yes	Yes	Yes (Sec. 6)	Yes
Incentive to small δ	✓	✗	✓	✓
Average-case metric	✓	✓	✗	✓
Adaptive attacks	✗	✓	✓	✓

1.3 Recent Improvements on Security Bounds for Encodings Only

In light of the previous drawbacks listed so far, Duc *et al.* conjectured at Eurocrypt 2015 that the weaknesses (1-3) were actually proof artifacts [24]. More precisely, it would be possible to prove a masking security bound in terms of MI with tight noise requirement, and tight amplification rates, while covering the leakage of the full block cipher. In a recent line of works, Ito *et al.* [33], Masure *et al.* [40], and Béguinot *et al.* [14] have been able to prove a reduced version of Duc *et al.*'s conjectured security bound, for the leakage of one encoding *only*. While these works represent a first milestone, they were limited in that they did not cover the leakage coming from the *computations*, and Duc *et al.*'s conjecture remained to be proven for the leakage of a full block cipher.

1.4 Our Contribution

In this paper, we prove new masking security bounds stated in the noisy leakage model, in the same setting as the one of the previous works discussed so far — namely Rivain-Prouff's masking scheme, with leak-free refreshings [45]. To this end, we revisit Prouff and Rivain's approach, by showing that some drawbacks of their results can be circumvented.

– **A tight bound with respect to the noise parameter δ.** We leverage the recent results of Ito *et al.* [33], Masure *et al.* [40] and Béguinot *et al.* [14], to bound the amount of informative leakage of computations coming from a full block cipher, masked with an I.S.W.-like masking scheme. As a result, our noise requirement is tight [31], while carrying a much higher incentive to noisier leakage than in the previous works.
– **A security bound with low dependency on the field size.** With the previous contribution alone, our final security bound would still carry a constant factor scaling *quadratically* with the size of the field over which the block cipher operates, regardless of the number of shares. While this is much better than Prouff & Rivain's bound and competitive with Dziembowski *et al.*'s bound, this still sounds unnatural, as it does not perfectly fit Duc *et al.*'s conjecture [24], and might be fatal for block ciphers operating over large fields. To tackle this problem, we show how a careful scrutiny of the implementation, under mild assumptions on the Sbox, can allow us to make this constant factor *quasi-linear* with the field size. We even show how this constant factor overhead can further be made almost independent of the field size, by combining the Rivain-Prouff masking scheme with *blinding*, a well-known counter-measure in asymmetric cryptography.
– **Security Bound with Average Metric.** In our masking security proof, any metric, be it the baseline noise δ or the final security bound ϵ, is expressed in MI. This contrasts with Prouff & Rivain's work where the parameters δ and ϵ are not expressed in the same metric. Since MI is an averaged metric, it is quite easy to estimate by evaluators when characterizing the behavior of the target device in worst-case evaluations [4].
– **Attacks with Chosen Messages.** Eventually, we argue how our security bounds stated for random plaintext attacks can be extended to the case of chosen plaintext attacks, using a similar argument as the one stated by Dziembowski *et al.* in their follow-up work at Tcc 2016 [26].

Overall, our work is the first to state a masking proof with meaningful security bounds, *i.e.*, for which the desired security level can be reached with a reasonable amount of masking shares, and requiring a reasonable amount of noise from the device. Therefore, our masking security bound can be practically used by an SCA evaluator to upper bound current state-of-the-art SCA adversaries. This suggests that masking proofs directly stated in the noisy leakage model can be seen as complementary to the more generic proofs in other threat models. The only shortcoming of our proof, in line of the previous works, concerns the use of leak-free refreshings. We hope future works may allow to relax this assumption, and thereby provide a comparable setting with masking security proofs in the indirect approach taking advantage of reductions between models.

2 Preliminaries

In this paper, we denote sets by calligraphic letters, *e.g.*, $\mathcal{X}$. In particular, the letter $\mathcal{Y}$ denotes a finite field $(\mathcal{Y}, \oplus, \times)$ of characteristic two. Upper-case letters

are used to denote random variables, while lower-case letters denote observations of random variables. In this paper, we adopt the following convention: A, B stand for independent random variables uniformly distributed over $\mathcal{Y}$, while G, H denote random variables that are not necessarily uniform over $\mathcal{Y}$, nor assumed to be independent. The letter L will be used to denote a random function $\mathcal{Y} \to \mathcal{L}$, were the set $\mathcal{L}$ is assumed without loss of generality to be discrete. When the context does not carry any ambiguity, we will often denote the random variable $\mathrm{L}(\mathrm{Y})$ by omitting the reference to Y. Finally, bold letters denote vectors of random variables.

Mutual Information. Let $\mathrm{Y} \in \mathcal{Y}$ be a discrete random variable. The *entropy* of Y, denoted by $\mathsf{H}(\mathrm{Y})$, defined by: $\mathsf{H}(\mathrm{Y}) = -\sum_{s \in \mathcal{Y}} \Pr(\mathrm{Y} = s) \log_2 \Pr(\mathrm{Y} = s)$. Moreover, we define MI between two discrete random variables Y and L as:

$$\mathsf{MI}(\mathrm{Y}; \mathrm{L}) = \mathsf{H}(\mathrm{Y}) - \mathop{\mathbb{E}}_{l} \left[\mathsf{H}(\mathrm{Y} \mid \mathrm{L} = l) \right] \ .$$

2.1 Model of Noisy Leaking Computation

We describe hereafter the frame in which Prouff and Rivain's result is established, that is mostly adapted from their seminal work [45].[2]

Block Cipher. A block cipher over a finite field $\mathcal{Y}$ is defined by a pair of inputs $\mathbf{K}, \mathbf{P}$ seen as vectors of $\mathcal{Y}$, and by a sequence of T *elementary calculations* $(\mathsf{C}_i)_{1 \leq i \leq T}$ defined either over $\mathcal{Y}$ or $\mathcal{Y} \times \mathcal{Y}$. More precisely, since $\mathcal{Y}$ is assumed to be a finite field, we consider the elementary calculations to be either an addition $\oplus$ or a field multiplication $\times$, whether the operands are constant or random variables.[3]

Leakage and SCA Adversary. When processed on some input Y (resp. a pair of inputs A, B), an elementary calculation C_i reveals $\mathrm{L}_i(\mathrm{Y})$ (resp. $\mathrm{L}(\mathrm{A}, \mathrm{B})$) to the adversary, for some *noisy leakage* function L_i, that depends both on Y (resp. A, B), and on some internal randomness assumed to be drawn independently each time L_i leaks. Whenever the context does not carry any ambiguity, we may simply denote the leakage $\mathrm{L}_i(\mathrm{Y})$ by L_i. In this paper, we consider an adversary having access to the full leakage induced by each elementary calculation and trying to recover a chunk of secret key.

Definition 1 (SCA key recovery adversary). *An* SCA adversary *for a block cipher defined over $\mathcal{Y}$ is an algorithm that, upon a sequence of N_a plaintexts $\mathbf{P} = (P_1, \ldots, P_{N_a})$, takes as an input a sequence $\{(L_i)_{1 \leq i \leq T}\}_{1 \leq j \leq N_a}$ of leakages*

[2] The interested reader may also refer to Rivain's habilitation thesis for a thorough discussion about the leakage model [46].

[3] As argued by Prouff & Rivain, any mapping over a finite field can be decomposed as a sequence of additions and multiplications, using Lagrange interpolation.

induced by each elementary calculation of a block cipher, and that returns a guess $\widehat{K}$ of one chunk $K \in \mathcal{Y}$ of the secret key $\mathbf{K}$. We say that the adversary is random-plaintext *if $\mathbf{P}$ is chosen randomly and uniformly over $\mathcal{Y}^{N_a}$, whereas we say that the adversary is* chosen-plaintext *if the adversary can arbitrarily choose the sequence $\mathbf{P}$ — possibly adaptively.*

Notice that $\widehat{\mathrm{K}}$ depends on the plaintexts used by the adversary (and on the internal randomness of the leakage functions). Accordingly, the accuracy of the key guessing is expected to increase with the number N_a of queries. We formalize this in the definition hereafter.

Definition 2 (Success Rate). *The* success rate *of an SCA key recovery adversary is the quantity*

$$\mathsf{SR}(N_a) = \Pr\left(\widehat{K} = K\right) \quad . \tag{1}$$

Similarly, for any probability threshold $\frac{1}{|\mathcal{Y}|} \leq \beta \leq 1$, we define the efficiency *$N_a^\star(\beta)$ of an SCA key recovery adversary as the minimal amount of queries necessary to get a success rate higher than β.*

MI-Noisy Leakage. The success of an SCA key recovery adversary depends on how informative the leakage is about the underlying secret data processed. To measure this, we assume that the evaluator may determine how *noisy* any leakage function is. To this end, we formally define hereafter the concept of MI-noisy leakage.

Definition 3 (Noisy leakage for unary gates). *Let $\mathsf{C} : \mathcal{Y} \to \mathcal{Y}$ be an elementary calculation associated with the leakage function L. L is said to be* δ-MI-noisy*, for some $\delta \geq 0$, if for any input random variable A of C, uniformly distributed over $\mathcal{Y}$,*

$$\mathsf{MI}(A; L(A)) \leq \delta \quad .$$

Definition 4 (Noisy leakage for binary gates). *Let $\mathsf{C} : \mathcal{Y}^2 \to \mathcal{Y}$ be an elementary calculation associated with the leakage function L. L is said to be* δ-MI-noisy*, for some $\delta \geq 0$, if for any input random variables A, B of C, uniformly distributed over $\mathcal{Y}$,*

$$\mathsf{MI}(A, B; L(A, B)) \leq \delta \quad .$$

We chose the MI as a metric of reference in our proof, because it is at the core of Prouff & Rivain's security bound that we revisit in this paper, and also because we can therefore rely on the recent improvement of Ito *et al.* [33], Masure *et al.* [40] and Béguinot *et al.* [14]. Moreover, the MI is known to be tightly linked to the complexity of Differential Power Analysis (DPA) attacks [37,38,39,22,17], and "generally carries more intuition (see, *e.g.*, [5] in the context of linear cryptanalysis)" [24]. We discuss this choice of metric in section 5.

2.2 Rivain-Prouff's Masking Scheme

We recall hereafter the definition of masking, mostly taken from Prouff and Rivain's paper [45, Def. 2].

Definition 5. *Let d be a positive integer. The* d-encoding of $\mathrm{Y} \in \mathcal{Y}$ *is a* $(d+1)$-*tuple* $(Y_i)_{0 \leq i \leq d}$ *satisfying* $\bigoplus_{i=0}^{d} Y_i = Y$ *and such that for any strict subset* $\mathcal{I}$ *of* $[\![0, d]\!]$, $(Y_i)_{\mathcal{I}}$ *is uniformly distributed over* $\mathcal{Y}^{|\mathcal{I}|}$.

The parameter d in Definition 5 refers here to the security parameter of the counter-measure. In their paper, Prouff and Rivain explain how to turn any block cipher into a d-order secure implementation — *i.e.* such that any intermediate computation depending on a secret has a $(d+1)$-encoding [45]. First, the plaintext and the secret key are split into $d+1$ shares. Then, each elementary calculation of the block cipher is transformed as follows. If the elementary calculation is linear with respect to its inputs, then it is replaced by the sequence of elementary calculations listed in Algorithm 1. If the elementary calculation

Algorithm 1 Linear gadget in Prouff & Rivain's proof.

Require: $\mathbf{A}$: $(d+1)$-sharing of A, C: elementary calculation linear with its input.
Ensure: $\mathbf{B}$: $(d+1)$-sharing of $\mathsf{C}(\mathrm{A})$.
1: **for** $i = 0, \ldots, d$ **do**
2: $\quad \mathrm{B}_i \leftarrow \mathsf{C}(\mathrm{A}_i)$ ▷ Type 1 or 2
3: **end for**
4: $\mathbf{B} \leftarrow \mathsf{Refresh}(\mathbf{B})$ ▷ Assumed to be leak-free
5: $\mathbf{A} \leftarrow \mathsf{Refresh}(\mathbf{A})$ ▷ Only if A used subsequently.

is an Sbox, then it can first be decomposed as a sequence of linear calculations and field multiplications. Then the linear calculations can be processed as in Algorithm 1, and the field multiplications can be replaced by the procedure listed in Algorithm 2. It is a variant of the actual I.S.W. scheme revisited by Rivain and Prouff at Ches 2010, up to a permutation between independent operations, so it does not change the amount of informative leakage. Overall, Rivain-Prouff's masked implementation can be decomposed as subsequences of any of the following types:

1. $(z_i \leftarrow g(x_i))_{0 \leq i \leq d}$, with g being a linear function (of the block-cipher);
2. $(z_i \leftarrow g(x_i))_{0 \leq i \leq d}$, with g being an affine function (within an Sbox evaluation);
3. $(v_{i,j} \leftarrow a_i \times b_j)_{0 \leq i,j \leq d}$ (cross-products computation step in multiplication);
4. $(t_{i,j} \leftarrow t_{i,j-1} \oplus v_{i,j})_{0 \leq i,j \leq d}$ (compression step multiplication).

For concreteness, we list two examples of schemes of the AES Sbox (at least its non-linear part) with this method in Algorithms 3 and 4. Algorithm 3 is the one initially proposed by Rivain and Prouff at Ches 2010. Recently, Cardoso *et al.* proposed at Cardis 2022 an alternative exponentiation scheme depicted

Algorithm 2 Multiplication gadget in Prouff & Rivain's proof.

Require: $\mathbf{A}, \mathbf{B}$: $(d+1)$-sharing of A, B.
Ensure: C : $(d+1)$-sharing of $\mathrm{A} \times \mathrm{B}$.
1: **for** $i = 0, \ldots, d$ **do**
2: **for** $j = 0, \ldots, d$ **do**
3: $\mathrm{V}_{i,j} \leftarrow \mathrm{A}_i \times \mathrm{B}_j$ ▷ Cross products (type 3)
4: **end for**
5: **end for**
6: $\mathbf{V} \leftarrow \mathsf{Refresh}(\mathbf{V})$ ▷ Assumed to be leak-free
7: **for** $i = 0, \ldots, d$ **do**
8: $\mathrm{C}_i = 0$
9: **for** $j = 0, \ldots, d$ **do**
10: $\mathrm{C}_i \leftarrow \mathrm{C}_i \oplus \mathrm{V}_{i,j}$ ▷ Compression (type 4)
11: **end for**
12: **end for**
13: $\mathbf{C} \leftarrow \mathsf{Refresh}(\mathbf{C})$ ▷ Assumed to be leak-free
14: $\mathbf{A}, \mathbf{B} \leftarrow \mathsf{Refresh}(\mathbf{A}), \mathsf{Refresh}(\mathbf{B})$ ▷ Only if A, B used subsequently.

in Algorithm 4 which, combined with other implementation tricks, improved upon Rivain-Prouff's exponentiation [48]. Both exponentiations contain the same number of I.S.W. multiplications.[4]

Algorithm 3 R&P's Exp254 [47].

Require: $\mathbf{X}$: $(d+1)$-sharing of X
Ensure: $\mathbf{C}$: $(d+1)$-sharing of X^{254}
1: $\mathbf{Z} \leftarrow \mathsf{SecLin}(s \mapsto s^2, \mathbf{X})$ ▷ $\mathrm{Z} = \mathrm{X}^2$
2: $\mathbf{X} \leftarrow \mathsf{Refresh}(\mathbf{X})$
3: $\mathbf{Y} \leftarrow \mathsf{SecMult}(\mathbf{Z}, \mathbf{X})$ ▷ $\mathrm{Y} = \mathrm{X}^3$
4: $\mathbf{V} \leftarrow \mathsf{SecLin}(s \mapsto s^4, \mathbf{Y})$ ▷ $\mathrm{V} = \mathrm{X}^{12}$
5: $\mathbf{V} \leftarrow \mathsf{Refresh}(\mathbf{V})$
6: $\mathbf{Y} \leftarrow \mathsf{SecMult}(\mathbf{Y}, \mathbf{V})$ ▷ $\mathrm{Y} = \mathrm{X}^{15}$
7: $\mathbf{Y} \leftarrow \mathsf{SecLin}(s \mapsto s^{16}, \mathbf{Y})$ ▷ $\mathrm{Y} = \mathrm{X}^{240}$
8: $\mathbf{Y} \leftarrow \mathsf{SecMult}(\mathbf{Y}, \mathbf{W})$ ▷ $\mathrm{Y} = \mathrm{X}^{252}$
9: $\mathbf{C} \leftarrow \mathsf{SecMult}(\mathbf{Y}, \mathbf{Z})$ ▷ $\mathrm{C} = \mathrm{X}^{254}$

Algorithm 4 Cardoso's Exp254 [48].

Require: $\mathbf{X}$: $(d+1)$-sharing of X
Ensure: $\mathbf{C}$: $(d+1)$-sharing of X^{254}
1: $\mathbf{Z} \leftarrow \mathsf{SecLin}(s \mapsto s^2, \mathbf{X})$ ▷ $\mathrm{Z} = \mathrm{X}^2$
2: $\mathbf{Z} \leftarrow \mathsf{Refresh}(\mathbf{Z})$
3: $\mathbf{Y} \leftarrow \mathsf{SecMult}(\mathbf{Z}, \mathbf{X})$ ▷ $\mathrm{Y} = \mathrm{X}^3$
4: $\mathbf{Z} \leftarrow \mathsf{SecLin}(s \mapsto s^2, \mathbf{Y})$ ▷ $\mathrm{Y} = \mathrm{X}^6$
5: $\mathbf{Y} \leftarrow \mathsf{SecMult}(\mathbf{Z}, \mathbf{X})$ ▷ $\mathrm{Y} = \mathrm{X}^7$
6: $\mathbf{Z} \leftarrow \mathsf{SecLin}(s \mapsto s^2, \mathbf{Y})$ ▷ $\mathrm{Z} = \mathrm{X}^{14}$
7: $\mathbf{Y} \leftarrow \mathsf{SecMult}(\mathbf{Z}, \mathbf{X})$ ▷ $\mathrm{Y} = \mathrm{X}^{15}$
8: $\mathbf{Y} \leftarrow \mathsf{SecLin}(s \mapsto s^{16}, \mathbf{Y})$ ▷ $\mathrm{Y} = \mathrm{X}^{240}$
9: $\mathbf{C} \leftarrow \mathsf{SecMult}(\mathbf{Y}, \mathbf{Z})$ ▷ $\mathrm{C} = \mathrm{X}^{254}$

3 Revisiting Prouff and Rivain's Bound

We are now ready to revisit Prouff and Rivain's formal security proof in this section. To this end we briefly recall the outline of their proof — that we follow as well — based on three steps. First, they leverage the assumption that refresh gadgets are leak-free in order to reduce the MI of a sequence of elementary computations to the sum of the MIs between the secret and each subsequence of leakage. Second, some of these elementary computations — *e.g.*, the non-linear operations of the Sbox — may process non-uniform secrets. That is why the

[4] There are other generic methods to securely compute an Sbox with masking [30], which are out of the scope of this paper.

authors make an intermediate reduction to the case where every elementary computation processes uniform secrets — and mutually independent as well, in the case of binary gates. Finally, the authors apply some noise amplification lemma from the literature. Our revisited proof applies the same outline. We now dig into the details of these steps.

3.1 Step 1: Decomposition into Subsequences

We first recall that the MI of a sequence of mutually independent leakages can be bounded by the sum of MIs of each leakage.

Theorem 1 (Subsequence decomposition [45]). *Let Y be a random variable over a finite set $\mathcal{Y}$, not necessarily uniform. Let $\mathbf{L} = (L_1, \ldots, L_t)$ be t random variables such that the random variables $(L_i \mid Y = y)_i$ are mutually independent for every $y \in \mathcal{Y}$. Then, we have*

$$\mathsf{MI}(Y; \mathbf{L}) \leq \sum_{i=1}^{t} \mathsf{MI}(Y; L_i) \quad . \tag{2}$$

Although we do not claim any improvement in this first step, we reproduce the proof in section B for completeness.

3.2 Step 2(a): Reduction to Uniform Secrets for Unary Gates

We now revisit the second step of Prouff and Rivain's work, namely the reduction from non-uniform secrets to uniform secrets. To this end, we will split our results into two cases. The first case processed in this subsection deals with non-uniform inputs of unary calculations, such as Line 4 in Algorithm 3. The second case deals with non-uniform and non-independent inputs of binary calculations, such as Line 6 in Algorithm 3, and will be deferred in subsection 3.3.

The results presented in this section aim at bounding the MI between $\mathsf{C}(\mathrm{Y})$, where $\mathsf{C} : \mathcal{Y} \to \mathcal{Y}$ and its corresponding leakage. We first state the following theorem that relies on a technical lemma from Shulman and Feder [49].

Theorem 2 (Generic Bound for Non-Uniform Secrets [49, p. 1360]). *Let $L : \mathcal{Y} \to \mathcal{L}$ be a random function denoting a leakage, and let Y be uniformly distributed over $\mathcal{Y}$. Then, there exists a constant α such that for all random variables G arbitrarily distributed over $\mathcal{Y}$, the following inequality holds true:*

$$\mathsf{MI}(G; L(G)) \leq \alpha \cdot |\mathcal{Y}| \cdot \mathsf{MI}(Y; \mathbf{L}(Y)) \quad . \tag{3}$$

Moreover, the smallest value α such that Equation 3 holds true belongs to the interval $\alpha \in \left[\frac{\log_2(e)}{e}, 1 - e^{-1}\right] \approx [0.53, 0.63]$.

Theorem 2 introduces an overhead scaling with $|\mathcal{Y}|$, which could decrease the final security level by one or several orders of magnitude (*e.g.*, for the AES, $|\mathcal{Y}| = 2^8$). Note that Equation 3 is nearly tight in the general case, in the sense that the range of α is narrow. Shulman and Feder exhibit an example of worst case leakage function, such that Equation 3 becomes an equality, for $\alpha \approx 0.53$ [49].

The Power Map Trick. However, such worst-case C functions are not likely to be used in cryptographic primitives, since, *e.g.*, the input and output of Sbox are expected to be uniformly distributed, for cryptographic reasons. That is why we refine hereafter the generic statement of Theorem 2, and we present some examples where this refinement could remove the dependency on the field size. To this end, we revisit Theorem 2 by relying on an intermediate result of Shulman and Feder's proof.

Lemma 1 ([49, Lemma 6]). *Given a leakage function L and two random variables Y, Y' distributed (non-necessarily uniformly) over the finite set $\mathcal{Y}$, and such that the support of $\Pr\big(Y'\big)$ contains the support of $\Pr(Y)$. Then, the following inequality holds:*

$$\frac{\mathsf{MI}(Y; L(Y))}{\mathsf{MI}\big(Y'; L(Y')\big)} \geq \min_{y \in \mathcal{Y}} \frac{\Pr(Y = y)}{\Pr\big(Y' = y\big)} \ .$$

As a result, we straightforwardly get the following corollary.

Corollary 1. *In the same setting as in Lemma 1, if now the support of $\Pr(Y)$ contains the support of $\Pr\big(Y'\big)$, the following inequality holds true:*

$$\frac{\mathsf{MI}\big(Y'; L(Y')\big)}{\mathsf{MI}(Y; L(Y))} \leq \max_{y \in \mathcal{Y}} \frac{\Pr\big(Y' = y\big)}{\Pr(Y = y)} \ . \tag{4}$$

Proof. Straightforward, using Lemma 1 and the identity $\max_{x \in \mathcal{X}} x = \frac{1}{\min_{x \in \mathcal{X}} \frac{1}{x}}$, for some finite ordered set $\mathcal{X}$. □

We will leverage Corollary 1 in the case where the Sbox is a monomial, *i.e.* is of the shape $y \mapsto y^k$. Admittedly, this makes our proof slightly more specific than Prouff and Rivain's one, as the latter one can handle any Sbox expressed as a polynomial. Nevertheless, this assumption remains mild, as it covers many Sboxes used in practical ciphers, including the AES, and will allow us to remove a constant factor equal to the field size.

We have seen in Algorithms 3 and 4 that the monomial $y \mapsto y^k$ can be computed in the Rivain-Prouff masking scheme by computing intermediate power maps $y \mapsto y^{k'}$ for some $k' \leq k$, through some square-and-multiply schemes [47]. The bound on the leakage induced by such an intermediate computation is handled by the following corollary.

Corollary 2. *Let Y be a uniform random variable over a finite field $\mathcal{Y}$ of size $M \geq 2$. For any $k \in [\![1, M-1]\!]$, define the function $\mathsf{C} : y \in \mathcal{Y} \mapsto y^k$. Let $L : \mathcal{Y} \to \mathcal{L}$ be a δ-MI-noisy leakage. Then:*

$$\mathsf{MI}\Big(Y; L\big(Y^k\big)\Big) \leq \frac{M}{M-1} \cdot \gcd\{k, M-1\} \cdot \delta \ . \tag{5}$$

Proof. Using the Data Processing Inequality (DPI) (stated in Lemma 2 in Appendix A), we are reduced to upper bound $\mathsf{MI}\Big(\mathrm{Y}^k; \mathrm{L}\big(\mathrm{Y}^k\big)\Big)$. To this end, we shall compute the p.m.f. of Y^k. The result will then follow from Lemma 1 and

Lemma 2. First, notice that by definition of a field, $y^k = 0$ if and only if (i.f.f.) $y = 0$, so $\Pr\Big(\mathrm{Y}^k = 0\Big) = \frac{1}{M}$. Second, notice that since $(\mathcal{Y}, \oplus, \times)$ is a finite field, the group $(\mathcal{Y}^*, \times)$ is cyclic, hence isomorphic with $\mathbb{Z}_{M-1}$. As a result, for any $s \neq 0$ for which there exists $y \in \mathcal{Y}$ verifying $y^k = s$, we have

$$\Pr\Big(\mathrm{Y}^k = s\Big) = \frac{\gcd(k, M-1)}{M-1} \ ,$$

and $\Pr\Big(\mathrm{Y}^k = s\Big) = 0$ otherwise. To summarize, for all $s \in \mathcal{Y}$, we have

$$\frac{\Pr\Big(\mathrm{Y}^k = s\Big)}{\Pr(\mathrm{Y} = s)} \leq \frac{M}{M-1} \cdot \gcd(k, M-1) \ . \tag{6}$$

□

Comparing the universal bound of Equation 3 to the specific bound in Equation 5, we can see that we replaced a factor $0.63 \cdot M$ by a factor $\gcd(k, M-1)$ (ignoring the factor $\frac{M}{M-1}$ for large values of M). As an example Table 2 reports the different constant factors induced by Equation 5 for the exponentiation scheme of Algorithms 3 and 4, and how they compare to the generic bound of Equation 3. Our power-map-specific bound is between one and two orders of magnitude lower than the generic bound in Equation 3.

Table 2: Factor overheads from Equation 5, and ratio between the generic bound of Equation 3 and the refined bound of Equation 5.

Scheme	k	$\gcd(k, 255)$	$\frac{(1-e^{-1})\cdot 256}{\gcd(k,255)}$
Rivain-Prouff [47]	$2, 3, 12,$ $15, 240, 252$	$1, 3, 3,$ $15, 15, 3$	$\mathbf{161.3}, 53.8, 53.8,$ $\mathbf{10.8}, 10.8, 53.8$
Cardoso *et al.* [48]	$2, 3, 6, 7,$ $14, 15, 240, 252$	$1, 3, 3, 1,$ $1, 15, 15$	$\mathbf{161.3}, 53.8, 53.8, 161.3,$ $161.3, \mathbf{10.8}, 10.8, 53.8$

Admittedly, the numbers reported in Table 2 depend on the exponentiation scheme, and thereby depend on the underlying power-map we aim at computing — which may differ for other block ciphers with power-map-based Sbox beyond the AES. We may therefore wonder how $\gcd(k, M-1)$ generally scales when M grows. It is not hard to find some integer k such that $\gcd(k, M-1)$ scales linearly with M,[5] so our improved bound could marginally improve the one from Equation 3 in some worst-case exponentiation schemes. Still, the following theorem suggests that this is not likely to happen.

[5] As an example, for the AES field $M - 1 = 255$, which is divided by 3 so there exists some k, *e.g.*, $k = 85$, such that $\gcd(k, M-1) = \frac{M-1}{3}$.

Theorem 3 ([12, Thm. 3.2]). *Let $M > 2$ be an integer. Then, for all $\epsilon > 0$, we have $\mathop{\mathbb{E}}_{k}[\gcd(k, M)] = \mathcal{O}(M^{\epsilon})$, where the expectation is taken with respect to k uniformly distributed in $[\![1, M]\!]$.*

The practical interpretation of Theorem 3 is that if a given exponentiation scheme gives high constant factors, then it should not be hard to modify it, in order to make the constant factor in the right hand-side of Equation 5 arbitrarily low. As a consequence, we may treat the right hand-side of Equation 5 as asymptotically independent of M with high probability. That is why in the remaining of this paper, we will abuse notation by denoting any gcd factor as scaling as $\mathcal{O}(M^{\epsilon})$ — which is confirmed on our implementations of interest by Table 2.

3.3 Step 2(b): Reduction to Uniform Secrets for Binary Gates

We have shown in subsection 3.2 how to significantly decrease the loss in the reduction from non-uniform secrets to uniform secrets for leakage coming from unary gates dealing with power maps. In order to have a complete toolbox for reductions to uniform secrets, we also need to deal with leakages coming from gadgets with two input operands, *e.g.*, I.S.W. multiplications. Hereupon, Theorem 2 straightforwardly applies, although spanning a loss of $0.63\,|\mathcal{Y}|^2$ in the reduction.

That is why we may naturally think of extending the power map trick introduced before. But contrary to Theorem 2, Corollary 2 does not extend as straightforwardly for binary gates. Indeed, calculations with more than one operand add another difficulty: not only the operands may not be uniformly distributed, but they might also be non-independent. This results in the following corollary.

Corollary 3. *Let Y be a random variable uniformly distributed over the finite field $\mathcal{Y}$. For $p, q \in [\![2, M-2]\!]$, let $\mathbf{Z} = (Y^p, Y^q)$. Let $L : \mathcal{Y}^2 \to \mathcal{L}$ be a δ-MI-noisy leakage. Then,*

$$\mathsf{MI}(Y; L(\mathbf{Z})) \leq \frac{M}{M-1} \cdot \min\left\{\gcd(p, M-1), \gcd(q, M-1)\right\} \cdot M \cdot \delta \ . \tag{7}$$

Proof. We apply Lemma 1 for the random vector $\mathbf{Z}' = (\mathrm{Y}, \mathrm{Y}')$, where Y' is an independent copy of Y. For any $x, y \in \mathcal{Y}$, the total probability formula implies that

$$\frac{\Pr(\mathrm{Y}^p = x, \mathrm{Y}^q = y)}{\Pr\big(\mathrm{Y} = x, \mathrm{Y}' = y\big)} \leq \frac{\sum_{y'} \Pr(\mathrm{Y}^p = x, \mathrm{Y}^q = y')}{\Pr\big(\mathrm{Y} = x, \mathrm{Y}' = y\big)} = \frac{\Pr(\mathrm{Y}^p = x)}{\Pr(\mathrm{Y} = x)\Pr\big(\mathrm{Y}' = y\big)} \ .$$

Using Equation 6, we get that

$$\frac{\Pr(\mathrm{Y}^p = x, \mathrm{Y}^q = y)}{\Pr(\mathrm{Y} = x)\Pr\big(\mathrm{Y}' = y\big)} \leq \frac{M}{M-1} \cdot \gcd(p, M-1) \cdot M \ . \tag{8}$$

By symmetry, we can obtain the same bound by permuting the roles of p and q, which gives Equation 7. □

Remark 1. Note that the inequality in Equation 8 is tight, *e.g.*, if p divides q, or inversely. Likewise, we argued that Equation 3 is generally tight — unless considering further assumptions on the prior distribution. Nevertheless, both facts do not necessarily imply that Equation 7 is tight. Whether the latter inequality could be refined for binary gates with non-independent operands remains an open-question that we will briefly discuss in subsection 3.4.

3.4 Step 3: The Amplification Theorems

We now revisit the third step of Prouff & Rivain's approach. To this end, like in subsection 3.2 and subsection 3.3, we make a discrepancy between the unary gates and the binary gates.

For Unary Gates. The following amplification theorem is at the core of our direct proof in the noisy leakage model, and holds the name of Mrs. Gerber's Lemma (MGL). It has initially been stated by Wyner and Zyv [53] for binary random variables, and has been recently extended by Jog and Anantharam to random variables in Abelian groups whose size is a power of two [34]. This result has recently been pointed out to the SCA community by Béguinot *et al.* at Cosade 2023 [14].

Theorem 4 (Mrs. Gerber's Lemma (MGL) [14, Cor. 1]). *Let $|\mathcal{Y}| = 2^n$ for some bit-size n and d be a positive integer. Let $Y_0, \ldots, Y_d$ be a $(d+1)$-encoding of the uniform random variable Y over $\mathcal{Y}$, and $\mathbf{L} = (L_0, \ldots, L_d)$ be such that, conditionally to Y_i, the variable L_i is independent of the others. Assume that for all $i \in [\![0, d]\!]$, $\mathsf{MI}(Y_i; L_i) \leq \delta_i$ for some parameter $0 \leq \delta_i \leq 1$. Then*

$$\mathsf{MI}(Y; \mathbf{L}) \leq \mathrm{f}_{\mathsf{MI}}(\delta_0, \ldots, \delta_d) \quad , \tag{9}$$

where $\mathrm{f}_{\mathsf{MI}}(\cdot)$ is Mrs. Gerber's function.

We refer to the works of Béguinot *et al.* for more details about Mrs. Gerber's function [14]. In our context, we only need the properties summarized hereafter.

Proposition 1 (The MGL function [14, Thm. 1, Prop. 3]). *The Mrs. Gerber's Lemma (MGL) function $\mathrm{f}_{\mathsf{MI}}(\cdot)$ is concave with respect to any of its variables, when the remaining ones are kept fixed. Let $\eta = (2 \log 2)^{-1} \approx 0.72$. Then for all $\delta_0, \ldots, \delta_d \in [0, 1]$, we have*

$$\mathrm{f}_{\mathsf{MI}}(\delta_0, \ldots, \delta_d) \leq \eta \prod_{i=0}^{d} \frac{\delta_i}{\eta} \quad . \tag{10}$$

For Binary Gates. We now extend Béguinot *et al.*'s Theorem 4 to the case of binary gates, as stated hereafter by the following theorem that we prove in Appendix B.1, following a similar outline as Prest *et al.* [44, Thm. 6].

Theorem 5. *Let A, B be two independent and uniform random variables, over a finite field $\mathcal{Y}$. Let $(A_i)_{0\leq i\leq d}, (B_j)_{0\leq j\leq d}$ be d-encodings of A and B respectively. Let $L_{i,j} : A_i, B_j \mapsto L_{i,j}(A_i, B_j)$ be a family of randomized and mutually independent leakage functions such that for every i, j, $\mathsf{MI}(A_i, B_j; L_{i,j}(A_i, B_j)) \leq \delta_{i,j}$, for some $\delta_{i,j} \in [0, 1]$. Denote the concatenation of the leakages $\{L_{i,j}\}_{0\leq i,j\leq d}$ by $\mathbf{L}$. Then,*

$$\mathsf{MI}(A, B; \mathbf{L}) \leq \mathrm{f}_{\mathsf{MI}}\left(\sum_{j=0}^{d} \delta_{0,j}, \ldots, \sum_{j=0}^{d} \delta_{d,j}\right) + \mathrm{f}_{\mathsf{MI}}\left(\sum_{i=0}^{d} \delta_{i,0}, \ldots, \sum_{i=0}^{d} \delta_{i,d}\right) \quad . \quad (11)$$

3.5 Security Bound for each Type of Subsequence

In this section, we leverage the noise amplification result to bound the amount of leakage in each subsequence.

Type 1 subsequences occur for linear elementary calculations over uniform secrets, and are already covered by Theorem 4, which is a straightforward application of the MGL.

Corollary 4 (Type 1 subsequences). *Let Y be a uniform random variable over a finite field $\mathcal{Y}$ and $(Y_i)_{0\leq i\leq d}$ be a d-encoding of Y. Let $\delta \geq 0$ and $L_0, \ldots, L_d$ be δ-MI-noisy leakage functions over $\mathcal{Y}$. Denote $(L_0(Y_0), \ldots, L_d(Y_d))$ by $\mathbf{L}$. Then we have:*

$$\mathsf{MI}(Y; \mathbf{L}) \leq \eta \cdot \left(\frac{\delta}{\eta}\right)^{d+1} \quad . \quad (12)$$

Likewise, type 2 subsequences cover linear elementary calculations over non-uniform secrets, *e.g.*, occurring inside Sboxes. Such subsequences are covered by the following corollary.

Corollary 5 (Type 2 subsequences). *Let Y be a uniform random variable over a finite field $\mathcal{Y}$. Let k, d be positive integers and $(G_i)_{0\leq i\leq d}$ be a $(d+1)$-sharing of Y^k. Let $0 \leq \delta \leq 1$ and let $L_0(G_0), \ldots, L_d(G_d)$ be δ-MI-noisy leakages. Denote the concatenation of the leakages $\{L_i\}_{0\leq i\leq d}$ by $\mathbf{L}$. Then, we have:*

$$\mathsf{MI}(Y; \mathbf{L}) \leq \frac{|\mathcal{Y}|}{|\mathcal{Y}| - 1} \cdot \gcd(k, |\mathcal{Y}| - 1) \cdot \eta \cdot \left(\frac{\delta}{\eta}\right)^{d+1} \quad . \quad (13)$$

Proof. Straightforward, by combining Theorem 4 with Corollary 2. □

We now focus on the more involved type of subsequences, namely type 3, which is a binary gate. It occurs in the cross-products of the I.S.W. multiplication.

Corollary 6 (Type 3 subsequences). *Let Y be a uniform random variable over a finite field $\mathcal{Y}$, let d, p, q be positive integers. Let $(G_i)_i$, $(H_j)_j$ be $d+1$-additive sharings of Y^p, Y^q respectively. Let $0 \leq \delta$, and $\{G_i, H_j \mapsto L_{i,j}(G_i, H_j)\}_{i,j}$ be δ-MI-noisy leakage functions. Let us denote the concatenation of the leakages*

$\{L_{i,j}\}_{0\leq i,j\leq d}$ *by* $\mathbf{L}$*, and denote* $\varphi(p,q,M) = \min(\gcd(p,M-1), \gcd(q,M-1))$. *Then we have:*

$$\mathsf{MI}(Y;\mathbf{L}) \leq 2\cdot|\mathcal{Y}|\cdot\frac{|\mathcal{Y}|}{|\mathcal{Y}|-1}\cdot\varphi(p,q,|\mathcal{Y}|)\cdot\eta\cdot\left(\frac{(d+1)\cdot\delta}{\eta}\right)^{d+1} \quad . \tag{14}$$

Proof. Combining Theorem 5 with Corollary 3. □

It now remains to give some upper bounds for type 4 subsequences. These subsequences can be observed in the compression phase of I.S.W. multiplications (after cross-products and refreshings). This is the aim of the following result that we prove in Appendix B.1.

Theorem 6. *Let* $Y_0,\ldots,Y_d$ *be* $d+1$ *independent uniformly random variables over a finite set* $\mathcal{Y}$*. Let* $L_1,\ldots,L_d$ *be a family of* δ_i*-MI leakage functions, defined over* $\mathcal{Y}\times\mathcal{Y}$*, for some* $0\leq\delta_i\leq 1$*. We have:*

$$\mathsf{MI}(Y_d; L_1(Y_0,Y_1),\ldots,L_d(Y_{d-1},Y_d)) \leq \delta_d \ . \tag{15}$$

Corollary 7 (Type 4 subsequences). *Let* Y *be a secret, such that for* $p,q\in\mathbb{N}$ *the product of the multiplication* $Y^p\times Y^q$ *is processed by an I.S.W. gadget. For* $0\leq i,j\leq d$ *and for* $T_{i,j}, V_{i,j}\in\mathcal{Y}$*, let* $\mathbf{L} = \{L_{i,j}(T_{i,j-1},V_{i,j})\}_{0\leq i,j\leq d}$ *denote the corresponding type 4 leakages such that for all* i,j*, the leakage* $L_{i,j}(T_{i,j-1},V_{i,j})$ *is* $\delta_{i,j}$*-MI-noisy, for* $\delta_{i,j}\leq\delta\leq 1$*. Then the following inequality holds true:*

$$\mathsf{MI}(Y;\mathbf{L}_{i,j}(T_{i,j-1},V_{i,j})_{0\leq i,j\leq d}) \leq \frac{|\mathcal{Y}|}{|\mathcal{Y}|-1}\cdot\gcd(p+q,M-1)\cdot\eta\cdot\left(\frac{\delta}{\eta}\right)^{d+1} \ . \tag{16}$$

Proof. Using Corollary 2, we reduce to the case where $\mathrm{Y}^p\times\mathrm{Y}^q$ is uniformly distributed over $\mathcal{Y}$, inducing a $\gcd(p+q,M-1)$ factor overhead. Then, by gathering the leakages $\mathbf{L}_{i,j}$ sharing the same index i by batches, we may notice that each batch of index only depends on one share of Y. We may therefore invoke Theorem 4 as follows:

$$\mathsf{MI}(\mathrm{Y};\mathbf{L}) \leq \mathrm{f}(\delta'_0,\ldots,\delta'_d) \ , \tag{17}$$

where $\delta'_i = \mathsf{MI}\Big(\mathrm{Y}_i;\{\mathrm{L}_{i,j}(\mathrm{T}_{i,j-1},\mathrm{V}_{i,j})\}_{0\leq j\leq d}\Big)$. Finally, we can upper bound each δ'_i by $\delta_{i,d}$ using Theorem 6. □

3.6 From Subsequences to a Complete Computation.

We can now combine the three previous steps to state the main result, in a similar way as Prouff and Rivain [45, Thm. 4] and as Prest *et al.* [44, Sec. 6.3].

Theorem 7. *Consider a* $\mathcal{Y}$*-block cipher with monomial Sboxes, where a sequence of elementary calculations depends on a random variable* Y *uniformly distributed. Assume that these elementary calculations are protected by a* d*-encoding masking*

scheme as described in subsection 2.2, resulting in T elementary calculations giving access to the leakage $\mathbf{L} = (L_i)_{1 \leq i \leq T}$, where each leakage function L_i is assumed to be δ-MI-noisy. Then, the following inequality is verified:

$$\mathsf{MI}(Y; \mathbf{L}) \leq t_{\mathsf{3}} \cdot \eta \cdot \left(\frac{(d+1)\delta}{\eta} \right)^{d+1} + t_{\mathsf{1,2,4}} \cdot \eta \cdot \left(\frac{\delta}{\eta} \right)^{d+1} \quad ,$$

such that

$$t_{\mathsf{3}} = \sum_{(p,q) \in \mathcal{M}} \varphi(p, q, |\mathcal{Y}|) \, , \;\; t_{\mathsf{1,2,4}} = \sum_{(p,q) \in \mathcal{M}} \phi(p, q, |\mathcal{Y}|) + \sum_{k \in \mathcal{S}} \psi(k, |\mathcal{Y}|) \;\; , \qquad (18)$$

where $\mathcal{M}$ is the sequence of pairs (p, q) of exponents in the operands of the I.S.W. multiplication gadgets, $\mathcal{S}$ is the sequence of exponents (k) of operands over which a linear transformation is applied, and

- $\varphi(p, q, M) = 2 \cdot M \cdot \frac{M}{M-1} \cdot \min(\gcd(p, M-1), \gcd(q, M-1))$,
- $\phi(p, q, M) = \frac{M}{M-1} \cdot \gcd(p + q, M - 1)$,
- $\psi(k, M) = \gcd(k, M - 1)$.

Proof. We apply Theorem 1 to decompose the MI into a sum of MIs for each subsequence. Since by assumption Y is uniformly distributed over $\mathcal{Y}$, Corollaries 4, 5, 6, 7 directly apply to bound each term in the sum. □

Note that in (18), $t_{\mathsf{3}} = \mathcal{O}\Big(|\mathcal{Y}|^{1+\epsilon} \cdot |\mathcal{M}|\Big)$, and $t_{\mathsf{1,2,4}} = \mathcal{O}(|\mathcal{Y}|^{\epsilon} \cdot (|\mathcal{M}| + |\mathcal{S}|))$.

Corollary 8. *For any random-plaintext SCA key recovery adversary targeting a $\mathcal{Y}$-block cipher protected by the masking scheme described in subsection 2.2, the efficiency verifies the following bound:*

$$N_a^{\star}(\mathsf{SR}) \geq \frac{f(\mathsf{SR}, |\mathcal{Y}|)}{t_{\mathsf{3}} + t_{\mathsf{1,2,4}}} \cdot \frac{1}{\eta} \cdot \left(\frac{\eta}{(d+1)\delta} \right)^{d+1} \quad ,$$

where $f(\mathsf{SR}, M) = \log_2(M) - (1 - \mathsf{SR}) \log_2(M - 1) - \mathsf{H}_2(\mathsf{SR})$, where H_2 is the binary entropy function, and where the constants t_{3} and $t_{\mathsf{1,2,4}}$ are the ones defined in Theorem 7.

Proof. Chérisey *et al.*'s security bound allows to link the SCA key recovery efficiency to the MI between $\mathrm{Y} = \mathrm{K} \oplus \mathrm{P}$ and the corresponding leakage:

$$N_a^{\star}(\beta) \geq \frac{f(\mathsf{SR}, |\mathcal{Y}|)}{\mathsf{MI}(\mathrm{Y}; \mathbf{L})} \;\; .$$

Plugging Theorem 7 into the latter inequality gives the result. □

In other words, any random plaintext attack on the masked implementation will require at least $\Omega\left(|\mathcal{Y}|^{-(1+\epsilon)} \cdot \log |\mathcal{Y}| \cdot \left(\frac{\eta}{(d+1)\delta} \right)^{d+1} \right)$ queries to the target device.

4 A Tweaked ISW Gadget with Tight Security Bounds

In Remark 1 we have discussed the fact that the bound in Equation 7 might not be tight, despite the different inequalities used to reach this result are individually tight. Therefore, whether Equation 7 could be further tightened remains an open question. To tackle this challenge, one should directly state an amplification result similar to Theorem 5 without going through the intermediate reduction to uniform and independent operands. Unfortunately, to the best of our knowledge, all the amplification lemmata used so far in direct proofs in the noisy leakage model [45, Thm. 1], [26, Thm. 2], [44, Lemma 6] always assume the shares to be mutually independent and *uniformly* distributed, in order to prove the noise amplification.[6] To the best of our knowledge, there is no amplification result for non-uniform secret yet. So the question our challenge opens may be seen as the challenge of finding such amplification results for non-uniform secrets.

4.1 The Blinding Counter-Measure to the Rescue

Rather than trying to tackle the challenge raised in section 4, we propose hereafter to circumvent it: if we cannot improve the bounds, we may still tweak the implementation. We instantiate this idea by proposing in Algorithm 5 a tweaked variant of the I.S.W. multiplication, relying on a similar idea as the so-called *blinding* counter-measure for asymmetric cryptography [38, p. 225].

Algorithm 5 "Blinded" I.S.W.

Require: $\mathbf{G}, \mathbf{H}$: $(d+1)$-sharing of $g(\mathrm{Y}), h(\mathrm{Y})$.
Ensure: $\mathbf{I}$: $(d+1)$-sharing of $g(\mathrm{Y}) \times h(\mathrm{Y})$.

1: $\mathbf{R} \leftarrow \$(1^{d+1})$
2: $\mathbf{G}' \leftarrow \mathbf{G} \oplus \mathbf{R}$
3: $\mathbf{G}', \mathbf{R} \leftarrow \mathsf{Refresh}(\mathbf{G}'), \mathsf{Refresh}(\mathbf{R})$
4: $\mathbf{M} \leftarrow \mathsf{ISW}_1(\mathbf{G}', \mathbf{H})$ ▷ $\mathrm{G}' \perp\!\!\!\perp \mathrm{H}$
5: $\mathbf{H} \leftarrow \mathsf{Refresh}(\mathbf{H})$
6: $\mathbf{H}' \leftarrow \mathsf{ISW}_2(\mathbf{H}, \mathbf{R})$ ▷ $\mathrm{H} \perp\!\!\!\perp \mathrm{R}$
7: $\mathbf{M}, \mathbf{H}' \leftarrow \mathsf{Refresh}(\mathbf{M}), \mathsf{Refresh}(\mathbf{H}')$
8: $\mathbf{I} \leftarrow \mathbf{M} \oplus \mathbf{H}'$ ▷ M, H' linked
9: $\mathbf{I} \leftarrow \mathsf{Refresh}(\mathbf{I})$
10: $\mathbf{G}, \mathbf{H} \leftarrow \mathsf{Refresh}(\mathbf{G}), \mathsf{Refresh}(\mathbf{H})$

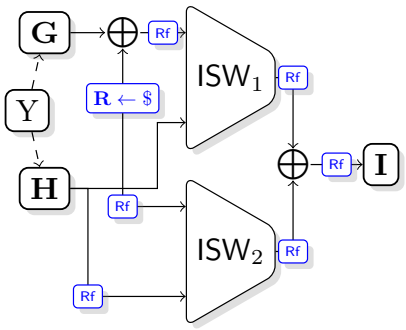


Fig. 1: "Blinded" I.S.W.. "Rf" denote the Refresh gadgets.

The idea, depicted in Figure 1, is to enforce the input operands of the I.S.W. multiplication gadget to be independent. This is done by *blinding* one operand of the ISW_1 gadget in Figure 1, by adding it with the $(d+1)$-sharing of a random nonce. Using a second I.S.W. gadget, we can keep the overall output correct by

[6] Proofs via reduction to the random probing model do not require the underlying secret to be uniformly distributed, as the reduction is applied to the leakage for each share anyway [23, Lemma 2]. Hence, it ensures that the d-encoding is uniformly distributed over $\mathcal{Y}^{d+1}$, which corresponds to a uniform secret.

leveraging the identity: $\mathrm{G} \times \mathrm{H} = (\mathrm{G} \oplus \mathrm{R}) \times \mathrm{H} \oplus (\mathrm{R} \times \mathrm{H})$. At first glance, one may think that this trick somewhat shifts the problem without fixing it, since the input operands of both I.S.W. are independent, but now the input operands of the final Xor in line 8 of Algorithm 5 are no longer independent. Surprisingly, the prior joint distribution of the outputs $(\mathrm{M}, \mathrm{H}')$ of the two I.S.W. multiplications has a much lower bias with respect to the joint uniform distribution, compared to the bias of the joint distribution of (G, H). This is formalized in the following theorem, proven in Appendix B.2.

Theorem 8. *Let $Y \in \mathcal{Y}$ be uniformly distributed, and let $\mathbf{L}$ corresponding to the leakage of Algorithm 5. Then, assuming leak-free refreshings, and that $g(Y) = Y^p$ and $h(Y) = Y^q$, for p, q positive integers, the following inequality is satisfied:*

$$\mathsf{MI}(Y; \mathbf{L}) \leq \varphi(p, q, |\mathcal{Y}|) \cdot \eta \cdot \left(\frac{(d+1)\delta}{\eta}\right)^{d+1} + \phi(p, q, |\mathcal{Y}|) \cdot \eta \cdot \left(\frac{\delta}{\eta}\right)^{d+1} \quad ,$$

where

- $\varphi(p, q, M) = 4 \cdot \frac{M}{M-1} \cdot \gcd(q, M-1)$,
- $\phi(p, q, M) = 4 + \frac{M}{M-1} \cdot \gcd(p, M-1) + \max\Big(2, \frac{M}{M-1} \cdot \gcd(p+q, M-1)\Big)$.

Note that in Theorem 8, both φ and ϕ are almost independent of the field size, whereas t_3 in Theorem 7 scales at least linearly with the field size. From Theorem 8 follows the corollary stated hereafter.

Corollary 9. *In the same setting as in Corollary 8, if the I.S.W. multiplication gadgets are replaced by the scheme in Algorithm 5, then*

$$N_a^\star(\mathsf{SR}) \geq \Omega\left(|\mathcal{Y}|^{-\epsilon} \cdot \log|\mathcal{Y}| \cdot \left(\frac{\eta}{(d+1)\delta}\right)^{d+1}\right) \quad .$$

Proof. The proof follows the one of Corollary 8, by updating the functions φ and ϕ in Equation 18 with the new values in Theorem 8. □

5 Discussion

We have established our main results in section 3 and section 4. We propose hereafter to discuss some features of our results, and to compare them to previous works. To this aim, we first compare in subsection 5.1 our bounds to previous works. We then discuss in subsection 5.2 how we can extend our results to security bounds in terms of chosen plaintext attacks. We conclude this section by discussing the advantages and drawbacks of the blinded I.S.W. gadget presented in section 4.

5.1 Comparison with Related Works

We compare in this section our security bounds with related works. To this end, we first discuss the noise requirements in the different security bounds in the literature. We synthesize in Table 3 the different noise requirements of masking security bounds. We can see that our security bound gets a similar noise requirement as the proofs of Dziembowski *et al.* [25] and Prest *et al.* [44], although stated in different metrics. Notice that the dependency of our noise requirement in d is tight, since it depicts the potential ability of an adversary to increase its success of recovering each share through *horizontal attacks*, as argued by Battistello *et al.* [7] and Grosso and Standaert [31]. Nevertheless, it is still possible to relax this dependency by using other multiplication gadgets [1,3,2,8,28,29].

Moreover, we also extend Prest *et al.*'s case study on the exemplary leakage distribution in which each intermediate calculation is assumed to leak its Hamming weight with an additive Gaussian noise of standard deviation σ [44, Table 1]. We complete Table 3 with our new result, by using the fact that for such a leakage model, $\mathsf{MI} = \Theta\left(\frac{\log(M)}{\sigma^2}\right)$. It can be noticed that on this particular leakage distribution, our requirement on the minimal noise level is now the weakest of all security proofs based on the I.S.W. masking scheme.

Table 3: Noise requirements, and illustration on a case study on a Hamming weight leakage model with additive Gaussian noise.

Work (year)	Noise requirement	Equivalent Gaussian noise
[45] (2013)	$\mathsf{EN} \leq \mathcal{O}\left(\frac{1}{dM^3}\right)$	$\sigma \geq \Omega\left(dM^{5/2}\sqrt{\log(M)}\right)$
[23] (2014)	$\mathsf{SD} \leq \mathcal{O}\left(\frac{1}{dM^2}\right)$	$\sigma \geq \Omega\left(dM^{2}\sqrt{\log(M)}\right)$
[25] (2015)	$\mathsf{SD} \leq \mathcal{O}\left(\frac{1}{d}\right)$	$\sigma \geq \Omega\left(d\sqrt{\log(M)}\right)$
[44] (2019)	$\mathsf{RE} \leq \mathcal{O}\left(\frac{1}{d}\right)$	$\sigma \geq \Omega(d\log(M))$[7]
This work	$\mathsf{MI} \leq \mathcal{O}\left(\frac{1}{d}\right)$	$\sigma \geq \Omega\left(\sqrt{d\log(M)}\right)$

At first glance, Table 3 suggests that Prest *et al.*'s RE-based security bound remains quite competitive with the other works based on the noise requirements. However, we emphasize that the RE is a *worst-case* metric, whereas all the other metrics in Table 3 are *averaged* metrics. Estimating worst-case metrics may not always be efficiently tractable by practitioners, especially for high-dimensional leakage. In addition, worst-case metrics are by definition much more conservative than averaged metrics, which contrasts with the concrete SCA security metrics like the GE or the SR [50] that are also averaged metrics. To illustrate this, let us

[7] As explained by Prest *et al.* [44, Remark 2], the RE is not even formally defined for leakage models with Gaussian noise, unless requiring to a tail-cut argument that adds another constant factor hidden in the $\Omega(\cdot)$ notation.

consider another example of leakage distribution, namely the now famous *random probing* model considered by Duc *et al.* in their groundbreaking work [23]. In this leakage model, the adversary can recover the exact value of any intermediate calculation, each with probability $0 \leq \kappa \leq 1$, where the parameter κ denotes the baseline noise level here. It can be verified that the MI of this leakage model is $\log |\mathcal{Y}| \cdot \kappa$, whereas its RE is always fixed to $|\mathcal{Y}| - 1$ regardless of the value of κ, so that the MI can be set arbitrarily close to zero — by setting κ accordingly — while the RE remains constant. In other words, the random probing model can *never* be proven secure with masking by using a security bound involving the RE, whereas our masking security bound remains meaningful for the random probing leakage model, as long as $\kappa \leq \mathcal{O}\left(\frac{1}{\log|\mathcal{Y}| \cdot d}\right)$.[8]

As a result, the only security bound comparable with ours in terms of noise requirements remains Dziembowski *et al.*'s bound [25]. Their bound is obtained using Chernoff-like concentration inequalities [11]. Although this approach has been fruitful in Duc *et al.*'s elegant reduction to the probing model [23] due to its genericity, it has a major drawback since the convergence rate of the security bound no longer depends on the actual baseline noise level δ. This is highlighted in their final bound [25, Eq. (42)]: it can be verified that the bound is even always increasing for values of d between 0 and 8, and becomes non-trivial — *i.e.*, lower than one — only for $d \geq 142$ if $|\mathcal{Y}| = 256$. On the opposite, our security bounds do not suffer from this caveat, since they depend on the actual baseline noise level δ, which makes our bounds non-trivial for arbitrarily small value of d, provided that δ is small enough as we will depict later in Figure 2.

5.2 Beyond Random Plaintext Attacks

One may argue that the latter comparison with the works of Dziembowski *et al.* is not completely fair, since their bound is stated for SCA adversary with chosen plaintext. Hereupon, the authors stated later at Tcc 2016 that by leveraging a reduction from non-uniform secrets to uniform secrets [26, Lemma 2],

> *"The cryptographic interpretation of [reductions from non-uniform to uniform secrets] is that it suffices to consider only random-plaintext attacks, instead of chosen-plaintext attacks"* [26, p. 297].

We notice that our Theorem 2 actually represents such a reduction. Accordingly, our main results Theorem 7 and Theorem 8 can be extended to cover adversaries with chosen plaintexts, by multiplying the constant factors by $(1 - e^{-1}) \cdot |\mathcal{Y}|$, as pointed out in the following Corollary 10.

Corollary 10. *Let Y be a random variable* arbitrarily *distributed over $\mathcal{Y}$, and protected by a masking scheme with $d + 1$ shares as described in subsection 2.2 resulting in T elementary calculations. Assume that the scheme protects $|\mathcal{S}|$ linear*

[8] This condition could even be relaxed to $\kappa \leq \mathcal{O}\left(\frac{1}{d}\right)$ in the particular case of leakage in the random probing model, if one would directly state a security bound for this leakage model, *e.g.*, by extending Eq. (9) of Duc *et al.* [24].

operations, and $|\mathcal{M}|$ I.S.W. multiplications that are part of a monomial Sbox, and protected according to Algorithm 5. Let $\mathbf{L} = (L_i)_{1 \leq i \leq T}$ be the random vector denoting the leakage of the full masking scheme, and let $\delta \geq 0$ be such that every L_i is δ-MI-noisy. Then, the inequality in Theorem 7 is verified for:

$$t_3 = (1 - e^{-1}) \cdot |\mathcal{Y}| \cdot \sum_{(p,q) \in \mathcal{M}} \varphi(p, q, |\mathcal{Y}|) \quad ,$$

$$t_{1,2,4} = (1 - e^{-1}) \cdot |\mathcal{Y}| \cdot \left(\sum_{(p,q) \in \mathcal{M}} \phi(p, q, |\mathcal{Y}|) + \sum_{k \in \mathcal{S}} \psi(k, |\mathcal{Y}|) \right) \quad ,$$

where φ and ϕ are the functions defined in Theorem 8, and ψ is the function defined in Theorem 7.

5.3 Beyond Monomial Sboxes

Likewise, we can extend our previous results to random or chosen plaintext attacks on block ciphers whose Sbox is not a monomial, as stated by Corollary 11.

Corollary 11. *Let Y be a random variable* arbitrarily *distributed over $\mathcal{Y}$, and protected by a masking scheme with $d + 1$ shares as described in subsection 2.2, resulting in T elementary calculations. Assume that the scheme protects $|\mathcal{S}|$ linear operations, and $|\mathcal{M}|$ I.S.W. multiplications. Let $\mathbf{L} = (L_i)_{1 \leq i \leq T}$ be the random vector denoting the leakage of the full masking scheme, and let $\delta \geq 0$ be such that every L_i is δ-MI-noisy. Then, the inequality of Theorem 7 is verified for:*

$$t_3 = 2 \cdot \left(1 - e^{-1}\right) \cdot |\mathcal{Y}|^2 \cdot |\mathcal{M}| \,, \;\; t_{1,2,4} = \quad \left(1 - e^{-1}\right) \cdot |\mathcal{Y}| \cdot (|\mathcal{S}| + |\mathcal{M}|) \,.$$

Proof. We apply Theorem 1, then we group the type 1, 2, and 4 subsequences together and we apply the reduction to uniform secrets using Theorem 2. Likewise, we apply Theorem 2 for type 3 subsequences over the domain $\mathcal{Y} \times \mathcal{Y}$. We can then directly apply Theorem 4 and Theorem 5 respectively. □

Notice that the only difference in the assumptions of Corollary 10 and Corollary 11 is that we no longer need any particular assumption on the Sbox in the latter case. Whether we could leverage further assumptions on the Sbox for arbitrarily distributed secrets is left as an open question for further works.

Table 4 synthesizes the different constant factors t_3, whether the SCA adversary is assumed to operate with random or chosen plaintexts, or whether the blinded I.S.W. multiplication gadget is used or not. Likewise, we may notice that the constant factor of Corollary 11 scaling quadratically with the field size seems at first glance worse than the one of Dziembowski *et al.* [25, Thm. 1], whereas their security bound only scales linearly with the field size $|\mathcal{Y}|$. Nevertheless, our work considers the paradigm where the leakage comes from the *computations* [41], where Dziembowski *et al.*'s result considers the simulation paradigm where the leakage comes from the *wires*. Duc *et al.* argue that security bounds stated

Table 4: Constant factor overhead, depending on the attack scenario, and on the multiplication gadget used.

Sbox \ Plaintext	Random	Chosen
Any Sbox	$\mathcal{O}(\lvert\mathcal{Y}\rvert^2)$	$\mathcal{O}(\lvert\mathcal{Y}\rvert^2)$
Monomial Sbox	$\mathcal{O}(\lvert\mathcal{Y}\rvert^{1+\epsilon})$	$\mathcal{O}(\lvert\mathcal{Y}\rvert^2)$
Monomial Sbox + Blinding	$\mathcal{O}(\lvert\mathcal{Y}\rvert^{\epsilon})$	$\mathcal{O}(\lvert\mathcal{Y}\rvert^{1+\epsilon})$

with the simulation paradigm with leakage from the wires can be converted into security bounds in the "leakage from computations" paradigm by considering wires defined over $\mathcal{Y} \times \mathcal{Y}$ rather than $\mathcal{Y}$ [23, Sec. 5.5]. This would convert the $|\mathcal{Y}|$ constant factor in Dziembowski *et al.*'s result into $|\mathcal{Y}|^2$.[9]

5.4 On the Tweaked ISW Gadget

We finally discuss some aspects of our blinded I.S.W. multiplication gadget. For concreteness, we present hereafter in Table 5 a comparison of the constant factors in Theorem 7 and Theorem 8, for the AES Sbox exponentiation only. We can

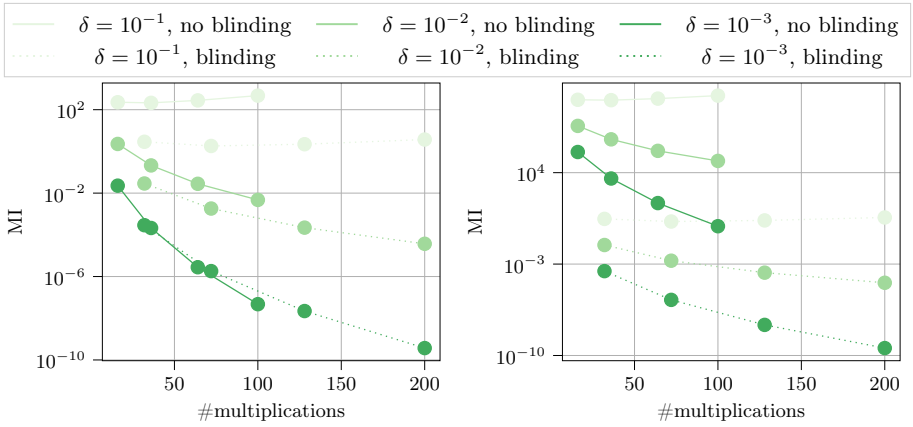
Table 5: Constant factors for the whole AES Sbox exponentiation.

Scheme	Corollary 10		Theorem 7		Theorem 8	
	t_3	$t_{1,2,4}$	t_3	$t_{1,2,4}$	t_3	$t_{1,2,4}$
Rivain-Prouff [47]	**331, 413**	1,132	**4096**	41.1	**32.1**	80.3
Cardoso *et al.* [48]	**331, 413**	1,294	**2048**	40.1	**16.1**	78.2

observe that while using the blinded I.S.W. gadget doubles the $t_{1,2,4}$ constant factor, it decreases the t_3 constant factor by a factor of $|\mathcal{Y}|/2$, which is of at least two orders of magnitude for the AES field. Interestingly, the t_3 constant factor in Cardoso *et al.*'s scheme is even close to 16, which is the tightest possible, given that their exponentiation scheme contains four multiplications. This is because each multiplication in Cardoso *et al.*'s scheme involves either Y or Y^{14}, but both 1 and 14 are coprime with $|\mathcal{Y}| - 1 = 255$.

We end this subsection by discussing whether blinding would have a significant practical interest for current masked implementations of AES. Admittedly, the significant gain in the constant factor comes though with an increased cost in terms of field multiplications and fresh randomness (by a factor two). Figure 2 compares the security bounds for the whole AES Sbox Rivain-Prouff scheme (stated in bits) with respect to the number of field multiplications. We can see on Figure 2a, that for the AES field and $\delta = 10^{-2}$, the dotted curve is below the plain curve, by one order of magnitude. This means that the blinded I.S.W.

[9] We also recall that their bound is stated in terms of SD, whereas ours is stated in terms of MI.



(a) $|\mathcal{Y}| = 2^8$.

(b) $|\mathcal{Y}| = 2^{32}$.

Fig. 2: Security bounds without and with blinding, with respect to the number of field multiplications, for $d \in [1, 5]$. The number of multiplications is calculated over the Rivain-Prouff scheme for the AES Sbox.

multiplication implies tighter security bounds for a comparable implementation cost. However, this gain vanishes for noisier implementations, *e.g.*, for $\delta = 10^{-3}$. The advantage of our blinded I.S.W. becomes more significant when working on larger fields, as depicted on Figure 2b where the field is of size 2^{32}. Whether the advantage of our blinded I.S.W. in terms of *provable* security also translates in terms of actual practical security gains remains an open questions and is let for further investigations in the future.

5.5 Perspectives

The main limitation of our work remains the leak-free assumption for the mask refreshings, like in the previous works [45,25,44]. It remains an open problem whether this assumption could be relaxed. Likewise, our masking security proof only covers the I.S.W. masking scheme, as in the previous works, whereas the generic approach through the probing model can cover any type of masking scheme. Nevertheless, we do not see any prior reason why our security proof could not be used to extend over different masking gadgets, beyond the I.S.W. multiplication gadget, and in particular for table-based masking schemes [18,20], that are known to be efficiently secure in the probing model, but much less in the noisy leakage [51,13]. Overall, this leaves the door open for good opportunities of improvement in the next few years.

Acknowledgments. François-Xavier Standaert is a Senior Associate Researcher 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).

A Utilitary Lemma

Proposition 2 (MGL properties). *Let* $\mathrm{f}(\cdot)$ *be the MGL function defined in Equation 9, and let* $\delta_0, \ldots, \delta_d$ *be* $d+1$ *mutually independent random variables. Then the following inequality holds:*

$$\mathop{\mathbb{E}}_{\delta_0,\ldots,\delta_d}\left[\mathrm{f}(\delta_0,\ldots,\delta_d)\right] \leq \mathrm{f}\Big(\mathbb{E}\left[\delta_0\right],\ldots,\mathbb{E}\left[\delta_d\right]\Big) \quad . \tag{19}$$

Lemma 2. *Let* $Y \in \mathcal{Y}$ *be a discrete random variable, and let* $g : Y \mapsto g(Y)$ *be a mapping* $\mathcal{Y} \to \mathcal{Y}$*. Let* $\mathbf{L} : \mathcal{Y} \to \mathcal{L}$ *be a random variable. Then:*

$$\mathsf{MI}(Y; \mathbf{L}(g(Y))) = \mathsf{MI}(g(Y); \mathbf{L}(g(Y))) \quad .$$

Proof of Lemma 2. First, notice that we have the two following Markov chains:

$$\begin{aligned} \mathrm{Y} &\to g(\mathrm{Y}) \to \mathbf{L}(g(\mathrm{Y})) \ , \\ g(\mathrm{Y}) &\leftarrow \mathrm{Y} \to \mathbf{L}(g(\mathrm{Y})) \ . \end{aligned}$$

By the DPI [21, Sec. 2.8] on the first two chains, we have $\mathsf{MI}(\mathrm{Y}; \mathbf{L}(g(\mathrm{Y}))) \lesseqgtr \mathsf{MI}(g(\mathrm{Y}); \mathbf{L}(g(\mathrm{Y})))$, hence:

$$\mathsf{MI}(\mathrm{Y}; \mathbf{L}(g(\mathrm{Y}))) = \mathsf{MI}(g(\mathrm{Y}); \mathbf{L}(g(\mathrm{Y}))) \quad .$$

□

Lemma 3. *Let* Y, R *be two independent random variables, uniformly distributed over a field* $\mathcal{Y}$ *of size* M*. For all* $a, b \in \mathcal{Y}$*, we have*

$$\Pr\big(Y^{p+q} \oplus R Y^q = a, R Y^q = b\big) \leq \frac{1}{M^2} \cdot \max\left\{2, \frac{M}{M-1} \cdot \gcd(p+q, M-1)\right\} \quad . \tag{20}$$

Proof. Denote by p the left hand-side of Equation 20. Notice that we may restate p as follows:

$$\mathsf{p} = \Pr\big(\mathrm{Y}^{p+q} = a \oplus b \ \mid \ \mathrm{R}\mathrm{Y}^q = b\big) \Pr(\mathrm{R}\mathrm{Y}^q = b)$$

Let us distinguish the following cases, in which we will show that p is always upper bounded by the right hand-side of Equation 20.

Case $a = b = 0$. Here, $\Pr\big(\mathrm{Y}^{p+q} = 0 \ \mid \ \mathrm{R}\mathrm{Y}^q = 0\big) = \Pr\big(\mathrm{Y}^{p+q} = 0\big) = \frac{1}{M}$, and

$$\begin{aligned} \Pr(\mathrm{R}\mathrm{Y}^q = 0) &= \Pr(\mathrm{R} = 0 \cup \mathrm{Y}^q = 0) \\ &= \Pr(\mathrm{R} = 0) + \Pr(\mathrm{Y}^q = 0) - \Pr(\mathrm{R} = 0 \cap \mathrm{Y}^q = 0) \\ &= \Pr(\mathrm{R} = 0) + \Pr(\mathrm{Y}^q = 0) - \Pr(\mathrm{R} = 0) \cdot \Pr(\mathrm{Y}^q = 0) \quad , \end{aligned}$$

where the first equality comes from the property of the field multiplication, the second equality is an application of the inclusion/exclusion formula, and the last equality comes from the independence between R and Y. Therefore, it comes that

$$\Pr(\mathrm{R}\mathrm{Y}^q = 0) = \frac{2}{M} - \frac{1}{M^2} \leq \frac{2}{M} \quad . \tag{21}$$

Hence, $\mathsf{p} \leq \frac{2}{M^2}$.

Case $a \neq 0, b = 0$. Using Bayes' theorem, we may restate p as follows:

$$\begin{aligned}\mathsf{p} &= \Pr\big(\mathrm{RY}^q = 0 \mid \mathrm{Y}^{p+q} = a\big) \cdot \Pr\big(\mathrm{Y}^{p+q} = a\big) \\ &= \Pr\big(\mathrm{R} = 0 \cup \mathrm{Y} = 0 \mid \mathrm{Y}^{p+q} = a\big) \cdot \Pr\big(\mathrm{Y}^{p+q} = a\big) \\ &= \Pr(\mathrm{R} = 0) \cdot \Pr\big(\mathrm{Y}^{p+q} = a\big) \\ &= \frac{1}{M} \cdot \Pr\big(\mathrm{Y}^{p+q} = a\big) \ ,\end{aligned}$$

where the third equality comes from R and Y being independent, and necessarily $\mathrm{Y} \neq 0$. Using Equation 6 and Equation 21, it comes that

$$\mathsf{p} \leq \frac{\gcd(p+q, M-1)}{M(M-1)} \leq 2\frac{\gcd(p+q, M-1)}{M^2} \ .$$

Case $b \neq 0$. Here, RY^q is uniformly distributed over the non-zero values of $\mathcal{Y}$, so $\Pr(\mathrm{RY}^q = b) \leq \frac{1}{M-1}$, and is independent of Y. As a consequence, we have

$$\Pr\big(\mathrm{Y}^{p+q} = a \oplus b \mid \mathrm{RY}^q = b\big) = \Pr\big(\mathrm{Y}^{p+q} = a \oplus b\big) \ ,$$

so $\mathsf{p} = \Pr\big(\mathrm{Y}^{p+q} = a \oplus b\big) \cdot \Pr(\mathrm{RY}^q = b)$. It remains to bound the first factor of p using Equation 6, and we get $\mathsf{p} \leq \frac{\gcd(p+q, M-1)}{(M-1)^2}$. Finally, using the inequality $\frac{1}{M-1} \leq \frac{2}{M}$, for $M \geq 2$, we obtain $\mathsf{p} \leq 2\frac{\gcd(p+q, M-1)}{M^2}$. □

B Proofs of Main Results

Proof of Theorem 1. By definition, we have

$$\mathsf{H}(\mathbf{L} \mid \mathrm{Y}) = \mathop{\mathbb{E}}_{y}\big[\mathsf{H}(\mathrm{L}_1, \ldots, \mathrm{L}_t \mid \mathrm{Y} = y)\big] \ . \tag{22}$$

By assumption, all the leakages, conditioned to $\mathrm{Y} = y$ are mutually independent so

$$\mathsf{H}(\mathbf{L} \mid \mathrm{Y} = y) = \sum_{i=1}^{t} \mathsf{H}(\mathrm{L}_i \mid \mathrm{Y} = y) \ .$$

Hence, combining with Equation 22, $\mathsf{H}(\mathbf{L} \mid \mathrm{Y}) = \sum_{i=1}^{t} \mathsf{H}(\mathrm{L}_i \mid \mathrm{Y})$. Thereby,

$$\mathsf{MI}(\mathbf{L}; \mathrm{Y}) \leq \sum_{i=1}^{t} \mathsf{MI}(\mathrm{L}_i; \mathrm{Y})$$

□

Proof of Theorem 2. Now, we can see $\mathbf{L}$ as an — undesired — communication channel. By definition of the capacity C of the channel $\mathbf{L}$, and using Lemma 2, we get that

$$\mathsf{MI}(\mathrm{Y}; \mathbf{L}(g(\mathrm{Y}))) = \mathsf{MI}(g(\mathrm{Y}); \mathbf{L}(g(\mathrm{Y}))) \leq \max_{\Pr(\mathrm{Z})} \mathsf{MI}(\mathrm{Z}; \mathbf{L}(\mathrm{Z})) = C \ .$$

Using [49, Thm. 1, Eq. (17)], we get that

$$\frac{C}{\mathsf{MI}(\mathrm{Y};\mathbf{L}(\mathrm{Y}))} \leq |\mathcal{Y}| \cdot \min\left\{2^{-C}, 1-e^{-1}\right\} \leq |\mathcal{Y}| \cdot \left(1-e^{-1}\right)$$

□

B.1 Proof of Theorem 5 and Theorem 6

Proof of Theorem 5. Using the chain rule of MI [21, Thm. 2.5.2], we have:

$$\mathsf{MI}((\mathrm{A},\mathrm{B});\mathbf{L}) = \mathsf{MI}(\mathrm{A};\mathbf{L}) + \mathsf{MI}(\mathrm{B};\mathbf{L} \mid \mathrm{A}) \quad . \tag{23}$$

Let us bound the first term of Equation 23. The bound on the second term will straightforwardly follow.

Bounding $\mathsf{MI}(A;\mathbf{L})$. Observe that since A and B are independent, it follows that A and $\mathbf{B}$ are also independent. As a result,

$$\mathsf{MI}(\mathrm{A};\mathbf{L}) \leq \mathsf{MI}(\mathrm{A};\mathbf{L} \mid \mathbf{B}) = \mathop{\mathbb{E}}_{\mathbf{b}}\left[\mathsf{MI}(\mathrm{A};\mathbf{L} \mid \mathbf{B}=\mathbf{b})\right] \quad . \tag{24}$$

Let $\mathbf{b} = (b_0, \ldots, b_d)$ be fixed for now, and let us bound $\mathsf{MI}(\mathrm{A};\mathbf{L} \mid \mathbf{B}=\mathbf{b})$. To this end, notice that we may now gather the leakages $\mathbf{L}_{i,j}$ by batches sharing the same index i as follows:

$$\mathsf{MI}(\mathrm{A};\mathbf{L} \mid \mathbf{B}=\mathbf{b}) = \mathsf{MI}\Big(\mathrm{A};\{\mathbf{L}_{0,j}(\mathrm{A}_0,b_j)\}_{0\leq j\leq d},\ldots,\{\mathbf{L}_{d,j}(\mathrm{A}_d,b_j)\}_{0\leq j\leq d}\Big) \quad . \tag{25}$$

By assumption, each batch of leakages $\{\mathbf{L}_{i,j}(\mathrm{A}_i,b_j)\}_{0\leq j\leq d}$ only depends on the share A_i. Hence, we may use Theorem 4 to bound the right hand-side of Equation 25 as follows. Let us define $\mathsf{MI}\Big(\mathrm{A}_i;\{\mathbf{L}_{i,j}\}_{0\leq j\leq d}\Big) = \delta'_i$ — notice that δ'_i depends on b_i. Then we have

$$\mathsf{MI}\Big(\mathrm{A};\{\mathbf{L}_{0,j}\}_{0\leq j\leq d},\ldots,\{\mathbf{L}_{d,j}\}_{0\leq j\leq d}\Big) \underset{(9)}{\leq} \mathrm{f}_{2^n}(\delta'_0,\ldots,\delta'_d) \quad . \tag{26}$$

Substituting Equation 26 in Equation 25, and then plugging into Equation 24 gives

$$\mathsf{MI}(\mathrm{A};\mathbf{L}) \leq \mathop{\mathbb{E}}_{b}\left[\mathrm{f}_{2^n}(\delta'_0,\ldots,\delta'_d)\right] \leq \mathrm{f}_{2^n}\left(\mathop{\mathbb{E}}_{b_0}\left[\delta'_0\right],\ldots,\mathop{\mathbb{E}}_{b_d}\left[\delta'_d\right]\right) \quad , \tag{27}$$

where the second inequality comes from Proposition 2. We are then reduced to upper bound $\mathbb{E}\left[\delta'_i\right]$ for all $0 \leq i \leq d$. To this end, notice that for i fixed, the batch of leakages $\{\mathbf{L}_{i,j}(\mathrm{A}_i,b_j) \mid \mathrm{A}_i\}_{0\leq j\leq d}$ are mutually independent. Hence, we can now leverage Theorem 1 to upper bound δ', as follows:

$$\mathsf{MI}\Big(\mathrm{A}_i;\{\mathbf{L}_{i,j}(\mathrm{A}_i,b_j)\}_{0\leq j\leq d}\Big) \underset{(2)}{\leq} \sum_{j=0}^{d} \mathsf{MI}(\mathrm{A}_i;\mathbf{L}_{i,j}(\mathrm{A}_i,b_j)) \quad . \tag{28}$$

Using the chain rule of MI [21, Thm. 2.5.2] the other way around, we get that

$$\mathsf{MI}(\mathrm{A}_i; \mathbf{L}_{i,j}(\mathrm{A}_i, b_j)) \leq \mathsf{MI}(\mathrm{A}_i, \mathrm{B}_j; \mathbf{L}_{i,j}(\mathrm{A}_i, \mathrm{B}_j)) = \delta_{i,j} \ . \tag{29}$$

Hence, combining Equation 28 with Equation 29, and taking the expectation, we get that

$$\mathbb{E}\left[\delta'_i\right] \leq \sum_{j=0}^{d} \delta_{i,j} \ . \tag{30}$$

Finally, plugging Equation 30 into Equation 27 gives the first term in the right hand-side of Equation 11.

Bounding $\mathsf{MI}(B; \mathbf{L} \mid A)$. Using the chain rule of the MI again, we may bound $\mathsf{MI}(\mathrm{B}; \mathbf{L} \mid \mathrm{A})$ by conditioning on the d last shares of $\mathbf{A}$ (except the share of index 0):

$$\mathsf{MI}(\mathrm{B}; \mathbf{L} \mid \mathrm{A}) \leq \mathsf{MI}\Big(\mathrm{B}; \mathbf{L} \mid \mathrm{A}, \{\mathrm{A}_i\}_{1 \leq i \leq d}\Big)$$

Using the same argument as Dziembowski *et al.* [26, Lemma 3], we may notice that since A is assumed to be uniform:

$$\Big(\mathrm{A}, \{\mathrm{A}_i\}_{1 \leq i \leq d}\Big) \stackrel{d}{=} \left(\mathrm{A} \oplus \left(\bigoplus_{i=1}^{d} \mathrm{A}_i\right), \{\mathrm{A}_i\}_{1 \leq i \leq d}\right) \stackrel{d}{=} \{\mathrm{A}_i\}_{0 \leq i \leq d} \ ,$$

it implies that $\mathsf{MI}\Big(\mathrm{B}; \mathbf{L} \mid \mathrm{A}, \{\mathrm{A}_i\}_{1 \leq i \leq d}\Big) = \mathsf{MI}(\mathrm{B}; \mathbf{L} \mid \mathbf{A})$. By symmetry of the roles, the latter term can be bound in the same way as the right hand-side of Equation 24, by permuting the roles of the indices i and j. □

Proof of Theorem 6. Let $\mathbf{L} = (\mathrm{L}_1(\mathrm{Y}_0, \mathrm{Y}_1), \ldots, \mathrm{L}_d(\mathrm{Y}_{d-1}, \mathrm{Y}_d))$ for short. Expanding $\mathsf{MI}(\mathrm{Y}_d; \mathbf{L})$, we have

$$\begin{aligned}\mathsf{MI}(\mathrm{Y}_d; \mathbf{L}) = \mathsf{MI}(\mathrm{Y}_d; \mathrm{L}_d(\mathrm{Y}_{d-1}, \mathrm{Y}_d) \mid \mathrm{L}_{d-1}(\mathrm{Y}_{d-2}, \mathrm{Y}_{d-1}), \ldots, \mathrm{L}_1(\mathrm{Y}_0, \mathrm{Y}_1)) \\ + \mathsf{MI}(\mathrm{Y}_d; \mathrm{L}_{d-1}(\mathrm{Y}_{d-2}, \mathrm{Y}_{d-1}), \ldots, \mathrm{L}_1(\mathrm{Y}_0, \mathrm{Y}_1))\end{aligned}$$

Notice first that the second term in the right hand-side equals 0, since by assumption Y_d is independent of the $\{\mathrm{Y}_i\}_{0 \leq i \leq d-1}$. Likewise, the first term of the right hand-side can be upper bounded by $\mathsf{MI}(\mathrm{Y}_d; \mathrm{L}_d(\mathrm{Y}_{d-1}, \mathrm{Y}_d) \mid \mathrm{Y}_{d-1})$, which can in turn be upper bounded by δ_d. □

B.2 Proofs for the Blinded ISW Gadget

Proof of Theorem 8. We now show how to bound the MI between Y and the whole leakage. Notice that we may not directly use Theorem 1 to upper bound $\mathsf{MI}(\mathrm{Y}; \mathbf{L})$ by the sum of the MIs over all the elementary subsequences of the gadget in Figure 1, since they are not all independent due to the presence of R. This problem can be easily circumvented by using the DPI:

$$\mathsf{MI}(\mathrm{Y}; \mathbf{L}) \leq \mathsf{MI}(\mathrm{Y}, \mathrm{R}; \mathbf{L}) \ . \tag{31}$$

Since the encodings of each sharing is refreshed by the leak-free refresh oracle, all the subsequences in the blinded I.S.W. are now mutually independent, so we may now use Theorem 1:

$$\begin{aligned}\mathsf{MI}(\mathrm{Y},\mathrm{R};\mathbf{L}) = \mathsf{MI}(\mathrm{Y},\mathrm{R};\mathbf{L}_{\oplus_1}) + \mathsf{MI}(\mathrm{Y},\mathrm{R};\mathbf{L}_{\oplus_2}) \\ + \mathsf{MI}\big(\mathrm{Y},\mathrm{R};\mathbf{L}_{\mathsf{ISW}_{1,in}}\big) + \mathsf{MI}\big(\mathrm{Y},\mathrm{R};\mathbf{L}_{\mathsf{ISW}_{1,out}}\big) \\ + \mathsf{MI}\big(\mathrm{Y},\mathrm{R};\mathbf{L}_{\mathsf{ISW}_{2,in}}\big) + \mathsf{MI}\big(\mathrm{Y},\mathrm{R};\mathbf{L}_{\mathsf{ISW}_{2,out}}\big)\end{aligned} \tag{32}$$

We shall upper bound each term in the right hand-side of Equation 32.

Bounding $\mathsf{MI}(\mathbf{Y},\mathbf{R};\mathbf{L}_{\oplus_1})$. Recall that $\mathbf{L}_{\oplus_1}$ is the random variable denoting the leakage of the input operands of the first Xor in Figure 1, namely Y^p and R. Using Lemma 2, we get that $\mathsf{MI}(\mathrm{Y},\mathrm{R};\mathbf{L}_{\oplus_1}(\mathrm{Y}^p,\mathrm{R})) = \mathsf{MI}(\mathrm{Y}^p,\mathrm{R};\mathbf{L}_{\oplus_1}(\mathrm{Y}^p,\mathrm{R}))$. Observe that G and R are independent, so we may refine Corollary 3 . To this end, let A be uniformly distributed over $\mathcal{Y}$, and independent of R. Observe that for any $g, r \in \mathcal{Y}^2$,

$$\frac{\Pr(\mathrm{Y}^p = g, \mathrm{R} = r)}{\Pr(\mathrm{A} = g, \mathrm{R} = r)} = \frac{\Pr(\mathrm{Y}^p = g)\Pr(\mathrm{R} = r)}{\Pr(\mathrm{A} = g)\Pr(\mathrm{R} = r)} \leq \frac{|\mathcal{Y}|}{|\mathcal{Y}|-1} \cdot \gcd(p, |\mathcal{Y}|-1) \quad . \tag{33}$$

Therefore, injecting Equation 33 into Equation 4 gives

$$\mathsf{MI}(\mathrm{Y}^p,\mathrm{R};\mathbf{L}_{\oplus_1}(\mathrm{Y}^p,\mathrm{R})) \leq \frac{|\mathcal{Y}|}{|\mathcal{Y}|-1} \cdot \gcd(p, |\mathcal{Y}|-1) \cdot \mathsf{MI}(\mathrm{A},\mathrm{R};\mathbf{L}_{\oplus_1}(\mathrm{A},\mathrm{R})) \quad .$$

We are then reduced to bound $\mathsf{MI}(\mathrm{A},\mathrm{R};\mathbf{L}_{\oplus_1})$. Applying Corollary 4 on the pair of variables (A,R), we get that:

$$\mathsf{MI}(\mathrm{A},\mathrm{R};\mathbf{L}_{\oplus_1}) \leq \mathrm{f}(\delta,\ldots,\delta) \quad .$$

Putting everything together, we get that

$$\mathsf{MI}(\mathrm{Y}^p,\mathrm{R};\mathbf{L}_{\oplus_1}) \leq \frac{|\mathcal{Y}|}{|\mathcal{Y}|-1} \cdot \gcd(p, |\mathcal{Y}|-1) \cdot \mathrm{f}(\delta,\ldots,\delta) \quad . \tag{34}$$

Bounding $\mathsf{MI}(\mathbf{Y},\mathbf{R};\mathbf{L}_{\oplus_2})$. Using Lemma 2, we get that

$$\mathsf{MI}\big(\mathrm{Y},\mathrm{R};\mathbf{L}_{\oplus_2}(\mathrm{H}',\mathrm{M})\big) = \mathsf{MI}\big(\mathrm{H}',\mathrm{M};\mathbf{L}_{\oplus_2}(\mathrm{H}',\mathrm{M})\big) \quad .$$

Using Lemma 3, we deduce that

$$\mathsf{MI}\big(\mathrm{H}',\mathrm{M};\mathbf{L}_{\oplus_2}(\mathrm{H}',\mathrm{M})\big) \leq \max\left\{2, \frac{M}{M-1} \cdot \gcd(p+q, M-1)\right\} \cdot \mathsf{MI}(\mathrm{A},\mathrm{B};\mathbf{L}_{\oplus_2}(\mathrm{A},\mathrm{B})) \quad ,$$

where A,B are independent and uniformly distributed over $\mathcal{Y}$. We may then apply Corollary 4 to get

$$\mathsf{MI}(\mathrm{A},\mathrm{B};\mathbf{L}_{\oplus_2}(\mathrm{A},\mathrm{B})) \leq \mathrm{f}(\delta,\ldots,\delta) \quad .$$

Putting everything together, we get that

$$\mathsf{MI}(\mathrm{Y},\mathrm{R};\mathbf{L}_{\oplus_2}) \leq \max\left\{2, \frac{M}{M-1} \cdot \gcd(p+q, M-1)\right\} \cdot \mathrm{f}(\delta,\ldots,\delta) \quad . \tag{35}$$

Bounding $\mathsf{MI}\big(\mathrm{Y},\mathrm{R};\mathbf{L}_{\mathsf{ISW}_{1,in}}\big)$. Using Lemma 2, we get that

$$\mathsf{MI}\big(\mathrm{Y},\mathrm{R};\mathbf{L}_{\mathsf{ISW}_{1,in}}(\mathrm{G}',\mathrm{H})\big) = \mathsf{MI}\big(\mathrm{G}',\mathrm{H};\mathbf{L}_{\mathsf{ISW}_{1,in}}(\mathrm{G}',\mathrm{H})\big) \enspace .$$

Observe that the input operands of ISW_1 are independent, and that G' is uniformly distributed over $\mathcal{Y}$. Moreover, for all $h \in \mathcal{Y}$ we have $\Pr(\mathrm{H}=h) \leq \frac{\gcd(q,|\mathcal{Y}|-1)}{|\mathcal{Y}|-1}$, which implies that

$$\mathsf{MI}\big(\mathrm{G}',\mathrm{H};\mathbf{L}_{\mathsf{ISW}_{1,in}}(\mathrm{G}',\mathrm{H})\big) \leq \frac{|\mathcal{Y}|}{|\mathcal{Y}|-1}\cdot\gcd(q,|\mathcal{Y}|-1)\,\mathsf{MI}\big(\mathrm{G}',\mathrm{B};\mathbf{L}_{\mathsf{ISW}_{1,in}}(\mathrm{G}',\mathrm{B})\big) \enspace ,$$

where $\mathrm{B} \in \mathcal{Y}$ is uniform and independent of G'. We may then apply Theorem 5 to get

$$\mathsf{MI}\big(\mathrm{G}',\mathrm{B};\mathbf{L}_{\mathsf{ISW}_{1,in}}(\mathrm{G}',\mathrm{B})\big) \leq 2\,\mathrm{f}((d+1)\delta,\ldots,(d+1)\delta) \enspace .$$

Putting everything together, we have that

$$\mathsf{MI}\big(\mathrm{Y},\mathrm{R};\mathbf{L}_{\mathsf{ISW}_{1,in}}\big) \leq \frac{2\,|\mathcal{Y}|}{|\mathcal{Y}|-1}\cdot\gcd(q,|\mathcal{Y}|-1)\cdot\mathrm{f}((d+1)\delta,\ldots,(d+1)\delta) \enspace . \quad (36)$$

Bounding $\mathsf{MI}\big(\mathbf{Y},\mathbf{R};\mathbf{L}_{\mathsf{ISW}_{2,in}}\big)$. In this case, we get exactly the same bound as in Equation 36, by changing G' with R. The same arguments then apply.

Bounding $\mathsf{MI}\big(\mathbf{Y},\mathbf{R};\mathbf{L}_{\mathsf{ISW}_{1,out}}\big)$. Using Lemma 2, we get that

$$\mathsf{MI}\big(\mathrm{Y},\mathrm{R};\mathbf{L}_{\mathsf{ISW}_{1,out}}(\mathrm{M})\big) = \mathsf{MI}\big(\mathrm{M};\mathbf{L}_{\mathsf{ISW}_{1,out}}(\mathrm{M})\big) \enspace .$$

Then, notice that $\mathrm{M} = (\mathrm{G}\oplus\mathrm{R})\mathrm{H}$, and that we argued that $\mathrm{G}\oplus\mathrm{R} = \mathrm{G}'$ is uniformly distributed, and independent of H. Therefore, M is uniformly distributed over the non-zero values of $\mathcal{Y}$, provided that $\mathrm{M} \neq 0$. If not, then we have $\Pr(\mathrm{M}=0) = \frac{1}{M}$. Overall, for all $y \in \mathcal{Y}$,

$$\frac{\Pr(\mathrm{M}=y)}{\Pr(\mathrm{Y}=y)} \leq 2 \enspace .$$

By virtue of Equation 6, we have that

$$\mathsf{MI}\big(\mathrm{M};\mathbf{L}_{\mathsf{ISW}_{1,out}}(\mathrm{M})\big) \leq 2\,\mathsf{MI}\big(\mathrm{Y};\mathbf{L}_{\mathsf{ISW}_{1,out}}(\mathrm{Y})\big) \enspace .$$

We can now apply the remaining of the proof of Corollary 7 (starting after the reduction from non-uniform to uniform secrets) to deduce that

$$\mathsf{MI}\big(\mathrm{Y};\mathbf{L}_{\mathsf{ISW}_{1,out}}(\mathrm{Y})\big) \leq \mathrm{f}(\delta,\ldots,\delta) \enspace .$$

Putting everything together, we have that

$$\mathsf{MI}\big(\mathrm{Y},\mathrm{R};\mathbf{L}_{\mathsf{ISW}_{1,out}}(\mathrm{M})\big) \leq 2\,\mathrm{f}(\delta,\ldots,\delta) \enspace . \quad (37)$$

Bounding $\mathsf{MI}\big(\mathbf{Y},\mathbf{R};\mathbf{L}_{\mathsf{ISW}_{2,out}}\big)$. Observing that $\mathrm{G}\oplus\mathrm{R}$ may be replaced by R in the previous case of $\mathbf{L}_{\mathsf{ISW}_{1,out}}$ without any loss of generality, we get the same bound as in Equation 37.

□

References

1. Ajtai, M.: Secure computation with information leaking to an adversary. In: Fortnow, L., Vadhan, S.P. (eds.) 43rd Annual ACM Symposium on Theory of Computing. pp. 715–724. ACM Press, San Jose, CA, USA (Jun 6–8, 2011). https://doi.org/10.1145/1993636.1993731 20
2. Ananth, P., Ishai, Y., Sahai, A.: Private circuits: A modular approach. In: Shacham, H., Boldyreva, A. (eds.) Advances in Cryptology – CRYPTO 2018, Part III. Lecture Notes in Computer Science, vol. 10993, pp. 427–455. Springer, Heidelberg, Germany, Santa Barbara, CA, USA (Aug 19–23, 2018). https://doi.org/10.1007/978-3-319-96878-0_15 20
3. Andrychowicz, M., Dziembowski, S., Faust, S.: Circuit compilers with $O(1/\log(n))$ leakage rate. In: Fischlin, M., Coron, J.S. (eds.) Advances in Cryptology – EUROCRYPT 2016, Part II. Lecture Notes in Computer Science, vol. 9666, pp. 586–615. Springer, Heidelberg, Germany, Vienna, Austria (May 8–12, 2016). https://doi.org/10.1007/978-3-662-49896-5_21 20
4. Azouaoui, M., Bellizia, D., Buhan, I., Debande, N., Duval, S., Giraud, C., Jaulmes, E., Koeune, F., Oswald, E., Standaert, F.X., Whitnall, C.: A systematic appraisal of side channel evaluation strategies. Cryptology ePrint Archive, Report 2020/1347 (2020), https://eprint.iacr.org/2020/1347 5
5. Baignères, T., Junod, P., Vaudenay, S.: How far can we go beyond linear cryptanalysis? In: Lee, P.J. (ed.) Advances in Cryptology – ASIACRYPT 2004. Lecture Notes in Computer Science, vol. 3329, pp. 432–450. Springer, Heidelberg, Germany, Jeju Island, Korea (Dec 5–9, 2004). https://doi.org/10.1007/978-3-540-30539-2_31 7
6. Barthe, G., Belaïd, S., Dupressoir, F., Fouque, P.A., Grégoire, B., Strub, P.Y., Zucchini, R.: Strong non-interference and type-directed higher-order masking. In: Weippl, E.R., Katzenbeisser, S., Kruegel, C., Myers, A.C., Halevi, S. (eds.) ACM CCS 2016: 23rd Conference on Computer and Communications Security. pp. 116–129. ACM Press, Vienna, Austria (Oct 24–28, 2016). https://doi.org/10.1145/2976749.2978427 3
7. Battistello, A., Coron, J.S., Prouff, E., Zeitoun, R.: Horizontal side-channel attacks and countermeasures on the ISW masking scheme. In: Gierlichs, B., Poschmann, A.Y. (eds.) Cryptographic Hardware and Embedded Systems – CHES 2016. Lecture Notes in Computer Science, vol. 9813, pp. 23–39. Springer, Heidelberg, Germany, Santa Barbara, CA, USA (Aug 17–19, 2016). https://doi.org/10.1007/978-3-662-53140-2_2 20
8. Belaïd, S., Coron, J.S., Prouff, E., Rivain, M., Taleb, A.R.: Random probing security: Verification, composition, expansion and new constructions. In: Micciancio, D., Ristenpart, T. (eds.) Advances in Cryptology – CRYPTO 2020, Part I. Lecture Notes in Computer Science, vol. 12170, pp. 339–368. Springer, Heidelberg, Germany, Santa Barbara, CA, USA (Aug 17–21, 2020). https://doi.org/10.1007/978-3-030-56784-2_12 3, 20
9. Belaïd, S., Rivain, M., Taleb, A.R.: On the power of expansion: More efficient constructions in the random probing model. In: Canteaut, A., Standaert, F.X. (eds.) Advances in Cryptology – EUROCRYPT 2021, Part II. Lecture Notes in Computer Science, vol. 12697, pp. 313–343. Springer, Heidelberg, Germany, Zagreb, Croatia (Oct 17–21, 2021). https://doi.org/10.1007/978-3-030-77886-6_11 3
10. Belaïd, S., Rivain, M., Taleb, A.R., Vergnaud, D.: Dynamic random probing expansion with quasi linear asymptotic complexity. In: Tibouchi, M., Wang, H. (eds.)

Advances in Cryptology – ASIACRYPT 2021, Part II. Lecture Notes in Computer Science, vol. 13091, pp. 157–188. Springer, Heidelberg, Germany, Singapore (Dec 6–10, 2021). https://doi.org/10.1007/978-3-030-92075-3_6 3
11. Boucheron, S., Lugosi, G., Massart, P.: Concentration inequalities: A nonasymptotic theory of independence. Oxford university press (2013) 21
12. Broughan, K.A.: The gcd-sum function. J. Integer Seq. **4**(2.2) (2001) 13
13. Bruneau, N., Guilley, S., Najm, Z., Teglia, Y.: Multivariate high-order attacks of shuffled tables recomputation. Journal of Cryptology **31**(2), 351–393 (Apr 2018). https://doi.org/10.1007/s00145-017-9259-7 24
14. Béguinot, J., Cheng, W., Guilley, S., Liu, Y., Masure, L., Rioul, O., Standaert, F.X.: Removing the field size loss from duc et al.'s conjectured bound for masked encodings. Cryptology ePrint Archive, Paper 2022/1738 (2022), https://eprint.iacr.org/2022/1738, https://eprint.iacr.org/2022/1738 4, 5, 7, 14
15. Cassiers, G., Faust, S., Orlt, M., Standaert, F.X.: Towards tight random probing security. In: Malkin, T., Peikert, C. (eds.) Advances in Cryptology – CRYPTO 2021, Part III. Lecture Notes in Computer Science, vol. 12827, pp. 185–214. Springer, Heidelberg, Germany, Virtual Event (Aug 16–20, 2021). https://doi.org/10.1007/978-3-030-84252-9_7 3
16. Chari, S., Jutla, C.S., Rao, J.R., Rohatgi, P.: Towards sound approaches to counteract power-analysis attacks. In: Wiener [52], pp. 398–412. https://doi.org/10.1007/3-540-48405-1_26 2
17. Cheng, W., Liu, Y., Guilley, S., Rioul, O.: Attacking masked cryptographic implementations: Information-theoretic bounds. In: IEEE International Symposium on Information Theory, ISIT 2022, Espoo, Finland, June 26 - July 1, 2022. pp. 654–659. IEEE (2022). https://doi.org/10.1109/ISIT50566.2022.9834556, https://doi.org/10.1109/ISIT50566.2022.9834556 7
18. Coron, J.S.: Higher order masking of look-up tables. In: Nguyen and Oswald [43], pp. 441–458. https://doi.org/10.1007/978-3-642-55220-5_25 24
19. Coron, J.S., Prouff, E., Rivain, M., Roche, T.: Higher-order side channel security and mask refreshing. In: Moriai [42], pp. 410–424. https://doi.org/10.1007/978-3-662-43933-3_21 3
20. Coron, J.S., Rondepierre, F., Zeitoun, R.: High order masking of look-up tables with common shares. IACR Transactions on Cryptographic Hardware and Embedded Systems **2018**(1), 40–72 (2018). https://doi.org/10.13154/tches.v2018.i1.40-72, https://tches.iacr.org/index.php/TCHES/article/view/832 24
21. Cover, T.M., Thomas, J.A.: Elements of information theory (2. ed.). Wiley (2006) 25, 27, 28
22. de Chérisey, E., Guilley, S., Rioul, O., Piantanida, P.: Best information is most successful. IACR Transactions on Cryptographic Hardware and Embedded Systems **2019**(2), 49–79 (2019). https://doi.org/10.13154/tches.v2019.i2.49-79, https://tches.iacr.org/index.php/TCHES/article/view/7385 7
23. Duc, A., Dziembowski, S., Faust, S.: Unifying leakage models: From probing attacks to noisy leakage. In: Nguyen and Oswald [43], pp. 423–440. https://doi.org/10.1007/978-3-642-55220-5_24 3, 18, 20, 21, 23
24. Duc, A., Faust, S., Standaert, F.X.: Making masking security proofs concrete - or how to evaluate the security of any leaking device. In: Oswald, E., Fischlin, M. (eds.) Advances in Cryptology – EUROCRYPT 2015, Part I. Lecture Notes in Computer Science, vol. 9056, pp. 401–429. Springer, Heidelberg, Germany, Sofia, Bulgaria (Apr 26–30, 2015). https://doi.org/10.1007/978-3-662-46800-5_16 4, 5, 7, 21

25. Dziembowski, S., Faust, S., Skorski, M.: Noisy leakage revisited. In: Oswald, E., Fischlin, M. (eds.) Advances in Cryptology – EUROCRYPT 2015, Part II. Lecture Notes in Computer Science, vol. 9057, pp. 159–188. Springer, Heidelberg, Germany, Sofia, Bulgaria (Apr 26–30, 2015). https://doi.org/10.1007/978-3-662-46803-6_6 3, 4, 20, 21, 22, 24
26. Dziembowski, S., Faust, S., Skórski, M.: Optimal amplification of noisy leakages. In: Kushilevitz, E., Malkin, T. (eds.) TCC 2016-A: 13th Theory of Cryptography Conference, Part II. Lecture Notes in Computer Science, vol. 9563, pp. 291–318. Springer, Heidelberg, Germany, Tel Aviv, Israel (Jan 10–13, 2016). https://doi.org/10.1007/978-3-662-49099-0_11 5, 18, 21, 28
27. Goubin, L., Patarin, J.: DES and differential power analysis (the "duplication" method). In: Koç, Çetin Kaya., Paar, C. (eds.) Cryptographic Hardware and Embedded Systems – CHES'99. Lecture Notes in Computer Science, vol. 1717, pp. 158–172. Springer, Heidelberg, Germany, Worcester, Massachusetts, USA (Aug 12–13, 1999). https://doi.org/10.1007/3-540-48059-5_15 2
28. Goudarzi, D., Joux, A., Rivain, M.: How to securely compute with noisy leakage in quasilinear complexity. In: Peyrin, T., Galbraith, S. (eds.) Advances in Cryptology – ASIACRYPT 2018, Part II. Lecture Notes in Computer Science, vol. 11273, pp. 547–574. Springer, Heidelberg, Germany, Brisbane, Queensland, Australia (Dec 2–6, 2018). https://doi.org/10.1007/978-3-030-03329-3_19 20
29. Goudarzi, D., Prest, T., Rivain, M., Vergnaud, D.: Probing security through input-output separation and revisited quasilinear masking. IACR Transactions on Cryptographic Hardware and Embedded Systems **2021**(3), 599–640 (2021). https://doi.org/10.46586/tches.v2021.i3.599-640, https://tches.iacr.org/index.php/TCHES/article/view/8987 20
30. Goudarzi, D., Rivain, M.: How fast can higher-order masking be in software? In: Coron, J.S., Nielsen, J.B. (eds.) Advances in Cryptology – EUROCRYPT 2017, Part I. Lecture Notes in Computer Science, vol. 10210, pp. 567–597. Springer, Heidelberg, Germany, Paris, France (Apr 30 – May 4, 2017). https://doi.org/10.1007/978-3-319-56620-7_20 9
31. Grosso, V., Standaert, F.X.: Masking proofs are tight and how to exploit it in security evaluations. In: Nielsen, J.B., Rijmen, V. (eds.) Advances in Cryptology – EUROCRYPT 2018, Part II. Lecture Notes in Computer Science, vol. 10821, pp. 385–412. Springer, Heidelberg, Germany, Tel Aviv, Israel (Apr 29 – May 3, 2018). https://doi.org/10.1007/978-3-319-78375-8_13 5, 20
32. Ishai, Y., Sahai, A., Wagner, D.: Private circuits: Securing hardware against probing attacks. In: Boneh, D. (ed.) Advances in Cryptology – CRYPTO 2003. Lecture Notes in Computer Science, vol. 2729, pp. 463–481. Springer, Heidelberg, Germany, Santa Barbara, CA, USA (Aug 17–21, 2003). https://doi.org/10.1007/978-3-540-45146-4_27 2, 3
33. Ito, A., Ueno, R., Homma, N.: On the success rate of side-channel attacks on masked implementations: Information-theoretical bounds and their practical usage. In: Yin, H., Stavrou, A., Cremers, C., Shi, E. (eds.) ACM CCS 2022: 29th Conference on Computer and Communications Security. pp. 1521–1535. ACM Press, Los Angeles, CA, USA (Nov 7–11, 2022). https://doi.org/10.1145/3548606.3560579 4, 5, 7
34. Jog, V.S., Anantharam, V.: The entropy power inequality and mrs. gerber's lemma for groups of order 2^n. IEEE Trans. Inf. Theory **60**(7), 3773–3786 (2014). https://doi.org/10.1109/TIT.2014.2317692, https://doi.org/10.1109/TIT.2014.2317692 14

35. Kocher, P.C.: Timing attacks on implementations of Diffie-Hellman, RSA, DSS, and other systems. In: Koblitz, N. (ed.) Advances in Cryptology – CRYPTO'96. Lecture Notes in Computer Science, vol. 1109, pp. 104–113. Springer, Heidelberg, Germany, Santa Barbara, CA, USA (Aug 18–22, 1996). https://doi.org/10.1007/3-540-68697-5_9 1
36. Kocher, P.C., Jaffe, J., Jun, B.: Differential power analysis. In: Wiener [52], pp. 388–397. https://doi.org/10.1007/3-540-48405-1_25 1
37. Mangard, S.: Hardware countermeasures against DPA – A statistical analysis of their effectiveness. In: Okamoto, T. (ed.) Topics in Cryptology – CT-RSA 2004. Lecture Notes in Computer Science, vol. 2964, pp. 222–235. Springer, Heidelberg, Germany, San Francisco, CA, USA (Feb 23–27, 2004). https://doi.org/10.1007/978-3-540-24660-2_18 7
38. Mangard, S., Oswald, E., Popp, T.: Power analysis attacks - revealing the secrets of smart cards. Springer (2007) 7, 18
39. Mangard, S., Oswald, E., Standaert, F.: One for all - all for one: unifying standard differential power analysis attacks. IET Inf. Secur. **5**(2), 100–110 (2011). https://doi.org/10.1049/iet-ifs.2010.0096, https://doi.org/10.1049/iet-ifs.2010.0096 7
40. Masure, L., Rioul, O., Standaert, F.X.: A nearly tight proof of duc et al.'s conjectured security bound for masked implementations. Cryptology ePrint Archive, Paper 2022/600 (2022), https://eprint.iacr.org/2022/600, https://eprint.iacr.org/2022/600 4, 5, 7
41. Micali, S., Reyzin, L.: Physically observable cryptography (extended abstract). In: Naor, M. (ed.) TCC 2004: 1st Theory of Cryptography Conference. Lecture Notes in Computer Science, vol. 2951, pp. 278–296. Springer, Heidelberg, Germany, Cambridge, MA, USA (Feb 19–21, 2004). https://doi.org/10.1007/978-3-540-24638-1_16 22
42. Moriai, S. (ed.): Fast Software Encryption – FSE 2013, Lecture Notes in Computer Science, vol. 8424. Springer, Heidelberg, Germany, Singapore (Mar 11–13, 2014) 32, 35
43. Nguyen, P.Q., Oswald, E. (eds.): Advances in Cryptology – EUROCRYPT 2014, Lecture Notes in Computer Science, vol. 8441. Springer, Heidelberg, Germany, Copenhagen, Denmark (May 11–15, 2014) 32
44. Prest, T., Goudarzi, D., Martinelli, A., Passelègue, A.: Unifying leakage models on a Rényi day. In: Boldyreva, A., Micciancio, D. (eds.) Advances in Cryptology – CRYPTO 2019, Part I. Lecture Notes in Computer Science, vol. 11692, pp. 683–712. Springer, Heidelberg, Germany, Santa Barbara, CA, USA (Aug 18–22, 2019). https://doi.org/10.1007/978-3-030-26948-7_24 3, 4, 14, 16, 18, 20, 24
45. Prouff, E., Rivain, M.: Masking against side-channel attacks: A formal security proof. In: Johansson, T., Nguyen, P.Q. (eds.) Advances in Cryptology – EUROCRYPT 2013. Lecture Notes in Computer Science, vol. 7881, pp. 142–159. Springer, Heidelberg, Germany, Athens, Greece (May 26–30, 2013). https://doi.org/10.1007/978-3-642-38348-9_9 2, 3, 4, 6, 8, 10, 16, 18, 20, 24
46. Rivain, M.: On the provable security of cryptographic implementations : Habilitation thesis. Personal website (2022), https://www.matthieurivain.com/hdr.html, https://www.matthieurivain.com/hdr.html 6
47. Rivain, M., Prouff, E.: Provably secure higher-order masking of AES. In: Mangard, S., Standaert, F.X. (eds.) Cryptographic Hardware and Embedded Systems – CHES 2010. Lecture Notes in Computer Science, vol. 6225, pp. 413–427. Springer, Heidelberg, Germany, Santa Barbara, CA, USA (Aug 17–20, 2010). https://doi.org/10.1007/978-3-642-15031-9_28 3, 9, 11, 12, 23

48. Cardoso dos Santos, L., Gérard, F., Großschädl, J., Spignoli, L.: Rivain-Prouff on steroids: Faster and stronger masking of the AES. In: Buhan, I., Schneider, T. (eds.) Smart Card Research and Advanced Applications. pp. 123–145. Springer International Publishing, Cham (2023) 9, 12, 23
49. Shulman, N., Feder, M.: The uniform distribution as a universal prior. IEEE Trans. Inf. Theory **50**(6), 1356–1362 (2004). https://doi.org/10.1109/TIT.2004.828152, https://doi.org/10.1109/TIT.2004.828152 10, 11, 27
50. Standaert, F.X., Malkin, T., Yung, M.: A unified framework for the analysis of side-channel key recovery attacks. In: Joux, A. (ed.) Advances in Cryptology – EUROCRYPT 2009. Lecture Notes in Computer Science, vol. 5479, pp. 443–461. Springer, Heidelberg, Germany, Cologne, Germany (Apr 26–30, 2009). https://doi.org/10.1007/978-3-642-01001-9_26 3, 20
51. Tunstall, M., Whitnall, C., Oswald, E.: Masking tables - an underestimated security risk. In: Moriai [42], pp. 425–444. https://doi.org/10.1007/978-3-662-43933-3_22 24
52. Wiener, M.J. (ed.): Advances in Cryptology – CRYPTO'99, Lecture Notes in Computer Science, vol. 1666. Springer, Heidelberg, Germany, Santa Barbara, CA, USA (Aug 15–19, 1999) 32, 34
53. Wyner, A.D., Ziv, J.: A theorem on the entropy of certain binary sequences and applications-i. IEEE Trans. Inf. Theory **19**(6), 769–772 (1973). https://doi.org/10.1109/TIT.1973.1055107, https://doi.org/10.1109/TIT.1973.1055107 14