Improved Alpha-Information Bounds for Higher-Order Masked Cryptographic Implementations

Yi Liu*, Julien Béguinot*, Wei Cheng†*, Sylvain Guilley†*, Loïc Masure‡, Olivier Rioul*, and François-Xavier Standaert‡

*LTCI, Télécom Paris, Institut Polytechnique de Paris, 91 120, Palaiseau, France
†Secure-IC S.A.S., 75 014, Paris, France
‡ICTEAM Institute, Université catholique de Louvain, Louvain-la-Neuve, Belgium

***Abstract*—Embedded cryptographic devices are usually protected against side-channel attacks by masking strategies. In this paper, the security of protected cryptographic implementations is evaluated for *any* masking order, using alpha-information measures. Universal upper bounds on the probability of success of *any* type of side-channel attack are derived. These also provide lower bounds on the minimum number of queries required to achieve a given success rate. An important issue, solved in this paper, is to remove the loss factor due to the masking field size.**

I. Introduction

When a cryptographic device is operating, any kind of unintended leakage (time, power, electromagnetic, etc.) can be exploited by an attacker. By querying the device multiple times, measuring the corresponding leakages, and correlating them with internal sensitive values, the attacker is able to guess the secret key with a given success probability.

Therefore, evaluating the security of cryptographic devices against side-channel attacks has become a major concern. Information-theoretic metrics turn out to be effective and has been used in many studies: Using classical metrics such as mutual information and Fano inequality, de Chérisey et al. [6] established several universal bounds on the probability of success and minimum number of queries required to achieve success. This approach has been extended to conditional α-informational quantities in [15]. Both [6] and [15], however, were restricted to unprotected cryptographic devices.

Masking is one of the most well-established protection with provable security. Some research [4], [7], [13], [16] was conducted to evaluate the security of masked implementations against side-channel attacks. To review the state-of-the-art, we follow the framework and notations from [4], [6], [12].

A. Background and Notations

Let K be the secret key and T be a public variable (usually, plain or cypher text) known to the attacker. Both K and T are n-bit variables, uniformly distributed, and independent of each other. The *field size* is $M = 2^n$. The cryptographic algorithm operates on K and T to compute an n-bit sensitive variable $V = f(K, T)$. In a masking scheme of order d, the sensitive variable is randomly split into $d + 1$ *shares* and cryptographic operations are performed on each share separately. Thus $V = X_0 \oplus X_1 \oplus \cdots \oplus X_d$, where each share X_i is a n-bit variable and $\oplus$ is the additive operation in the underlying field (or Abelian group). A typical example is "Boolean masking," for which $\oplus$ is the bitwise XOR operation. During computation, side-channel information $\boldsymbol{X} = (X_0, X_1, \ldots, X_d)$ is leaking and can be measured as a noisy "trace" by the attacker, denoted by $\boldsymbol{Y} = (Y_0, Y_1, \ldots, Y_d)$. We assume that $\boldsymbol{Y}$ is the output of a memoryless side-channel with input $\boldsymbol{X}$. Since masking shares are drawn uniformly and independently, both and $\boldsymbol{X}$ and $\boldsymbol{Y}$ are i.i.d. sequences.

The attacker measures m traces $\boldsymbol{Y}^m = (\boldsymbol{Y}_1, \boldsymbol{Y}_2, \ldots, \boldsymbol{Y}_m)$ corresponding to the independent text sequence $T^m = (T_1, T_2, \ldots, T_m)$—assumed independent of the secret K—and exploits her knowledge of $\boldsymbol{Y}^m$ and T^m to estimate the secret key $\hat{K}$. Again, since the side channel is memoryless, $\boldsymbol{X}^m$ and $\boldsymbol{Y}^m$ are i.i.d. sequences. Let $\mathbb{P}_s = \mathbb{P}(K = \hat{K})$ be the probability of success of the attack upon observing T^m and $\boldsymbol{Y}^m$. In theory, maximum success is obtained by the MAP (maximum a posteriori probability) rule with success probability denoted by $\mathbb{P}_s = \mathbb{P}_s(K|\boldsymbol{Y}^m, T^m)$. The whole process is illustrated in Fig. 1.

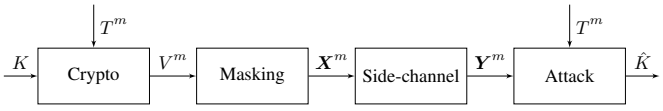


Fig. 1. Side-channel analysis as a (unintended) "communication" channel.

B. State-of-the-art

Duc et al. [7] derived a lower bound on the minimum number m of queries required to achieve a given probability of success $\mathbb{P}_s$:

$$m \geq \frac{\log(1 - \mathbb{P}_s)}{\log\left(1 - (\frac{M}{\sqrt{2}})^{d+1} \prod_{i=0}^{d} I(X_i; Y_i)\right)} \tag{1}$$

where $d + 1$ is the number of shares, M is the field size, and $I(X_i; Y_i)$ is the mutual information between each share and its corresponding leakage. They also showed that this bound was quite loose in practice and conjectured that when the leakage of shares is sufficiently noisy (and independent among shares), the lower bound on m should take the approximate form

$$m \gtrsim \frac{\beta(\mathbb{P}_s)}{\prod_{i=0}^{d} I(X_i; Y_i)} \tag{2}$$

where β is a "small constant depending on $\mathbb{P}_s$" [8, p. 1279].

The bound (1) was improved recently in [16]:

$$m \geq \frac{d(\mathbb{P}_s \| \frac{1}{M})}{\log\Big(1 + M \cdot \prod_{i=0}^{d} (2\log 2) I(X_i; Y_i)\Big)}. \tag{3}$$

A very similar bound was derived independently in [13]. Although this greatly improves (1) for small M, when the field size M is large, the M factor in the denominator loosens the bound by an substantial amount. Therefore, an important issue is to find out whether this factor M can be removed.

C. Outline

In this paper, we have two main contributions. First, we generalize the "linear bound" $d(\mathbb{P}_s \| \frac{1}{M}) \leq m I(V; \boldsymbol{Y})$ in [6] to α-informational quantities where the usual linear bound is recovered by letting $\alpha \to 1$. Second, we derive the following novel bound which removes the loss caused by the field size:

$$m \geq \frac{d_2(\mathbb{P}_s \| \frac{1}{M})}{\log\Big(1 + \prod_{i=0}^{d} (e^{I_2^R(X_i;Y_i)} - 1)\Big)}. \tag{4}$$

Here, instead of using usual Kullback–Leibler divergence and mutual information, we consider the α-divergence and the Rényi α-mutual information for $\alpha = 2$: d_2 and I_2^R. This particular value of α allows one to link $\mathbb{P}_s$ to α-information *via* a quadratic version of the total variation distance.

Our bounds are particularly useful under the usual "high noise assumption," that is, when the side channel of Fig. 1 has low capacity. Then, values of $I_2^R(X_i; Y_i)$ will be small, and the lower bound on m is approximately equal to:

$$m \gtrsim \frac{d_2(\mathbb{P}_s \| \frac{1}{M})}{\prod_{i=0}^{d} I_2^R(X_i; Y_i)}. \tag{5}$$

This is very similar to the conjectured bound (2), except for the use of I_2^R instead of I. Additionally, we show that when M is large, the numerator does not lose tightness compared that of (3).

In the remainder of the paper, we first recall some definitions and properties of α-informational quantities in Section II, and then derive the α-extension of the main inequality ("linear bound") in Section III. The main result is then derived in Section IV and illustrated by numerical simulations. Section V gives some perspectives.

II. α-Information Measures

A. α-Entropy and α-Divergence

Assume that either $0 < \alpha < 1$ or $1 < \alpha < +\infty$ (the limiting values $0, 1, +\infty$ will be obtained by taking limits). We consider probability distributions P, Q with a dominating measure, with respect to which they follow densities denoted by the corresponding lower-case letters p, q.

We follow the notations of [15] in the following

Definition 1 (Rényi α-Entropy and α-Divergence)**.**

$$H_\alpha(P) = \tfrac{\alpha}{1-\alpha} \log \|p\|_\alpha \tag{6}$$

$$D_\alpha(P\|Q) = \tfrac{1}{\alpha-1} \log \langle p \| q \rangle_\alpha^\alpha \tag{7}$$

with the following special notation:

$$\|p\|_\alpha = \Big(\int |p|^\alpha d\mu \Big)^{1/\alpha} \tag{8}$$

$$\langle p \| q \rangle_\alpha = \Big(\int p^\alpha q^{1-\alpha} d\mu \Big)^{1/\alpha} \tag{9}$$

The usual entropy and Kullback-Leibler divergence are recovered by letting $\alpha \to 1$.

B. Conditional α-Entropy

Many different definitions of conditional α-entropy $H_\alpha(X|Y)$ were proposed in the literature (see, e.g., [9]). Any reasonable definition should at least yield the classical definition of conditional entropy as $\alpha \to 1$, and satisfy the property that *conditioning reduces entropy* (CRE): $H_\alpha(X|Y) \leq H_\alpha(X)$, where equality holds if and only if X and Y are independent. At least four definitions are often used:

1) $\tilde{H}_\alpha^{(o)}(X|Y) = H_\alpha(X, Y) - H_\alpha(Y)$
2) $\tilde{H}_\alpha^{(i)}(X|Y) = \mathbb{E}_Y H_\alpha(X|Y = y)$
3) $\tilde{H}_\alpha^{(ii)}(X|Y) = \frac{1}{1-\alpha} \log \mathbb{E}_Y \|P_{X|Y}\|_\alpha^\alpha$
4) $\tilde{H}_\alpha(X|Y) = \frac{\alpha}{1-\alpha} \log \mathbb{E}_Y \|P_{X|Y}\|_\alpha$

The first two definitions appear in [14, § 2.2] (see also [10, equation (2.10)]) and in [3, equation (2.15)]. However, both violate the CRE property [9]. The last two definitions were proposed by Hayashi [11] and Arimoto [1] respectively. Both satisfy the CRE property. In the sequel, we use Arimoto's definition which we simply denote as $H_\alpha(X|Y)$.

C. α-Information

Again, many different definitions of α-informtion $I_\alpha(X; Y)$ were proposed in the literature. Any reasonable definition should at least yield the classical definition of mutual information as $\alpha \to 1$, and possibly also satisfy the following useful properties:

- *independence*: $I_\alpha(X; Y) \geq 0$ with equality if and only if X and Y are independent;
- *data post-processing inequality (post-DPI)*: if $X - Y - Z$ forms a Markov chain, then post-processing cannot increase the information, i.e., $I_\alpha(X; Z) \leq I_\alpha(X; Y)$;
- *data pre-processing inequality (pre-DPI)*: if $X - Y - Z$ forms a Markov chain, then pre-processing cannot increase the information, i.e., $I_\alpha(X; Z) \leq I_\alpha(Y; Z)$.
- *monotonicity*: $I_\alpha(X; Y)$ is nondecreasing as α increases;
- *closed-form expression* amenable to efficient numerical estimation.

At least four definitions are used in the literature:

1) $I_\alpha^A(X; Y) = H_\alpha(X) - H_\alpha(X|Y)$
2) $I_\alpha^C(X; Y) = \min_{Q_Y} \mathbb{E}_X(D_\alpha(P_{Y|X} \| Q_Y))$
3) $I_\alpha^R(X; Y) = D_\alpha(P_{XY} \| P_X \times P_Y)$
   $= \frac{1}{\alpha-1} \log \mathbb{E}_Y \langle p_{X|Y} \| p_X \rangle_\alpha^\alpha.$
4) $I_\alpha(X; Y) = \min_{Q_Y} D_\alpha(P_{XY} \| P_X \times Q_Y)$
   $= \frac{\alpha}{\alpha-1} \log \mathbb{E}_Y \langle p_{X|Y} \| p_X \rangle_\alpha.$

which somehow parallel the corresponding ones for conditional entropy. The first definition was proposed by Arimoto [1]. It

is easily seen to satisfy both the *independence* and *post-DPI* property because of the CRE property of Arimoto's conditional entropy. However, it does not satisfy *monotonicity* because $I_\alpha^A(X;X) = H_\alpha(X)$ can be decreasing in α. The second definition is from Csiszár [5]. It does not seem to admit a closed-form expression, and the minimization is hard to solve analytically even in simple examples [21]. However, one can prove *monotonicity* and the *independence* property, based on the properties of the α-divergence.

The third definition requires no minimization and appears in [20, equation (50)]. We call it *Rényi's α-mutual information* because it is a natural definition from Rényi's divergence, just as in the classical case $\alpha = 1$. Also, it is *mutual* in the sense that $I_\alpha^R(X;Y) = I_\alpha^R(Y;X)$. From the nonnegativity of α-divergence: $D_\alpha(P\|Q) \geq 0$ with equality if and only if $P = Q$, it is easily seen that $I_\alpha^R(X;Y)$ satisfies the *independence* property. From the monotonicity property of α-divergence, it also satisfies monotonicity. One can also check *post-DPI* and *pre-DPI* properties, by same reasoning line as in the proof of [15, Property 12], replacing $Q_{Y|T}, Q_{Z|T}$ by $P_{Y|T}, P_{Z|T}$, respectively.

Finally, the fourth definition is due to Sibson [19] (see also [21]). In contrast to Rényi α-mutual information, symmetry does not hold in general: $I_\alpha(X;Y) \neq I_\alpha(Y;X)$. However, it is known to satisfy the independence property, monotonicity, and the pre and post-DPI [17] (see also [18]). See Table I for a summary of all properties. In the sequel, we often use Sibson's definition, which we simply denote as $I_\alpha(X;Y)$

TABLE I
SUMMARY OF PROPERTIES FOR VARIOUS DEFINITIONS OF α-INFORMATION.

Def.	Independence	Post-DPI	Pre-DPI	Monotonicity	Closed-form
I_α^A	yes	yes	—	no	yes
I_α^C	yes	—	—	yes	no
I_α^R	yes	yes	yes	yes	yes
I_α	yes	yes	yes	yes	yes

Remark 1. *Since* $\min_{Q_Y} D_\alpha(P_{XY}\|P_X \times Q_Y) \leq D_\alpha(P_{XY}\|P_X \times P_Y)$*, Sibson's α-information can not exceed Rényi mutual information:*

$$I_\alpha(X;Y) \leq I_\alpha^R(X;Y), \tag{10}$$

III. Lower Bound on Sibson's α-Information

The first result of this paper is based on the following generalized Fano inequality [18]. Assume K is discrete and estimated from Y using the MAP rule, with (maximal) probability of success $\mathbb{P}_s = \mathbb{P}_s(K|Y) = \mathbb{E}\sup_k p_{K|Y}(k|Y)$. Also let $\mathbb{P}_s(K) = \sup p_K$ be the probability of success when guessing K without even knowing Y.

Lemma 1 (Generalized Fano Inequality [18, Thm. 1])**.**

$$d_\alpha\big(\mathbb{P}_s(K|Y)\|\mathbb{P}_s(K)\big) \leq I_\alpha(K;Y) \tag{11}$$

where $d_\alpha(p\|q)$ is the binary α-divergence:

$$d_\alpha(p\|q) = \tfrac{1}{\alpha-1}\log\big(p^\alpha q^{1-\alpha} + (1-p)^\alpha(1-q)^{1-\alpha}\big). \tag{12}$$

A. Bound Probability of Success by Sibson's α-Information

In Fig. 1, the sensitive variable V^m is a function of K and T^m; $\hat{K}$ is a function of $(\boldsymbol{Y}^m, T^m)$. It is easily seen from the figure that the following Markov chains hold:

$$K \longleftrightarrow (\boldsymbol{Y}^m, T^m) \longleftrightarrow \hat{K} \tag{13}$$

$$(K, T^m) \longleftrightarrow V^m \longleftrightarrow \boldsymbol{Y}^m \tag{14}$$

The probability of success of the side-channel attack is $\mathbb{P}_s = \mathbb{P}_s(K|\boldsymbol{Y}^m, T^m)$. Using Lemma 1, one has $d_\alpha\big(\mathbb{P}_s\|\frac{1}{M}\big) \leq I_\alpha(K;\boldsymbol{Y}^m, T^m)$. Now, the following lemma is proved in Appendix A:

Lemma 2. $I_\alpha(K;\boldsymbol{Y}^m, T^m) \leq I_\alpha(K, T^m;\boldsymbol{Y}^m)$. (15)

It follows that the generalized Fano inequality implies

$$d_\alpha\big(\mathbb{P}_s\|\frac{1}{M}\big) \leq I_\alpha(K, T^m;\boldsymbol{Y}^m). \tag{16}$$

Because $(K, T^m) \leftrightarrow V^m \leftrightarrow \boldsymbol{Y}^m$ forms a Markov chain, using the DPI of Sibson's α-information we have

$$I_\alpha(K, T^m;\boldsymbol{Y}^m) \leq I_\alpha(V^m;\boldsymbol{Y}^m). \tag{17}$$

Also, when T^m is not observed, each component of V^m is i.i.d., and since the side-channel is memoryless, $(V^m;\boldsymbol{Y}^m)$ is an i.i.d. sequence. It easily follows from the definition that

$$I_\alpha(V^m;\boldsymbol{Y}^m) = mI_\alpha(V;\boldsymbol{Y}). \tag{18}$$

From (16), (17), and (18), we arrive at the main result of this section:

Theorem 1. $d_\alpha\big(\mathbb{P}_s\|\frac{1}{M}\big) \leq mI_\alpha(V;\boldsymbol{Y})$. (19)

Note that since $d_\alpha\big(p\|q\big)$ is increasing in p when $p \geq q$, Theorem 1 gives an upper bound on $\mathbb{P}_s$.

B. Comparison with the Classical Bound

A natural question is to compare (19) with the classical bound for $\alpha = 1$, especially in terms of how it depends on M. Since d_α and I_α are non-decreasing in α, a precise answer is not obvious. One can argue as follows. Assume $\mathbb{P}_s$ is fixed in $(0,1)$. For $\alpha = 1$, one has at first order

$$d\big(\mathbb{P}_s\|\frac{1}{M}\big) = \log M - (1-\mathbb{P}_s)\log(M-1) - h(\mathbb{P}_s) \approx \mathbb{P}_s \log M \tag{20}$$

where $h(\mathbb{P}_s)$ is the binary entropy function. For $\alpha < 1$, $d_\alpha\big(\mathbb{P}_s\|\frac{1}{M}\big) \leq d\big(\mathbb{P}_s\|\frac{1}{M}\big)$ does not grow faster than $O(\log M)$. For $\alpha > 1$, one has at first order

$$d_\alpha\big(\mathbb{P}_s\|\frac{1}{M}\big) = \log M + \frac{1}{\alpha-1}\log\Big(\mathbb{P}_s^\alpha + \frac{(1-\mathbb{P}_s)^\alpha}{(M-1)^{\alpha-1}}\Big) \approx \log M \tag{21}$$

Thus the $O(\log M)$ term applies for any α, and the lower bound in (19) will not become less tight as the classical bound as the field size M increases.

IV. Upper Bound on Rényi Mutual Information

A. Euclidean Distance to the Uniform

In the field of cryptography, the *total variation distance* $\|P - U\|_1$ of a given M-ary distribution P to the uniform distribution $U \sim \mathcal{U}(M)$ is a common criterion to evaluate randomness. For $\alpha \neq 1$ we have the following

Definition 2 (α-Distance)**.** *Let X be an M-ary random variable. The "α-distance" between P_X and a uniform distribution $U \sim \mathcal{U}(M)$ is defined as*

$$\|P_X - U\|_\alpha = \Big(\sum_x \Big|p_X(x) - \frac{1}{M}\Big|^\alpha\Big)^{\frac{1}{\alpha}}. \tag{22}$$

In this section we focus on the Euclidean distance ($\alpha = 2$) because of the following

Lemma 3. *With the same notations, one has*

$$D_2(P_X\|U) = \log(1 + M \cdot \|P_X - U\|_2^2) \tag{23}$$

Proof. One has $\|P_X - U\|_2^2 = \sum_x (p_X(x) - \frac{1}{M})^2 = \sum_x p_X^2(x) - \frac{1}{M}$. Since $D_2(P_X\|U) = \log(M \cdot \sum_x p_X^2(x))$, the result follows. □

The following important Lemma is known as the XOR Lemma in the case of Boolean Masking [16]. The general proof is given in Appendix B:

Lemma 4 (Group Lemma)**.** *Let X_1, X_2 be independent random variables over a finite Abelian group $\mathcal{X}$ of size M, and $U \sim \mathcal{U}(\mathcal{X})$. Let $V = X_1 \oplus X_2$, where $\oplus$ denotes the group operator in $\mathcal{X}$. One has*

$$\|P_V - U\|_2^2 \leq M \cdot \|P_{X_1} - U\|_2^2 \cdot \|P_{X_2} - U\|_2^2. \tag{24}$$

By finite induction, if V is split into $d + 1$ independent shares: $V = X_0 \oplus X_1 \oplus \cdots \oplus X_d$, one has

$$\|P_V - U\|_2^2 \leq M^d \|P_{X_0} - U\|_2^2 \|P_{X_1} - U\|_2^2 \cdots \|P_{X_d} - U\|_2^2. \tag{25}$$

Using Lemma 3 this can easily be written as

$$e^{D_2(P_V\|U)} \leq 1 + \prod_{i=0}^{d} (e^{D_2(P_{X_i}\|U)} - 1). \tag{26}$$

B. Upper Bound of Rényi 2-Information for Each Share

Since Sibson's α-information does not exceed Rényi mutual information (inequality (10)), Theorem 1 implies

$$d_\alpha\big(\mathbb{P}_s\|\frac{1}{M}\big) \leq m I_\alpha^R(V;\boldsymbol{Y}). \tag{27}$$

We now upper bound $I_\alpha^R(V;\boldsymbol{Y})$ by noting that, by definition since V is uniformly distributed,

$$e^{I_2^R(V;\boldsymbol{Y})} = \mathbb{E}_{\boldsymbol{Y}} e^{D_2(P_{V|\boldsymbol{Y}}\|U)}. \tag{28}$$

Since $\{X_i, Y_i\}_{i=0,\ldots,d}$ are mutually independent, (26) applies for $V|\boldsymbol{Y}$ and we have

$$e^{I_2^R(V;\boldsymbol{Y})} \leq \mathbb{E}_{\boldsymbol{Y}}\Big(1 + \prod_{i=0}^{d} (e^{D_2(P_{X_i|Y_i}\|U)} - 1)\Big) \tag{29}$$

$$= 1 + \prod_{i=0}^{d} (\mathbb{E}_{Y_i} e^{D_2(P_{X_i|Y_i}\|U)} - 1) \tag{30}$$

$$= 1 + \prod_{i=0}^{d} (e^{I_2^R(X_i;Y_i)} - 1) \tag{31}$$

Putting all inequalities together yields the main result of this paper:

Theorem 2 (**Main Result**)**.** *The number of traces m can be lower bounded by*

$$m \geq \frac{d_2(\mathbb{P}_s\|\frac{1}{M})}{\log\Big(1 + \prod_{i=0}^{d} (e^{I_2^R(X_i;Y_i)} - 1)\Big)}. \tag{32}$$

Note that from Subsection III-B with $\alpha = 2$, the numerator does not lose tightness compared the case $\alpha = 1$ (compare (3)).

C. Numerical Results

In this subsection, we validate our results by simulation. The side-channel settings of § I-A are as follows:

- the field of variables is the AES (Advanced Encryption Standard) field with $n = 8$, thus $M = 256$;
- side-channel information is generated by taking the Hamming weight leakage model and additive white Gaussian noise (one of the most commonly adopted models);
- the Boolean masking is considered with orders $d \in 0, 1, 2$.

The Shannon and Rényi mutual informations (MI) are evaluated by Monte-Carlo simulation In particular, we compare Rényi MI in (31) with the following

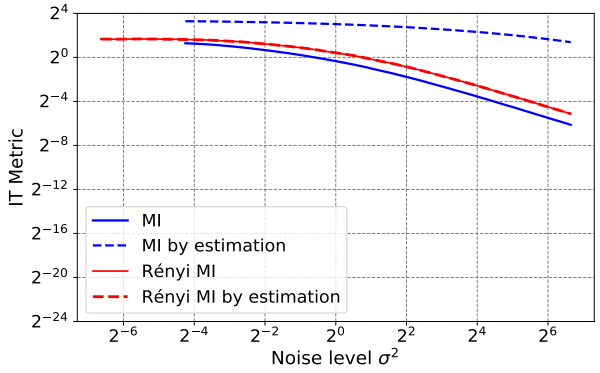
$$I(V;\boldsymbol{Y}) \leq \log\big(1 + M \cdot \prod_{i=0}^{d} (2\log 2 \cdot I(X_i;Y_i))\big) \tag{33}$$

used in (3). Fig. 2 compares MI and Rényi MI for $d = 0, 1, 2$. Our result based on Rényi MI significantly narrows the gap between the direct evaluation and the estimation. This leads to more accurate prediction of number of queries m to achieve certain success rate $\mathbb{P}_s$.

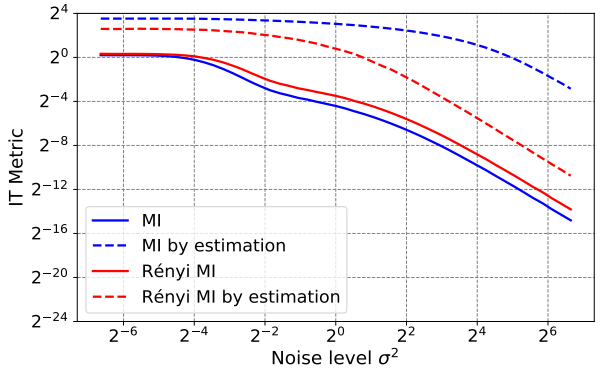
Fig. 3 confirms this on the performance bounds on the success rate as a function of m, for $d = 1$ and 2. Our new bounds are significantly more accurate than the state-of-the-art: For $\mathbb{P}_s = 80\%$ and $d = 1$, the ML attack gives about $m \geq 60$, our new bound gives $m \geq 25$, while (3) gives only $m \geq 1$. Much improvement can also be observed for $d = 2$.

V. Perspective

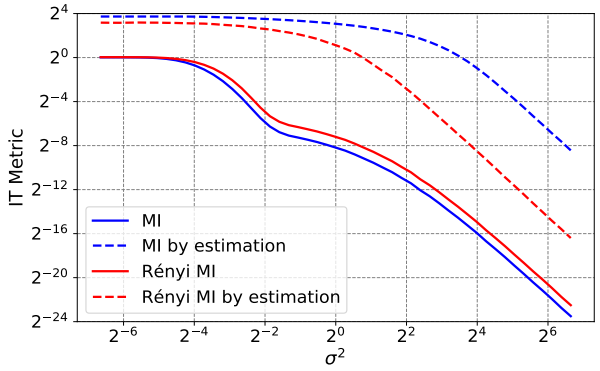
Similar improved bounds (removing the field size loss) can also be obtained in the cases of Boolean masking and arithmetic masking modulo a power of two, using "Mrs. Gerber's lemma", see [2]. Extending this work to α-information is left for future work.



(a) $d = 0$ without masking.



(b) $d = 1$.



(c) $d = 2$.

Fig. 2. Comparison of various bounds for $M = 256$ under Hamming weight leakages with Gaussian noise. The plain curves show the direct evaluation of $I(V;\boldsymbol{Y})$ and $I_2^R(V;\boldsymbol{Y})$; dash curves show the corresponding estimations in (33) and (31), respectively.

APPENDIX

A. Proof of Lemma 2

$$
\begin{aligned}
I_\alpha(K;\boldsymbol{Y}^m,T^m) &= \frac{\alpha}{\alpha-1}\log \mathbb{E}_{\boldsymbol{Y}^m,T^m}\langle p_{K|\boldsymbol{Y}^m,T^m}\|p_K\rangle_\alpha \\
&= \frac{\alpha}{\alpha-1}\log \mathbb{E}_{T^m}\int_{\boldsymbol{Y}^m} p_{\boldsymbol{Y}^m|T^m}\Big(\sum_k p^\alpha_{K|\boldsymbol{Y}^m,T^m}p_K^{1-\alpha}\Big)^{\frac{1}{\alpha}} \\
&= \frac{\alpha}{\alpha-1}\log \mathbb{E}_{T^m}\int_{\boldsymbol{Y}^m}\Big(\sum_k p^\alpha_{K,\boldsymbol{Y}^m|T^m}p_K^{1-\alpha}\Big)^{\frac{1}{\alpha}}
\end{aligned}
$$

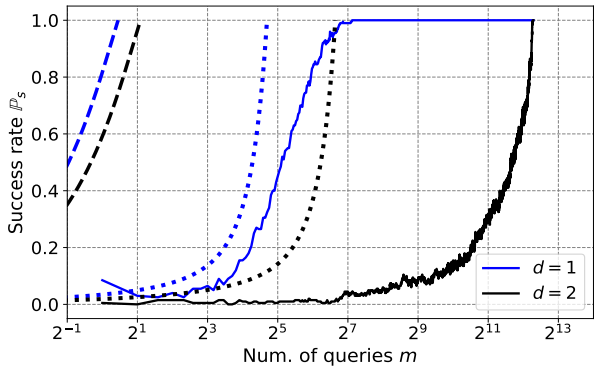


Fig. 3. $\mathbb{P}_s$ vs m in attacks and the corresponding bounds for noise variance $\sigma^2 = 8$. The plain curves show the results of direct maximum likelihood (ML) attacks [12]; the *dotted* curves show the predictions by Theorem 2; the *dashed* curves are for the state-of-the-art bound (3).

$$
\begin{aligned}
&\overset{(\star)}{=} \frac{\alpha}{\alpha-1}\log \mathbb{E}_{T^m}\int_{\boldsymbol{Y}^m}\Big(\sum_k p^\alpha_{\boldsymbol{Y}^m|K,T^m}p_{K|T^m}\Big)^{\frac{1}{\alpha}} \\
&\overset{(\star\star)}{\leq} \frac{\alpha}{\alpha-1}\log \int_{\boldsymbol{Y}^m}\Big(\mathbb{E}_{T^m}\sum_k p^\alpha_{\boldsymbol{Y}^m|K,T^m}p_{K|T^m}\Big)^{\frac{1}{\alpha}} \\
&= \frac{\alpha}{\alpha-1}\log \int_{\boldsymbol{Y}^m}\Big(\sum_{k,t^m} p^\alpha_{\boldsymbol{Y}^m|K,T^m}p_{K,T^m}\Big)^{\frac{1}{\alpha}} \\
&= \frac{\alpha}{\alpha-1}\log \int_{\boldsymbol{Y}^m} p_{\boldsymbol{Y}^m}\Big(\sum_{k,t^m} p^\alpha_{K,T^m|\boldsymbol{Y}^m}p^{1-\alpha}_{K,T^m}\Big)^{\frac{1}{\alpha}} \\
&= \frac{\alpha}{\alpha-1}\log \mathbb{E}_{\boldsymbol{Y}^m}\langle p_{K,T^m|\boldsymbol{Y}^m}\|p_{K,T^m}\rangle_\alpha \\
&= I_\alpha(K,T^m;\boldsymbol{Y}^m)
\end{aligned}
$$

where $(\star)$ holds since $p_K = p_{K|T^m}$ (K and T^m are independent) and $p^\alpha_{K,\boldsymbol{Y}^m|T^m}p^{-\alpha}_{K|T^m} = p^\alpha_{\boldsymbol{Y}^m|K,T^m}$; $(\star\star)$ is Jensen's inequality: when $\alpha > 1$, $x^{\frac{1}{\alpha}}$ is concave and $\frac{\alpha}{\alpha-1}$ is positive; when $0 < \alpha < 1$, $x^{\frac{1}{\alpha}}$ is convex and $\frac{\alpha}{\alpha-1}$ is negative. In both cases the inequality holds in the same direction. □

B. Proof of Lemma 4

Let $\ominus$ denote the inverse operation of $\oplus$ in the Abelian group. By independence of X_1, X_2, one has

$$
\begin{aligned}
\|P_V - U\|_2^2 &= \sum_v \Big|p_V(v) - \frac{1}{M}\Big|^2 \\
&= \sum_v \Big|\sum_{x_1} p_{X_1}(x_1)p_{X_2}(v\ominus x_1) - \frac{1}{M}\Big|^2 \\
&= \sum_v \Big|\sum_{x_1}(p_{X_1}(x_1) - \frac{1}{M})(p_{X_2}(v\ominus x_1) - \frac{1}{M})\Big|^2 \\
&\overset{(\star)}{\leq} \sum_v \Big(\sum_{x_1}(p_{X_1}(x_1) - \frac{1}{M})^2\Big)\Big(\sum_{x_1}(p_{X_2}(v\ominus x_1) - \frac{1}{M})^2\Big) \\
&= \Big(\sum_{x_1}(p_{X_1}(x_1) - \frac{1}{M})^2\Big)\sum_v\Big(\sum_{x_1}(p_{X_2}(v\ominus x_1) - \frac{1}{M})^2\Big) \\
&= \|P_{X_1} - U\|_2^2 \cdot M \cdot \|P_{X_2} - U\|_2^2
\end{aligned}
$$

where $(\star)$ is the Cauchy-Schwarz inequality. □

REFERENCES

[1] S. Arimoto, "Information Measures and Capacity of Order α for Discrete Memoryless Channels," in *Topics in Information Theory, Proc. 2nd Colloq. Math. Societatis János Bolyai*, A. Joux, Ed., vol. 16, 1975, pp. 41–52.

[2] J. Beguinot, W. Cheng, S. Guilley, Y. Liu, L. Masure, O. Rioul, and F. Standaert, "Removing the Field Size Loss from Duc et al.'s Conjectured Bound for Masked Encodings," *IACR Cryptol. ePrint Arch.*, pp. 1–18, 2022. [Online]. Available: https://eprint.iacr.org/2022/1738

[3] C. Cachin, "Entropy Measures and Unconditional Security in Cryptography," Ph.D. dissertation, ETH Zurich, 1997.

[4] W. Cheng, Y. Liu, S. Guilley, and O. Rioul, "Attacking Masked Cryptographic Implementations: Information-theoretic Bounds," in *2022 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2022, pp. 654–659.

[5] I. Csiszar, "Generalized Cutoff Rates and Renyi's Information Measures," *IEEE Transactions on Information Theory*, vol. 41, no. 1, pp. 26–34, 1995.

[6] E. de Chérisey, S. Guilley, O. Rioul, and P. Piantanida, "Best Information is Most Successful Mutual Information and Success Rate in Side-Channel Analysis," {*IACR*} *Trans. Cryptogr. Hardw. Embed. Syst.*, vol. 2019, pp. 49–79, 2019. [Online]. Available: https://tches.iacr.org/index.php/TCHES/article/view/7385/6557

[7] A. Duc, S. Faust, and F. Standaert, "Making Masking Security Proofs Concrete - Or How to Evaluate the Security of Any Leaking Device," in *Advances in Cryptology - EUROCRYPT 2015 - 34th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Sofia, Bulgaria, April 26-30, 2015, Proceedings, Part I*, ser. Lecture Notes in Computer Science, E. Oswald and M. Fischlin, Eds., vol. 9056. Springer, 2015, pp. 401–429. [Online]. Available: https://doi.org/10.1007/978-3-662-46800-5_16

[8] ——, "Making Masking Security Proofs Concrete (Or How to Evaluate the Security of Any Leaking Device), Extended Version," *J. Cryptol.*, vol. 32, no. 4, pp. 1263–1297, 2019. [Online]. Available: https://doi.org/10.1007/s00145-018-9277-0

[9] S. Fehr and S. Berens, "On the Conditional Rényi Entropy," *IEEE Transactions on Information Theory*, vol. 60, pp. 6801–6810, 2014.

[10] L. Golshani, E. Pasha, and G. Yari, "Some Properties of Rényi Entropy and Rényi Entropy Rate," *Information Sciences*, vol. 179, no. 14, pp. 2426–2433, 2009, including Special Section – Linguistic Decision Making. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0020025509001145

[11] M. Hayashi, "Exponential Decreasing Rate of Leaked Information in Universal Random Privacy Amplification," *IEEE Transactions on Information Theory*, vol. 57, no. 6, pp. 3989–4001, 2011.

[12] A. Heuser, O. Rioul, and S. Guilley, "Good is Not Good Enough Deriving Optimal Distinguishers from Communication Theory." Springer, 9 2014, pp. 55–74.

[13] A. Ito, R. Ueno, and N. Homma, "On the Success Rate of Side-Channel Attacks on Masked Implementations: Information-Theoretical Bounds and Their Practical Usage," in *Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security, CCS 2022, Los Angeles, CA, USA, November 7-11, 2022*, H. Yin, A. Stavrou, C. Cremers, and E. Shi, Eds. ACM, 2022, pp. 1521–1535. [Online]. Available: https://doi.org/10.1145/3548606.3560579

[14] P. Jizba and T. Arimitsu, "The world according to Rényi: thermodynamics of multifractal systems," *Annals of Physics*, vol. 312, no. 1, pp. 17–59, 2004.

[15] Y. Liu, W. Cheng, S. Guilley, and O. Rioul, "On Conditional Alpha-Information and its Application to Side-Channel Analysis," in *IEEE Information Theory Workshop, ITW 2021, Kanazawa, Japan, October 17-21, 2021*. IEEE, 2021, pp. 1–6. [Online]. Available: https://doi.org/10.1109/ITW48936.2021.9611409

[16] L. Masure, O. Rioul, and F. Standaert, "A Nearly Tight Proof of Duc et al.'s Conjectured Security Bound for Masked Implementations," *IACR Cryptol. ePrint Arch.*, p. 600, 2022. [Online]. Available: https://eprint.iacr.org/2022/600

[17] Y. Polyanskiy and S. Verdú, "Arimoto Channel Coding Converse and Rényi Divergence," in *2010 48th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, 2010, pp. 1327–1333.

[18] O. Rioul, "A Primer on Alpha-Information Theory with Application to Leakage in Secrecy Systems," in *5th conference on Geometric Science of Information (GSI'21), Paris, France, 21-23 July 2021*, ser. Lecture Notes in Computer Science, 2021. [Online]. Available: https://perso.telecom-paristech.fr/rioul/publis/202102rioul.pdf

[19] R. Sibson, "Information Radius," *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, vol. 14, pp. 149–160, 1969.

[20] M. Tomamichel and M. Hayashi, "Operational Interpretation of Rényi Information Measures via Composite Hypothesis Testing against Product and Markov Distributions," *IEEE Transactions on Information Theory*, vol. 64, no. 2, pp. 1064–1082, 2017.

[21] S. Verdú, "α-Mutual Information." IEEE, 2015, pp. 1–6. [Online]. Available: https://doi.org/10.1109/ITA.2015.7308959