HAL
open science

On Conditional alpha-Information and its Application to Side-Channel Analysis

Yi Liu, Wei Cheng, Sylvain Guilley, Olivier Rioul

To cite this version:

Yi Liu, Wei Cheng, Sylvain Guilley, Olivier Rioul. On Conditional alpha-Information and its Application to Side-Channel Analysis. 2021 IEEE Information Theory Workshop (ITW2021), Oct 2021, Kanazawa, Japan. hal-03323522

HAL Id: hal-03323522

https://telecom-paris.hal.science/hal-03323522v1

Submitted on 15 Sep 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

On Conditional Alpha-Information and its Application to Side-Channel Analysis

Yi Liu*, Wei Cheng*, Sylvain Guilley†*, and Olivier Rioul*

*LTCI, Télécom Paris, Polytechnique de Paris, 91120, Palaiseau, France, firstname.lastname@telecom-paris.fr
†Secure-IC S.A.S., 75015, Paris, France, sylvain.guilley@secure-ic.com

***Abstract*—A conditional version of Sibson's α-information is defined using a simple closed-form "log-expectation" expression, which satisfies important properties such as consistency, uniform expansion, and data processing inequalities. This definition is compared to previous ones, which in contrast do not satisfy all of these properties. Based on our proposal and on a generalized Fano inequality, we extend the case $\alpha = 1$ of previous works to obtain sharp universal upper bounds for the probability of success of any type side-channel attack, particularly when $\alpha = 2$.**

I. Introduction

Mutual information as a theoretical tool to analyse the capability of an attacker to perform side-channel analysis has been advocated since Standaert et al. [14]. The communication channel model for this problem was used in [7] to optimize the side-channel attack distinguishers for any given type of leakage model. Recently, Chérisey et al. [2], [3] used such information-theoretic tools to establish some *universal* inequalities between the probability of success of a side-channel attack and the minimum number of queries to reach a given success rate. Such inequalities are "universal" in the sense that they can apply to any type of attack and depend only on the leakage model.

In this paper, we aim at extending the approach of [2], [3] to Rényi information quantities depending on a parameter α. For that we need the following ingredients that were crucial in the derivation steps of [2], [3]:

- a *closed-form expression* of conditional mutual information, amenable to efficient numerical estimation;
- a *data processing inequality* of conditional mutual information over a "conditional" Markov chain for a given plain or cypher text T (known to the attacker);
- a *expansion property* of conditional mutual information, i.e., its decomposition as a difference between conditional entropies, valid at least when the secret is assumed *uniformly distributed*;
- a *Fano inequality* which yields a lower bound on mutual information that depends on the probability of success (or equivalently on the probability of error).

Our aim, therefore, is to establish all of these properties for a suitably defined conditional Rényi version of mutual information of order $\alpha > 0$.

The rest of this paper is organized as follows. Section II reviews some useful definitions and properties of Rényi informational quantities. Emphasis is made on consistency, uniform expansion and data processing inequalities. Section III then proposes a natural definition of conditional α-information satisfying the required properties and Section IV makes a detailed comparison to previous proposals. Section V presents the main result applied to side-channel analysis, which is then validated using simulations.

II. Background and Definitions

A. α-Entropy and α-Divergence

Rényi entropy and divergence are well-known generalizations of Shannon's entropy and Kullback-Leibler divergence:

Definition 1. Assume that either $0 < \alpha < 1$ or $1 < \alpha < +\infty$ (the limiting values $0, 1, +\infty$ being obtained by taking limits).

The *α-entropy* of a probability distribution P and *α-divergence* of P from Q are defined as

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

$$D_\alpha(P\|Q) = \tfrac{1}{\alpha-1} \log \langle p\|q\rangle_\alpha^\alpha \tag{2}$$

where we have used the special notation:

$$\|p\|_\alpha = \Big(\int p^\alpha d\mu\Big)^{1/\alpha} \tag{3}$$

$$\langle p\|q\rangle_\alpha = \Big(\int p^\alpha q^{1-\alpha} d\mu\Big)^{1/\alpha} \tag{4}$$

with the following convention: All considered probability distributions P, Q possess a dominating measure μ such that $P \ll \mu$ and $Q \ll \mu$, the corresponding lower-case letters p, q are densities of P, Q with respect to μ.

Remark 1. When μ is a counting measure we obtain the classical definitions for discrete random variables; when μ is the Lebesgue measure we obtain the corresponding definitions for continuous variables. While it is easily seen that the definition of α-divergence does not depend on the chosen dominating measure μ, that of α-entropy does.

A link between these two quantities is the following *uniform expansion property* (UEP). Let $U \sim \mathcal{U}(M)$ be uniformly distributed over a set of finite μ-measure M. (In the discrete case U simply takes M equiprobable values.) Since $u \equiv \frac{1}{M}$ we have $\langle p\|u\rangle_\alpha = M^{\frac{\alpha-1}{\alpha}} \|p\|_\alpha$, hence

Property 1 (UEP of α-Divergence [16])**.** $D_\alpha(P\|U) = H_\alpha(U) - H_\alpha(P) = \log M - H_\alpha(P)$.

Another important property is the *data processing inequality* (DPI). A random transformation given by a conditional distribution $P_{Y|X}$ is noted $P_X \to \boxed{P_{Y|X}} \to P_Y$ if a random variable $X \sim P_X$ is input and the output distribution P_Y satisfies $p_Y(y) = \int p_{Y|X}(y|x)p_X(x)\,d\mu(x)$. Similarly for $Q_X \to \boxed{P_{Y|X}} \to Q_Y$ we have $q_Y(y) = \int_{\mathcal{X}} p(y|x)q_X(x)\,d\mu(x)$.

Property 2 (DPI for α-Divergence [10], [11])**.** *Any transformation can only reduce α-divergence: $D_\alpha(P_X\|Q_X) \geq D_\alpha(P_Y\|Q_Y)$.*

B. Conditional α-Entropy and α-Divergence

Both definitions of α-divergence and α-entropy have been extended to *conditional* versions, in a fairly natural way:

Definition 2. The conditional α-divergence is defined as [17]

$$D_\alpha(P_{Y|X}\|Q_{Y|X}|P_X) = D_\alpha(P_{Y|X}P_X\|Q_{Y|X}P_X) \tag{5}$$

This definition is consistent with the unconditional one:

Property 3 (Consistency of Conditional α-Divergence w.r.t. α-Divergence)**.** *If $X \equiv 0$ then $D_\alpha(P_{Y|X}\|Q_{Y|X}|P_X) = D_\alpha(P_Y\|Q_Y)$.*

Here following Shannon [12] we have noted $X \equiv 0$ for any random variable independent of everything else considered (e.g., a constant variable).

In Definition 2 we remark that the *expectation* over the conditioned variable is only taken *inside the logarithm* in the α-divergence's expression:

$$D_\alpha(P_{Y|X}\|Q_{Y|X}|P_X) = \tfrac{1}{\alpha-1}\log \mathbb{E}_X\langle p_{Y|X}\|q_{Y|X}\rangle_\alpha^\alpha \tag{6}$$

A similar "log-expectation" definition holds for the following preferred form of the conditional α-entropy (a.k.a. Arimoto's conditional entropy). Considering the expression $H_\alpha(X) = H_\alpha(P_X) = \frac{\alpha}{1-\alpha}\log\|p_X\|_\alpha$ and taking the expectation over a conditioned variable inside the logarithm yields the following

Definition 3. The conditional α-entropy of X given Y is defined as [1], [6]

$$H_\alpha(X|Y) = \frac{\alpha}{1-\alpha}\log \mathbb{E}_Y\|p_{X|Y}\|_\alpha \tag{7}$$

Among other variations of conditional α-entropy [6] it is this definition that enjoys all three important properties: *consistency, UEP and DPI*. Consistency is obvious from the definition:

Property 4 (Consistency of Conditional α-Entropy w.r.t. α-Entropy)**.** *If $Y \equiv 0$ then $H_\alpha(X|Y) = H_\alpha(X)$.*

As in the case of the α-entropy, since $\langle p_{X|Y}\|u\rangle_\alpha = M^{\frac{\alpha-1}{\alpha}}\|p_{X|Y}\|_\alpha$, we have the following

Property 5 (UEP)**.** *If $U \sim \mathcal{U}(M)$ is uniform independent of X, $D_\alpha(P_{Y|X}\|U|P_X) = H_\alpha(U) - H_\alpha(Y|X) = \log M - H_\alpha(Y|X)$.*

Property 6 (DPI for Conditional α-Entropy [6], [11])**.** *If $X - Y - Z$ forms a Markov chain, then $H_\alpha(X|Y) \leq H_\alpha(X|Z)$.*

In particular for $Z \equiv 0$, conditioning reduces α-entropy: $H_\alpha(X|Y) \leq H_\alpha(X|0) = H_\alpha(X)$. More generally one has [6] $H_\alpha(X|YY') \leq H_\alpha(X|Y')$.

C. α-Information

Sibson's α-information is perhaps the preferred generalization of Fano's classical mutual information and has found various applications [4], [5], [10], [11], [15], [17]:

Definition 4. The α-information [13], [17] of X from Y is defined as

$$I_\alpha(X;Y) = \frac{\alpha}{\alpha-1}\log \mathbb{E}_Y\langle p_{X|Y}\|p_X\rangle_\alpha \tag{8}$$

This is again a "log-expectation" expression where one takes the expectation over Y inside the logarithm in the expression of the divergence

$$D_\alpha(P_{X|Y=y}\|P_X) = \frac{\alpha}{\alpha-1}\log\langle p_{X|Y=y}\|p_X\rangle_\alpha$$

Remark 2. This construction focuses on the distribution of X, conditioned on Y or not. In contrast to the classical case $\alpha = 1$, the resulting definition of information is not symmetric: $I_\alpha(X;Y) \neq I_\alpha(Y,X)$. Therefore, α-information is no longer "mutual" when $\alpha \neq 1$.

As in the case of the conditional α-entropy, since $\langle p_{U|Y}\|u\rangle_\alpha = M^{\frac{\alpha-1}{\alpha}}\|p_{U|Y}\|_\alpha$, we have the following

Property 7 (UEP for α-Information [11], [16])**.** *If $U \sim \mathcal{U}(M)$ is uniformly distributed, then $I_\alpha(U;Y) = H_\alpha(U) - H_\alpha(U|Y) = \log M - H_\alpha(U|Y)$.*

Property 8 (DPI for α-Information [10], [11])**.** *If $W - X - Y - Z$ forms a Markov chain, then $I_\alpha(X;Y) \geq I_\alpha(W;Z)$.*

Proof (for completeness). Let $P_{X,Y} \to \boxed{P_{X,Z|X,Y}} \to P_{X,Z} \to \boxed{P_{W,Z|X,Z}} \to P_{W,Z}$. By the Markov condition, one has $P_{X,Z|X,Y} = P_{X|X}P_{Z|X,Y} = P_{X|X}P_{Z|Y}$ where $P_{X|X}$ is the identity operator; similarly $P_{W,Z|X,Z} = P_{W|X,Z}P_{Z|Z} = P_{W|X}P_{Z|Z}$. Thus if $Q_Y \to \boxed{P_{Z|Y}} \to Q_Z$, we find $P_XQ_Y \to \boxed{P_{X,Z|X,Y}} \to P_XQ_Z \to \boxed{P_{W,Z|X,Z}} \to P_WQ_Z$. Now by the data processing inequality for α-divergence (Property 2), $D_\alpha(P_{X,Y}\|P_XQ_Y) \geq D_\alpha(P_{W,Z}\|P_WQ_Z) \geq I_\alpha(W;Z)$. Minimizing over Q_Y gives the announced DPI. □

Remark 3. Because of the non-symmetric nature of α-information, the DPI corresponds to two separate statements of pre- and post-processing inequalities [10].

We remark that the Lapidoth-Pfister *mutual* information, which is symmetric, $J_\alpha(X;Y) = J_\alpha(Y;X)$ does also enjoy data processing inequalities but unfortunately does not seem to possess a closed-form expression [8].

D. Sibson's identity

An important property of α-information is *Sibson's identity*. It is straightforward to compute

$$\langle p_{XY}\|p_Xq_Y\rangle_\alpha^\alpha = \iint p_Y^\alpha p_{X|Y}^\alpha p_X^{1-\alpha}q_Y^{1-\alpha} \tag{9}$$

$$= \left\langle p_Y\langle p_{X|Y}\|p_X\rangle_\alpha \,\|\, q_Y\right\rangle_\alpha^\alpha. \tag{10}$$

Defining the (suitably normalized) distribution $q_Y^* = p_Y \langle p_{X|Y} \| p_X \rangle_\alpha / \mathbb{E}_Y \langle p_{X|Y} \| p_X \rangle_\alpha$, substituting and taking the logarithm gives the following

Proposition 1 (Sibson's identity [13], [17])**.** *One has*

$$D_\alpha(P_{XY} \| P_X Q_Y) = D_\alpha(Q_Y^* \| Q_Y) + I_\alpha(X;Y), \tag{11}$$

hence the following alternate minimizing definition:

$$I_\alpha(X;Y) = \min_{Q_Y} D_\alpha(P_{XY} \| P_X Q_Y). \tag{12}$$

E. Generalized Fano's Inequality

Assume X is discrete and estimated from Y using the MAP rule, with (maximal) probability of success $\mathbb{P}_s = \mathbb{P}_s(X|Y) = \mathbb{E} \sup_x p_{X|Y}(x|Y)$. Also let $\mathbb{P}_s(X) = \sup p_X$ be the probability of success when guessing X without even knowing Y. Using the DPI for α-information and α-divergence, we have the following

Lemma 1 (Rioul's Generalized Fano Inequality [11, Thm. 1])**.**

$$I_\alpha(X;Y) \geq d_\alpha\big(\mathbb{P}_s(X|Y) \| \mathbb{P}_s(X)\big) \tag{13}$$

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

$$d_\alpha(p\|q) = \tfrac{1}{\alpha-1} \log\big(p^\alpha q^{1-\alpha} + (1-p)^\alpha (1-q)^{1-\alpha}\big). \tag{14}$$

III. Conditional α-Information

A. Definition as a Log-Expectation Expression

As a natural continuation of the definitions in the preceding section, we define the conditional α-information with a "log-expectation" closed-form expression, obtained by taking the expectation over the conditional variable inside the logarithm in the expression of Sibson's (unconditional) α-information (8):

Definition 5 (Conditional α-Information, Closed-Form Definition)**.**

$$\begin{aligned} I_\alpha(X;Y|Z) &= \frac{\alpha}{\alpha-1} \log \mathbb{E}_Z \mathbb{E}_{Y|Z} \langle p_{X|YZ} \| p_{X|Z} \rangle_\alpha \\ &= \frac{\alpha}{\alpha-1} \log \mathbb{E}_{YZ} \langle p_{X|YZ} \| p_{X|Z} \rangle_\alpha \end{aligned} \tag{15}$$

To the best of our knowledge, this definition has not been considered elsewhere.

B. Basic Properties

Our definition enjoys three important properties: *consistency, UEP and DPI.*

Property 9 (Consistency of Conditional α-Information w.r.t. α-Information)**.** *If Z is independent of (X,Y) then $I_\alpha(X;Y|Z) = I_\alpha(X;Y)$.*

Proof. Obvious from the definitions. □

Property 10 (UEP for Conditional α-Information)**.** *If $U \sim \mathcal{U}(M)$ is uniformly distributed independent of Z, then $I_\alpha(U;Y|Z) = H_\alpha(U) - H_\alpha(U|YZ) = \log M - H_\alpha(U|YZ)$.*

Proof. Similarly as for the preceding UEPs, we have $\langle p_{U|YZ} \| u \rangle_\alpha = M^{\frac{\alpha-1}{\alpha}} \| p_{U|YZ} \|_\alpha$. Averaging over (Y,Z) and taking the logarithm gives the announced formula. □

We say that a sequence of random variables forms a *conditional Markov chain* given some random variable T if it is Markov for any $T = t$.

Property 11 (DPI for Conditional α-Information)**.** *If $W - X - Y - Z$ forms a conditional Markov chain given T, then $I_\alpha(X;Y|T) \geq I_\alpha(W;Z|T)$.*

Proof. By Property 8, $I_\alpha(X;Y|T=t) \geq I_\alpha(W;Z|T=t)$ for any t. From Definition 4 this gives $\langle p_{X|Y,T} \| p_{X|T} \rangle_\alpha \geq \langle p_{W|Z,T} \| p_{W|T=t} \rangle_\alpha$ for $\alpha > 1$ and the opposite inequality for $0 < \alpha < 1$. This in turn from Definition 5 gives the announced inequality for any α. □

C. Conditional Sibson's Identity

Proposition 2 (Conditional Sibson's Identity)**.** *One has*

$$D_\alpha(P_{XYZ} \| P_{X|Z} Q_{YZ}) = D_\alpha(Q_{YZ}^* \| Q_{YZ}) + I_\alpha(X;Y|Z), \tag{16}$$

hence the following alternate minimizing definition:

$$I_\alpha(X;Y|Z) = \min_{Q_{YZ}} D_\alpha(P_{XYZ} \| P_{X|Z} Q_{YZ}) \tag{17}$$

Proof. Similarly as in the case of α-information, it is straightforward to compute

$$\langle p_{XYZ} \| p_{X|Z} q_{YZ} \rangle_\alpha^\alpha = \iiint p_{YZ}^\alpha p_{X|YZ}^\alpha p_{X|Z}^{1-\alpha} q_{YZ}^{1-\alpha} \tag{18}$$

$$= \big\langle p_{YZ} \langle p_{X|YZ} \| p_{X|Z} \rangle_\alpha \,\|\, q_{YZ} \big\rangle_\alpha^\alpha \tag{19}$$

Defining the (suitably normalized) distribution $q_{YZ}^* = p_{YZ} \langle p_{X|YZ} \| p_{X|Z} \rangle_\alpha / \mathbb{E}_{YZ} \langle p_{X|YZ} \| p_{X|Z} \rangle_\alpha$, substituting and taking the logarithm gives the announced identity. □

IV. Comparison to Previous Definitions

A. Various Other Definitions

All previous definitions of conditional α-information we are aware of are variations of the form (17) where α-divergence is minimized with respect to different probability measures $Q_{X|Z}$, $Q_{Y|Z}$, Q_Z or combinations. There are exactly $2^3 = 8$ possibilities:

(o) $I_\alpha^{000}(X;Y|Z) = D_\alpha(P_{XYZ} \| P_{X|Z} P_{Y|Z} P_Z)$.
(i) $I_\alpha^{001}(X;Y|Z) = \min_{Q_Z} D_\alpha(P_{XYZ} \| P_{X|Z} P_{Y|Z} Q_Z)$.
(ii) $I_\alpha^{010}(X;Y|Z) = \min_{Q_{Y|Z}} D_\alpha(P_{XYZ} \| P_{X|Z} Q_{Y|Z} P_Z)$.
(iii) $I_\alpha^{011}(X;Y|Z) = \min_{Q_{YZ}} D_\alpha(P_{XYZ} \| P_{X|Z} Q_{YZ})$.
(iv) $I_\alpha^{100}(X;Y|Z) = \min_{Q_{X|Z}} D_\alpha(P_{XYZ} \| Q_{X|Z} P_{Y|Z} P_Z)$.
(v) $I_\alpha^{101}(X;Y|Z) = \min_{Q_{XZ}} D_\alpha(P_{XYZ} \| Q_{XZ} P_{Y|Z})$.
(vi) $I_\alpha^{110}(X;Y|Z) = \min_{Q_{X|Z} Q_{Y|Z}} D_\alpha(P_{XYZ} \| Q_{X|Z} Q_{Y|Z} P_Z)$.
(vii) $I_\alpha^{111}(X;Y|Z) = \min_{Q_{X|Z} Q_{YZ}} D_\alpha(P_{XYZ} \| Q_{X|Z} Q_{YZ})$.

Definition (o) is mentioned in [15, Eq. (70)]. Definition (i) is the main proposal of Esposito et al. [5]. Definition (ii) is discussed by Tomamichel and Hayashi [15, Eq. (74)] and

is equivalent to definition (iv) by permuting the roles of X and Y: $I_\alpha^{100}(X;Y|Z) = I_\alpha^{010}(Y;X|Z)$. Our definition (17) is definition (iii), and is equivalent to definition (v) by permuting the roles of X and Y: $I_\alpha^{101}(X;Y|Z) = I_\alpha^{011}(Y;X|Z)$. Finally, definitions (vi) and (vii) seem new and related to a conditional version of the Lapidoth-Pfister mutual information [8]: $J_\alpha(X;Y) = \min_{Q_X Q_Y} D_\alpha(P_{XY} \parallel Q_X Q_Y)$. Thus we need only to compare our definition to (o), (i), (ii), (vi) and (vii).

We now discuss various properties for these definitions, by decreasing order of importance: The fact that they admit or not a closed-form expression in terms of the involved probability densities; their consistency with respect to α-information $I_\alpha(X;Y|0) = I_\alpha(X;Y)$; the existence of a uniform expansion of the form $I_\alpha(U;Y|Z) = \log M - H_\alpha(U|YZ)$ when $U \sim \mathcal{U}(M)$ is independent of Z; and the fact that they satisfy data processing inequalities for conditional Markov chains.

B. Closed-Form and Consistency

Definition (o) is by itself a closed-form expression but is clearly *inconsistent* with respect to Sibson's α-information since $I_\alpha^{000}(X;Y|0) = D_\alpha(P_{XY}\|P_X P_Y)$ which by (12) is $\geq I_\alpha(X;Y)$ where the inequality is, in general, strict.

Definition (i) of Esposito et al. does admit a closed-form expression [5, Thm. 2]. In fact, since

$$\begin{aligned}\langle p_{XYZ}\|p_{X|Z}p_{Y|Z}q_Z\rangle_\alpha^\alpha &= \iiint p_Z^\alpha p_{XY|Z}^\alpha (p_{X|Z}p_{Y|Z})^{1-\alpha} q_Y^{1-\alpha} \\ &= \left\langle p_Z \langle p_{XY|Z}\|p_{X|Z}p_{Y|Z}\rangle_\alpha \parallel q_Z \right\rangle_\alpha^\alpha,\end{aligned}$$

letting $q_Z^* = p_Z\langle p_{X|YZ}\|p_{X|Z}p_{Y|Z}\rangle_\alpha / \mathbb{E}_Z\langle p_{X|YZ}\|p_{X|Z}p_{Y|Z}\rangle_\alpha$ and taking the logarithm gives the following variation of Sibson's identity (whose existence is mentioned but does not explicitly appear in [5]):

Proposition 3.

$$D_\alpha(P_{XYZ}\|P_{X|Z}P_{Y|Z}Q_Z) = D_\alpha(Q_Z^*\|Q_Z) + I_\alpha^{001}(X;Y|Z), \tag{20}$$

with the following closed-form expression:

$$I_\alpha^{001}(X;Y|Z) = \tfrac{\alpha}{\alpha-1}\log \mathbb{E}_Z\langle p_{XY|Z}\|p_{X|Z}p_{Y|Z}\rangle_\alpha. \tag{21}$$

However, I_α^{001} is *inconsistent* (with respect to Sibson's α-information) for the same reason as in the case of I_α^{000}: From (21) we have $I_\alpha^{001}(X;Y|0) = D_\alpha(P_{XY}\|P_X P_Y) \geq I_\alpha(X;Y)$.

Definition (ii) of Tomamichel and Hayashi also admits a closed-form expression [15, Eq. (75)]. In fact by the (unconditional) Sibson identity (11) applied to all variables conditioned on $Z = z$ for any z, one easily sees that $D_\alpha(P_{XYZ}\|P_{X|Z}Q_{Y|Z}P_Z)$ achieves its minimum when for $q_{Y|Z} = q_{Y|Z}^* = p_{Y|Z}\langle p_{X|YZ}\|p_X\rangle_\alpha / \mathbb{E}_{Y|Z}\langle p_{X|YZ}\|p_{X|Z}\rangle_\alpha$ as given above in the proof of (11), which gives

$$I_\alpha^{010}(X;Y|Z) = \tfrac{1}{\alpha-1}\log \mathbb{E}_Z(\mathbb{E}_{Y|Z}\langle p_{X|YZ}\|p_{X|Z}\rangle_\alpha)^\alpha. \tag{22}$$

From this it follows that $I_\alpha^{010}(X;Y|0) = I_\alpha(X;Y)$, proving that I_α^{010} is *consistent*.

Finally, definitions (vi) and (vii) are neither closed-form nor consistent; for when $Z \equiv 0$, the definitions reduce to the Lapidoth-Pfister mutual information: $J_\alpha(X;Y) = \min_{Q_X Q_Y} D_\alpha(P_{XY} \parallel Q_X Q_Y)$ which already does not admit a closed-form expression, and for which $J_\alpha(X;Y) \leq I_\alpha(X;Y)$ where the inequality is, in general, strict [8]. In the following we focus on the other definitions which admit closed-form expressions.

C. Uniform Expansion Property

The uniform expansion property (UEP) is a crucial requirement in our subsequent derivations (Theorem 1). It is naturally satisfied for α-information (Property 7) and it is important that it is also satisfied for its conditional version.

Using the above closed-form expressions it is easy to check the UEP when $U \sim \mathcal{U}(M)$ is independent of Z, neither $I_\alpha^{000}(U;Y|Z)$, nor $I_\alpha^{001}(U;Y|Z)$, nor $I_\alpha^{010}(U;Y|Z)$ equals $\log M - H_\alpha(U|YZ)$. This is not surprising since in general, from the different minimizations of α-divergence,

$$\begin{aligned}I_\alpha(X;Y|Z) &= I_\alpha^{011}(X;Y|Z) \\ &\leq \min\{I_\alpha^{001}(X;Y|Z), I_\alpha^{010}(X;Y|Z)\} \\ &\leq I_\alpha^{000}(X;Y|Z)\end{aligned} \tag{23}$$

where inequalities are, in general, strict. Hence the only case where the UEP (which is crucial in our subsequent derivations) holds is for the definition (iii) proposed in this paper.

D. Data Processing Inequality

Finally, since definitions (o) and (i) are inconsistent with $I_\alpha^{000}(X;Y|0) = I_\alpha^{001}(X;Y|0) = D_\alpha(P_{XY}\|P_X P_Y)$, they do not even satisfy data processing inequalities for a unconditional Markov chain. Therefore, the only remaining candidate for DPI is definition (ii).

Property 12 (DPI for $I_\alpha^{010}(X;Y|Z)$)**.** *If $W-X-Y-Z$ forms a conditional Markov chain given T, then $I_\alpha^{010}(X;Y|T) \geq I_\alpha^{010}(W;Z|T)$.*

Proof. We mirror the proof of Property 8. Let $P_{X,Y,T} \to \boxed{P_{X,Z,T|X,Y,T}} \to P_{X,Z,T} \to \boxed{P_{W,Z,T|X,Z,T}} \to P_{W,Z,T}$. By the conditional Markov condition, we have $P_{X,Z,T|X,Y,T} = P_{X,T|X,T}P_{Z|X,Y,T} = P_{X,T|X,T}P_{Z|Y,T}$ where $P_{X,T|X,T}$ is the identity operator; similarly $P_{W,Z,T|X,Z,T} = P_{W|X,Z,T}P_{Z,T|Z,T} = P_{W|X,T}P_{Z,T|Z,T}$. Thus if $Q_{Y|T} \to \boxed{P_{Z|Y,T}} \to Q_{Z|T}$, we find $P_{X|T}Q_{Y|T}P_T \to \boxed{P_{X,Z,T|X,Y,T}} \to P_{X|T}Q_{Z|T}P_T \to \boxed{P_{W,Z,T|X,Z,T}} \to P_{W|T}Q_{Z|T}P_T$. By the data processing inequality for α-divergence (Property 2), $D_\alpha(P_{X,Y,T}\|P_{X|T}Q_{Y|T}P_T) \geq D_\alpha(P_{W,Z,T}\|P_{W|T}Q_{Z|T}P_T) \geq I_\alpha(W;Z|T)$. Minimizing over $Q_{Y|T}$ gives the announced DPI. □

Table I summarizes the comparison between properties of (o)–(vii).

V. Application to Side-Channel Analysis

A. Theoretical Derivation

We follow the framework and notations from [2], [3] and [7]. Let K be a secret key and T be a plain text known

TABLE I
COMPARISON OF SOME PROPERTIES FOR THE VARIOUS DEFINITIONS.

Definition	Ref.	Closed-form	Consistency	UEP	DPI
o	[15]	yes	no	no	no
i	[5]	yes	no	no	no
ii,iv	[15]	yes	yes	no	yes
iii,v	(this paper)	**yes**	**yes**	**yes**	**yes**
vi,vii	—	no	no		

to the attacker. During cryptographic processing, X is leaked from the implementation and measured as a "trace" Y by the attacker at the output of some noisy measurement channel. The secret key K can take M equiprobable values and is evidently independent of the text T. The leakage function $X = f(K, T)$ is unknown, but deterministic. The attacker then exploits his knowledge of T and Y to estimate the secret $\hat{K}$ and we let $\mathbb{P}_s = \mathbb{P}(\hat{K} = K)$ be the probability of success. The communication channel model is depicted in Fig. 1.

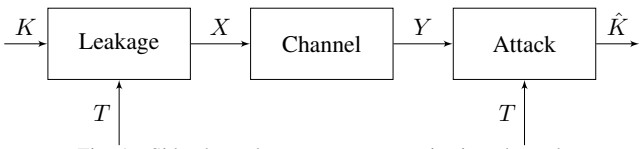

Fig. 1. Side-channel seen as a communication channel.

One might think at first glance that this model is similar to a *wiretap channel model* [9] where K would play the role of the "message" and the leakage that of an (unknown) encoder. However, in total contrast to the wire-tap model, here K should always be kept as secret and is never to be transmitted reliably to any destination (there is no legitimate receiver). The actual message (plain or cyphertext T) is assumed known to the eavesdropper, and X is not an encoded version of it but just some information leaked when the device executing cryptographic operations (power consumption, electromagnetic leaks, etc.).

Theorem 1. *One has the following upper bound on the probability of success $\mathbb{P}_s$:*

$$I_\alpha(X, Y|T) \geq d_\alpha(\mathbb{P}_s \| \frac{1}{M}) \tag{24}$$

Proof. The chain $K - X - Y$ is Markov given T by assumption but since $X = f(K, T)$, the chain $X - K - Y$ is also Markov given T. Therefore, by the conditional DPI (Property 11), $I_\alpha(X, Y|T) = I_\alpha(K, Y|T)$ (inequalities in both directions). Now since $K - Y - \hat{K}$ is also Markov given T, we have $I_\alpha(K; Y|T) \geq I_\alpha(K; \hat{K}|T)$. Since K is equiprobable independent of T, by the UEP (Property 10), $I_\alpha(K; \hat{K}|T) = \log M - H_\alpha(K|\hat{K}, T) \geq \log M - H_\alpha(K|\hat{K}) = I_\alpha(K; \hat{K})$. Finally, using Lemma 1, $I_\alpha(K; \hat{K}) \geq d_\alpha(\mathbb{P}_s(K|Y) \| \mathbb{P}_s(K)) = d_\alpha(\mathbb{P}_s \| \frac{1}{M})$, which proves (24). □

Remark 4. From (14), $d_\alpha(p, q)$ is increasing in p when $p \geq q$. Hence (24) gives an upper bound on $\mathbb{P}_s$ (which is obviously $\geq 1/M$ since $\mathbb{P}_s = 1/M$ corresponds to a blind guess when the attacker does not know Y).

B. Numerical Simulations

We consider an implementation of the AES with a large number q of measurement traces. Here $M = 256$ and the most commonly used leakage model is

$$Y_i = w_H(S(T_i \oplus K)) + N_i \qquad (i = 1, 2, \ldots, q) \tag{25}$$

where w_H denotes the Hamming weight, S denotes a S-box permutation and $\mathbf{N}_i$ are i.i.d $\sim \mathcal{N}(0, \sigma^2)$. Letting $\mathbf{X} = (X_i)_i$, $\mathbf{Y} = (Y_i)_i$, $\mathbf{T} = (T_i)_i$, we can compute $I_\alpha(\mathbf{X}, \mathbf{Y}|\mathbf{T}) = I_\alpha(K, \mathbf{Y}|\mathbf{T})$ using Monte-Carlo simulation similarly as in [2].

The numerical results on the success probability *upper bounds* for $\alpha = 1/2$, 1, and 2. are shown in Fig. 2, which compares them to the average performance of the optimal ML attack (with error bars). Since $I_\alpha(\mathbf{X}, \mathbf{Y}|\mathbf{T})$ increases with q, these in turn allows us to derive *lower bounds* on the number of traces $q_{\min}$ which are needed to achieve a given success rate $\mathbb{P}_s$. This is illustrated in Fig. 3.

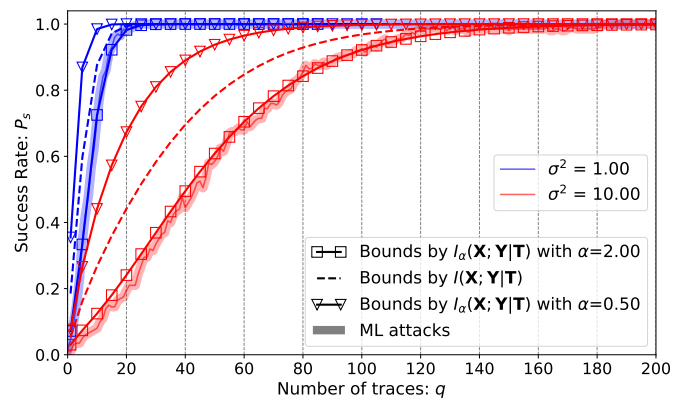

Fig. 2. Comparison of upper bounds on success rate P_s given α-information $I_\alpha(\mathbf{X}, \mathbf{Y}|\mathbf{T})$ for different values of α, for a Hamming weight leakage model in an AES-256 implementation.

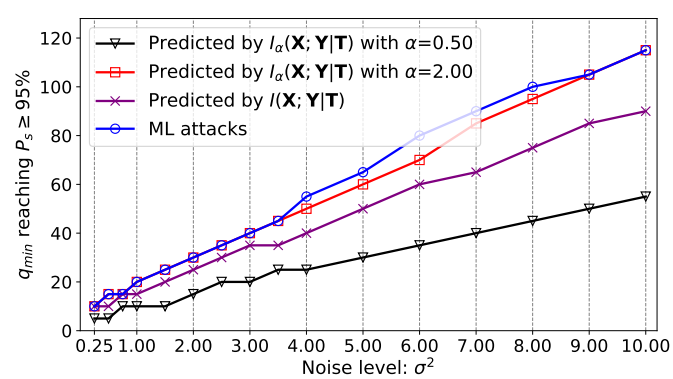

Fig. 3. Comparison of lower bounds on the number of traces $q_{\min}$ required to reach $\mathbb{P}_s \geq 95\%$ success rate.

It is quite remarkable to see that the case $\alpha = 2$, corresponding to a *collision* entropy $H_\alpha(K|\hat{K})$, gives a very sharp bound in our setting, which improves the results of [2], [3] for $\alpha = 1$ very much.

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] É. 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, no. 2, pp. 49–79, 2019. [Online]. Available: https://doi.org/10.13154/tches.v2019.i2.49-79

[3] ——, "An information-theoretic model for side-channel attacks in embedded hardware," in *2019 IEEE International Symposium on Information Theory (ISIT 2019)*, July 2019.

[4] A. R. Esposito, M. Gastpar, and I. Issa. Generalization error bounds via rényi-, f-divergences and maximal leakage. [Online]. Available: https://arxiv.org/abs/1912.01439

[5] A. R. Esposito, D. Wu, and M. Gastpar, "On conditional sibson's α-mutual information," *CoRR*, vol. abs/2102.00720, 2021. [Online]. Available: https://arxiv.org/abs/2102.00720

[6] S. Fehr and S. Berens, "On the conditional Rényi entropy," *IEEE Trans. Inf. Theory*, vol. 60, no. 11, pp. 6801–6810, 2014. [Online]. Available: https://doi.org/10.1109/TIT.2014.2357799

[7] A. Heuser, O. Rioul, and S. Guilley, "Good is not good enough - Deriving optimal distinguishers from communication theory," in *Cryptographic Hardware and Embedded Systems - CHES 2014 - 16th International Workshop, Busan, South Korea, September 23-26, 2014. Proceedings*, ser. Lecture Notes in Computer Science, L. Batina and M. Robshaw, Eds., vol. 8731. Springer, 2014, pp. 55–74. [Online]. Available: https://doi.org/10.1007/978-3-662-44709-3_4

[8] A. Lapidoth and C. Pfister, "Two measures of dependence," in *2016 IEEE International Conference on the Science of Electrical Engineering (ICSEE)*, 2016, pp. 1–5. [Online]. Available: https://ieeexplore.ieee.org/document/7806035

[9] Y. Liang, H. V. Poor, and S. Shamai, "Information theoretic security," *Found. Trends Commun. Inf. Theory*, vol. 5, no. 4-5, pp. 355–580, 2009. [Online]. Available: https://doi.org/10.1561/0100000036

[10] 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.

[11] 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

[12] C. E. Shannon, "The lattice theory of information," *Transactions of the IRE Professional Group on Information Theory*, vol. 1, no. 1, pp. 105–107, Feb. 1953.

[13] R. Sibson, "Information radius," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, vol. 14, no. 2, pp. 149–160, 1969. [Online]. Available: https://doi.org/10.1007/BF00537520

[14] F. Standaert, T. Malkin, and M. Yung, "A unified framework for the analysis of side-channel key recovery attacks," in *Advances in Cryptology - EUROCRYPT 2009, 28th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Cologne, Germany, April 26-30, 2009. Proceedings*, ser. Lecture Notes in Computer Science, A. Joux, Ed., vol. 5479. Springer, 2009, pp. 443–461. [Online]. Available: https://doi.org/10.1007/978-3-642-01001-9_26

[15] M. Tomamichel and M. Hayashi, "Operational interpretation of Rényi information measures via composite hypothesis testing against product and markov distributions," *IEEE Trans. Inf. Theory*, vol. 64, no. 2, pp. 1064–1082, 2018. [Online]. Available: https://doi.org/10.1109/TIT.2017.2776900

[16] T. van Erven and P. Harremoës, "Rényi divergence and Kullback-Leibler divergence," *IEEE Trans. Inf. Theory*, vol. 60, no. 7, pp. 3797–3820, 2014. [Online]. Available: https://doi.org/10.1109/TIT.2014.2320500

[17] S. Verdú, "α-mutual information," in *2015 Information Theory and Applications Workshop, ITA 2015, San Diego, CA, USA, February 1-6, 2015*. IEEE, 2015, pp. 1–6. [Online]. Available: https://doi.org/10.1109/ITA.2015.7308959