A Primer on Alpha-Information Theory with Application to Leakage in Secrecy Systems

Olivier Rioul

A Primer on Alpha-Information Theory with Application to Leakage in Secrecy Systems

Olivier Rioul[0000−0002−8681−8916]

LTCI, Télécom Paris, Institut Polytechnique de Paris, 91120 Palaiseau, France
olivier.rioul@telecom-paris.fr

Abstract. We give an informative review of the notions of Rényi's α-entropy and α-divergence, Arimoto's conditional α-entropy, and Sibson's α-information, with emphasis on the various relations between them. All these generalize Shannon's classical information measures corresponding to $\alpha = 1$. We present results on data processing inequalities and provide some new generalizations of the classical Fano's inequality for any $\alpha > 0$. This enables one to α-information as a information theoretic metric of leakage in secrecy systems. Such metric can bound the gain of an adversary in guessing some secret (any potentially random function of some sensitive dataset) from disclosed measurements, compared with the adversary's prior belief (without access to measurements).

Keywords: Rényi entropy and divergence · Arimoto conditional entropy · Sibson's information · Data processing inequalities · Fano's inequality · Information leakage · Side-Channel analysis.

1 Introduction

Shannon's information theory is based on the classical notions of entropy $H(X)$, relative entropy $D(P\|Q)$ a.k.a. divergence, conditional entropy $H(X|Y)$ and mutual information $I(X;Y)$. The fundamental property that makes the theory so powerful is that all these informational quantities satisfy *data processing inequalities*[1].

An increasingly popular generalization of entropy is *Rényi entropy* of order α, or α-entropy $H_\alpha(X)$. Many compatible generalizations of relative and conditional entropies and information have been proposed, yet the only suitable quantities that do satisfy the correct data processing inequalities are Arimoto's conditional entropy $H_\alpha(X|Y)$ and Sibson's α-information $I_\alpha(X;Y)$. In this paper, we first review the corresponding α-information theory that beautifully generalizes Shannon's classical information theory (which is recovered as the limiting case $\alpha \to 1$). For $\alpha \neq 1$, however, α-information is no longer *mutual*: $I_\alpha(X;Y) \neq I_\alpha(Y;X)$ in general.

The classical *Fano inequality* $H(X|Y) \leq h(\mathbb{P}_e) + \mathbb{P}_e \log(M-1)$ relating conditional entropy and probability of error $\mathbb{P}_e$ for a M-ary variable X can then be appropriately generalized using the data processing inequality for α-divergence. We present an appealing application to side-channel analysis.

[1] See [11] for a review of several data processing results.

2 A Primer on α-Information Theory

2.1 Notations

Probability Distributions In this paper, probability distributions such as P, Q are such that $P \ll \mu$ and $Q \ll \mu$ where μ is a given reference (σ-finite) measure. The corresponding lower-case letters denote the Radon-Nykodym derivatives $p = \frac{\mathrm{d}P}{\mathrm{d}\mu}$, $q = \frac{\mathrm{d}Q}{\mathrm{d}\mu}$. The probability distribution P of random variable X is sometimes noted P_X. The reference measure is then noted μ_X and the Radon-Nykodym derivative is $p_X = \frac{\mathrm{d}P_X}{\mathrm{d}\mu_X}$.

Discrete and Continuous Random Variables If X is a discrete random variable (taking values in a discrete set $\mathcal{X}$), then μ_X can be taken as the counting measure on $\mathcal{X}$; any integral over μ_X then reduces to a discrete sum over $x \in \mathcal{X}$, and we have $p_X(x) = \mathbb{P}(X{=}x)$. When X is a continuous random variable taking values in $\mathbb{R}^n$, μ_X is the Lebesgue measure on $\mathbb{R}^n$ and $p_X(x)$ is the corresponding probability density function. The notation $\mathbb{E}_x$ denotes expectation with respect to μ_X: $\mathbb{E}_x f(x) = \mathbb{E}\, f(X) = \int_{\mathcal{X}} f(x) p_X(x) \mathrm{d}\mu_X(x)$.

Uniform distributions As an example, we write $X \sim \mathcal{U}(M)$ if X is uniformly distributed over a set $\mathcal{X}$ of finite measure $M = \int_{\mathcal{X}} \mathrm{d}\mu$. The corresponding density is $p_X(x) = \frac{1_{\mathcal{X}}(x)}{M}$. In the discrete case, this means that X takes M equiprobable values in the set $\mathcal{X}$ of cardinality M.

Escort distributions For any distribution $p = \frac{\mathrm{d}P}{\mathrm{d}\mu}$, its *escort* distribution P_α of exponent α is given by the normalized α-power:

$$p_\alpha(x) = \frac{p^\alpha(x)}{\int_{\mathcal{X}} p^\alpha(x) \mathrm{d}\mu(x)}. \tag{1}$$

Joint Distributions A joint distribution $P_{X,Y}$ of two random variables X, Y is such that $P_{X,Y} \ll \mu_X \otimes \mu_Y$ where μ_X is the reference measure for X and μ_Y is the reference measure for Y. In this paper, all functions of (x, y) integrated over $\mu_X \otimes \mu_Y$ will always be measurable and nonnegative so that all the integrals considered in this paper exist (with values in $[0, +\infty]$) and Fubini's theorem always applies. The conditional distributions $P_{Y|X}$ and $P_{X|Y}$ are such that $p_{X,Y}(x, y) = p_X(x) p_{Y|X}(y|x) = p_Y(y) p_{X|Y}(x|y)$ $(\mu_X \otimes \mu_Y)$-a.e.. We write $P_{X,Y} = P_X P_{Y|X} = P_{X|Y} P_Y$. In particular, $P_X \otimes P_Y$ is simply noted $P_X P_Y$.

Random Transformations A random transformation $P_{Y|X}$ applies to any input distribution P_X and provides an output distribution P_Y, which satisfies $p_Y(y) = \int p_{Y|X}(y|x) p_X(x) \mathrm{d}\mu_X(x)$. We write $P_X \to \boxed{P_{Y|X}} \to P_Y$. The same random transformation can be applied to another distribution Q_X. We then write $Q_X \to \boxed{P_{Y|X}} \to Q_Y$ where the corresponding output distribution Q_Y is such that $q_Y(y) = \int p_{Y|X}(y|x) q_X(x) \mathrm{d}\mu_X(x)$.

Deterministic Transformations Any deterministic function $Y = f(X)$ taking discrete values can be seen as a particular case of a random transformation $P_{Y|X}$ where $p_{Y|X}(y|x) = \delta_{f(x)}(y)$.

2.2 Definitions

Throughout this paper we consider a Rényi order $\alpha \in (0,1) \cup (1,+\infty)$. In the following definitions, the corresponding quantities for $\alpha = 0, 1$ and $+\infty$ will be obtained by taking limits.

Definition 1 (Rényi Entropy [10]). *The α-entropy of $X \sim p$ is*

$$H_\alpha(X) = \frac{1}{1-\alpha} \log \int_{\mathcal{X}} p^\alpha(x) \, \mathrm{d}\mu(x). \tag{2}$$

It can also be noted $H_\alpha(P)$ or $H_\alpha(p)$. When X is binary with distribution $(p, 1-p)$, the α-entropy reduces to

$$h_\alpha(p) = \frac{1}{1-\alpha} \log\big(p^\alpha + (1-p)^\alpha\big). \tag{3}$$

Definition 2 (Rényi Divergence [10,5]). *The α-divergence or relative α-entropy of P and Q is*

$$D_\alpha(P\|Q) = \frac{1}{\alpha-1} \log \int_{\mathcal{X}} p^\alpha(x) q^{1-\alpha}(x) \, \mathrm{d}\mu(x). \tag{4}$$

For binary distributions $(p, 1-p)$ and $(q, 1-q)$, it reduces to

$$d_\alpha(p\|q) = \frac{1}{\alpha-1} \log\big(p^\alpha q^{1-\alpha} + (1-p)^\alpha (1-q)^{1-\alpha}\big). \tag{5}$$

Definition 3 (Arimoto-Rényi Entropy [1,7]). *The conditional α-entropy of X given Y is*

$$H_\alpha(X|Y) = \frac{\alpha}{1-\alpha} \log \int_{\mathcal{Y}} p_Y(y) \Big(\int_{\mathcal{X}} p^\alpha_{X|Y}(x|y) \, \mathrm{d}\mu_X(x) \Big)^{1/\alpha} \mathrm{d}\mu_Y(y). \tag{6}$$

Definition 4 (Sibson's mutual information [14,4,15]). *The α-mutual information (or simply α-information) of X and Y is*

$$I_\alpha(X;Y) = \frac{\alpha}{\alpha-1} \log \int_{\mathcal{Y}} \Big(\int_{\mathcal{X}} p_X(x) p^\alpha_{Y|X}(y|x) \, \mathrm{d}\mu_X(x) \Big)^{1/\alpha} \mathrm{d}\mu_Y(y) \tag{7}$$

$$= \frac{\alpha}{\alpha-1} \log \int_{\mathcal{Y}} p_Y(y) \Big(\int_{\mathcal{X}} p^\alpha_{X|Y}(x|y) p_X^{1-\alpha}(x) \, \mathrm{d}\mu_X(x) \Big)^{1/\alpha} \mathrm{d}\mu_Y(y). \tag{8}$$

Notice that $I_\alpha(X;Y) \neq I_\alpha(Y;X)$ in general.

By continuous extension of the above quantities, we recover the classical entropy $H_1(X) = H(X)$, divergence $D_1(P\|Q) = D(P\|Q)$, conditional entropy $H_1(X|Y) = H(X|Y)$, and mutual information $I_1(X;Y) = I(X;Y)$.

2.3 Basic Properties

Several known properties of α-divergence, α-entropies, and α-information are needed in the sequel. For completeness we list them as lemmas along with proof sketches.

Lemma 1 (α-information inequality [5, Thm. 8][2]).

$$D_\alpha(P\|Q) \geq 0 \tag{9}$$

with equality if and only if $P = Q$.

Proof. Apply Hölder's inequality $\int p^\alpha q^{1-\alpha} \leq (\int p)^\alpha (\int q)^{1-\alpha}$ for $\alpha < 1$, or the reverse Hölder inequality for $\alpha > 1$. Equality holds when $p = q$ μ-a.e., that is, $P = Q$. An alternative proof uses Jensen's inequality applied to $x \mapsto x^\alpha$, which is concave for $\alpha < 1$ and convex for $\alpha > 1$.

Lemma 2 (link between α-divergence and α-entropy [5, Eq. (2)]). *Letting $U = \mathcal{U}(M)$ be the uniform distribution,*

$$D_\alpha(P\|U) = \log M - H_\alpha(P) \tag{10}$$

Proof. Set $Q = U$ and $q = 1/M$ in (4).

Lemma 2 holds for discrete or continuous distributions. In particular from (9), (10) implies that if $M = \int_{\mathcal{X}} \mathrm{d}\mu < \infty$ then

$$H_\alpha(X) \leq \log M \tag{11}$$

with equality if and only if $X \sim \mathcal{U}(M)$.

The following is the natural generalization for $\alpha \neq 1$ of the well-known Gibbs inequality $H(X) \leq -\mathbb{E} \log q(X)$.

Lemma 3 (α-Gibbs inequality [12, Thm 1]). *Let $X \sim p$. For any probability distribution $\phi(x)$,*

$$H_\alpha(X) \leq \frac{\alpha}{1-\alpha} \log \mathbb{E}\big[\phi^{\frac{\alpha-1}{\alpha}}(X)\big] \tag{12}$$

with equality if and only if $\phi = p_\alpha$, the escort distribution (1) of p. For any family of conditional densities $\phi(x|y)$,

$$H_\alpha(X|Y) \leq \frac{\alpha}{1-\alpha} \log \mathbb{E}\big[\phi^{\frac{\alpha-1}{\alpha}}(X|Y)\big] \tag{13}$$

with equality if and only if $\phi(x|Y) = p_\alpha(x|Y)$ μ_Y-a.e. where $p_\alpha(x|y)$ is the escort distribution of $p(x|y) = p_{X|Y}(x|y)$.

For uniform $q(x) \sim \mathcal{U}(M)$ in (12) we recover (11).

[2] We name this "information inequality" after Cover and Thomas which used this terminology for the usual divergence $D_1(P\|Q) = D(P\|Q)$ [3, Theorem 2.6.3].

Proof. Let $q = \phi_{1/\alpha}$ so that $\phi = q_\alpha$. An easy calculation gives $\frac{\alpha}{1-\alpha}\log \mathbb{E}\big[\phi^{\frac{\alpha-1}{\alpha}}(X)\big] = D_{1/\alpha}(P_\alpha\|Q_\alpha) + H_\alpha(X)$. The first assertion then follows from Lemma 1.

Now for fixed $y \in \mathcal{Y}$, one has $H_\alpha(X|Y=y) \le \frac{\alpha}{1-\alpha}\log \mathbb{E}\big[\phi^{\frac{\alpha-1}{\alpha}}(X|Y=y)\big]$ with equality if and only if $\phi(x|y) = p_\alpha(x|y)$. In both cases $\alpha < 1$ and $\alpha > 1$, it follows that $H_\alpha(X|Y) = \frac{\alpha}{1-\alpha}\log \mathbb{E}_y \exp\frac{1-\alpha}{\alpha} H_\alpha(X|Y=y) \le \frac{\alpha}{1-\alpha}\log \mathbb{E}\big[\phi^{\frac{\alpha-1}{\alpha}}(X|Y)\big]$.

Lemma 4 (conditioning reduces α-entropy [1,7]).

$$H_\alpha(X|Y) \le H_\alpha(X) \tag{14}$$

with equality if and only if X and Y are independent.

Proof. Set $\phi(x|y) = p_\alpha(x)$ in (13), with equality iff $p_\alpha(x|Y) = p_\alpha(x)$ a.e., i.e., X and Y are independent. (An alternate proof [7] uses Minkowski's inequality.)

Lemma 5 (α-information and conditional α-entropy [13, Eq. (47)]). *If $X \sim \mathcal{U}(M)$,*

$$I_\alpha(X;Y) = \log M - H_\alpha(X|Y). \tag{15}$$

Proof. Set $p_X(x) = 1/M$ in (8).

It is not true in general that $I_\alpha(X;Y) = H_\alpha(X) - H_\alpha(X|Y)$ for nonuniform X [15].

Lemma 6 (α-information and α-divergence). *Sibson's identity[14, Thm. 2.2], [4, Eq. (20)], [15, Thm. 1]:*

$$D_\alpha(P_{X,Y}\|P_X Q_Y) = I_\alpha(X;Y) + D_\alpha(Q_Y^*\|Q_Y) \tag{16}$$

for any distribution Q_Y, where Q_Y^* is given by

$$q_Y^*(y) = \frac{\big(\int_{\mathcal{X}} p_X(x) p_{Y|X}^\alpha(y|x)\,\mathrm{d}\mu(x)\big)^{1/\alpha}}{\int_{\mathcal{Y}}\big(\int_{\mathcal{X}} p_X(x) p_{Y|X}^\alpha(y|x)\,\mathrm{d}\mu(x)\big)^{1/\alpha}\mathrm{d}\mu_Y(y)} \tag{17}$$

In particular from (9),

$$I_\alpha(X;Y) = \min_{Q_Y} D_\alpha(P_{X,Y}\|P_X Q_Y) = D_\alpha(P_{X,Y}\|P_X Q_Y^*) \tag{18}$$

hence $I_\alpha(X;Y) \ge 0$ with equality if and only if X and Y are independent.

Proof. $D_\alpha(P_{X,Y}\|P_X Q_Y) = \frac{1}{\alpha-1}\log\int p_X(x)\Big(\int p_{Y|X}^\alpha(y|x) q_Y^{1-\alpha}(y)\mathrm{d}\mu_Y(y)\Big)\mathrm{d}\mu_X(x)$ $= \frac{1}{\alpha-1}\log\int q_Y^{1-\alpha}(y)\Big(\int p_X(x) p_{Y|X}^\alpha(y|x)\mathrm{d}\mu_X(x)\Big)\mathrm{d}\mu_Y(y)$ where we applied Fubini's theorem in the second equality. Substituting the expression inside parentheses using (17) gives (16). Using Lemma 1 it follows as expected that $I_\alpha(X;Y) \ge 0$ with equality if and only if X and Y are independent (in which case $Q_Y^* = P_Y$).

Lemma 7 (Convexity of α-divergence [4, Appendix], [5, Thm. 12]). *$Q \mapsto D_\alpha(P\|Q)$ is convex: For any $\lambda \in (0,1)$,*

$$D_\alpha(P\|\lambda Q_1 + (1-\lambda)Q_2) \le \lambda D_\alpha(P\|Q_1) + (1-\lambda) D_\alpha(P\|Q_2). \tag{19}$$

Notice that $D_\alpha(P\|Q)$ is not convex in P in general (when $\alpha > 1$) [4,5].

2.4 Data Processing Inequalities

Lemma 8 (Data processing reduces α-divergence [9], [5, Thm. 1]). *For random transformations $P_X \to \boxed{P_{Y|X}} \to P_Y$ and $Q_X \to \boxed{P_{Y|X}} \to Q_Y$, we have the data processing inequality:*

$$D_\alpha(P_X\|Q_X) \geq D_\alpha(P_Y\|Q_Y) \tag{20}$$

with equality if and only if $P_{X|Y} = Q_{X|Y}$, where $P_{X|Y}P_Y = P_{Y|X}P_X$ and $Q_{X|Y}Q_Y = P_{Y|X}Q_X$.

In particular, for any μ_X-measurable set $A \subset \mathcal{X}$,

$$D_\alpha(P_X\|Q_X) \geq d_\alpha(p_A\|q_A) \tag{21}$$

where $p_A = \mathbb{P}(X \in A)$, $q_A = \mathbb{Q}(X \in A)$, and d_α is the binary α-divergence (5) *. Equality in* (21) *holds if and only if $P_{X|X\in A} = Q_{X|X\in A}$ and $P_{X|X\notin A} = Q_{X|X\notin A}$.*

Proof. Write $D_\alpha(P_X\|Q_X) = D_\alpha(P_XP_{Y|X}\|Q_XP_{Y|X}) = D_\alpha(P_YP_{X|Y}\|Q_YQ_{X|Y}) = \frac{1}{\alpha-1}\log\int p_Y^\alpha q_Y^{1-\alpha}\Big(\int p_{X|Y}^\alpha q_{X|Y}^{1-\alpha}\mathrm{d}\mu_X\Big)\mathrm{d}\mu_Y \geq D_\alpha(P_Y\|Q_Y)$ where we used (9) in the form $\int p^\alpha q^{1-\alpha} \leq (\int p)^\alpha(\int q)^{1-\alpha}$ for $\alpha < 1$ and the opposite inequality for $\alpha > 1$, applied to $p = p_{X|Y}$ and $q = q_{X|Y}$. The equality condition in (9) implies $P_{X|Y} = Q_{X|Y}$, which in turns implies equality in (20). Applying the statement to the deterministic transformation $Y = 1_{X\in A}$ gives (21).

Lemma 9 (Data processing reduces α-information [9, Thm. 5 (2)]). *For any Markov chain[3] $W - X - Y - Z$,*

$$I_\alpha(X;Y) \geq I_\alpha(W;Z) \tag{22}$$

Proof. 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, $P_{X,Z|X,Y} = P_{X|X}P_{Z|X,Y} = P_{X|X}P_{Z|Y}$ 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 $P_XQ_Y \to \boxed{P_{X,Z|X,Y}} \to P_XQ_Z \to \boxed{P_{W,Z|X,Z}} \to P_WQ_Z$. By the data processing inequality for α-divergence (20), $D_\alpha(P_{X,Y}\|P_XQ_Y) \geq D_\alpha(P_{W,Z}\|P_WQ_Z) \geq I_\alpha(W;Z)$. Minimizing over Q_Y gives (22).

Lemma 10 (Data processing increases conditional α-entropy [7, Cor. 1]). *For any Markov chain $X - Y - Z$,*

$$H_\alpha(X|Y) \leq H_\alpha(X|Z) \tag{23}$$

with equality if and only if $X - Z - Y$ forms a Markov chain.

When $X \sim \mathcal{U}(M)$, inequality (23) also follows from (15) and the data processing inequality for α-information (22) where $W = X$.

[3] We do not specify a direction since reversal preserves the Markov chain property: $X_1 - X_2 - \cdots - X_n$ is Markov if and only if $X_n - X_{n-1} - \cdots - X_1$ is Markov.

Proof. Since $X-Y-Z$ forms a Markov chain, $H_\alpha(X|Y) = H_\alpha(X|Y,Z)$. Now set $\phi(x|y,z) = p_\alpha(x|z)$ in (13) to obtain $H_\alpha(X|Y,Z) \le H_\alpha(X|Z)$, with equality iff $p_\alpha(x|Y,Z) = p_\alpha(x|Z)$ a.e., i.e., $X - Z - Y$ is Markov. (Alternate proof: see [7].)

Lemma 11 (Data processing inequalities for binary α-divergence). *Suppose $0 \le p \le q \le r \le 1$ (or that $1 \ge p \ge q \ge r \ge 0$). Then*

$$d_\alpha(p\|r) \ge d_\alpha(p\|q) \quad \textit{and} \quad d_\alpha(p\|r) \ge d_\alpha(q\|r). \tag{24}$$

Proof. Consider an arbitrary binary channel with with parameters $\delta, \epsilon \in [0,1]$ and transition matrix $P_{Y|X} = \left(\begin{smallmatrix} 1-\delta & \delta \\ \epsilon & 1-\epsilon \end{smallmatrix}\right)$. By Lemma 8 applied to $P_X = (1-p, p)$ and $Q_X = (1-q, q)$, one obtains $d_\alpha(p\|q) \ge d_\alpha(\delta(1-p)+(1-\epsilon)p \,\|\, \delta(1-q)+(1-\epsilon)q)$ for any $p, q \in [0,1]$. Specializing to values of δ, ϵ such that $\epsilon p = \delta(1-p)$ or $\epsilon q = \delta(1-q)$ gives (24).

3 Fano's Inequality Applied to a Side-Channel Attack

We now apply α-information theory to any Markov chain X–Y–$\hat{X}$, modeling a side-channel attack using leaked information in a secrecy system. Here X is a sensitive data that depends on some secret (cryptographic key or password), input to a side channel $P_X \to \boxed{P_{Y|X}} \to P_Y$ through which information leaks, and Y is disclosed to the attacker (e.g., using some sniffer or probe measurements), and the attack provides $\hat{X}$ as a function of Y estimating X using the maximum a posteriori probability (MAP) rule so as to maximize the probability of success $\mathbb{P}_s = \mathbb{P}_s(X|Y) = \mathbb{P}(\hat{X} = X \,|\, Y)$, or equivalently, minimize the probability of error $\mathbb{P}_e = \mathbb{P}_e(X|Y) = \mathbb{P}(\hat{X} \ne X \,|\, Y)$.

The classical Fano inequality [6] then writes $H(X|Y) \le h(\mathbb{P}_e) + \mathbb{P}_e \log(M-1)$ when $X \sim \mathcal{U}(M)$. It was generalized by Han and Verdú [8] as a lower bound on the mutual information $I(X;Y) \ge d(\mathbb{P}_s(X|Y)\|\mathbb{P}_s(X)) = d(\mathbb{P}_e(X|Y)\|\mathbb{P}_e(X))$ where $d(p\|q)$ denotes discrete divergence, and where $\mathbb{P}_s(X) = 1 - \mathbb{P}_e(X)$ corresponds to the case where X (possibly nonuniform) is guessed without even knowing Y. Using the MAP rule it can be easily seen that

$$\begin{cases} \mathbb{P}_s(X|Y) = \mathbb{E}\big[\max\limits_{x\in\mathcal{X}} p_X(x|Y)\big] = \exp\big(-H_\infty(X|Y)\big) \\ \mathbb{P}_s(X) = \sup_{x\in\mathcal{X}}\ p_X(x) = \exp\big(-H_\infty(X)\big). \end{cases} \tag{25}$$

We now generalize the (generalized) Fano inequality to any value of $\alpha > 0$.

Theorem 1 (Generalized Fano's Inequality for α-Information).

$$I_\alpha(X;Y) \ge d_\alpha(\mathbb{P}_s(X|Y)\|\mathbb{P}_s(X)) = d_\alpha(\mathbb{P}_e(X|Y)\|\mathbb{P}_e(X)) \tag{26}$$

Proof. By the data processing inequality for α-information (Lemma 9), $I_\alpha(X;Y) \ge I_\alpha(X;\hat{X})$. Then by (18), $I_\alpha(X;\hat{X}) = D_\alpha(P_{X,\hat{X}}\|P_X Q^*_{\hat{X}}) \ge d_\alpha(\mathbb{P}_s(X|Y)\|\mathbb{P}'_s)$ where we have used the data processing inequality for α-divergence (Lemma 8, inequality (21)) to the event $A = \{\hat{X} = X\}$. Here $\mathbb{P}(\hat{X} = X) = \mathbb{P}_s(X|Y)$ by definition and $\mathbb{P}'_s = \sum_x p_X(x) q^*_{\hat{X}}(x) \le \max_x p_X(x) = \mathbb{P}_s(X)$. Now by the binary data processing inequality (Lemma 11), $d_\alpha(\mathbb{P}_s(X|Y)\|\mathbb{P}'_s) \ge d_\alpha(\mathbb{P}_s(X|Y)\|\mathbb{P}_s(X))$.

Our main Theorem 1 states that α-information $I_\alpha(X;Y)$ bounds the gain $d_\alpha(\mathbb{P}_s(X|Y)\|\mathbb{P}_s(X))$ of any adversary in guessing secret X from disclosed measurements Y with success $\mathbb{P}_s(X|Y)$, compared with the adversary's prior belief without access to measurements, with lower success $\mathbb{P}_s(X)$. It additionally provides an implicit upper bound on $\mathbb{P}_s(X|Y)$ (or lower bound on $\mathbb{P}_e(X|Y)$) as a function of α-information—which can be loosened to obtain explicit bounds on success or error by further lower bounding the binary α-divergence.

Also, bounding α-information (by some "α-capacity" of the side channel) one can obtain bounds on the success of any possible attack for a given secrecy system based on some leakage model, in a similar fashion as what was made in the classical case $\alpha = 1$ in [2]. This is particularly interesting for the designer who needs to evaluate the robustness of a given implementation to any type of side-channel analysis, regardless of the type of attacker.

References

1. Arimoto, S.: Information measures and capacity of order α for discrete memoryless channels. In Topics in Information Theory, Proc. 2nd Colloq. Math. Societatis János Bolyai. North Holland, Keszthely, Hungary, 1975, **2**(16), 41–52 (1977)
2. de Chérisey, E., Guilley, S., Piantanida, P., Rioul, O.: Best information is most successful: Mutual information and success rate in side-channel analysis. IACR Transactions on Cryptographic Hardware and Embedded Systems (TCHES'19) **2019**(2), 49–79 (2019)
3. Cover, T., Thomas, J.: Elements of Information Theory. J. Wiley & Sons (2006)
4. Csiszár, I.: Generalized cutoff rates and Rényi's information measures. IEEE Transactions on Information Theory **41**(1), 26–34 (Jan 1995)
5. van Erven, T., Harremoës, P.: Rényi divergence and Kullback-Leibler divergence. IEEE Transactions on Information Theory **60**(7), 3797–3820 (Jul 2014)
6. Fano, R.M.: Class notes for course 6.574: Transmission of Information. MIT, Cambridge, MA (1952)
7. Fehr, S., Berens, S.: On the conditional Rényi entropy. IEEE Transactions on Information Theory **60**(11), 6801–6810 (Nov 2014)
8. Han, T.S., Verdú, S.: Generalizing the Fano inequality. IEEE Transactions on Information Theory **40**(4), 1247–1251 (Jul 1994)
9. Polyanskiy, Y., Verdú, S.: Arimoto channel coding converse and Rényi divergence. In: Proc. Forty-Eighth Annual Allerton Conference. Allerton House, UIUC, IL, 1327–1333 (Oct 2010)
10. Rényi, A.: On measures of entropy and information. In: Proc. 4th Berkeley Symp. Math. Stat. Prob., Univ. Calif. Press, Berkeley, CA. **1**, 547–561 (1961)
11. Rioul, O.: Information theoretic proofs of entropy power inequalities. IEEE Transactions on Information Theory **57**(1), 33–55 (Jan 2011)
12. Rioul, O.: Rényi entropy power inequalities via normal transport and rotation. Entropy **20**(9, 641), 1–17 (Sept 2018)
13. Sason, I., Verdú, S.: Arimoto-Rényi conditional entropy and Bayesian M-ary hypothesis testing. IEEE Transactions on Information Theory **64**(1), 4–25 (Jan 2018)
14. Sibson, R.: Information radius. Zeitschrift Wahrscheinlichkeitstheorie Verwandte Gebiete **14**, 149–160 (Jun 1969)
15. Verdú, S.: α-mutual information. In: Proc. Information Theory and Applications Workshop. La Jolla, CA (Feb 2015)