HAL
open science

Information theoretic comparison of side-channel distinguishers: Inter-class distance, confusion, and success

Annelie Heuser, Olivier Rioul, Sylvain Guilley, Jean-Luc Danger

To cite this version:

Annelie Heuser, Olivier Rioul, Sylvain Guilley, Jean-Luc Danger. Information theoretic comparison of side-channel distinguishers: Inter-class distance, confusion, and success. Bernard Candaele; Dimitrios Soudris; Iraklis Anagnostopoulos. Trusted Computing for European Embedded Systems, Springer, pp.187-225, 2015, 978-3-319-09419-9. 10.1007/978-3-319-09420-5_10 . hal-02286944

HAL Id: hal-02286944

https://telecom-paris.hal.science/hal-02286944v1

Submitted on 21 Aug 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.

Information Theoretic Comparison of Side-channel Distinguishers

Inter-class Distance, Confusion, and Success

Annelie Heuser, Olivier Rioul, Sylvain Guilley and Jean-Luc Danger

1 Side-channel Analysis

Side-channel analysis (SCA) constitutes a serious threat against modern cryptographic implementations. They exploit unintentionally emitted physical leakage—such as power consumption or electromagnetic emanation—in order to reveal secret information. The introduction of differential power analysis by Kocher *et al.* [15] gave rise to many developments of new attacks, countermeasures and models for the evaluation of physical security. In particular, a large variety of *distinguishers* have been proposed as statistical tests in order to discriminate the correct key. To overcome limitations such as the restriction to linear dependency between the leakage and the assumed leakage model, new types of distinguishers have been proposed.

First, mutual information analysis (MIA) was proposed by Gierlichs *et al.* [10]. It uses mutual information (MI) to measure the total dependency between the measurements and the leakage model. Extensive previous work [2, 17, 39, 23, 24, 21, 41] has shown that this distinguisher is indeed able to cope with non linearities between the leakage model and the measurements.

Second, to avoid explicit density estimations as required for MIA, the Kolmogorov-Smirnov (KS) test was proposed by Veyrat-Charvillon and Standaert [39] and the

Annelie Heuser
Telecom ParisTech, Institut Mines-Telecom, CNRS LTCI, Department Comelec,46 rue Barrault, 75 634 Paris Cedex 13, FRANCE, e-mail: annelie.heuser@telecom-paristech.fr

Olivier Rioul
Telecom ParisTech, Institut Mines-Telecom, CNRS LTCI, Department Comelec, 46 rue Barrault, 75 634 Paris Cedex 13, FRANCE, e-mail: olivier.rioul@telecom-paristech.fr

Sylvain Guilley
Telecom ParisTech, Institut Mines-Telecom, CNRS LTCI, Department Comelec, 46 rue Barrault, 75 634 Paris Cedex 13, FRANCE, e-mail: sylvain.guilley@telecom-paristech.fr

Jean-Luc Danger
Telecom ParisTech, Institut Mines-Telecom, CNRS LTCI, Department Comelec, 46 rue Barrault, 75 634 Paris Cedex 13, FRANCE, e-mail: jean-luc.danger@telecom-paristech.fr

corresponding Kolmogorov-Smirnov analysis (KSA) was studied in [41, 43, 45]. Although it has been highlighted in [43] that KSA may have disadvantages compared to MIA, a recent study [45] has identified variants of KSA that may perform better than MIA in some circumstances.

In [18], the authors suggested an alternative *inter-class* Kolmogorov-Smirnov analysis (IKSA) that compares the conditional distributions between themselves instead of comparing them with the global distribution of the leakage. This novel approach is shown to be of a different nature (non equivalent), and can outperform KSA in terms of success rate.

Similar ideas have also emerged in the literature: The single-bit DPA [16] can already be seen as a comparison of (means of) different classes without referring to the marginal distribution. Moreover, in [1] a cluster approach has been introduced that compares the inter- and intra-class variance of conditional classes. Also, in [40] a copula-based distinguisher has been introduced that compares each conditional distribution internally without referring directly to the global leakage distribution.

It is important to note that in general, a distinguisher's performance also depends on the choice of the leakage model. As pointed out in [44] a distinguisher would fail to distinguish if the model consists of a bijective function of the secret and plaintext. Therefore, in this paper, we restrict ourselves to leakage models for which the studied distinguishers are able to distinguish.

Because so many types of side-channel distinguishers have become available, their fair evaluation and comparison is an important topic. One cannot rely on one single experiment carried out on raw leakage measurements from one particular device to draw unequivocal conclusions about the relative efficiency of competing distinguishers (see e.g., the discussion in [34]). Therefore, we seek to compare statistical procedures and methodologies in *ideal* scenarios with clearly defined and fixed leakage models, where in particular the signal-to-noise ratio can be varied as a parameter.

Now, there has been two distinct evaluation frameworks considered in the literature so far:

1. *A theoretical framework* proposed by Whitnall and Oswald [41] that uses the exact values of the distinguishers to evaluate the capability to recover the correct key hypothesis. One relevant metric is the so-called *relative margin*, that computes a normalized distance between the distinguisher's value for the correct key guess to that of its nearest rival.
2. *An empirical framework* proposed by Standaert *et al.* [35] in which the distinguishers are estimated on empirical data. The performance evaluation can be typically carried out using one of the following two metrics: the *success rate*, which estimates the probability of ranking the correct hypothesis first, and the *guessing entropy,* which estimates the average ranking of the correct hypothesis.

It should be emphasized that the theoretical framework is based on the *exact* computation of the distinguisher to evaluate its intrinsic distinguishing power—as if it was estimated on a infinite number of samples. In contrast, the empirical framework uses simulations or measurements to evaluate the ability of a distinguisher to succeed

efficiently: it depends not only on the choice of the theoretical distinguisher, but also on the efficiency of its estimation. Roughly speaking, it can be said that the empirical framework encompasses the theoretical one plus the estimation algorithm. For this reason, it appears to be more practical. On the other hand, the theoretical framework is more amenable to a mathematical analysis, since it only involves the distinguisher's values. So far, no link between the theoretical and empirical outcomes of a given distinguisher has been shown in the literature.

1.1 Our Contributions

1.1.1 Interclass Distinguisher

As a first contribution we introduce a new information-theoretic metric, referred to as *inter-class information*, that compares conditional probability density functions between themselves. Before applying it to side-channel analysis, we first carry out a detailed mathematical study on the metric itself. In particular, we show that inter-class information (II) shares similar properties with mutual information (MI). Interestingly, both can be computed from the *same* probability density estimates. But we also prove that the two metrics are *not equivalent* with a precise definition of the term.

Next, we extend the inter-class information to the scenario of side-channel analysis and refer to the corresponding distinguisher as *inter-class information analysis* (IIA). We continue our mathematical investigation by proving *soundness* of IIA. Finally, we use the above-mentioned frameworks to investigate the efficiency of both MIA and IIA. From the theoretical framework we select the *relative distinguishing margin* as the relevant metric. From the empiral framework we select the *success rate* as the relevant metric. The results from both frameworks agree: IIA is shown to outperform MIA for the theoretical *and* empirical metric.

1.1.2 Success Metric

Second, we introduce a new metric, called *success metric* (SM), which evaluates estimated distinguishers while providing more feedback about the efficiency. Therefore, the SM is more suitable when comparing and evaluating distinguishers than the currently state-of-the-art. In fact, SM relies on the estimation parameters of the distinguisher affecting the theoretical success rate. To be precise, the key features of the success metric are:

- Monotony with the success rate (theoretically and empirically);
- Quantification of the relationship between the distinguishing value of the correct key and its nearest rival;
- Consideration of the noise probability distribution function (e.g., its variance), number of measurements, and estimation method

- Simplicity of the closed-form expressions for additive distinguisher (e.g., DPA, CPA) compared to the success rate;
- Ability to derive a closed-form expression for MIA when estimated with histograms, which has not been shown for any other metric before.

Furthermore, we show further benefits of the closed-form expression of SM: We are able to connect the closed-form of the success metric for DPA/ CPA with properties of the sbox in case of a practical signal-to-noise ratio. Remarkably, unlike previous works [11, 22] we first not only derive bounds but achieve direct links, and second utilize a success rate/metric instead of only using properties of a distinguisher. However, our new metric, *transparency metric*, follows the same intuition as the transparency order introduced in [22], but is more reasonable and simple. Additionally, we are able to answer the question how the size of the keyspace is influencing the success metric and therefore the success rate.

1.2 Side-channel Analysis Model

Calligraphic letters (e.g., $\mathscr{X}$) denote finite sets, capital letters (e.g., X) denote random variables taking values in these sets, and the corresponding lowercase letters (e.g., x) denote their realizations. We write $\mathbb{P}\{X = x\}$ or $p(x)$ for the probability that $X = x$ and $p(x|y) = \mathbb{P}\{X = x \,\big|\, Y = y\}$ for conditional probabilities.

Let k^* denote the secret cryptographic key, k any possible key hypothesis from the keyspace $\mathscr{K}$, and let T be the input or cipher text of the cryptographic algorithm. The mapping $g : (\mathscr{T}, \mathscr{K}) \to \mathscr{I}$ maps the input or cipher text and a key hypothesis $k \in \mathscr{K}$ to an internally processed variable in some space $\mathscr{I}$ that is assumed to relate to the leakage X. Usually, $\mathscr{T}, \mathscr{K}, \mathscr{I}$ are taken as $\mathbb{F}_2^n$, where n is the number of bits (for AES $n = 8$).

Generally it is assumed that f is known to the attacker. A common consideration is $g(T,k) = \texttt{Sbox}[T \oplus k]$ where `Sbox` is a substitution box. The measured leakage X can then be written as

$$X = \psi(g(T,k^*)) + N, \tag{1}$$

where N denotes an independent additive noise. ψ is a device-specific deterministic function, which we assume to be known to the attacker in this contribution. For any key guess $k \in \mathscr{K}$ the attacker computes the *sensitive variable*

$$Y(k) = \psi(g(T,k)). \tag{2}$$

Without loss of generality we may assume that Y is centered and normalized, i.e., $\mathbb{E}\{Y\} = 0$ and $\mathrm{Var}\{Y\} = 1$, and that the values in $\mathscr{Y}$ are regularly spaced with step Δy. For ease of notation, we let $Y^* = Y(k^*)$ and $Y = Y(k)$.

2 A New Distinguisher Based on Intraclass Information

In this section, we introduce a new information-theoretic metric, referred to as *inter-class information*, that compares conditional probability density functions between themselves. Before applying it to side-channel analysis, we first carry out a detailed mathematical study on the metric itself. In particular, we show that inter-class information (II) shares similar properties with mutual information (MI). Interestingly, both can be computed from the *same* probability density estimates. But we also prove that the two metrics are *not equivalent* with a precise definition of the term.

Next, we extend the inter-class information to the scenario of side-channel analysis and refer to the corresponding distinguisher as *inter-class information analysis* (IIA). We continue our mathematical investigation by proving *soundness* of IIA. Finally, we use the above-mentioned frameworks to investigate the efficiency of both MIA and IIA.

We review some information-theoretic tools to evaluate the dependence between two random variables X and Y, and refer to [6] for more details. We focus in this section on metrics and postpone the application to side-channel analysis to Sect. 4. However, since for this application one random variable (X) is continuous and the other (Y) is discrete, we adopt this convention whenever it is possible.

Let $p(x)$ be the probability density function of the continuous random variable X and $p(y) = \mathbb{P}\{Y = y\}$ be the probability mass function of the discrete random variable Y. The corresponding *expectations* are $\mathbb{E}(X) = \int_{-\infty}^{\infty} x \cdot p(x)\,\mathrm{d}x$ and $\mathbb{E}(Y) = \sum_y y \cdot p(y)$, respectively. The *variance* is defined as $\sigma_X^2 = \mathbb{E}\{(X - \mathbb{E}(X))^2\}$, and similarly for Y. Let $p(x|y) = p(x|Y = y)$ be the conditional probability distribution of X knowing that $Y = y$ and $p(x,y)$ be the joint probability distribution of X and Y. Notice that the marginal distribution $p(x)$ becomes the average over Y of the conditional distribution $p(x|y)$:

$$p(x) = \sum_y p(x,y) = \sum_y p(y)p(x|y) = \mathbb{E}(p(x|Y)). \tag{3}$$

2.1 Information Divergence

Definition 1 (Kullback-Leibler divergence [6]). Let $q(x)$ be another probability distribution defined over the same space as $p(x)$. The *Kullback-Leibler divergence* of q with respect to p is defined as

$$D_{\mathsf{KL}}[p \parallel q] = \int_{-\infty}^{\infty} p(x) \cdot \log \frac{p(x)}{q(x)}\,\mathrm{d}x. \tag{4}$$

It is well known that $D_{\mathsf{KL}}[p \parallel q] \geq 0$ and equals zero if and only if $p(x)$ and $q(x)$ coincide. The divergence is sometimes termed "distance" in the literature although it is not a distance in the mathematical sense of the word, because it is not symmetric:

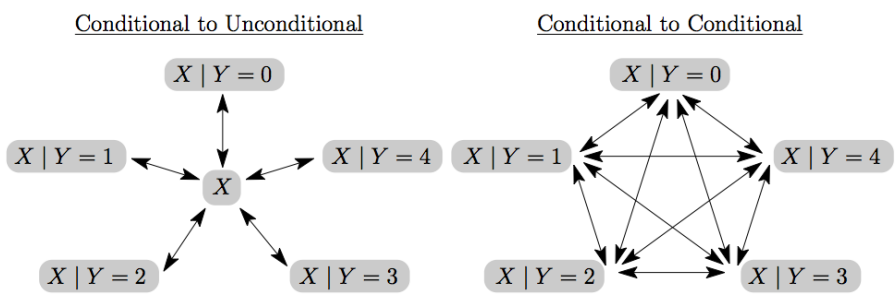
$D_{\mathsf{KL}}[p \parallel q] \neq D_{\mathsf{KL}}[q \parallel p]$ and the triangle inequality is not satisfied in general. To achieve symmetry, Kullback and Hajek made the following definition:

Definition 2 (Symmetric Kullback-Leibler divergence). The *symmetric divergence* between distributions p and q is defined as

$$\delta_{\mathsf{KL}}(p \parallel q) = \frac{D_{\mathsf{KL}}[p \parallel q] + D_{\mathsf{KL}}[q \parallel p]}{2} \tag{5}$$

$$= \frac{1}{2}\int_{-\infty}^{\infty} (p(x) - q(x)) \cdot \log \frac{p(x)}{q(x)} \, \mathrm{d}x. \tag{6}$$

2.2 Conditional-to-Unconditional Metric



(a) Conditional vs Unconditional. (b) Conditional vs Conditional.

Fig. 1: Illustration of comparing probability distributions (the "distance" is depicted with an arrow).

To evaluate the dependence between X and Y, one possibility is to compute the distance between conditional probabilities $p(x|y)$ and the unconditional probability $p(x) = \mathbb{E}(p(x|Y))$ (see Fig. 1a). Using Kullback-Leibler divergence, we obtain

$$I(X;Y) = \mathbb{E}\{D_{\mathsf{KL}}[p(x|Y) \parallel p(x)]\} \tag{7}$$

$$= \sum_y p(y) D_{\mathsf{KL}}[p(x|y) \parallel p(x)] \tag{8}$$

$$= \sum_y \int_{-\infty}^{\infty} p(x,y) \cdot \log \frac{p(x|y)}{p(x)} \, \mathrm{d}x. \tag{9}$$

This is well-known as the *mutual information* between the two random variables X and Y. Mutual information can also be written as

$$I(X;Y) = H(X) - H(X|Y) \tag{10}$$

where

$$H(X) = -\int_{-\infty}^{\infty} p(x) \cdot \log p(x)\,\mathrm{d}x \tag{11}$$

is the (differential) *entropy* of X and

$$H(X|Y) = \sum_y p(y) \cdot H(X|Y=y) \tag{12}$$

$$= -\sum_y \int_{-\infty}^{\infty} p(x,y) \cdot \log p(x|y)\,\mathrm{d}x \tag{13}$$

is the *conditional entropy* of X knowing Y. Note that unlike the (discrete) entropy [6], differential entropy can be negative and hence should *not* be interpreted as a measure of uncertainty[1]. For more details on the relationship between differential and discrete entropy and the absolute entropy we refer to [6].

2.3 Conditional-to-Conditional Metric

As suggested in [18], instead of referring to the average distribution $p(x)$, a more direct strategy would be to consider *all* pairwise distances between conditional probabilities $p(x|y)$ (see Fig. 1b). Therefore, we may replace the Kullback-Leibler divergence of $p(x|y)$ with respect to the average distribution $p(x) = \mathbb{E}(p(x|Y))$ by all Kullback-Leibler divergences between conditional probabilities $p(X|Y=y)$ and $p(X|Y=y')$ for all pairs (y,y'). This yields to the following definition.

Definition 3 (Inter-class information). The *inter-class information* between random variables X and Y is defined as

$$II(X;Y) = \frac{1}{2}\,\mathbb{E}\big\{D_{\mathsf{KL}}[p(x|Y=y) \parallel p(x|Y=y')]\big\} \tag{14}$$

$$= \frac{1}{2}\sum_{y \neq y'} p(y)p(y')D_{\mathsf{KL}}[p(x|y) \parallel p(x|y')] \tag{15}$$

where the summation over $y = y'$ has disappeared because divergence vanishes for identical distributions.

Proposition 1. *The inter-class information can also be written in terms of the symmetric Kullback-Leibler divergence as*

$$II(X;Y) = \mathbb{E}\{\delta_{\mathsf{KL}}(p(x|Y) \parallel p(x))\} \tag{16}$$

$$= \tfrac{1}{2}\sum_y \int_{-\infty}^{\infty} (p(x,y) - p(x)p(y)) \log \frac{p(x,y)}{p(x)p(y)}\,\mathrm{d}x. \tag{17}$$

[1] Another reason is that differential entropy is not "coordinate-free" - it depends on the underlying coordinate system.

Proof. We show equivalence between Eq. (15) and Eq. (16).

$$
\begin{aligned}
&\frac{1}{2}\sum_{y\neq y'} p(y)p(y')D_{\mathsf{KL}}[p(x|y) \parallel p(x|y')] \\
&\quad = \frac{1}{2}\sum_{y,y'} p(y)p(y')\int p(x|y)\log\frac{p(x|y)}{p(x|y')}\,\mathrm{d}x && (18)\\
&\quad = \frac{1}{2}\sum_{y}\int\sum_{y'} p(y)p(y')p(x|y)\log\frac{p(x|y)}{p(x)}\,\mathrm{d}x \\
&\qquad + \frac{1}{2}\sum_{y'}\int\sum_{y} p(y')p(y)p(x|y)\log\frac{p(x)}{p(x|y')}\,\mathrm{d}x && (19)\\
&\quad = \frac{1}{2}\sum_{y} p(y)\int p(x|y)\log\frac{p(x|y)}{p(x)}\,\mathrm{d}x \\
&\qquad + \frac{1}{2}\sum_{y'} p(y')\int p(x)\log\frac{p(x)}{p(x|y')}\,\mathrm{d}x && (20)\\
&\quad = \frac{1}{2}\big(\mathbb{E}\{D_{\mathsf{KL}}[p(x|Y) \parallel p(x)]\} + \mathbb{E}\{D_{\mathsf{KL}}[p(x) \parallel p(x|Y)]\}\big) && (21)\\
&\quad = \mathbb{E}\big\{\delta_{\mathsf{KL}}(p(x|Y) \parallel p(x))\big\} && (22)
\end{aligned}
$$

Equation (17) then follows easily from definition 1. □

Interestingly, Eq. (16) is similar to Eq. (7) where the divergence (definition 1) is replaced by the symmetric divergence (definition 2). The latter is also sometimes referred to as *inter-class divergence* (see e.g., [31]).

Moreover, similarly as for mutual information, it can be expressed in terms of entropies as shown in the following proposition.

Proposition 2. *Let*

$$H'(X \mid Y) = -\sum_{y}\int_{-\infty}^{\infty} p(x)p(y)\cdot\log p(x|y)\,\mathrm{d}x, \tag{23}$$

be the conditional cross-entropy *of X knowing Y. The inter-class information can be expressed as*

$$II(X;Y) = \frac{H'(X \mid Y) - H(X|Y)}{2}. \tag{24}$$

Proof. We show the equivalence between Eq. (17) and Eq. (24). Since

$$\sum_{y} p(x,y) - p(x)p(y) = 0, \tag{25}$$

we can remove $p(x)$ inside the logarithm in (17). Furthermore, since $\frac{p(x,y)}{p(y)} = p(x|y)$, we can write

$$\frac{1}{2}\sum_{x,y}(p(x,y)-p(x)p(y))\log\frac{p(x,y)}{p(x)p(y)}$$

$$= \frac{1}{2}\sum_{x,y}(p(x,y)-p(x)p(y))\log p(x|y) \tag{26}$$

$$= \frac{H'(X \mid Y)-H(X|Y)}{2} \tag{27}$$

□

It is important to notice that cross-entropy is, contrary to Eq. (13), averaged over the product distribution $p(x)p(y)$ instead of the joint distribution $p(x|y)p(y) = p(x,y)$.

3 Theoretical Analysis

Inter-class information has some important properties that are similar to well-known properties of mutual information. These are summarized in the following proposition.

Proposition 3. *For any two random variables X,Y:*

(a) (Symmetry) $II(X;Y) = II(Y;X)$
(b) (Independence) $II(X;Y) = 0$ *if and only if X, Y are independent*
(c) (Markov Chain Inequality) *For any Markov chain $X-Y-Z$, the following hold: $II(X;Y) \geq II(X;Z)$ and $II(Y;Z) \geq II(X;Z)$*
(d) (Relation to Mutual Information)

$$2II(X;Y) = \mathbb{E}\{D_{\mathsf{KL}}[p(x|Y) \parallel p(x)]\} + \mathbb{E}\{D_{\mathsf{KL}}[p(x) \parallel p(x|Y)]\} \tag{28}$$

$$= I(X;Y) + \mathbb{E}\{D_{\mathsf{KL}}[p(x) \parallel p(x|Y)]\} \tag{29}$$

It follows in particular that $II(X;Y) \geq \frac{1}{2}I(X;Y)$.

Proof. The symmetry is obvious from Eq. (17). Independency is an obvious consequence of the following well-known property of (symmetric) divergence: $D_{\mathsf{KL}}[p \parallel q] \geq 0$ and $D_{\mathsf{KL}}[p \parallel q] = 0$ if and only if $p = q$ [6]. Markov Chain Inequality: Recall that $X \to Y \to Z$ forms a Markov chain if $p(z|x,y) = p(z|y)$ for all x; in other words X and Z are independent given Y [6]. Since $X \to Y \to Z$ is a Markov chain if and only if $Z-Y-X$ is a Markov chain [6], it is sufficient to prove the first inequality $II(X;Y) \geq II(X;Z)$. Furthermore we already have $I(X;Y) \geq I(X;Z)$ from the corresponding property for mutual information. Since the latter is equivalent to the inequality $H(X|Y) \leq H(X|Z)$, thanks to Proposition 2, it is sufficient to prove the inequality $H'(X|Y) \geq H'(X|Z)$ for cross-entropies.

Now since $p(x|y) = p(x|y,z)$ by the Markov chain condition, it is easily checked that

$$H'(X|Y) = -\sum_{y,z} \int p(x)p(y,z) \log p(x|y,z)\,\mathrm{d}x = H'(X|Y,Z) \tag{30}$$

which can be rewritten as

$$H'(X|Y,Z) = \sum_{z} \int p(x)p(z) \sum_{y} p(y|z) \log \frac{1}{p(x|y,z)}\,\mathrm{d}x. \tag{31}$$

By the strict concavity of the logarithm, we have the following inequality

$$H'(X|Y,Z) \geq \sum_{z} \int p(x)p(z) \log \frac{1}{\sum_y p(y|z)p(x|y,z)}\,\mathrm{d}x \tag{32}$$

$$= \sum_{z} \int p(x)p(z) \log \frac{1}{p(x|z)}\,\mathrm{d}x = H'(X|Z) \tag{33}$$

Finally, the relation to mutual information is obvious from the definition. □

3.1 A Normal Example

In order to illustrate the difference between MI and II, we give the exact expression of $I(X;Y)$ and $II(X;Y)$ for two jointly normal random variables[2].

Proposition 4. *Let the two random variables X,Y be identically distributed, zero-mean and jointly normal, with covariance matrix $\sigma^2\left(\begin{smallmatrix}1 & \rho\\ \rho & 1\end{smallmatrix}\right)$, where $|\rho| \leq 1$ is the correlation coefficient and σ^2 is the common variance of X and Y. One finds*

$$I(X;Y) = \frac{1}{2}\log\frac{1}{1-\rho^2} \tag{34}$$

$$II(X;Y) = \frac{\log e}{2}\frac{\rho^2}{1-\rho^2}. \tag{35}$$

Proof. Since X follows the normal density $\mathcal{N}(0,\sigma^2)$, its differential entropy is easily computed as [6]

$$H(X) = -\mathbb{E}\{\log p(X)\} \tag{36}$$

$$= \log\sqrt{2\pi\sigma^2} + (\log e)\mathbb{E}\{X^2/2\sigma^2\} \tag{37}$$

$$= \frac{1}{2}\log(2\pi e\sigma^2). \tag{38}$$

Now for every y, X given $Y=y$ follows the density $p(x|y) = \frac{p(x,y)}{p(y)}$ which is easily seen to be the normal $\mathcal{N}(\rho y, \sigma^2(1-\rho^2))$. It follows that

[2] Note that, unlike in our previous definitions, the random variable Y is also continuous in this example. Thus sums have to be replaced by integrals.

$$H(X|Y) = \frac{1}{2}\log(2\pi e\sigma^2(1-\rho^2)). \tag{39}$$

Subtracting Eq. (39) from Eq. (38) yields the announced expression for $I(X;Y) = H(X) - H(X|Y)$.

To calculate inter-class information, we use Eq. (24). The conditional cross-entropy can be similarly computed as

$$H'(X|Y) = -\int_{-\infty}^{\infty} p(y)\cdot\mathbb{E}\{\log p(X|y)\}\,\mathrm{d}y \tag{40}$$

$$= \frac{1}{2}\log(2\pi\sigma^2(1-\rho^2)) + (\log e)\int_{-\infty}^{\infty} p(y)\cdot\mathbb{E}\Big\{\frac{(X-\rho y)^2}{2\sigma^2(1-\rho^2)}\Big\}\,\mathrm{d}y. \tag{41}$$

Using (39) and expanding $\mathbb{E}\{(X-\rho y)^2\} = \mathbb{E}(X^2) + \rho^2 y^2 - 0$ inside the integral, we obtain

$$H'(X|Y) = \Big(H(X|Y) - \frac{\log e}{2}\Big) + (\log e)\frac{\sigma^2 + \rho^2\mathbb{E}\{Y^2\}}{2\sigma^2(1-\rho^2)} \tag{42}$$

$$= H(X|Y) + (\log e)\Big(\frac{\sigma^2+\rho^2\sigma^2}{2\sigma^2(1-\rho^2)} - \frac{1}{2}\Big) \tag{43}$$

$$= H(X|Y) + (\log e)\frac{\rho^2}{1-\rho^2} \tag{44}$$

Subtracting $H(X|Y)$ and dividing by 2 yields the desired expression for $II(X;Y) = \frac{1}{2}(H'(X|Y) - H(X|Y))$. □

The limit case $\rho = 0$ corresponds to independent random variables X, Y in this example, while $\rho = 1$ corresponds to total dependency. From proposition 4, both mutual and inter-class informations vanish when $\rho = 0$ in accordance with proposition 3 (b). However, when $\rho \to 1^-$, $II(X;Y)$ is increasing to infinity much faster than $I(X;Y)$. This shows that $II(X;Y)$ is more sensitive in the dependency of the random variables. We found that this behavior is quite general for many probability distributions including the case of discrete random variables. This gives a first intuition, confirmed in the next section, why II may be more efficient than MI as a side-channel distinguisher.

3.2 Non-Equivalence of Mutual and Inter-Class Informations

Since $I(X;Y)$ and $II(X;Y)$ share similar properties (see proposition 3 (a)-(c)), and since we aim to compare these two informations as side-channel distinguishers to measure dependency between the measurements and the leakage model, it is important to assert generally whether $I(X;Y)$ and $II(X;Y)$ are equivalent or not. Although this does not reflect the ability to distinguish in the context of side-channel

analysis, it would give a necessary condition whether $II(X;Y)$ could be applicable. For this we need a clear definition of equivalent metrics (see e.g., [30]).

Definition 4 (Equivalence). Two distances $\mathscr{D}(p,q)$ and $\mathscr{D}'(p,q)$ are said to be *equivalent* if there exist finite constants $\alpha > 0$ and $\beta > 0$ such that for any p,q,

$$\mathscr{D}(p,q) \;\leq\; \alpha \cdot \mathscr{D}'(p,q) \text{ and } \mathscr{D}'(p,q) \;\leq\; \beta \cdot \mathscr{D}(p,q). \tag{45}$$

In particular, whenever one of two distances becomes small, so does the other and mathematically speaking, both "distances" define the same "topology"[3].

Just to illustrate the usefulness of Definition 4 we provide the following example.

Example 1. Consider the linear correlation coefficient

$$\rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y} \tag{46}$$

versus mutual information $I(X;Y)$. Although correlation implies dependence, it is possible that X and Y are linearly uncorrelated while still being dependent—take e.g., $Y = X^2$ where $X \sim \mathscr{N}(0,1)$. It follows that an inequality of the form $I(X;Y) \leq \alpha \cdot \rho(X,Y)$ *cannot* hold. Therefore, correlation and mutual information are *not* equivalent. The same conclusion goes unchanged if linear correlation is replaced by higher-order or nonlinear correlation—take e.g. $X \sim \mathscr{N}(0,1)$ and $Y = \pm X$ where the random sign is uniformly distributed and independent of X. This explains why correlation power analysis (CPA) and MIA are not equivalent.

Regarding IIA vs. MIA, proposition 3 (d) shows the inequality in one direction: $I(X;Y) \leq 2 \cdot II(X;Y)$. However, we have the following.

Proposition 5. *Mutual information $I(X;Y)$ and inter-class information $II(X;Y)$ are* not *equivalent.*

Proof. It is sufficient to give the following counterexample. Consider X,Y as in § 3.1. Letting $\lambda = \frac{1}{1-\rho^2}$ we have

$$2I(X;Y) = \log\lambda \quad \text{and} \quad 2II(X;Y) = (\lambda - 1)\log e. \tag{47}$$

Because the fraction $\frac{\lambda-1}{\log\lambda}$ is unbounded as $\lambda \to \infty$, letting $\rho \to 1^-$ shows that *no* inequality of the form $II(X;Y) \leq \alpha \cdot I(X;Y)$ may hold for some finite constant $\alpha > 0$. □

The fact that mutual and inter-class informations are *not* equivalent and at the same time require the estimations of the *same* conditional probability distributions $p(x|y)$ for their computation motivates for a formal comparison study in the context of side-channel analysis. This is investigated in the next section.

[3] Note that this equivalence of metrics is not the same as the equivalence between distinguishers stated in [7].

4 Side-Channel Analysis Scenario and Soundness

4.1 Side-Channel Scenario

There exists some necessary conditions on $Y(k)$ for MIA—and hence IIA—to be able to distinguish. In particular, [23, 44] show that there should be at least one $k \in \mathscr{K}$ such that $Y(k)$ is not an injective function of Z. Hence, if for all k, $f(\cdot,k)$ is injective the attacker has to choose φ to be non-injective. In the following, we assume that these necessary conditions are satisfied. As in [23, 24] we deduce the following scenario for wrong or correct key assumptions.

4.1.1 Wrong Key Assumption

The conditional distribution $p(x|y)$ of the measured leakage X knowing the predicted leakage Y is given by

$$p(x|y) = \sum_{y^*} p(y^*|y) \cdot p(x|y,y^*) \tag{48}$$

$$= \sum_{y^*} p(y^*|y) \cdot p(x-y^*|y) \tag{49}$$

$$= \sum_{y^*} p(y^*|y) \cdot p_N(x-y^*), \tag{50}$$

where p_N denotes the noise pdf and Eq. (48) follows from the law of total probability. The equivalence between Eq. (49) and Eq. (50) follows from the fact that N is independent of the leakage predictions Y. Thus, as proved in [25], if the key guess is incorrect we have a *nontrivial* linear mixture of shifted noise densities, whose coefficients depend on the relationship between Y and Y^*.

4.1.2 Correct Key Assumption

In contrast, if the key guess is correct, one obtains a Kronecker symbol $p(y^*|y) = \delta_{y,y^*}$ so that the density mixture simplifies to

$$p(x|y) = p_N(x-y^*), \tag{51}$$

which is simply identically distributed as $N + y^*$.

4.2 Soundness Proofs

Recall the following definition.

Definition 5 (Soundness). A given distinguisher $\mathscr{D}$ is said to be *sound* if the value of the distinguisher for the correct key k^* is strictly greater than for all other keys $k \neq k^*$:

$$\mathscr{D}(k^*) > \mathscr{D}(k) \qquad (\forall k \neq k^*) \tag{52}$$

Under this condition, it is an easy consequence of the law of large numbers that the corresponding success rate tends to 1 as the number of measurements increases indefinitely. For mutual information used as a side-channel distinguisher [10]: $\mathscr{D}(k) = I(X;Y(k))$, the soundness condition is expressed by the strict inequality $I(X;Y^*) > I(X;Y)$ for all $k \neq k^*$.

Proposition 6. *Mutual information analysis is sound for arbitrary (not necessarily Gaussian) noise.*

Proof. Moradi *et al.* [21] proved that $I(X;Y^*) \geq I(X;Y)$ which relies on the fact that $Y \to Y^* \to X$ forms a Markov chain [6, Thm 2.8.1]. Their paper [21] was written (as the title states) "under a Gaussian [noise] assumption" but the argument goes unchanged for non-Gaussian noise; in fact, the Markov chain condition $p(x|y,y^*) = p(x|y^*)$ relies only on the fact that N and Y are independent and not on the Gaussian nature of the noise.

To prove strict inequality, we use the fact that X given $Y = y$ is a *nontrivial* linear mixture of densities $p_N(x - y^*)$ of the same entropy as $H(N)$. Since the entropy is *strictly* concave in the probability density function [6, Thm 2.7.3][4] we have the strict inequality

$$H(X \mid Y = y) > \sum_{y^*} p(y^*|y) H(N + y^*) = H(N) \tag{53}$$

for all y. Taking expectations over Y yields $H(X|Y) > H(N) = H(X|Y^*)$, that is, $I(X;Y^*) > I(X;Y)$. □

For inter-class information used as a side-channel distinguisher: $\mathscr{D}(k) = II(X;Y(k))$, soundness is similarly expressed by the strict inequality $II(X;Y^*) > II(X;Y)$ for all $k \neq k^*$.

Proposition 7. *IIA is sound for arbitrary noise.*

Proof. Let $k \neq k^*$. By strict concavity of the logarithm (or by strict convexity of function $x \mapsto \log(1/x)$):

[4] A well-known information-theoretic property commonly referred to as "mixing increases entropy".

$$
\begin{aligned}
H'(X \mid Y) = & \sum_{y,y'} p(y)p(y') \sum_{y'^*} p(y'^*|y') \\
& \times \int p_N(x-y'^*) \log \frac{1}{\sum_{y^*} p(y^*|y) p_N(x-y^*)} \, \mathrm{d}x && (54) \\
< & \sum_{y,y'} p(y)p(y') \sum_{y^*,y'^*} p(y'^*|y')p(y^*|y) \\
& \times \int p_N(x-y'^*) \log \frac{1}{p_N(x-y^*)} \, \mathrm{d}x && (55) \\
= & \sum_{y^*,y'^*} p(y'^*)p(y^*) \\
& \times \int p_N(x-y'^*) \log \frac{1}{p_N(x-y^*)} \, \mathrm{d}x && (56) \\
= & H'(X \mid Y^*). && (57)
\end{aligned}
$$

Now as in the proof of Proposition 6, we still have $H(X|Y) > H(X|Y^*)$. Combining the two strict inequalities yields

$$
\begin{aligned}
II(X;Y) = & \frac{H'(X \mid Y) - H(X|Y)}{2} && (58) \\
< & \frac{H'(X \mid Y^*) - H(X|Y^*)}{2} = II(X;Y^*), && (59)
\end{aligned}
$$

which is the required soundness statement for IIA. □

5 Why Inter-class Information Analysis is more Discriminating than Mutual Information Analysis

In this section, we theoretically compare MIA and IIA under a Gaussian noise assumption using the scenario and the hypothesis of Sect. 4. We start by a theoretical investigation of $I(X;Y^*)$ and $II(X;Y^*)$, which is then extended with the help of some numerical calculation to $I(X;Y)$ and $II(X;Y)$.

5.1 Theoretical Comparison of $I(X;Y^*)$ and $II(X;Y^*)$

A key feature of IIA is that inter-class information is no less than mutual information for the correct key guess[5].

[5] Interestingly, it is not true that $II(X;Y) \geq I(X;Y)$ for general random variables X and Y. For example, we can find a counterexample when X,Y are binary variables with small $p(x|y)$ for all $x,y \neq 0$.

Proposition 8. *Let X be as in Eq.* (1) *with Gaussian noise $N \sim \mathcal{N}(0,\sigma^2)$. One has*

$$II(X;Y^*) = \frac{\log e}{2} \cdot \frac{\sigma^2_{Y^*}}{\sigma^2} \tag{60}$$

and

$$I(X;Y^*) \leq II(X;Y^*)\ . \tag{61}$$

Proof. To proof Eq. (60) we evaluate $II(X;Y^*)$ using Eq. (24). Conditional cross-entropy can be written as

$$H'(X \mid Y) = \sum_y p(y) \int p(x) \log \frac{1}{p(x \mid y)}\, \mathrm{d}x. \tag{62}$$

Plugging the expressions $p(x) = \sum_y p(y)p(x|y)$ and $p(x|y) = \sum_{y^*} p(y^*|y)p_N(x-y^*)$ yields

$$H'(X \mid Y) = \sum_{y,y'} p(y)p(y') \sum_{y'^*} p(y'^*|y') \cdot \tag{63}$$

$$\int p_N(x-y'^*) \log \frac{1}{\sum_{y^*} p(y^*|y)p_N(x-y^*)}\, \mathrm{d}x. \tag{64}$$

For $k = k^*$ this boils down to

$$H'(X \mid Y^*) = \sum_{y^*,y'^*} p(y^*)p(y'^*) \underbrace{\int_x p_N(x-y'^*) \log \frac{1}{p_N(x-y^*)}\, \mathrm{d}x}_{(*)}. \tag{65}$$

Substituting $\xi = x - y'^*$ in $(*)$ and assuming $N \sim \mathcal{N}(0,\sigma^2)$ results in

$$\begin{aligned}
&\int p_N(\xi) \log \frac{1}{p_N(\xi + y'^* - y^*)}\, \mathrm{d}\xi \\
&\quad = \frac{1}{2}\log(2\pi\sigma^2) + \frac{\log(e)}{2\sigma^2}\mathbb{E}\big\{(N + y^* - y'^*)^2\big\} && (66)\\
&\quad = \frac{1}{2}\log(2\pi\sigma^2) + \frac{\log(e)}{2\sigma^2}\big(\sigma^2 + (y^* - y'^*)^2\big) && (67)\\
&\quad = H(N) + \frac{\log(e)}{2\sigma^2}(y^* - y'^*)^2. && (68)
\end{aligned}$$

So, by letting Y'^* denote a random variable independent and identically distributed as Y^*,

$$\begin{aligned}
H'(X \mid Y^*) &= H(N) + \frac{\log(e)}{2\sigma^2} \sum_{y^*,y'^*} p(y^*)p(y'^*)(y^* - y'^*)^2 && (69)\\
&= H(N) + \frac{\log(e)}{2\sigma^2}\mathbb{E}((Y^* - Y'^*)^2) && (70)
\end{aligned}$$

where

$$\mathbb{E}((Y^* - Y'^*)^2) = 2\mathbb{E}((Y^* - \mathbb{E}(Y^*))^2) \tag{71}$$
$$= 2\sigma^2_{Y^*}. \tag{72}$$

Combining using Eq. (24) and that fact that $H(X|Y^*) = H(N)$ for $k = k^*$ gives the announced formula:

$$II(X;Y^*) = \frac{H'(X \mid Y^*) - H(N)}{2} = \frac{\log e}{2} \cdot \frac{\sigma^2_{Y^*}}{\sigma^2}. \tag{73}$$

To prove Eq. (61) we use the fact that the differential entropy is maximum for normal densities [6]:

$$H(X) \leq \frac{1}{2}\log(2\pi e \sigma^2_X) \tag{74}$$

Since X given Y^* is normal, we obtain

$$I(X;Y^*) = H(X) - H(X|Y^*) \tag{75}$$
$$\leq \frac{1}{2}\log(2\pi e \sigma^2_X) - \frac{1}{2}\log(2\pi e \sigma^2_{X|Y^*}) \tag{76}$$
$$= \frac{1}{2}\log\frac{\sigma^2_X}{\sigma^2_{X|Y^*}} \tag{77}$$
$$= \frac{1}{2}\log\frac{\sigma^2_{Y^*} + \sigma^2}{\sigma^2} \tag{78}$$
$$\leq \frac{\log e}{2}\frac{\sigma^2_{Y^*}}{\sigma^2} = II(X;Y^*) \tag{79}$$

where we have used the well-known inequality $\log x \leq (\log e)(x-1)$. □

5.2 Distinguishability of $I(X;Y)$ and $II(X;Y)$

We now investigate the ability to distinguish between the correct key k^* and the incorrect keys $k \neq k^*$ for MIA and for IIA. For this purpose, we use the theoretical metric given by the *relative distinguishing margin* introduced in the SCA evaluation framework in [41] and defined by

$$\texttt{RelMarg}(\mathscr{D}) = \frac{\mathscr{D}(k^*) - \max_{k \neq k^*} \mathscr{D}(k)}{\sqrt{\mathrm{Var}\,\mathscr{D}(K)}}. \tag{80}$$

where K is the random variable uniformy distributed in the keyspace $\mathscr{K}$.

The theoretical evaluation for both MIA and IIA involves the determination of the Gaussian density mixture of the leakage X given each possible input Z, with mean value y^* and variance σ^2. That of the conditional densities of $p(x|y)$ follow similarly

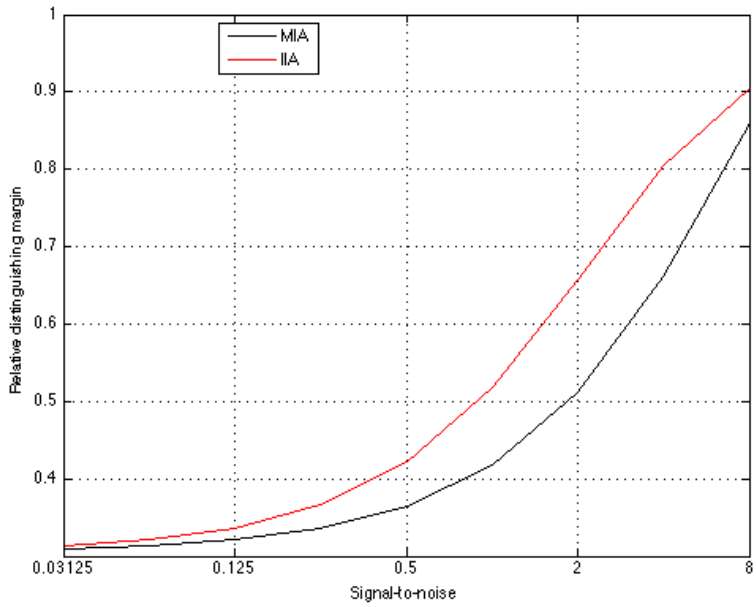


Fig. 2: Relative distinguishing margin for MIA (black) and IIA (red) for various SNRs

for all possible values of y. Given the expressions for $p(x)$ and $p(x|y)$, we are able to compute the required entropies given in Eq. (11), Eq. (13) and Eq. (23) with the help of numerical integration with arbitrary precision. To compute Eq. (80) we have chosen the following practical side-channel scenario:

$$Y(k) = HW(\mathtt{SBox}_P^{-1}[Z \oplus k^*]) \tag{81}$$

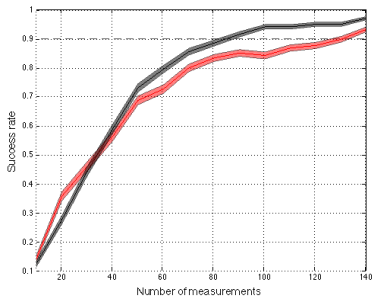
$$X = Y(k^*) + N, \tag{82}$$

where $\mathtt{SBox}_P^{-1}$ is the inverse substitution box operation in PRESENT ($\mathbb{F}_2^4 \to \mathbb{F}_2^4$), HW is the Hamming weight, and $N \sim \mathcal{N}(0, \sigma^2)$.

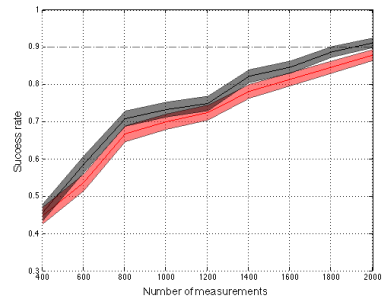
Figure 2 displays the relative distinguishing margin for various signal-to-noise ratios (SNR), defined as

$$\mathrm{SNR} = \frac{\mathrm{Var}(Y^*)}{\mathrm{Var}(N)} = \frac{2}{\sigma^2}. \tag{83}$$

It is clearly observed that `RelMarg`(IIA) lies essentially *above* `RelMarg`(MIA) for high SNR while at smaller SNR the two curves tend to the same asymptote.



(a) Success rate for MIA (red) and IIA (black) with error bars using $\sigma = 1$.



(b) Success rate for MIA (red) and IIA (black) with error bars using $\sigma = 4$.

6 Simulation Results

In order to compare the practical and theoretical evaluations, we consider the same leakage scenario as before (Eq. (81) and Eq. (82)). Again $N \sim \mathcal{N}(0, \sigma^2)$ with $\sigma = \{1, 4\}$ in our simulations. Although the assumption of additive white Gaussian noise may not be always realistic, it is common in numerous works in the community.

The maximum distinguisher's value gives the key prediction $\widehat{k}^*$, *viz.*,

$$\widehat{k}^* = \arg\max_k I(X;Y) \text{ or } \widehat{k}^* = \arg\max_k II(X;Y). \tag{84}$$

To compare the performance of MIA and IIA empirically we used the first-order *success rate* (SR), which we computed over a set of 230 independent experiments for $\sigma = 1$ and 120 experiments for $\sigma = 4$, where the secret key is chosen randomly for each experiment. In order to guarantee a fair comparison, we choose the same data set for both MIA and IIA.

We used the kernel density estimation to estimate the required probability densities. The parameters were chosen as recommended in previous publications (see e.g., [39, 24, 2]). To be specific, the bandwidth was chosen according to *normal scale rule* [32] and we used the *normal* kernel.

Moreover as suggested in [18], we highlight the standard deviation of the SR by computing *error bars*. More precisely, since SR follows a binomial distribution for multiple retries R with variance $\sqrt{\frac{\mathrm{SR}(1-\mathrm{SR})}{R}}$, we obtain confidence intervals

$$\left[\mathrm{SR} - \sqrt{\frac{\mathrm{SR}(1-\mathrm{SR})}{R}}, \mathrm{SR} + \sqrt{\frac{\mathrm{SR}(1-\mathrm{SR})}{R}}\right]$$

that are drawn as error bars to provide a fair comparison.

Figures 3a shows the success rate with error bars for $\sigma = 1$. One can see that IIA reaches the threshold of the SR of 0.9 before MIA. The success rate for $\sigma = 4$ is

displayed in Figures 3b, which again highlights the same classification for MIA and IIA. Interestingly, one can see that the difference between MIA and IIA is smaller for low SNR than for high SNR. Thus, the empirical results confirm our theoretical results and mathematical study made in the previous sections.

7 Comparing Side-Channel Distinguishers

7.1 Existing Evaluation Metrics

7.1.1 Comparing Empirical Distinguishers

The success rate (SR) is a classical evaluation metric when comparing empirical side-channel distinguishers $\widehat{\mathscr{D}}_m(K)$. In most publications, SR is derived empirically as defined in Def. 6 (e.g. in [19, 18, 7]). Moreover, in [35] the authors tackled the essential question *how to compare two implementations?* or *how to compare two side-channel adversaries?* by presenting an empirical framework including the empirical success rate.

Definition 6 (*Empirical* success rate). Let $\hat{k} = \arg\max_k \widehat{\mathscr{D}}_m(K)$ denote the key guess maximizing the experimental distinguisher $\widehat{\mathscr{D}}_m(K)$ for one experiment and let $\hat{\mathbf{k}} = [\hat{k}_1, \ldots, \hat{k}_r]$ define a vector of key guesses of r independent experiments. Then the *empirical success rate* is defined as

$$\widehat{SR}(\widehat{\mathscr{D}}_m) = \frac{1}{r}\sum_{i=1}^{r} \mathbb{1}_{k^*=\hat{k}_i}. \tag{85}$$

Even if the empirical success rate directly describes the practical outcome of a distinguisher, the given feedback is very limited. In particular, it only outputs the average probability of success without revealing influencing factors or quantifying how close the outcome of the correct key to its rivals is.

Apart from comparing the empirical SR, contributions tackled the questions on determining the *theoretical* success rate of distinguishers:

Definition 7 (*Theoretical* success rate). The *theoretical success rate* is defined as

$$SR(\widehat{\mathscr{D}}_m) = \mathbb{P}\Big(\widehat{\mathscr{D}}_m(X;Y(k^*)) > \widehat{\mathscr{D}}_m(X;Y(k)) \quad (\forall k \neq k^*)\Big) \tag{86}$$

$$= \mathbb{P}\Big(\widehat{\Delta}_m(k^*,k) > 0 \quad (\forall k \neq k^*)\Big). \tag{87}$$

In [28] Rivain determined the theoretical[6] SR for CPA and Bayesian attacks. Recently in [8], Fei et al. provided a *closed-form expression* for the theoretical success rate of DPA. Interestingly, their approach consists in estimating the theoretical success rate depending on the relationship between the correct and incorrect key hypothesis (named as *confusion*), the number of measurements and the SNR. Following this approach, Thillard et al. [38] extended the idea of confusion coefficients to the general case and reformulated the theoretical success rate of [28]. Thus, it is possible to determine the success rate without the need of measurements or simulations. Even more, the influencing factors of the success rate as the number of measurements, SNR and the confusion due to the leakage model are determinable. Unfortunately, the computation of the closed-form is not straightforward as mentioned in [28] and it again gives no quantification of the goodness of the distinguisher. Further, up to now only closed-forms for DPA and CPA exists.

7.1.2 Comparing Theoretical Distinguishers

A different approach to classify the efficiency of side-channel distinguishers has been presented in [42]. The authors aim at characterizing the behavior of theoretic distinguishers $\mathscr{D}(K)$ instead of $\widehat{\mathscr{D}}_m(K)$. Thus, the distinguisher is provided with full information about the leakage distribution without the need of estimation. The framework overall consists in six metrics, however, the most common metric is the *relative distinguishing margin* (RDM) that has been used as a reference in [41, 43][7]:

Definition 8 (Relative distinguishing margin [42]). Let $\mathscr{D}(k^*)$ be the theoretical distinguishing value of the correct key and $\mathscr{D}(k)$ the theoretical distinguishing value of any incorrect key hypotheses, then the *relative distinguishing margin* RDM is defined as

$$\mathrm{RDM}(\mathscr{D}) = \frac{\mathscr{D}(k^*) - \max\limits_{k \neq k^*} \mathscr{D}(k)}{\sqrt{\mathrm{Var}(\mathscr{D}(K))}} = \min_{k \neq k^*} \frac{\mathscr{D}(k^*) - \mathscr{D}(k)}{\sqrt{\mathrm{Var}(\mathscr{D}(K))}}. \tag{88}$$

The RDM gives a quantified feedback about the margin between the correct key $\mathscr{D}(k^*)$ and its nearest rival, unfortunately, no link between the outcome of an empirical and a theoretical distinguisher has been shown so far. Apart from this, the denominator in Eq. (88) is highly dependent on the number of key hypothesis used. For example, $\sqrt{\mathrm{Var}(\mathscr{D}(K))}$ with $\mathscr{K} = \mathbb{F}_2^8$ (8-bit key hypothesis) will be smaller than for $\mathscr{K} = \mathbb{F}_2^4$ (4-bit key hypothesis) and so RDM will be smaller for smaller key spaces than vice versa, which does not seem intuitive and we prove in Subsect. 8.2 the contrary. Thus, it is *not* possible to make reasonable comparisons between different cryptographic algorithms or implementations.

[6] In [28] the term exact instead of theoretical is used.

[7] Note that, in some publications, the relative distinguishing margin is also called *nearest-rival distinguishing score*.

7.2 A Novel Approach to Compare Distinguishers

As pointed out above, both state-of-the art approaches, the SR and the RDM, have significant drawbacks, which shows the need of a new metric. Our aim is to develop a novel metric that on the one hand coincides with the empirical outcome of distinguishers, like the SR, but on the other hand gives more quantified feedback as the RDM. Our new metric, called *success metric*, captures the relevant parameters of the theoretical success rate. We provide all necessary approximations from the theoretical success rate to the success metric. In particular, we first define the failure rate as the contrary to the success rate to apply the union bound. Following, we give two different approximations identifying the same relevant influencing factors with different convergence rate and, finally, we utilize a first order approximation to achieve the success metric in Def. 11.

7.2.1 Theoretical Foundation

Complementary to the theoretical success rate (see Def. 7) we define:

Definition 9 (Failure rate). The *failure rate* is defined as

$$FR(\widehat{\mathscr{D}}_m) = 1 - SR(\widehat{\mathscr{D}}_m) = \mathbb{P}\big(\exists k \neq k^* \ / \ \widehat{\Delta}_m(k) \leq 0\big). \tag{89}$$

We first use the *union bound* (Boole's inequality) to achieve an upper bound of the failure rate:

$$\mathbb{P}\big(\exists k \neq k^* \ / \ \widehat{\Delta}_m(k) \leq 0\big) \leq \sum_{k \neq k^*} \mathbb{P}\big(\widehat{\Delta}_m(k) \leq 0\big). \tag{90}$$

Next, we give two different approximations that both indicate the same properties but with different convergence rates and pre consumptions.

Definition 10. Let $X \sim \mathscr{N}(0,1)$. The *Q-function* is defined as

$$Q(x) = \frac{1}{2\pi} \int_x^\infty e^{-t^2/2} \, \mathrm{d}t \tag{91}$$

$$= \mathbb{P}(X > x). \tag{92}$$

Under the assumption of $\widehat{\Delta}_m(k^*,k) \sim \mathscr{N}(\Delta(k^*,k), \mathrm{EV}(k^*,k))$ we use the Q-function to approximate $P\big(\widehat{\Delta}_m(k^*,k) \leq 0\big)$, i.e.,

$$\mathbb{P}\big(\widehat{\Delta}_m(k^*,k) \leq 0\big) \tag{93}$$

$$= \mathbb{P}\Big(\frac{\widehat{\Delta}_m(k^*,k) - \mathbb{E}(\widehat{\Delta}_m(k^*,k))}{\sqrt{\mathrm{EV}(k^*,k)}} \leq -\frac{(\Delta(k^*,k) + \mathrm{EB}(k^*,k))}{\sqrt{\mathrm{EV}(k^*,k)}}\Big) \tag{94}$$

$$= Q\Big(\frac{\Delta(k^*,k) + \mathrm{EB}(k^*,k)}{\sqrt{\mathrm{EV}(k^*,k)}}\Big), \tag{95}$$

since $Q(x) = 1 - Q(-x)$. Accordingly, if $\mathrm{EB}(k^*,k)$ is small with respect to $\Delta(k^*,k)$, we have

$$\mathbb{P}(\widehat{\Delta}_m(k^*,k) \leq 0) \longrightarrow 0 \tag{96}$$

exponentially as

$$\frac{\Delta(k^*,k) + \mathrm{EB}(k^*,k)}{\sqrt{\mathrm{EV}(k^*,k)}} \longrightarrow \infty \tag{97}$$

increases for large m. We recall the Chebyshev bound [37]: Let $\rho > 0$, then

$$\mathbb{P}(X > \mathbb{E}(X) + \rho) \leq \mathbb{P}(|X - \mathbb{E}(X)| > \rho) \leq \frac{\mathrm{Var}(X)}{\rho^2}. \tag{98}$$

Accordingly, we achieve

$$FR = \mathbb{P}(\widehat{\Delta}_m(k^*,k) \leq 0) \tag{99}$$

$$= \mathbb{P}\big(\widehat{\Delta}_m(k^*,k) \leq \mathbb{E}\{\widehat{\Delta}_m(k^*,k)\} \underbrace{-\Delta(k^*,k) - \mathrm{EB}(k^*,k)}_{-\rho}\big) \tag{100}$$

$$\leq \frac{\mathrm{EV}(k^*,k)}{(\mathrm{EB}(k^*,k) + \Delta(k^*,k))^2}. \tag{101}$$

As $\rho \longrightarrow 0$ the term $\frac{\mathrm{EV}(k^*,k)}{(\mathrm{EB}(k^*,k)+\Delta(k^*,k))^2} \longrightarrow 0$ *exponentially*.

Note that, a similar usage of the Chernov bound [5] allows to prove exponentially convergence. Further, since we achieved exponentially convergence of $\mathbb{P}(\widehat{\Delta}_m(k^*,k) \leq 0)$ against 0, we use the following first order approximation

$$\sum_{k^* \neq k} \mathbb{P}(\widehat{\Delta}_m(k^*,k) \leq 0) \approx \max_{k \neq k^*} \mathbb{P}(\widehat{\Delta}_m(k^*,k) \leq 0). \tag{102}$$

Concluding, using the relationship between success and failure rate, we define the success metric as

Definition 11 (Success Metric (SM)).

$$\mathrm{SM}(\mathscr{D}, \widehat{\mathscr{D}}_m) = \min_{k \neq k^*} \frac{\Delta(k^*,k) + \mathrm{EB}(k^*,k)}{\sqrt{\mathrm{EV}(k^*,k)}} \tag{103}$$

$$= \min_{k \neq k^*} \frac{\mathbb{E}\{\widehat{\Delta}_m(k^*,k)\}}{\sqrt{\mathrm{Var}(\widehat{\Delta}_m(k^*,k))}} \quad . \tag{104}$$

Interestingly, the success metric includes the minimum distance between the correct key and its nearest-rival as the RDM, however, it is, of course, based on the estimated distinguisher and thus includes the variance of the estimated difference $\widehat{\Delta}_m(k^*,k)$ in the denominator.

Remark 1. From Appendix 7.2.1 one can see that SR can be approximated from SM. More precisely,

$$\mathrm{SR} \doteq 1 - \exp\left(-\frac{1}{2}\mathrm{SM}^2\right), \tag{105}$$

so SM is the *first order exponent* of SR regarding the following definition of equivalence [6, page 63, Eqn. (3.76)]:

Definition 12. The notation $a_m \doteq b_m$ means that

$$\lim_{m\to\infty} \frac{1}{m} \log \frac{a_m}{b_m} = 0. \tag{106}$$

Thus, $a_m \doteq b_m$ implies that a_m and b_m are equal *to the first order* in the exponent.

As the success rate, the success metric can be derived empirically from simulations/ measurements or theoretically from closed-form expressions. In the next subsection we develop closed-form expressions for additive distinguisher (e.g., DPA, CPA). Even more, in Subsect. 7.4 we derive a closed-form expression of the information theoretic distinguisher MIA for the success metric, which has not be done for any metric so far and cannot be straightforwardly extended to the success rate.

7.3 Closed-Form Expression for Additive Distinguishers

Definition 13 (Additive distinguisher). We call an estimated distinguisher $\widehat{\mathscr{D}}_m(k)$ *additive* if it is unbiased (i.e., $\mathrm{EB}(k^*,k) = 0$) and takes the form

$$\widehat{\mathscr{D}}_m(k) = \frac{1}{m}\sum_{i=1}^{m} \widehat{\mathscr{D}}(X_i, Y_i(k)), \tag{107}$$

where $\widehat{\mathscr{D}}(X_i, Y_i(k))$ is a deterministic function of the i.i.d. sequence $(X_i, Y_i(k))$ and, therefore

$$\mathbb{E}\{\widehat{\mathscr{D}}_m(k)\} = \mathscr{D}(k). \tag{108}$$

Remark 2. This definition implicitly assumes that the distribution of $Y(k)$ is identical for all $k \in \mathscr{K}$. In other words, knowing the distribution of $Y(k)$ does not give any evidence about the secret (see [13, 26] for similar assumptions). Thus, $\mathrm{Var}\{Y(k)\}$ is constant for all $k \in \mathscr{K}$. Furthermore, without loss of generality we assume that the sensitive variable Y is normalized such that $\mathbb{E}\{Y(k)\} = 0$ and $\mathrm{Var}\{Y(k)\} = \mathbb{E}\{Y(k)^2\} = 1$.

Proposition 9. *Considering Remark 2 one can simplify both* $\widehat{\mathscr{D}}_{m\mathsf{DPA}}$ *[16] and* $\widehat{\mathscr{D}}_{m\mathsf{CPA}}$ *[3] to*

$$\frac{1}{m}\sum_{i=1}^{m} X_i Y_i(k). \tag{109}$$

Proof. A proof for $\widehat{\mathscr{D}}_{m\mathsf{CPA}}$ is given in the following. As formalized in [7] $\widehat{\mathscr{D}}_{m\mathsf{DPA}}$ and $\widehat{\mathscr{D}}_{m\mathsf{CPA}}$ can be directly translated into each other. Recall the definition of CPA:

$$\widehat{\mathscr{D}}_{m\mathsf{CPA}}(k) = \frac{\frac{1}{m}\sum_{i=1}^{m}(X_i-\overline{X})(Y_i(k)-\overline{Y(k)})}{\sqrt{\frac{1}{m}\sum_{i=1}^{n}(X_i-\overline{X})^2}\sqrt{\frac{1}{m}\sum_{i=1}^{m}(Y_i(k)-\overline{Y(k)})^2}}, \tag{110}$$

where

$$\overline{X} = \frac{1}{m}\sum_{i=1}^{m} X_i \qquad \overline{Y(k)} = \frac{1}{m}\sum_{i=1}^{m} Y_i(k). \tag{111}$$

Due to Remark 2, (for large m) we have $\overline{Y(k)} = 0$ and $\frac{1}{m}\sum_{i=1}^{m}(Y_i(k)-\overline{Y(k)})^2 = 1$. Straightforward computation yields Proposition 9 for $\widehat{\mathscr{D}}_{m\mathsf{CPA}}(k)$. For more details on CPA (and side-channel distinguisher) we refer to [33, 36]. □

To formulate a closed-form expression for the success metric for any additive distinguisher, we extend the idea of confusion similar to [38], which we call *general 2-way confusion coefficients*.

Definition 14 (General 2-way confusion coefficients). For $k \neq k^*$ we define

$$\kappa(k^*,k) = \mathbb{E}\left\{\left(\frac{Y(k^*)-Y(k)}{2}\right)^2\right\}, \tag{112}$$

$$\kappa'(k^*,k) = \mathbb{E}\left\{Y(k^*)^2\left(\frac{Y(k^*)-Y(k)}{2}\right)^2\right\}. \tag{113}$$

Remark 3. The confusion coefficient introduced in [38] is defined as $\kappa^\circ(k^*,k) = \mathbb{E}\{Y(k^*)Y(k)\}$ and we obtain the following relationship

$$\kappa^\circ(k^*,k) = 1 - 2\kappa(k^*,k). \tag{114}$$

Note that, our definition is consistent and a natural extension of the work in [8]. We now precise our side-channel model from Eq. (1) and Eq. (2)) in case of additive distinguishers. As these distinguishers are most usually used when the leakage X is linearly depend on Y^*, we assume $X = \alpha Y^* + N$ [8].

Proposition 10 (SM for CPA). *Let $\varepsilon = 2\alpha$. The success metric for any additive distinguisher takes the closed-form expression*

$$\mathrm{SM}(\mathscr{D},\widehat{\mathscr{D}}_m) = \min_{k\neq k^*}\frac{\varepsilon\kappa(k^*,k)}{\sqrt{\varepsilon^2(\kappa'(k^*,k)-\kappa^2(k^*,k))+4\sigma^2\kappa(k^*,k)}}\sqrt{m}. \tag{115}$$

[8] Note that, a similar model was also implicitly used in [8, 38].

Proof. We first give the following proposition.

Proposition 11. *The first two moments of $\widehat{\Delta}_m(k^*,k)$ are given by*

$$\mathbb{E}\{\widehat{\Delta}_m(k^*,k)\} = 2\alpha\kappa(k^*,k), \tag{116}$$

$$\mathrm{Var}(\widehat{\Delta}_m(k^*,k)) = 4[\alpha^2(\kappa'(k^*,k) - \kappa^2(k^*,k)) + \sigma^2\kappa(k^*,k)]. \tag{117}$$

Proof. Recall

$$\widehat{\Delta}_m(k^*,k) = (\alpha Y(k^*) + N)(Y(k^*) - Y(k)).$$

Since $\mathbb{E}\{Y(k^*)^2\} = 1$ (see Remark 2), we obtain

$$\mathbb{E}\{Y(k^*)(Y(k^*) - Y(k))\} = 1 - \mathbb{E}\{Y(k^*)Y(k)\} \tag{118}$$

$$= 2\mathbb{E}\{\big(\frac{Y(k^*) - Y(k)}{2}\big)\} \tag{119}$$

$$= 2\kappa(k^*,k). \tag{120}$$

Because N is independent of $Y(k)$,

$$\mathbb{E}\{N \cdot (Y(k^*) - Y(k))\} = \mathbb{E}\{N\} \cdot \mathbb{E}\{Y(k^*) - Y(k)\} = 0. \tag{121}$$

Therefore we obtain

$$\mathbb{E}\{\widehat{\Delta}_m(k^*,k)\} = 2\alpha\kappa(k^*,k). \tag{122}$$

For the variance we obtain

$$\mathbb{E}\{\widehat{\Delta}_m(k^*,k)^2\} = \mathbb{E}\{(XY^* - XY)^2\} \tag{123}$$

$$= 2\mathbb{E}\{N^2(Y^* - Y)^2\} + \alpha^2\mathbb{E}\{Y^{*2}(Y^{*2} - Y)^2\} \tag{124}$$

$$= 4\sigma^2\kappa(k^*,k) + \alpha^2 4\kappa'(k^*,k), \tag{125}$$

since all cross terms with N vanish. Hence, we have

$$\mathrm{Var}(\widehat{\Delta}_m(k^*,k)) = \mathbb{E}\{\widehat{\Delta}_m(k^*,k)^2\} - \mathbb{E}\{\widehat{\Delta}_m(k^*,k)\}^2 \tag{126}$$

$$= 4[\alpha^2(\kappa'(k^*,k) - \kappa^2(k^*,k)) + \sigma^2\kappa(k^*,k)]. \tag{127}$$

□

Plugging Proposition 11 into the success metric given in Eq. (103) and considering the normalizing factor of the variance $\sqrt{m}$ (see Eq. (107)) directly derives Proposition 10. □

For DPA with one-bit variables $Y(k)$ we can further simplify the success metric such that it can be expressed directly through the SNR, number of measurements and 2-way confusion coefficient $\kappa(k^*,k)$:

Proposition 12 (SM for 1-bit DPA). *Let $\varepsilon = 2\alpha$, Y a one-bit variable (e.g., $Y \in \{\pm 1\}$) and $\widehat{\mathscr{D}}_m(k)$ an additive distinguisher, then*

$$\mathrm{SM}(\mathscr{D}, \widehat{\mathscr{D}}_m) = \frac{\sqrt{m}}{\sqrt{\max\limits_{k \neq k^*} \frac{1-\kappa(k^*,k)}{\kappa(k^*,k)} + \frac{1}{\kappa(k^*,k)\,\mathrm{SNR}}}}, \tag{128}$$

with $\mathrm{SNR} = \frac{\mathrm{Var}(signal)}{\mathrm{Var}(noise)} = \frac{\varepsilon^2}{\sigma^2}$*, since $\varepsilon = 2\alpha$ is the difference between X when $Y = 1$ and $Y = -1$.*

Proof. When $Y(k)\ \forall k \in \mathscr{K}$ is a one-bit variable, we achieve the following simplification:

$$\kappa(k^*,k) = \mathbb{E}\{\big(\frac{Y(k^*) - Y(k)}{2}\big)^2\} = \mathbb{E}\{Y(k^*)^2\big(\frac{Y(k^*) - Y(k)}{2}\big)^2\} = \kappa'(k^*,k). \tag{129}$$

From this, Proposition 12 follows directly. □

Remark 4. Estimating the success rate from confusion coefficients includes a computation of a multivariate normal cumulative distribution function [27] for which (contrary as stated in [8]) no closed-form expression exists. Moreover, we discovered that the calculated covariance matrices[9] that directly depend on the confusion coefficients are not of full rank. This effect was similarly discovered for CPA by Rivain in [28], where the author propose to use Monte-Carlo simulation to overcome this problem.

According to Remark 4, we stress that the computation of the success metric as a closed-form expression is more convenient than using the closed-form expression for the success rate for DPA and CPA, since only 2-way confusion coefficients $(\kappa(k^*,k), \kappa'(k^*,k))$ without multivariate distributions are involved.

Additionally, with the help of $\kappa(k^*,k)$ we can give a closed-form expression for RDM (see Eq. (88)) for any additive distinguisher:

Proposition 13. *For additive distinguisher the* $\mathrm{RDM}(\mathscr{D})$ *can be simplified as*

$$\mathrm{RDM}(\mathscr{D}) = \frac{\min\limits_{k \neq k^*} \kappa(k^*,k)}{\sqrt{\mathrm{Var}(\kappa(k^*,K))}}. \tag{130}$$

Proof Sketch: As the RDM takes as a input the theoretical value of a distinguisher $\mathscr{D}$, $\kappa(k^*,k)$ directly describes the difference between $\mathscr{D}(k^*)$ and $\mathscr{D}(k)$ for any $k \in K$. Thus, Prop. 13 directly follows. □

The comparison of the closed-form expressions of RDM in Eq. (130) and SM in Eq. (115) again highlights the different aspects of both metrics.

[9] Namely $[\kappa(k^*,i,j)]_{(i,j)\in\mathscr{K}\setminus\{0\}}$ and $[\kappa(k^*,i) \times \kappa(k^*,j)]_{(i,j)\in\mathscr{K}\setminus\{0\}}$.

7.4 Closed-form Expression for Mutual Information Analysis

Definition 15. The Mutual Information Analysis distinguisher (MIA) [10] between a continuous variable X and a discrete variable Y is defined by

$$I(X;Y) = H(X) - H(X|Y), \tag{131}$$

where $H(X) = -\int_{-\infty}^{\infty} f(x) \cdot \log f(x)\,\mathrm{d}x$ is the (differential) *entropy* of X and $H(X|Y) = \sum_y p(y) \cdot H(X|Y=y) = -\sum_y p(y) \int_{-\infty}^{\infty} f(x|y) \cdot \log f(x|y)\,\mathrm{d}x$ is the *conditional entropy* of X knowing Y.

In practice, $I(X;Y)$ has to be estimated, while unlike for CPA or DPA the estimation of MIA is a nontrivial problem. For a detailed evaluation of estimation methods of mutual information distinguishers we refer to [39]. In the following, we consider the estimation with histograms in order to formulate a closed-form expression. To estimate MIA with histograms (H-MIA), one has to partition the leakage X into h distinct bins b_i of width Δx with $i = 1, \ldots, h$. Note again that, Y is already discrete.

Definition 16. Let $\hat{p}(x) = \frac{\#b_i}{m}$ with x falling into bin b_i and let $\hat{p}(x|y)$ be the estimated probability knowing $Y = y$, then

$$\hat{I}_m(X;Y) = -\sum_x \hat{p}(x) \log \hat{p}(x) + \sum_y \hat{p}(y) \sum_x \hat{p}(x|y) \log \hat{p}(x|y). \tag{132}$$

For simplification, we consider in the following only the negative conditional entropy $-\hat{H}(X|Y)$ as a distinguisher, since $\hat{H}(X)$ does not depend on a key hypothesis. Additionally, we reasonably assume that the distribution of Y is know to the attacker and thus we use $p(y)$ instead of $\hat{p}(y)$. So, H-MIA simplifies to

$$\text{H-MIA}(X,Y) = \sum_y p(y) \sum_x \hat{p}(x|y) \log \hat{p}(x|y) + \log \Delta x. \tag{133}$$

Note that, since we estimate the differential entropy the additional term $\log \Delta x$ arises, which is eliminated in Eq. (132). For more information on differential entropy and mutual information we refer to [6].

First, we develop a closed-form expression for $\mathrm{EB}\{\widehat{\Delta}_m(k^*,k)\}$: Since Y is discrete the bias only arise due to the discretization of X and the limited number of measurements m. Thus, we utilize the approximations given for the bias of $\hat{H}(X)$ in [20] (3.14) to calculate $\mathbb{E}\{\widehat{\mathscr{D}}_m(k)\}$ and $\mathbb{E}\{\widehat{\Delta}_m(k^*,k)\}$ for H-MIA. To be specific, let h define the number of bins and Δx their width, then

$$\mathbb{E}\{\widehat{\mathscr{D}}_m(k)\} = -\mathbb{E}\{\hat{H}(X|Y)\} = -\sum_y p(y) \mathbb{E}\{\hat{H}(X|Y=y)\}, \tag{134}$$

$$\approx -\sum_y p(y) \left[H(X|Y=y) + \frac{\Delta x^2}{24} J(X|Y=y)\right] - \frac{h-1}{2m}, \tag{135}$$

$$\mathbb{E}\{\widehat{\Delta}_m(k^*,k)\} \approx \sum_y p(y) \left[H(X|Y=y) + \frac{\Delta x^2}{24} J(X|Y=y)\right]$$

$$-\Big(\sum_{y^*} p(y^*)\big[H(X|Y^*=y^*)+\frac{\Delta x^2}{24}J(X|Y^*=y^*)\big]\Big), \tag{136}$$

with $J(X|Y)=\sum_y p(y)J(X|Y=y)$ and $J(X|Y=y)$ being the Fisher information $\int_{-\infty}^{\infty}\frac{[\frac{\mathrm{d}}{\mathrm{d}x}p(x|y)]^2}{p(x|y)}\,\mathrm{d}x$ [9].

Next, to calculate $\mathrm{Var}\{\widehat{\mathscr{D}}_m(k)\}$ we use the law of total variance [14] (Eq. (137) $\Leftrightarrow$ Eq. (138)) and the approximations for the variance given in [20] (4.9) for Eq. (138) $\Rightarrow$ Eq. (139) and Eq. (140) $\Rightarrow$ Eq. (141):

$$\mathrm{Var}\{\widehat{\mathscr{D}}_m(k)\}=\mathrm{Var}\{\hat{H}(X|Y)\}\}=\mathrm{Var}\{\mathbb{E}\{\hat{H}(X|Y=y)\}\} \tag{137}$$

$$=\mathrm{Var}\{\hat{H}(X)\}-\mathbb{E}\{\mathrm{Var}\{\hat{H}(X|Y=y)\}\} \tag{138}$$

$$\approx \mathrm{Var}\{H(X)\}-\frac{1}{m}\sum_y p(y)\,\mathrm{Var}\{-\log f(x|y)\} \tag{139}$$

$$\mathrm{Var}\{\widehat{\Delta}_m(k^*,k)\}=\mathrm{Var}\{\mathbb{E}\{\hat{H}(X|Y=y\}\}-\mathrm{Var}\{\mathbb{E}\{\hat{H}(X|Y^*=y^*\}\} \tag{140}$$

$$-2\,\mathrm{Cov}(\mathbb{E}\{\hat{H}(X|Y=y\}\},\mathbb{E}\{\hat{H}(X|Y^*=y^*\}\})$$

$$\approx \frac{1}{m}\sum_y p(y)\,\mathrm{Var}\{-\log f(x|y)\}+\frac{1}{m}\sum_y p(y^*)\,\mathrm{Var}\{-\log f(x|y^*)\} \tag{141}$$

$$-2\,\mathrm{Cov}(\mathbb{E}\{\hat{H}(X|Y=y\}\},\mathbb{E}\{\hat{H}(X|Y^*=y^*\}\})$$

$$\leq \frac{1}{m}\Big(\sum_y p(y)\,\mathrm{Var}\{-\log f(x|y)\}+\sum_y p(y^*)\,\mathrm{Var}\{-\log f(x|y^*)\}\Big) \tag{142}$$

Using the closed-form expressions for $\mathrm{EB}\{\widehat{\Delta}_m(k^*,k)\}$ and $\mathrm{EV}\{\widehat{\Delta}_m(k^*,k)\}$ we formulate the following proposition.

Proposition 14 (SM for H-MIA).

$$\mathrm{SM}(\mathscr{D},\widehat{\mathscr{D}}_m)\lessapprox \min_{k^*\neq k}\frac{\big(\Delta(k^*,k)+\frac{\Delta x^2}{24}\big(J(X|Y)-J(X|Y^*)\big)\big)\sqrt{m}}{\sqrt{\sum_y p(y)\,\mathrm{Var}\{-\log f(x|y)\}+\sum_{y^*}p(y^*)\,\mathrm{Var}\{-\log f(x|y^*)\}}}, \tag{143}$$

with $\Delta(k^*,k)=H(X|Y)-H(X|Y^*)$, $J(X|Y)=\sum_y p(y)J(X|Y=y)$ while $J(X|Y=y)$ is the Fisher information $\int_{-\infty}^{\infty}\frac{[\frac{\mathrm{d}}{\mathrm{d}x}f(x|y)]^2}{f(x|y)}\,\mathrm{d}x$ [9].

Interestingly, the SM of MIA involves the number of traces as the $\sqrt{m}$ in the nominator like DPA and CPA, which seems reasonable.

Remark 5. If N is normal distributed with variance σ^2 we can further simplify $H(X|Y^*=y^*)=\frac{1}{2}\log(2\pi e\sigma^2)$ since $p(x|y^*)=p_N(x-y^*)$. Moreover, $J(X|Y^*=y)=\frac{1}{\sigma^2}$ and $\mathrm{Var}\{-\log f(x|y^*)\}=\frac{1}{2m}$.

Remark 6. Remarkably, the variance is approximately independent of the size of Δx. Only in extreme cases like $\Delta x = 1$ and $\Delta x \to \infty$ is affecting the variance. Also see [20] for more information. Interestingly, all linear terms have disappeared in the expression of the SM.

8 Features of SM Expressions

8.1 Linking the Success to Properties of the Sbox

All previous studies about the relationship between the sbox properties and side-channel analysis considered the direct link between a metric on a *distinguisher* itself and the sbox. In [11], Guilley et al. use as a metric the maximal value of the distinguisher divided by its standard deviation (SNR). The authors demonstrate that for DPA the SNR is lower bounded by quantities that are expected to be large for sboxes resisting against linear differential cryptanalyses. Prouff introduces in [22], an alternative metric for CPA, called the *transparency order*, that is defined as the difference between the maximal value of CPA and the average of all rivals. Besides, the power model is not the Hamming weight, but the Hamming distance; however, strangely enough, the sensitive variable is not the Hamming distance, but instead the average of the initial state exclusive-ored with all possible final states. This leakage model is, to our best knowledge, rather unusual in practice. In both previous works the relationship is only stated as an expected outcome but not proven. The results have been further investigated by Carlet in [4].

In the following, we not only bound but directly link the success metric and the sbox in case of low SNR (practical conditions). As DPA is a special case of CPA, we further concentrate on the closed-form expression of CPA and simplify Eq. (115) when $\sigma \gg \alpha$. More precisely,

$$\mathrm{SM}(\mathscr{D}, \widehat{\mathscr{D}}_m) \approx \min_{k \neq k^*} \sqrt{\frac{4\alpha^2 \kappa^2(k^*,k) m}{\sigma^2 4 \kappa(k^*,k)}} \tag{144}$$

$$= \sqrt{\mathrm{SNR}} \sqrt{m} \min_{k \neq k^*} \sqrt{\kappa(k^*,k)}. \tag{145}$$

From Eq. (112), $\kappa(k^*,k^*) = 0$ and $\kappa(k^*,k) \geq 0$, thus the argument of the square root in Eq. (145) is always positive. Besides, by the Cauchy-Schwarz theorem, we also have that $\kappa(k^*,k) \leq 1$. Now, the objective to minimizing $\min_{k \neq k^*} \sqrt{\kappa(k^*,k)}$ (i.e., making side-channel attacks as hard as possible) is tantamount to maximizing $\max_{k \neq k^*} \mathbb{E}(Y(k^*)Y(k))$. In the following, we assume that Y^* and Y explicitly depend on an sbox (or inverse sbox) and a Hamming weight (w_H) leakage model[10] as for example $w_H(\mathtt{Sbox}[T \oplus k])$, so $Y(k) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (-1)^{S_i(T \oplus k)} = \frac{1}{\sqrt{n}} (2 w_H(S(T \oplus k)) - n)$ and

[10] One can easily extend the calculation also for the Hamming distance model.

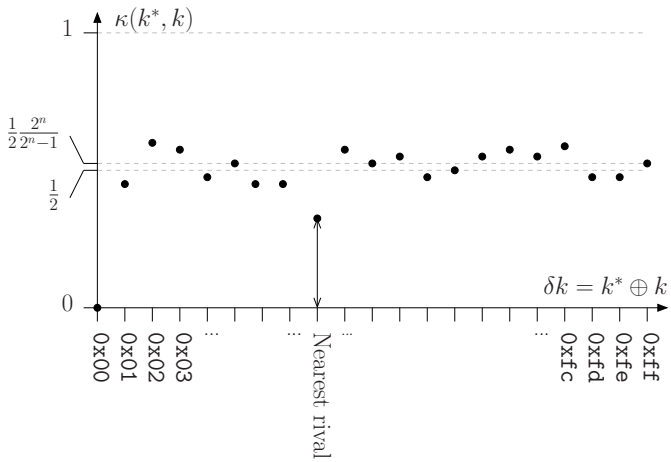


Fig. 3: Illustration of the confusion coefficients for CPA

$$\mathbb{E}\{Y(k^*)Y(k)\} = \frac{1}{n}\sum_{i,j=0}^{n}\frac{1}{2^n}\sum_{t\in\mathbb{F}_2^n}(-1)^{S_i(t\oplus k^*)\oplus S_j(t\oplus k)}. \tag{146}$$

As $\forall a \in \{0,1\}, (-1)^a = 1-2a$, the goal to make CPA difficult is to minimize the following quantity, that we call the *transparency metric*

$$\min_{k\neq k^*}\sum_{i,j=0}^{n}\sum_{t\in\mathbb{F}_2^n} S_i(t\oplus k^*)\oplus S_j(t\oplus k). \tag{147}$$

Remark 7. Note that, for single-bit attacks ($n=1$), the criteria of Eq. (147) simplifies to the *one-sided* criteria discovered in [12].

So, minimizing the objective on the sbox in Eq. (147) is equivalent to minimizing $\min_{k\neq k^*}\kappa(k^*,k)$, which can be understood intuitively on the illustration of Fig. 3. The key corresponding to the nearest rival, i.e., $\operatorname{argmin}_{k\neq k^*}\kappa(k^*,k)$, shall have a confusion coefficient as high as possible.

To further illustrate the transparency metric and show the relationship to the *transparency order* [22], we use the same three sboxes as in [12]: Let $\oplus$ and $\odot$ be respectively the inner addition and multiplication of the Galois field $\mathbb{F}_{2^8}$ of 256 elements, then the sboxes are given by

1. A "bad" `Sbox`$[\cdot]$, termed S_1, of equation $y \mapsto a\odot y \oplus b$,
2. An "average" `Sbox`$[\cdot]$, termed S_{101}, of equation $y \mapsto a\odot y^{101} \oplus b$,
3. A "good" `Sbox`$[\cdot]$, termed S_{254} and used in AES, of equation $y \mapsto a\odot y^{254} \oplus b$.

Fig. 4 displays the confusion coefficient for S_1, S_{101} and S_{254}. One can see, that the minimal $\min_{k\neq k^*}\kappa(k^*,k)$ is achieved by S_1, which is the hardest to attack with

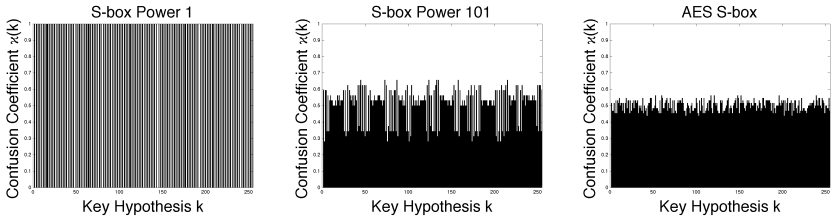


Fig. 4: Confusion coefficients for S_1, S_{101} and S_{254} (*courtesy of [12]*)

CPA, whereas S_{254} has the highest $\min_{k \neq k^*} \kappa(k^*, k)$ being the most vulnerable. Tab. 1 displays the transparency metric and order. The transparency metric is different from the transparency order, nonetheless, it remains consistent with it, meaning that the order of S_1, S_{101} and S_{254} is the same for both metrics and consistent with the rating through $\kappa(k^*, k)$.

Table 1: Comparison of side-channel metrics for sboxes

	Transparency Order [22]	**Transparency Metric (Eqn.** (147)**)**
S_1	5.84	7424
S_{101}	7.50	7936
S_{254}	7.86	8000

8.2 How does the Size of the Key Space Influence the SM/SR?

Hardware devices are known to leak approximately in Hamming distance. This makes leakage models complicated, because they involve two consecutive states of the cipher. Let us consider the example of an AES-128 computed one round per clock period. The plaintext is P, the cipher C, and the first (resp. last) round key K^1 (resp. K^{11}).

On the one hand, the uncentered and non-normalized leakage model at the first round for the byte at position 0 is:

$$Y^1(T, K^1) = w_H(T_0 \oplus \texttt{02} \cdot S(T_0 \oplus K_0^1) \oplus \texttt{01} \cdot S(T_5 \oplus K_5^1) \tag{148}$$

$$\oplus\, \texttt{01} \cdot S(T_{10} \oplus K_{10}^1) \oplus \texttt{03} \cdot S(T_{15} \oplus K_{15}^1)) \ , \tag{149}$$

where 01, 02 and 03 are the MixColumns constants, and S is the SubBytes operation. Clearly, a guess for this model requires an hypothesis on 4 bytes of the key K^1.

On the other hand, the uncentered and non-normalized leakage model at the last round for the byte at position 0 is:

$$Y^{10}(C,K^{10}) = w_H\left(C_0 \oplus S^{-1}(C_0 \oplus K_0^{10})\right) \quad , \tag{150}$$

where S^{-1} is the InvSubBytes operation. So, a guess for the model requires simply one hypothesis on a key byte (namely K_0^{10}). This is due to the absence of MixColumns at the last round.

The transparency order (resp. metric) of InvSubBytes is 7.85 (resp. 7964), meaning that it is very close to that of SubBytes. So, the confusion coefficient associated to Y^1 and to Y^{10} have similar distributions, meaning that the data complexity (the number of traces m) of the attack is similar at either end of the AES. Specifically, the minimal nonzero confusion coefficient for Y^1 is 0.468750, whereas it is 0.404297 for Y^{10}. The most crucial difference is the computational complexity, owing to the largest key space to explore at the first round.

Acknowledgement

Annelie Heuser is partly founded by the Google Doctoral European Fellowship in the field of privacy.

References

1. Lejla Batina, Benedikt Gierlichs, and Kerstin Lemke-Rust. Differential Cluster Analysis. In Christophe Clavier and Kris Gaj, editors, *Cryptographic Hardware and Embedded Systems – CHES 2009*, volume 5747 of *Lecture Notes in Computer Science*, pages 112–127, Lausanne, Switzerland, 2009. Springer-Verlag.
2. Lejla Batina, Benedikt Gierlichs, Emmanuel Prouff, Matthieu Rivain, François-Xavier Standaert, and Nicolas Veyrat-Charvillon. Mutual Information Analysis: a Comprehensive Study. *J. Cryptology*, 24(2):269–291, 2011.
3. Éric Brier, Christophe Clavier, and Francis Olivier. Correlation Power Analysis with a Leakage Model. In *CHES*, volume 3156 of *LNCS*, pages 16–29. Springer, August 11–13 2004. Cambridge, MA, USA.
4. Claude Carlet. On Highly Nonlinear S-Boxes and Their Inability to Thwart DPA Attacks. In *INDOCRYPT*, volume 3797 of *LNCS*, pages 49–62. Springer, december 2005. Bangalore, India. (PDF on SpringerLink; Complete version on IACR ePrint).
5. H. Chernoff. A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Ann. Math. Stat.*, 23:493–507, 1952.
6. Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. Wiley-Interscience, July 18 2006. ISBN-10: ISBN-10: 0471241954, ISBN-13: 978-0471241959, 2nd edition.
7. Julien Doget, Emmanuel Prouff, Matthieu Rivain, and François-Xavier Standaert. Univariate side channel attacks and leakage modeling. *J. Cryptographic Engineering*, 1(2):123–144, 2011.
8. Yunsi Fei, Qiasi Luo, and A. Adam Ding. A Statistical Model for DPA with Novel Algorithmic Confusion Analysis. In Emmanuel Prouff and Patrick Schaumont, editors, *CHES*, volume 7428 of *LNCS*, pages 233–250. Springer, 2012.

9. R. A. Fisher. *Statistical Methods for Research Workers*. Oliver and Boyd, Edinburgh, 1925.
10. Benedikt Gierlichs, Lejla Batina, Pim Tuyls, and Bart Preneel. Mutual information analysis. In *CHES, 10th International Workshop*, volume 5154 of *Lecture Notes in Computer Science*, pages 426–442. Springer, August 10-13 2008. Washington, D.C., USA.
11. Sylvain Guilley, Philippe Hoogvorst, and Renaud Pacalet. Differential Power Analysis Model and some Results. In Kluwer, editor, *Proceedings of WCC/CARDIS*, pages 127–142, Aug 2004. Toulouse, France. DOI: 10.1007/1-4020-8147-2$\underline{9}$.
12. Annelie Heuser, Sylvain Guilley, and Olivier Rioul. A Theoretical Study of Kolmogorov-Smirnov Distinguishers: Side-Channel Analysis vs. Differential Cryptanalysis. Cryptology ePrint Archive, Report 2014/008, 2014. `http://eprint.iacr.org/2014/008`.
13. Annelie Heuser, Michael Kasper, Werner Schindler, and Marc Stottinger. How a symmetry metric assists side-channel evaluation - a novel model verification method for power analysis. In *Proceedings of the 2011 14th Euromicro Conference on Digital System Design*, DSD '11, pages 674–681, Washington, DC, USA, 2011. IEEE Computer Society.
14. O.J.W.F. Kardaun. *Classical Methods of Statistics*. Springer, 2005.
15. Paul C. Kocher, Joshua Jaffe, and Benjamin Jun. Differential Power Analysis. In *Proceedings of CRYPTO'99*, volume 1666 of *LNCS*, pages 388–397. Springer-Verlag, 1999.
16. Paul C. Kocher, Joshua Jaffe, and Benjamin Jun. Differential Power Analysis. In *CRYPTO*, volume 1666 of *LNCS*, pages pp 388–397. Springer, 1999.
17. Thanh-Ha Le and Maël Berthier. Mutual Information Analysis under the View of Higher-Order Statistics. In Isao Echizen, Noboru Kunihiro, and Ryôichi Sasaki, editors, *IWSEC*, volume 6434 of *LNCS*, pages 285–300. Springer, 2010.
18. Houssem Maghrebi, Olivier Rioul, Sylvain Guilley, and Jean-Luc Danger. Comparison between Side Channel Analysis Distinguishers. In Tat Wing Chim and Tsz Hon Yuen, editors, *ICICS*, volume 7618 of *LNCS*, pages 331–340. Springer, October 29-31 2012. Hong Kong.
19. Stefan Mangard, Elisabeth Oswald, and François-Xavier Standaert. One for All - All for One: Unifying Standard DPA Attacks. *Information Security, IET*, 5(2):100–111, 2011. ISSN: 1751-8709 ; Digital Object Identifier: 10.1049/iet-ifs.2010.0096.
20. R. Moddemeijer. On estimation of entropy and mutual information of continuous distributions. *Signal Processing*, 16(3):233–248, March 1989.
21. Amir Moradi, Nima Mousavi, Christof Paar, and Mahmoud Salmasizadeh. A Comparative Study of Mutual Information Analysis under a Gaussian Assumption. In *WISA (Information Security Applications, 10th International Workshop)*, volume 5932 of *Lecture Notes in Computer Science*, pages 193–205. Springer, August 25-27 2009. Busan, Korea.
22. Emmanuel Prouff. DPA Attacks and S-Boxes. In *FSE*, volume 3557 of *LNCS*, pages 424–441. Springer-Verlag, february 2005. Paris, France.
23. Emmanuel Prouff and Matthieu Rivain. Theoretical and Practical Aspects of Mutual Information Based Side Channel Analysis. In Springer, editor, *ACNS*, volume 5536 of *LNCS*, pages 499–518, June 2-5 2009. Paris-Rocquencourt, France.
24. Emmanuel Prouff and Matthieu Rivain. Theoretical and practical aspects of mutual information-based side channel analysis. *International Journal of Applied Cryptography (IJACT)*, 2(2):121–138, 2010.
25. Emmanuel Prouff and Matthieu Rivain. Theoretical and Practical Aspects of Mutual Information Based Side Channel Analysis. In *IJACT*, 2010.
26. Emmanuel Prouff, Matthieu Rivain, and Régis Bevan. Statistical Analysis of Second Order Differential Power Analysis. *IEEE Trans. Computers*, 58(6):799–811, 2009.
27. C. R. Rao. *Linear Statistical Inference and its Applications*. J. Wiley and Sons, New York, 2nd edition, 1973.
28. Matthieu Rivain. On the Exact Success Rate of Side Channel Analysis in the Gaussian Model. In *Selected Areas in Cryptography*, volume 5381 of *LNCS*, pages 165–183. Springer, August 14-15 2008. Sackville, New Brunswick, Canada.
29. Phillip Rogaway, editor. *Advances in Cryptology - CRYPTO 2011 - 31st Annual Cryptology Conference, Santa Barbara, CA, USA, August 14-18, 2011. Proceedings*, volume 6841 of *Lecture Notes in Computer Science*. Springer, 2011.

30. Walter Rudin. *Principles of mathematical analysis*. McGraw-Hill Book Co., New York, third edition, 1976. International Series in Pure and Applied Mathematics.
31. George Saon and Mukund Padmanabhan. Minimum Bayes Error Feature Selection for Continuous Speech Recognition. In Todd K. Leen, Thomas G. Dietterich, and Volker Tresp, editors, *NIPS*, pages 800–806. MIT Press, 2000.
32. B. W. Silverman and P. J. Green. *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London, 1986.
33. François-Xavier Standaert. Introduction to Side-Channel Attacks Secure Integrated Circuits and Systems. In Ingrid M. R. Verbauwhede, editor, *Secure Integrated Circuits and Systems*, Integrated Circuits and Systems, chapter 2, pages 27–42. Springer US, Boston, MA, 2010.
34. François-Xavier Standaert, Philippe Bulens, Giacomo de Meulenaer, and Nicolas Veyrat-Charvillon. Improving the Rules of the DPA Contest. Cryptology ePrint Archive, Report 2008/517, December 8 2008. `http://eprint.iacr.org/2008/517`.
35. François-Xavier Standaert, Tal Malkin, and Moti Yung. A Unified Framework for the Analysis of Side-Channel Key Recovery Attacks. In *EUROCRYPT*, volume 5479 of *LNCS*, pages 443–461. Springer, April 26-30 2009. Cologne, Germany.
36. François-Xavier Standaert, Éric Peeters, Gaël Rouvroy, and Jean-Jacques Quisquater. An Overview of Power Analysis Attacks Against Field Programmable Gate Arrays. *Proceedings of the IEEE*, 94(2):383–394, February 2006. (Invited Paper).
37. Pafnouti Tchebichef. Des valeurs moyennes. *Journal de mathématiques pures et appliqués*, 12(2):177–184, 1867.
38. Adrian Thillard, Emmanuel Prouff, and Thomas Roche. Success through Confidence: Evaluating the Effectiveness of a Side-Channel Attack. In Guido Bertoni and Jean-Sébastien Coron, editors, *CHES*, volume 8086 of *Lecture Notes in Computer Science*, pages 21–36. Springer, 2013.
39. Nicolas Veyrat-Charvillon and François-Xavier Standaert. Mutual Information Analysis: How, When and Why? In *CHES*, volume 5747 of *LNCS*, pages 429–443. Springer, September 6-9 2009. Lausanne, Switzerland.
40. Nicolas Veyrat-Charvillon and François-Xavier Standaert. Generic Side-Channel Distinguishers: Improvements and Limitations. In Rogaway [29], pages 354–372.
41. Carolyn Whitnall and Elisabeth Oswald. A Comprehensive Evaluation of Mutual Information Analysis Using a Fair Evaluation Framework. In Rogaway [29], pages 316–334.
42. Carolyn Whitnall and Elisabeth Oswald. A Fair Evaluation Framework for Comparing Side-Channel Distinguishers. *J. Cryptographic Engineering*, 1(2):145–160, 2011.
43. Carolyn Whitnall, Elisabeth Oswald, and Luke Mather. An Exploration of the Kolmogorov-Smirnov Test as a Competitor to Mutual Information Analysis. In Emmanuel Prouff, editor, *CARDIS*, volume 7079 of *Lecture Notes in Computer Science*, pages 234–251. Springer, 2011.
44. Carolyn Whitnall, Elisabeth Oswald, and François-Xavier Standaert. The myth of generic DPA...and the magic of learning. Cryptology ePrint Archive, Report 2012/256, 2012. `http://eprint.iacr.org/2012/256`.
45. Hui Zhao, Yongbin Zhou, François-Xavier Standaert, and Hailong Zhang. Systematic Construction and Comprehensive Evaluation of Kolmogorov-Smirnov Test based Side-Channel Distinguishers. Cryptology ePrint Archive, Report 2013/091, 2013. `http://eprint.iacr.org/2013/091`.