HAL open science

Inter-Class vs. Mutual Information as Side-Channel Distinguishers

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

▶ **To cite this version:**

Olivier Rioul, Annelie Heuser, Sylvain Guilley, Jean-Luc Danger. Inter-Class vs. Mutual Information as Side-Channel Distinguishers. 2016 IEEE International Symposium on Information Theory (ISIT'16), Jul 2016, Barcelona, Spain. 10.1109/ISIT.2016.7541410 . hal-02287308

HAL Id: hal-02287308

https://telecom-paris.hal.science/hal-02287308v1

Submitted on 11 Aug 2022

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.

Inter-Class vs. Mutual Information as Side-Channel Distinguishers

Olivier Rioul*†, Annelie Heuser*, Sylvain Guilley*‡ and Jean-Luc Danger*‡
* LTCI, CNRS, Télécom ParisTech, Université Paris-Saclay, 75 013 Paris, France
Email: firstname.lastname@telecom-paristech.fr
† CMAP, Ecole Polytechnique, Université Paris-Saclay, 91 128 Palaiseau, France
Email: olivier.rioul@polytechnique.edu
‡ Secure-IC S.A.S., 15 Rue Claude Chappe, Bât. B, ZAC des Champs Blancs,
35 510 Cesson-Sévigné, France

***Abstract*—A novel "interclass information" side-channel distinguisher is compared to mutual information analysis. Interclass information possesses properties similar to mutual information but uses a different comparing strategy between the underlying conditional distributions. It is shown that interclass information can outperform mutual information in side-channel analysis, especially under low noise. The theoretical comparison is confirmed by simulations.**

I. Introduction

Side-channel analysis constitutes a serious threat against modern cryptographic implementations. A side-channel attack exploits unintentionally emitted physical leakage—such as power consumption or electromagnetic emanation—from an embedded device to retrieve secret information. The introduction of differential power analysis by Kocher *et al.* [1] gave rise to many developments, attacks and models for the evaluation of physical security. Prouff *et al.* [2], [3] made a careful theoretical description of side-channel analysis for various scenarios and attacks, including information-based attacks.

A typical side-channel scenario is as follows. A cryptographic algorithm like AES or PRESENT is implemented on a device using a secret sequence of key bytes. At a specific step in the algorithm, some given (plain or cypher) text byte t is combined with a specific secret key byte k^* through e.g., a XOR operation $k^* \oplus t$. From the attacker's viewpoint, k^* is fixed (deterministic) but unknown, while t is known but varies for each encryption request; hence it is seen as a realization of a uniformly distributed random variable T. The measured leakage—the side channel output—takes the form

$$X = f(k^* \oplus T) + Z \tag{1}$$

where Z is an additive noise independent of T, with density ϕ, and where f is a partially unknown, device-specific function. The attacker computes a key estimate $\hat{k}$ from an i.i.d. sequence of such leakage measurements X, the attack being successful if $\hat{k} = k^*$. The success rate gives a practical measure of the attack performance.

In [4] the authors show that the side-channel problem is equivalent to a communication problem in which the "message" is the secret key, T is a side information available at both emitter and receiver, and the encoder or the deterministic part of the memoryless channel is partially unknown. If f is completely known (e.g., thanks to a preliminary profiling phase), the best attack corresponds to a maximum likelihood decoder [4].

If, however, profiling is impossible as it would require an exact copy of the device, the attacker is led to carry out some statistical test in order to discriminate the correct key. The attacker computes a leakage model *class*:

$$Y(k) = \hat{f}(k \oplus T) \tag{2}$$

for any key hypothesis k, where $\hat{f}$ is a prediction function for the deterministic part of the leakage. The test usually takes the form of a maximization of a side-channel *distinguisher* [1]–[7]:

$$\hat{k} = \arg\max_k \widehat{\mathcal{D}}(k) \tag{3}$$

where $\widehat{\mathcal{D}}(k)$ is computed as an estimation of some theoretical distinguisher $\mathcal{D}(k)$. The estimation is obtained from the available sequence of realizations of X and $Y(k)$ for all k.

A large variety of distinguishers $\mathcal{D}(k)$ have been proposed in the literature. Some classical choices include difference-of-means [1], correlation [5], Kolmogorov-Smirnov [6], and mutual information [7]:

$$\mathcal{D}_{\text{MIA}}(k) = I(X; Y(k)). \tag{4}$$

Mutual information analysis (MIA) was proposed by Gierlichs *et al.* [7] to overcome limitations such as the restriction to linear dependency between the measured leakage and the assumed leakage model. It turns out to be implemented similarly as Goppa's maximum mutual information decoder [8].

A fair comparison between different distinguishers is desirable as it helps to better understand how efficient a side-channel attack can be and how to take appropriate countermeasures. Yet in general it is a difficult task because many factors come into play—in particular, intrinsic statistical properties and quality of estimation.

In this paper, we consider a new information-theoretic distinguisher:

$$\mathcal{D}_{\text{IIA}}(k) = \mathit{II}(X; Y(k)) \tag{5}$$

Annelie Heuser is a Google European fellow in the field of privacy and is partially founded by this fellowship.

and the corresponding *interclass information analysis* (IIA). Here *interclass information*, denoted by $I\!I(X;Y)$, is defined similarly as $I(X;Y)$ but compares the underlying conditional distributions differently. We investigate a theoretical comparative study of MIA and IIA which we validate by simulations.

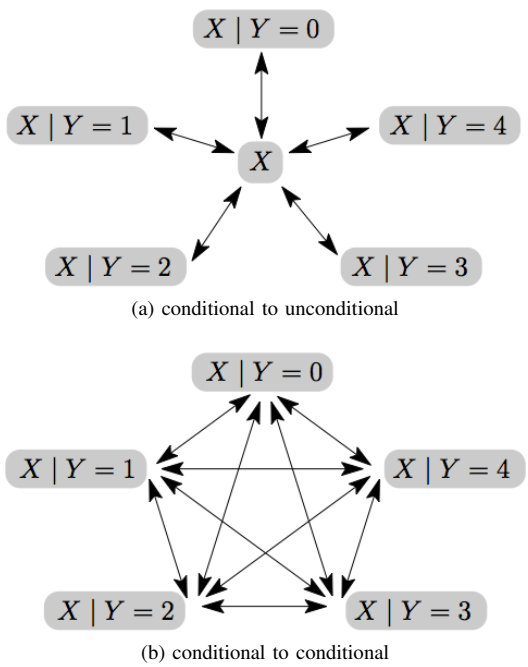
The remainder of this paper is organized as follows. Section II introduces a conditional-to-conditional comparison to define interclass information. Section III derives some of its properties in relation to mutual information. That both MIA and IIA are *sound* as side-channel attacks is shown in Section IV. Section V carries out the theoretical comparison between MIA and IIA, with the help of numerical computation. The results are confirmed by simulations in Section VI. Section VII concludes.

II. Inter-Class Information

It is well-known [9] that mutual information can be seen as an expected Kullback-Leibler distance

$$I(X;Y) = \mathbb{E}\{D\big(p(\cdot|Y)\|p(\cdot)\big)\} \tag{6}$$

between the conditional distribution $p(x|y)$ representative of the "class" $Y = y$ and the unconditional distribution $p(x) = \mathbb{E}\{p(x|Y)\}$, which is obtained as an expectation over *all* classes. The comparison strategy is illustrated in Fig. 1a.



(a) conditional to unconditional

(b) conditional to conditional

Fig. 1. Two methods of comparison between probability distributions (the "distance" being depicted as an arrow).

In [10], the authors suggested an alternative *inter-class* comparison strategy in which the conditional distributions are compared between themselves rather than with the global distribution of the leakage, as illustrated in Fig. 1b. Simulations have shown that a side-channel attack based on the more direct conditional-to-conditional strategy can be more efficient in the case of the Kolmogorov-Smirnov distance [10]. A detailed theoretical comparison is also carried out in [11] for the Kolmogorov-Smirnov test on one-bit leakages. In the case of mutual information, we obtain the following definition.

Definition 1. The *inter-class information* between random variables X and Y is defined as

$$I\!I(X;Y) = \frac{1}{2}\,\mathbb{E}\{D\big(p(\cdot|Y)\|p(\cdot|Y')\big)\} \tag{7}$$

where Y' is an independent copy of Y and the expectation is made over (Y, Y'). The $1/2$ factor makes up for double counts $(Y, Y') = (y, y')$ and (y', y).

From (1)–(2), $p(x|y)$ is a conditional density and $Y = Y(k)$ assumes discrete values, and we can write

$$I\!I(X;Y) = \frac{1}{2}\sum_y\sum_{y'} p(y)p(y')\int p(x|y)\log\frac{p(x|y)}{p(x|y')}\,\mathrm{d}x. \tag{8}$$

An alternative definition is that divergence in (6) is replaced by a symmetric version:

Proposition 1.

$$I\!I(X;Y) = \mathbb{E}\{\Delta\big(p(\cdot|Y)\|p(\cdot)\big)\} \tag{9}$$

where $\Delta(p\|q) = \frac{1}{2}\big(D(p\|q) + D(q\|p)\big)$ *is the symmetrized Kullback-Leibler or Jeffreys divergence*[1].

Proof: Write $\dfrac{p(x|y)}{p(x|y')} = \dfrac{p(x|y)}{p(x)} \times \dfrac{p(x)}{p(x|y')}$ in (8) and expand. ■

Note. Since $D(q\|p) \geq 0$, hence $\Delta(p\|q) \geq \frac{1}{2}D(p\|q)$ one has[2]

$$I\!I(X;Y) \geq \frac{1}{2}I(X;Y). \tag{10}$$

III. Information-Theoretic Properties

Just as mutual information can be written as a difference of two entropies, we have the following[3]

Proposition 2.

$$I\!I(X;Y) = \frac{1}{2}\big(H(X\|Y) - H(X|Y)\big) \tag{11}$$

where $H(X\|Y)$ *is a conditional* "cross-entropy" *defined as*

$$H(X\|Y) = -\int\sum_y p(x)p(y)\log p(x|y)\,\mathrm{d}x. \tag{12}$$

Proof: Since $\Delta(p\|q) = \frac{1}{2}\int(p-q)\log\frac{p}{q}$, from (9) we can write

$$I\!I(X;Y) = \frac{1}{2}\sum_y p(y)\int\big(p(x|y) - p(x)\big)\log\frac{p(x|y)}{p(x)} \tag{13}$$

$$= \frac{1}{2}\sum_y p(y)\int\big(p(x|y) - p(x)\big)\log p(x|y) \tag{14}$$

where the simplification is due to the fact that $\mathbb{E}\{p(x|Y)\} = p(x)$. The announced formula follows at once. ■

[1] Also known as the *interclass* divergence in cognitive and neural science [12].

[2] During the finalization of this paper we became aware that in fact $2I\!I(X;Y) - I(X;Y)$ is the lautum information [13].

[3] When X and Z follow densities, the entropies in question are differential entropies. We keep the notation H (in place of h) to encompass the discrete case for which the same formal relations hold.

Interclass information has other properties similar to well-known properties of mutual information.

Proposition 3.

(a) $I\!I(X;Y) \geq 0$ *with equality* $I\!I(X;Y) = 0$ *if and only if* X, Y *are independent;*

(b) Symmetry: $I\!I(X;Y) = I\!I(Y;X)$*;*

(c) Data processing inequality: if $X - Y - Z$ *is a Markov chain, then* $I\!I(X;Y) \geq I\!I(X;Z)$ *and* $I\!I(Y;Z) \geq I\!I(X;Z)$*.*

Proof: (a) is obvious from (10); (b) It is well-known that (6) is symmetric in X, Y, and easily seen that the same holds for the quantity $\mathbb{E}\{D\big(p(x)\|p(x|Y)\big)\}$. Hence (9) is also symmetric in X, Y. (c) Suppose $X - Y - Z$ is Markov. It is sufficient to prove that $I\!I(X;Y) \geq I\!I(X;Z)$, the other inequality being a consequence of the fact that $Z - Y - X$ is also Markov. We use Prop. 2. On one hand, by the (usual) data processing inequality [9, Thm 2.8.1], one has $I(X;Y) \geq I(X;Z)$ or $H(X|Y) \leq H(X|Z)$. On the other hand, using the Markov property $p(x|y) = p(x|y,z)$,

$$H(X\|Y) = -\,\mathbb{E}\int \sum_y p(x)p(y|Z)\log p(x|y,Z)\,\mathrm{d}x \quad (15)$$

$$\geq -\,\mathbb{E}\int p(x)\log\sum_y p(y|Z)p(x|y,Z)\,\mathrm{d}x \quad (16)$$

$$= -\,\mathbb{E}\int p(x)\log p(x|Z)\,\mathrm{d}x = H(X\|Z) \quad (17)$$

where the inequality comes from the concavity of the logarithm. Thus $H(X\|Y) \geq H(X\|Z)$, which combined with $H(X|Y) \leq H(X|Z)$ yields the announced inequality. ■

Although the above properties of interclass and mutual informations are similar, it is important to note that their behavior can be very different. This is best illustrated with an example.

Example 1. Assume that X, Y are zero-mean (jointly) Gaussian with variance σ^2 and correlation coefficient $\rho(X,Y) = \rho$. It is easily seen that $H(X|Y)$ is the differential entropy of a Gaussian of variance $= \sigma^2(1-\rho^2)$. On one hand, this gives the well-known formula for mutual information:

$$I(X;Y) = \frac{1}{2}\log\frac{1}{1-\rho^2}. \quad (18)$$

On the other hand, cross-entropy $H(X\|Y)$ can be similarly calculated:

$$H(X\|Y) = -\int p(y)\,\mathbb{E}\{\log p(X|y)\}\,\mathrm{d}y \quad (19)$$

$$= \frac{1}{2}\log\big(2\pi\sigma^2(1-\rho^2)\big) \quad (20)$$

$$+\,\frac{\log e}{2\sigma^2(1-\rho^2)}\int p(y)\,\mathbb{E}\{(X-\rho y)^2\}\,\mathrm{d}y$$

$$= H(X|Y) + (\log e)\Big(-\frac{1}{2} + \frac{\sigma^2 + \rho^2\,\mathbb{E}(Y^2)}{2\sigma^2(1-\rho^2)}\Big) \quad (21)$$

$$= H(X|Y) + (\log e)\frac{\rho^2}{1-\rho^2} \quad (22)$$

which from (11) gives the corresponding formula for interclass information:

$$I\!I(X;Y) = \frac{\log e}{2}\,\frac{\rho^2}{1-\rho^2}. \quad (23)$$

Both (18) and (23) vanish for $\rho = 0$ (independence) and are infinite for $|\rho| = 1$ (linear dependence). However, as $|\rho| \to 1$ interclass information is increasing much faster than mutual information. This shows, in particular, that *no* inequality of the form

$$I\!I(X;Y) \leq c \cdot I(X;Y)$$

can hold generally for some constant c—although the opposite type of inequality holds, see (10). In fact, as $|\rho| \to 1$, $\lambda = 1/(1-\rho^2) \to +\infty$ and the fraction $I\!I(X;Y)/I(X;Y) = (\log e)(\lambda - 1)/\log\lambda$ is unbounded.

This observation is similar to the well-known inequivalence between independence $I(X;Y) = 0$ and decorrelation $\rho(X,Y) = 0$, which means that no general inequality of the form $I(X;Y) \leq c \cdot \rho(X,Y)$ can hold. This suggests that interclass information can be somehow more sensitive to dependence than mutual information.

IV. Attack Soundness

We now come back to the side-channel scenario (1)–(5). Hereafter we use the convenient notations

$$Y = Y(k) \qquad \text{and} \qquad Y^* = Y(k^*). \quad (24)$$

To simplify the theoretical study of distinguishers, it is customary [14] to assume that $\hat{f} = f$ in (2). As a consequence we assume that $X = Y^* + Z$. It is then easily seen that for the correct key, the conditional distribution is simply $p(x|y^*) = \phi(x - y^*)$, while for any key $k \neq k^*$, by the law of total probability,

$$p(x|y) = \sum_{y^*} p(y^*|y)p(x|y,y^*) = \sum_{y^*} p(y^*|y)\phi(x-y^*) \quad (25)$$

is a *nontrivial* linear mixture of shifted noise densities [2], where at least one coefficient $p(y^*|y)$ is < 1.

Definition 2 (see e.g., [2])**.** A side-channel attack is *sound* if the corresponding (theoretical) distinguisher $\mathcal{D}(k)$ is maximum when k is the correct key:

$$\mathcal{D}(k) < \mathcal{D}(k^*) \qquad (\forall k \neq k^*). \quad (26)$$

Soundness is a basic prerequisite for the success of an attack (3). It implies that the success rate will eventually converge to 100% as the number of measurements increases indefinitely [2], [15]. MIA was proved sound in [16] under Gaussian noise using the data processing inequality. As seen below, the proof extends easily to any type of additive noise. That IIA is also sound is proved similarly using the data processing inequality for interclass information (Prop. 3).

Proposition 4. *Under the above assumptions, MIA and IIA are both sound.*

Proof: Assume $k \neq k^*$. From (1)–(2) it is easily seen that $Y - Y^* - X$ forms a Markov chain. By the data processing

inequalities, $I(X;Y^*) \geq I(X;Y)$ and $\mathit{II}(X;Y^*) \geq \mathit{II}(X;Y)$. It remains to prove that these inequalities are strict.

Since mixing increases entropy [9, Thm 2.7.3],

$$H(X|Y) > \mathbb{E}\sum_{y^*} p(y^*|Y)H(y^* + Z) = H(Z) = H(X|Y^*)$$

where the inequality is strict because the linear mixture (25) is nontrivial. This shows that $I(X;Y^*) > I(X;Y)$, i.e., MIA is sound. Similarly, in the proof of Prop. 3 (c) above (written for the Markov chain $X - Y^* - Y$), the fact that the linear mixture (25) is nontrivial implies that (16) is a strict inequality by the *strict* concavity of the logarithm. This proves that $H(X\|Y^*) > H(X|Y)$, which combined with $H(X|Y) > H(X|Y^*)$ shows that $\mathit{II}(X;Y^*) > \mathit{II}(X;Y)$, i.e., IIA is sound. ■

V. MIA vs. IIA under Gaussian Noise

In this section, we compare the performances of MIA vs. IIA under a Gaussian noise assumption $Z \sim \mathcal{N}(0, \sigma^2)$. We first compare $I(X;Y^*)$ and $\mathit{II}(X;Y^*)$.

Lemma 1.

$$\mathit{II}(X, Y^*) = \frac{\log e}{2}\,\frac{\sigma^2_{Y^*}}{\sigma^2} \tag{27}$$

where $\sigma^2_{Y^*}$ denotes the variance of Y^*.

Proof: Expanding the relation $p(x) = \sum_{y'} p(y')p(x|y')$ in (12) and plugging (25) we have

$$H(X\|Y^*) = -\,\mathbb{E}\sum_{y'} p(y')\int p(x|y')\log p(x|Y^*)\,\mathrm{d}x \tag{28}$$

$$= -\,\mathbb{E}\int \phi(x - Y'^*)\log\phi(x - Y^*)\,\mathrm{d}x \tag{29}$$

where Y'^* is an independent copy of Y^*, where (Y^*, Y'^*) is independent of Z. Making the change of variables $z = x - y'^*$, the integral becomes

$$\begin{aligned} -\,\mathbb{E}\int \phi(z)\log\phi(z + Y'^* - Y^*)\,\mathrm{d}z \\ = \frac{1}{2}\log(2\pi\sigma^2) + \frac{\log e}{2\sigma^2}\,\mathbb{E}\{(Z + Y^* - Y'^*)^2\} \\ = H(Z) + \frac{\log e}{2\sigma^2}\,\mathbb{E}\{(Y^* - Y'^*)^2\}. \end{aligned} \tag{30}$$

Since $H(X|Y^*) = H(Z)$ in (11) we obtain

$$\mathit{II}(X;Y^*) = \frac{\log e}{4\sigma^2}\,\mathbb{E}\{(Y^* - Y'^*)^2\} = \frac{\log e}{2\sigma^2}\sigma^2_{Y^*}. \tag{31}$$

■

Theorem 1. *For the correct key k^*, we have the inequality*

$$\mathit{II}(X, Y^*) \geq I(X, Y^*). \tag{32}$$

Proof: It is easily seen (as in the proof of Shannon's capacity formula) that

$$I(X;Y^*) = H(X) - H(X|Y^*) = H(X) - H(Z) \tag{33}$$

$$\leq \frac{1}{2}\log\big(\sigma^2_X/\sigma^2\big) = \frac{1}{2}\log\Big(1 + \frac{\sigma^2_{Y^*}}{\sigma^2}\Big). \tag{34}$$

The results follows at once from the well-known inequality $\log x \leq (\log e)(x - 1)$. ■

Thanks to Theorem 1 we can compare the performances of MIA and IIA. In view of (3)-(5), it is clear that difference values $\mathcal{D}(k^*) - \mathcal{D}(k)$ for $k \neq k^*$ play a important rôle. In [15], the authors show that the first order exponent of the success rate SR as the number of measurements increases is given by the so-called *success exponent* $\mathrm{SE} \sim -\log(1 - \mathrm{SR})$:

$$\mathrm{SE} = \min_{k \neq k^*} \frac{\mathcal{D}(k^*) - \mathcal{D}(k)}{2V(k, k^*)} \tag{35}$$

where $V(k, k^*) = \mathrm{Var}(\widehat{\mathcal{D}}(k^*) - \widehat{\mathcal{D}}(k))$ denotes the estimation variance of the distinguisher difference for keys k, k^*.

The numerator in (35) is simply $I(X, Y^*) - I(X;Y)$ for MIA and $\mathit{II}(X, Y^*) - \mathit{II}(X;Y)$ for IIA. Unfortunately, for $k \neq k^*$ the calculation of $I(X;Y)$ or $\mathit{II}(X;Y)$ is intricate as it requires the integration of Gaussian mixtures of the form (25). Therefore, we rely on numerical integration. We checked numerically that $X = Y(k^*) + Z$ will be much less dependent on $Y(k)$ than on $Y(k^*)$, due to the nonlinearity of f [17]. The dominant term is then $I(X, Y^*)$ for MIA and $\mathit{II}(X, Y^*)$ for IIA. Therefore, from Theorem 1, it appears that the contribution of the numerator of the success exponent is in favor of IIA.

However, due to the denominator in (35), a complete characterization of performance also depends on the method used to *estimate* mutual or interclass information from the available data. With the help of numerical computation, we have found that the normalizing factor $V(k, k^*)$ assumes very comparable values for MIA and IIA, for a given estimation method. Fig. 2 illustrates the resulting success exponent for the same kernel density estimation of the p.d.f.'s $p(x|y)$ and the same leakage model (1).

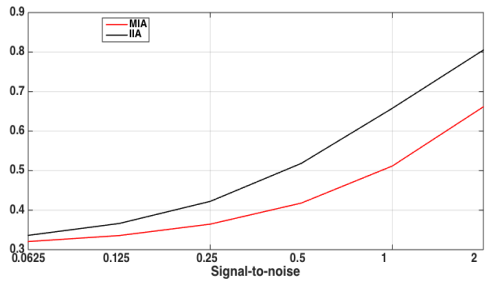


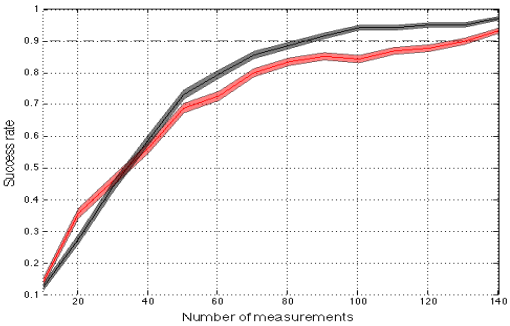
Fig. 2. Success exponent for MIA (red) and IIA (black) as a function of the SNR $1/\sigma^2$ for the PRESENT implementation. Here f in (1) is the composition of the inverse PRESENT substitution box and Hamming weight, a common leakage model in the side-channel analysis literature [3]–[5].

From the figure we expect IIA to outperform MIA for relatively low noise. For small SNR, however, the curves tend to the same asymptote.

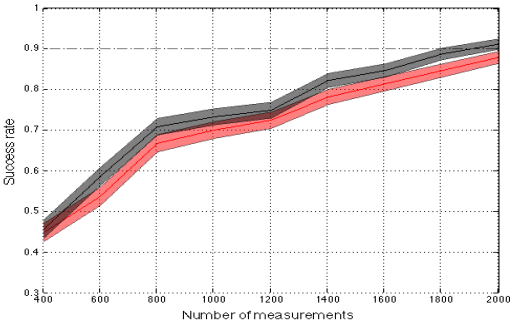
VI. Simulation Results

We carried out both attacks MIA and IIA and computed the resulting success rates over a set of 230 independent experiments for $\sigma = 1$ and 120 independent experiments

for $\sigma = 4$, where the secret key is chosen randomly in each experiment. The *same* leakage model, kernel density estimation of p.d.f.'s and data set were used in both cases to provide a fair comparison. We have also computed error bars as suggested in [10]: since the success rate SR follows a binomial distribution for multiple retries R with deviation $\delta = \sqrt{\frac{\mathrm{SR}(1-\mathrm{SR})}{R}}$, we obtain confidence intervals of the form $[\mathrm{SR} - \delta, \mathrm{SR} + \delta]$. that are drawn as vertical error bars in Fig. 3.



(a) $\sigma^2 = 1$



(b) $\sigma^2 = 16$

Fig. 3. Success rate for MIA (red) and IIA (black) for the PRESENT implementation (same leakage model as in Fig. 2).

Fig. 3 shows that IIA reaches the threshold of success rate $= 90\%$ before MIA. As predicted by the theoretical study of the preceding section, the difference between MIA and IIA is smaller for low SNR than for high SNR. Thus, the empirical results confirm the theoretical ones.

VII. Conclusion

A new information-theoretic side-channel distinguisher based on *interclass information* is investigated in relation to mutual information. A theoretical comparative study of MIA and IIA is carried out and validated by simulations. It is shown in particular that IIA can outperform MIA for large values of SNR. To our knowledge, this is the first time that a fairly complete theoretical comparison between two information-based side-channel distinguishers can be carried out.

Interclass information is similar to mutual information but uses a different comparing strategy between the underlying conditional distributions, which can make it more sensitive to dependence. As such, it appears as an interesting information-theoretic quantity on its own for which we are not aware of any previous application to a practical problem.

Acknowledgments

The authors would like to thank Houssem Maghrebi (Morpho SAS) for preliminary works on this subject.

References

[1] P. C. Kocher, J. Jaffe, and B. Jun, "Differential Power Analysis," in *Proceedings of CRYPTO'99*, ser. LNCS, vol. 1666. Springer-Verlag, 1999, pp. 388–397.

[2] E. Prouff and M. Rivain, "Theoretical and practical aspects of mutual information-based side channel analysis," *International Journal of Applied Cryptography (IJACT)*, vol. 2, no. 2, pp. 121–138, 2010.

[3] E. Prouff, M. Rivain, and R. Bevan, "Statistical Analysis of Second Order Differential Power Analysis," *IEEE Trans. Computers*, vol. 58, no. 6, pp. 799–811, 2009.

[4] A. Heuser, O. Rioul, and S. Guilley, "Good is Not Good Enough — Deriving Optimal Distinguishers from Communication Theory," in *CHES*, ser. Lecture Notes in Computer Science, L. Batina and M. Robshaw, Eds., vol. 8731. Springer, 2014. [Online]. Available: http://dx.doi.org/10.1007/978-3-662-44709-3

[5] É. Brier, C. Clavier, and F. Olivier, "Correlation Power Analysis with a Leakage Model," in *CHES*, ser. LNCS, vol. 3156. Springer, August 11–13 2004, pp. 16–29, Cambridge, MA, USA.

[6] C. Whitnall, E. Oswald, and L. Mather, "An Exploration of the Kolmogorov-Smirnov Test as a Competitor to Mutual Information Analysis," in *CARDIS*, ser. Lecture Notes in Computer Science, E. Prouff, Ed., vol. 7079. Springer, 2011, pp. 234–251.

[7] B. Gierlichs, L. Batina, P. Tuyls, and B. Preneel, "Mutual information analysis," in *CHES, 10th International Workshop*, ser. Lecture Notes in Computer Science, vol. 5154. Springer, August 10-13 2008, pp. 426–442, Washington, D.C., USA.

[8] V. D. Goppa, "Nonprobabilistic mutual information without memory," *Probl. Cont. Information Theory*, vol. 4, pp. 97–102, 1975.

[9] T. M. Cover and J. A. Thomas, *Elements of information theory*, 2nd ed. John Wiley & Sons, 1991, 2006.

[10] H. Maghrebi, O. Rioul, S. Guilley, and J.-L. Danger, "Comparison between Side Channel Analysis Distinguishers," in *ICICS*, ser. LNCS, T. W. Chim and T. H. Yuen, Eds., vol. 7618. Springer, October 29-31 2012, pp. 331–340, Hong Kong.

[11] A. Heuser, O. Rioul, and S. Guilley, "A Theoretical Study of Kolmogorov-Smirnov Distinguishers — Side-Channel Analysis vs. Differential Cryptanalysis," in *COSADE 2014, Paris, France, April 13-15, 2014.*, ser. Lecture Notes in Computer Science, E. Prouff, Ed., vol. 8622. Springer, 2014, pp. 9–28. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-10175-0_2

[12] G. Saon and M. Padmanabhan, "Minimum Bayes Error Feature Selection for Continuous Speech Recognition," in *Advances in Neural Information Processing Systems 13*. MIT Press, 2000, pp. 800–806.

[13] D. P. Palomar and S. Verdú, "Lautum information," *IEEE Trans. Inform. Theory*, vol. 54, no. 3, pp. 964–975, 2008.

[14] A. Thillard, E. Prouff, and T. Roche, "Success through Confidence: Evaluating the Effectiveness of a Side-Channel Attack," in *CHES*, ser. Lecture Notes in Computer Science, G. Bertoni and J.-S. Coron, Eds., vol. 8086. Springer, 2013, pp. 21–36.

[15] S. Guilley, A. Heuser, and O. Rioul, "A Key to Success – Success Exponents for Side-Channel Distinguishers," in *Progress in Cryptology - INDOCRYPT 2015 - 16th International Conference on Cryptology in India*, December 6-10 2015, Bangalore, India.

[16] A. Moradi, N. Mousavi, C. Paar, and M. Salmasizadeh, "A Comparative Study of Mutual Information Analysis under a Gaussian Assumption," in *WISA (Information Security Applications, 10th International Workshop)*, ser. Lecture Notes in Computer Science, vol. 5932. Springer, August 25-27 2009, pp. 193–205, Busan, Korea.

[17] E. Prouff, "DPA Attacks and S-Boxes," in *Fast Software Encryption*, ser. LNCS, vol. 3557. Springer-Verlag, february 2005, pp. 424–441, paris, France.