

Some Results about the Distinction of Side-Channel Distinguishers based on Distributions

Housseem MAGHREBI, Sylvain GUILLEY
Olivier RIOUL, Jean-Luc DANGER

<first_name.family_name@telecom-paristech.fr>

Institut Mines-Télécom / Télécom-ParisTech
CNRS – LTCI (UMR 5141)



Cryptarchi'12, St-Etienne France
Friday, June, 22th, 2012

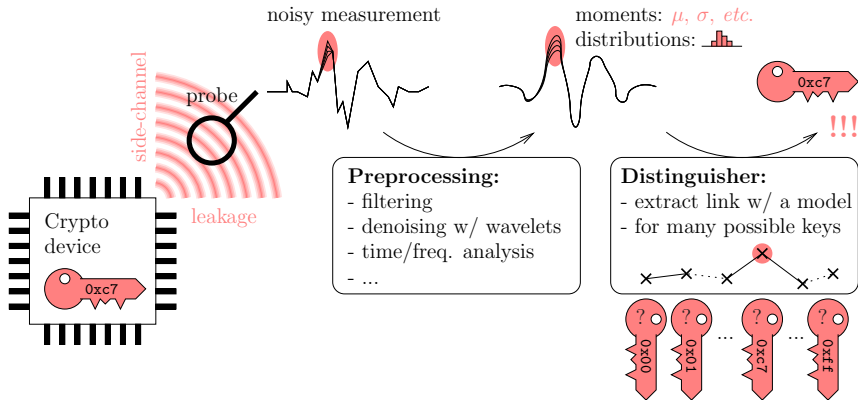
Presentation Outline

- 1 Introduction
- 2 Inter-Class Information Analysis (IIA)
- 3 Comparison of MIA and IIA: Theory
- 4 Comparison of MIA and IIA: Simulations
- 5 Conclusions and Perspectives

Presentation Outline

- 1 Introduction
- 2 Inter-Class Information Analysis (IIA)
- 3 Comparison of MIA and IIA: Theory
- 4 Comparison of MIA and IIA: Simulations
- 5 Conclusions and Perspectives

Side Channel Analysis



Classical distinguishers

- **Covariance** used for Differential Power Analysis (Kocher *et al.* [5])
- **Correlation** calculated with the Pearson's coefficient (Brier *et al.* [1])
- **Likelihood**, as in Template Attacks (Chari *et al.* [2])
- **Mutual Information Analysis** (Gierlichs *et al.* [4])

Information theoretic definitions

Notation : Let $p(X = x)$ be abridged $p(x)$

- Mutual information, a general measure of dependence:

$$I(X; Y) = H(X) - H(X | Y) \geq 0$$

- Shannon's entropy, a measure of information:

$$H(X) = - \sum_{i=1}^n p(x_i) \cdot \log p(x_i)$$

- Conditional entropy:

$$H(X | Y) = - \sum_{i=1}^n \sum_{j=0}^m p(x_i, y_j) \cdot \log p(x_i | y_j)$$

Probability density function estimation

- Non-parametric methods:
 - Histogram
 - Kernel density:

$$f(x) = \frac{1}{nh} \sum_{i=0}^n k\left(\frac{x - x_i}{h}\right),$$

where k is the kernel function and h is the "bandwidth"

- Parametric method:

$$f(x) = \sum_{i=0}^{n-1} w_i \cdot \mathcal{N}(x, \mu_i, \sigma_i),$$

where w_i , μ_i and σ_i are the weight, the mean and the standard deviations, respectively.

Presentation Outline

- 1 Introduction
- 2 **Inter-Class Information Analysis (IIA)**
- 3 Comparison of MIA and IIA: Theory
- 4 Comparison of MIA and IIA: Simulations
- 5 Conclusions and Perspectives

Conventional metrics

- The metric is an average distance \mathcal{D} over Y between $p(x|Y)$ and its average $\mathbb{E}(p(x|Y)) = p(x)$:

$$M_0 = \mathbb{E}(\mathcal{D}(p(x|Y), p(x))) = \sum_y p(y) \mathcal{D}(p(x|y), p(x))$$

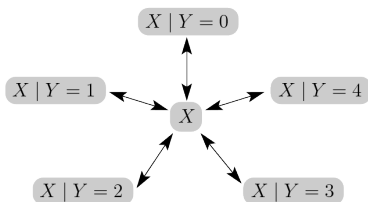
- Examples of distances \mathcal{D} are:
 - MIA : Kullback-Leibler divergence
 - Total variation $D(p, q) = \sum |p - q|$
 - KSA : Kolmogorov-Smirnov distance (cumulative distribution functions).

Inter-class metric

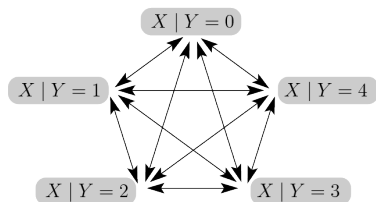
The inter-class is a distance \mathcal{D} between two classes $p(x|Y)$ and $p(x|Y')$

$$M_I = \frac{1}{2} \mathbb{E}(\mathcal{D}(p(x, Y), p(x, Y'))) = \frac{1}{2} \sum_{y, y'} p(y)p(y') \mathcal{D}(p(x|y), p(x|y'))$$

Mutual Analysis



Inter-class Analysis



Contributions

- 1 Inter-class Information Analysis (IIA), a new test for SCA proved to be efficient when exploiting several kinds of leakages
- 2 A theoretical study to compare the inter-class information with the mutual information
- 3 Some experiments in order to evaluate IIA (simulated attacks against unprotected and protected AES with Boolean masking)

First definition

$$I(X; Y) = \frac{1}{2} \mathbb{E}(D_{\text{KL}}[X | Y || X | Y'])$$

Where $D_{\text{KL}}[X | Y]$ is the Kullback Leibler divergence

Second definition



$$I(X; Y) = \frac{H'(X | Y) - H(X | Y)}{2}$$

- Recall the conditional entropy is averaged over the joint distribution: $H(X | Y) = \sum_{x,y} p(x,y) \log \frac{1}{p(x|y)}$
- The “cross-entropy” is averaged over the product distribution: $H'(X | Y) = \sum_{x,y} p(x)p(y) \log \frac{1}{p(x|y)}$

Properties

- **Relation to Mutual Information** Since $2I(X; Y) = \mathbb{E}(D_{\text{KL}}[X | Y || X] + D_{\text{KL}}[X || X | Y]) = I(X; Y) + \mathbb{E}(D_{\text{KL}}[X || X | Y])$

$$2I(X; Y) \geq I(X; Y)$$

- **Symmetry** Since

$$2I(X; Y) = I(X; Y) + \sum_{x,y} p(X=x)p(Y=y) \log \frac{p(X=x)p(Y=y)}{p(X=x, Y=y)}$$

$$I(X; Y) = I(Y; X) .$$

- **Independence** $I(X; Y) = 0$ implies $I(X; Y) = 0$ (independence)
Conversely, if X and Y are independent,
 $D_{\text{KL}}[X || X | Y] = D_{\text{KL}}[X | Y || X] = D_{\text{KL}}[X || X] = 0$, hence
 $I(X; Y) = 0$ So:

$$I(X; Y) = 0 \iff X, Y \text{ are independent} .$$

Soundness Proof

Let k^* the good key and k another key $k \neq k^*$,

The leakage X is continuous, the model Y is discrete. The PDF of $X = Y(k^*) + N = \phi(Z, k^*) + N$ (where N is an arbitrary RV modelling the noise)

For each y , $X|y$ is a PDF mixture whose coefficients depend on y , and whose modes y^* take values on a finite set

- **MI:** Moradi et al [6] relies on the fact that $Y \rightarrow Y^* \rightarrow X$ (where $k \neq k^*$) forms a Markov chain; then by the well-known information inequality for Markov chains, $I(X; Y^*) \geq I(X; Y)$.
- **II:** Since $H(X|Y) > H(N) = H(X|Y^*)$, we end up with

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

Presentation Outline

- 1 Introduction
- 2 Inter-Class Information Analysis (IIA)
- 3 Comparison of MIA and IIA: Theory**
- 4 Comparison of MIA and IIA: Simulations
- 5 Conclusions and Perspectives

IIA and MIA under the Gaussian mixture

When using the good key and for a Gaussian noise N , we have:

$$II(X; Y^*) = \frac{H'(X | Y^*) - H(N)}{2} = \frac{\log e}{2} \cdot \frac{\sigma_{Y^*}^2}{\sigma^2},$$

$$I(X; Y^*) = \frac{H(Y^* + N) - H(N)}{2} = \frac{1}{2} \log \frac{\sigma_{Y^*}^2 + \sigma^2}{\sigma^2}.$$

Key Inequality

$$I(X; Y^*) = \frac{1}{2} \log \frac{\sigma_{Y^*}^2 + \sigma^2}{\sigma^2} \leq \frac{\log e}{2} \frac{\sigma_{Y^*}^2}{\sigma^2} = II(X; Y^*)$$

The difference $II(X; Y^*) - I(X; Y^*)$ increases with the SNR $\frac{\sigma_{Y^*}^2}{\sigma^2}$.

- With the help of the Markov chain condition on $Y \rightarrow Y^* \rightarrow X$, we have:

$$\begin{cases} I(X; Y) \leq I(Y; Y^*) \ll I(X; Y^*) \\ H(X; Y) \leq H(Y; Y^*) \ll H(X; Y^*) \end{cases}$$

- Noting the nearest-rival distinguishability by ζ :

$$\begin{cases} \zeta_{\text{MIA}}(k) = I(X; Y(k^*)) - \max_{k \neq k^*} I(X; Y(k)) , \\ \zeta_{\text{IIA}}(k) = H(X; Y(k^*)) - \max_{k \neq k^*} H(X; Y(k)) \end{cases}$$

- By neglecting the second-order terms, we end up with:

$$\zeta_{\text{IIA}}(k) \gtrsim \zeta_{\text{MIA}}(k)$$

- For low SNR, the equality holds at the limit: both methods coincide.
- The convergence rate towards 0 of the failure probability is faster for IIA compared to MIA.

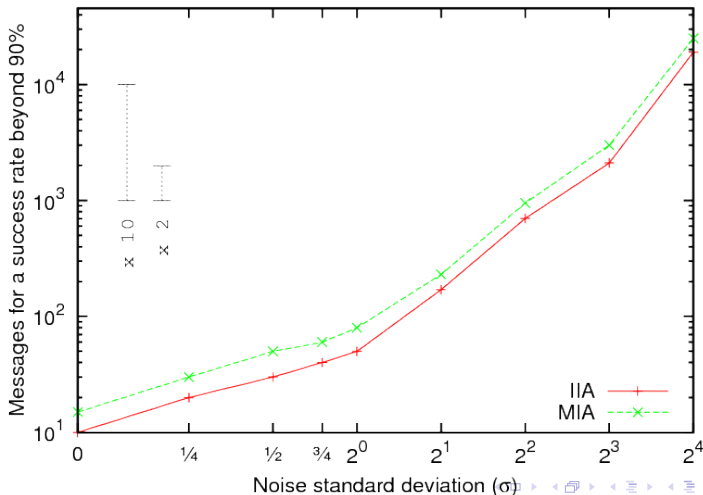
Presentation Outline

- 1 Introduction
- 2 Inter-Class Information Analysis (IIA)
- 3 Comparison of MIA and IIA: Theory
- 4 Comparison of MIA and IIA: Simulations**
- 5 Conclusions and Perspectives

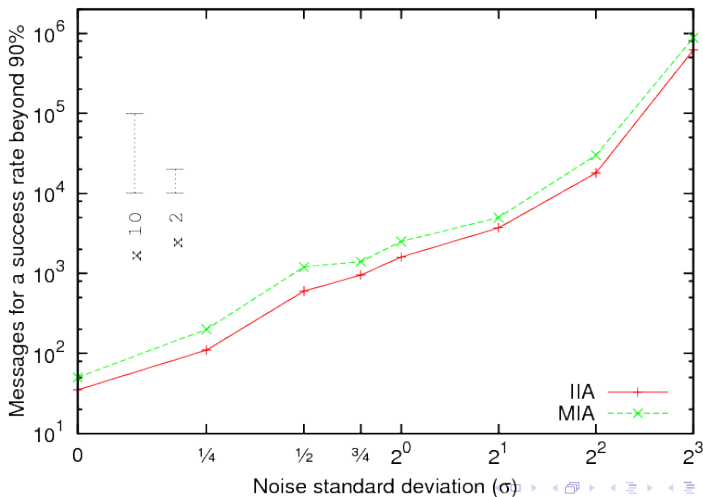
Evaluation Methodology

- *Leakage model*: Hamming distance
- *Target leakage*:
 - 1st-order leakage of unprotected implementation:
$$X = f(Y) + N = HW(Y) + N;$$
 - 2nd-order leakage of 1st-order Boolean masking scheme (Standaert *et al.* [7]):
$$X = f(Y, M) + N = HW(Y \oplus M) + HW(M) + N$$
- *Target secret*: AES S-box output, $Y = S(Z \oplus k)$
- *Adversary model*: known plaintext model by estimating $I(X; \psi(Y))$ and $II(X; \psi(Y))$
- *Prediction function*: Hamming Distance
- *PDF estimation method*: histogram-based PDF estimation

Unprotected implementation



Masked implementation



Presentation Outline

- 1 Introduction
- 2 Inter-Class Information Analysis (IIA)
- 3 Comparison of MIA and IIA: Theory
- 4 Comparison of MIA and IIA: Simulations
- 5 **Conclusions and Perspectives**

Conclusions

- New “Inter-class” concept to distinguish between various partitionings
- Mathematical analysis of the II metric and comparison with the MI metric
- Simulations confirm that IIA is more efficient than MIA in discriminating correct key assumptions:
 - For the usual HD leakage models
 - For unprotected and masked implementation

Perspectives

- Compare this distinguisher with an Inter-class KS analysis (IKSA)
- Try other kinds of masking (e.g., multiplicative or affine masking (Fumaroli et al [3]))

Thank you for your attention.
Any Questions?

References

- [1] 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.
- [2] Suresh Chari, Josyula R. Rao, and Pankaj Rohatgi.
Template Attacks.
In *CHES*, volume 2523 of *LNCS*, pages 13–28. Springer, August 2002.
San Francisco Bay (Redwood City), USA.
- [3] Guillaume Fumaroli, Ange Martinelli, Emmanuel Prouff, and Matthieu Rivain.
Affine Masking against Higher-Order Side Channel Analysis.
In Alex Biryukov, Guang Gong, and Douglas R. Stinson, editors, *Selected Areas in Cryptography*, volume 6544 of *LNCS*, pages 262–280. Springer, 2010.
- [4] 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.
- [5] Paul C. Kocher, Joshua Jaffe, and Benjamin Jun.
Differential Power Analysis.
In *CRYPTO*, volume 1666 of *LNCS*, pages pp 388–397. Springer, 1999.
- [6] Amir Moradi, Nima Mousavi, Christof Paar, and Mahmoud Salmasizadeh.
A Comparative Study of Mutual Information Analysis under a Gaussian Assumption.
In *WISA*, volume 5932 of *LNCS*, pages 193–205. Springer, August 25–27 2009.
Busan, Korea.
- [7] François-Xavier Standaert, Gaël Rouvroy, and Jean-Jacques Quisquater.

FPGA Implementations of the DES and Triple-DES Masked Against Power Analysis Attacks.
In *FPL*. IEEE, August 2006.
Madrid, Spain.