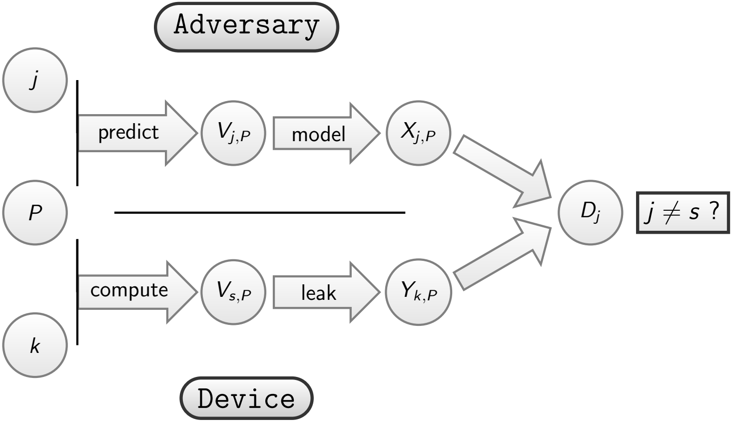
Mutual Information Analysis: How, When and Why?

Nicolas Veyrat-Charvillon and François-Xavier Standaert

Crypto Group
Université catholique de Louvain

CHES '09, September 3rd 2009

Side-channel analysis



Classical attacks

Classical solutions in non profiled SCA:

- Kocher's original DPA, at Crypto 1999
- Correlation attacks, at CHES 2004

So, what to do?

$X_0 =$

0	1	2	3	4
y_0	y_1	y_3	y_7	y_{15}
$\dots$	y_2	y_5	y_{11}	$\dots$
	y_4	y_6	y_{13}	
	y_8	y_9	y_{14}	
	$\dots$	y_{10}	$\dots$	
		y_{12}		
		$\dots$		

$X_1 =$

0	1	2	3	4
y_7	y_0	y_1	y_5	y_6
$\dots$	y_2	y_3	y_8	$\dots$
	y_{13}	y_4	y_{10}	
	y_{15}	y_9	y_{12}	
	$\dots$	y_{11}	$\dots$	
		y_{14}		
		$\dots$		

Pearson's correlation coefficient

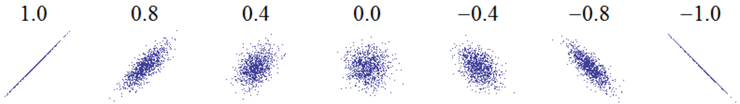
Measure of *linear* dependence between r.v.'s X and Y.

$$\rho(X, Y) = \frac{\mathrm{cov}(X, Y)}{\sigma_X \cdot \sigma_Y} = \frac{\mathrm{E}[XY] - \mathrm{E}[X] \cdot \mathrm{E}[Y]}{\sigma_x \cdot \sigma_Y}.$$

Pearson's correlation coefficient

Measure of *linear* dependence between r.v.'s X and Y.

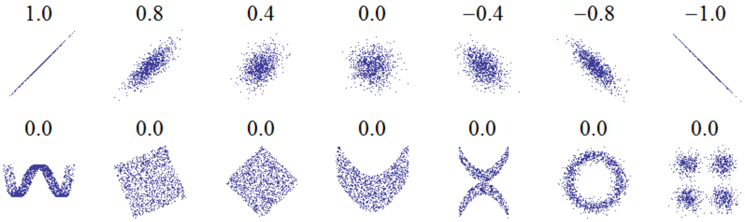
$$\rho(X,Y) = \frac{\mathrm{cov}(X,Y)}{\sigma_X \cdot \sigma_Y} = \frac{\mathrm{E}[XY] - \mathrm{E}[X]\cdot\mathrm{E}[Y]}{\sigma_x \cdot \sigma_Y}.$$



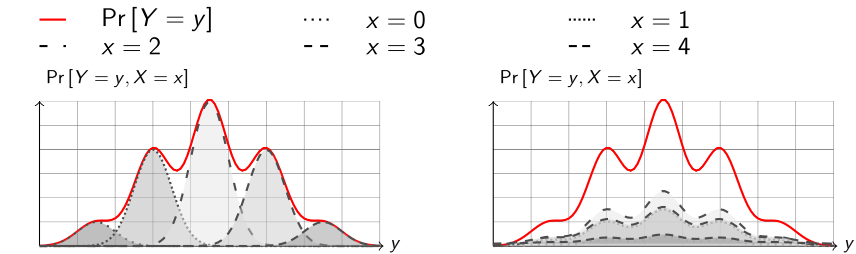
Pearson's correlation coefficient

Measure of *linear* dependence between r.v.'s X and Y.

$$\rho(X, Y) = \frac{\mathrm{cov}(X, Y)}{\sigma_X \cdot \sigma_Y} = \frac{\mathrm{E}[XY] - \mathrm{E}[X] \cdot \mathrm{E}[Y]}{\sigma_x \cdot \sigma_Y}.$$



So, what to do?



$X_0 =$ 0	1	2	3	4
y_0	y_1	y_3	y_7	y_{15}
...	y_2	y_5	y_{11}	...
	y_4	y_6	y_{13}	
	y_8	y_9	y_{14}	
	...	y_{10}	...	
		y_{12}		
		...		

$X_1 =$ 0	1	2	3	4
y_7	y_0	y_1	y_5	y_6
...	y_2	y_3	y_8	...
	y_{13}	y_4	y_{10}	
	y_{15}	y_9	y_{12}	
	...	y_{11}	...	
		y_{14}		
		...		

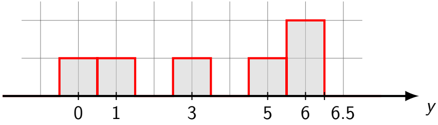
So, what to do?

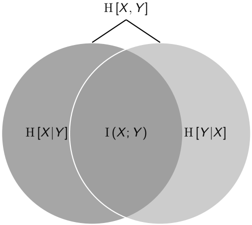
1. Estimate the probability density of the leakages
2. Test for a dependence between X and Y

Mutual information Analysis

Introduced at CHES 2008 by Gierlichs *& al.*

Aims at genericity: as little assumptions as possible about the leakage

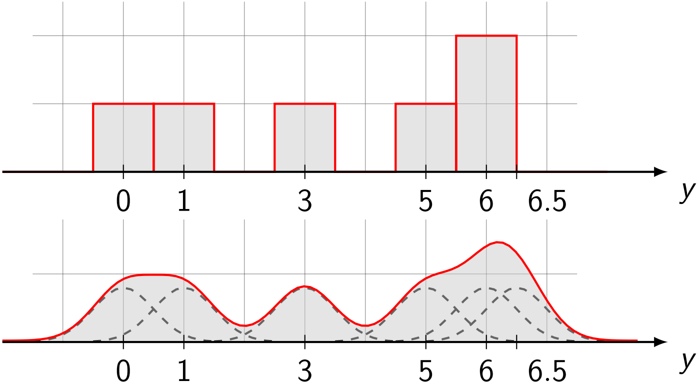




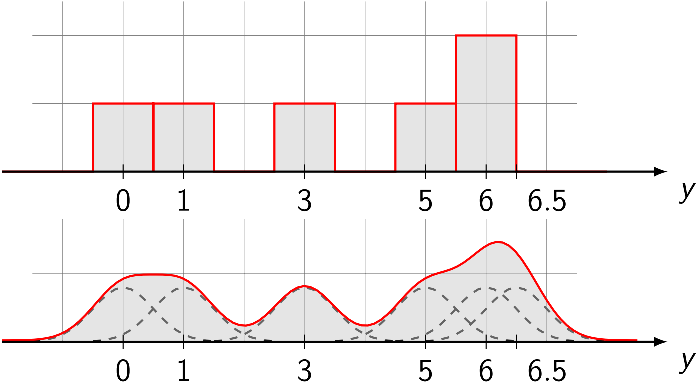
Outline

1. How to use MIA: the information theoretic toolbox
2. When to use it: MIA versus correlation
3. Why to use it: MIA as an evaluation metric

1 Estimation: Non-parametric methods



1 Estimation: Non-parametric methods



Well, non-parametric... bin width and bandwidth to choose

Information theoretic definitions

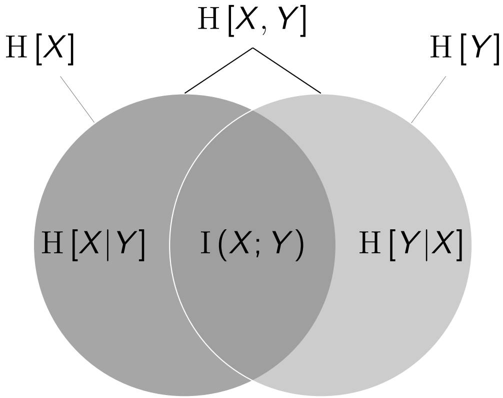
- Shannon's entropy, a measure of information

$$\mathrm{H}[X] = -\sum_{x \in \mathcal{X}} \Pr[X = x] \cdot \log(\Pr[X = x])$$

- Mutual information, a *general* measure of dependence

$$\mathrm{I}(X;Y) = \sum_{x \in \mathcal{X}, y \in \mathcal{Y}} \Pr[X = x, Y = y] \times \log\left(\frac{\Pr[X = x, Y = y]}{\Pr[X = x] \cdot \Pr[Y = y]}\right)$$

Information theoretic definitions



Information diagram

2 Test: Kullback-Leibler divergence

$$D_{\mathrm{KL}}\left(\mathrm{P}\|\mathrm{Q}\right) = \sum_{z \in \mathcal{Z}} \Pr\left[Z = z, Z \sim \mathrm{P}\right] \cdot \log \frac{\Pr\left[Z = z, Z \sim \mathrm{P}\right]}{\Pr\left[Z = z, Z \sim \mathrm{Q}\right]}$$

Relation to mutual information:

$$\begin{aligned} \mathrm{I}\left(X; Y\right) &= D_{\mathrm{KL}}\left(\Pr\left[X, Y\right] \| \Pr\left[X\right] \cdot \Pr\left[Y\right]\right) \\ &= \mathop{\mathrm{E}}_{x \in \mathcal{X}} \left(D_{\mathrm{KL}}\left(\Pr\left[Y | X = x\right] \| \Pr\left[Y\right]\right)\right) \end{aligned}$$

2 Test: F-divergences

$$\mathrm{I}_f(\mathrm{P},\mathrm{Q}) = \sum_{z \in \mathcal{Z}} \Pr[Z = z, Z \sim \mathrm{Q}] \cdot f\left(\frac{\Pr[Z = z, Z \sim \mathrm{P}]}{\Pr[Z = z, Z \sim \mathrm{Q}]}\right)$$

Different parameter functions f give different measures:

- Kullback-Leibler divergence $f(t) = t \log t$
- Inverse Kullback-Leibler $f(t) = -\log t$
- Pearson χ^2–divergence $f(t) = (t-1)^2$
- Hellinger distance $f(t) = 1 - \sqrt{t}$
- Total variation $f(t) = |t-1|$

1&2: Implicit pdf estimation

Empirical cumulative function:

$$F(x_t) = \frac{1}{n}\sum_{i=1}^{n} \chi_{x_i \leq x_t}, \text{ where } \chi_{x_i \leq x_t} = \begin{cases} 1 \text{ if } x_i \leq x_t \\ 0 \text{ otherwise.} \end{cases}$$

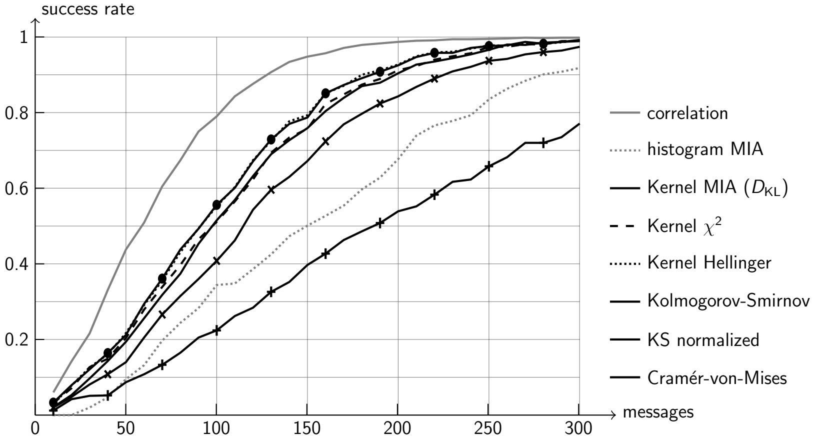
Two sample Kolmogorov-Smirnov test

$$D_{\mathrm{KS}}\left(P \| Q\right) = \sup_{x_t} \left| F_P(x_t) - F_Q(x_t) \right|$$

Two sample Cramér-von-Mises test

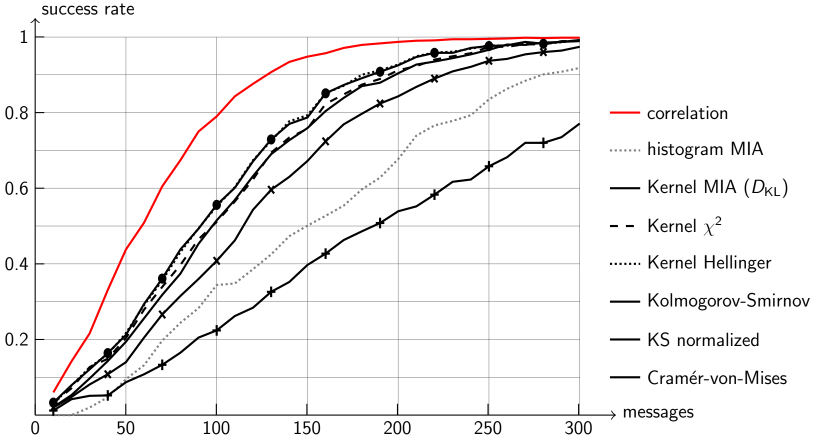
$$D_{\mathrm{CvM}}\left(P \| Q\right) = \int_{-\infty}^{+\infty} \left(F_P(x_t) - F_Q(x_t)\right)^2 dx_t$$

Experimental results



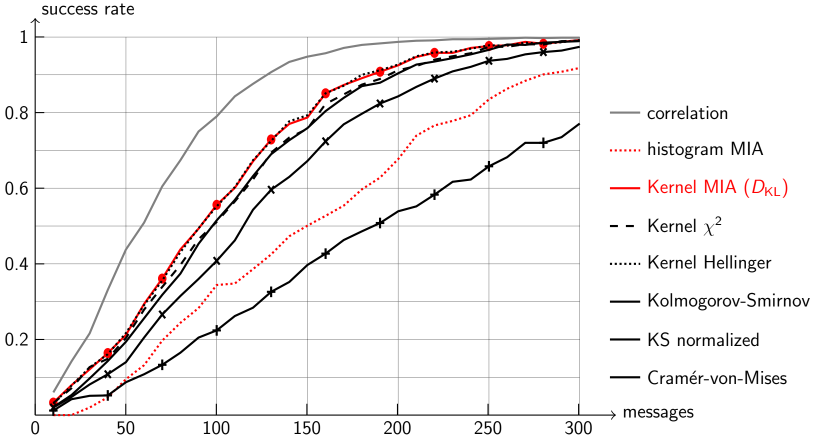
Success rate of different distinguishers

Experimental results



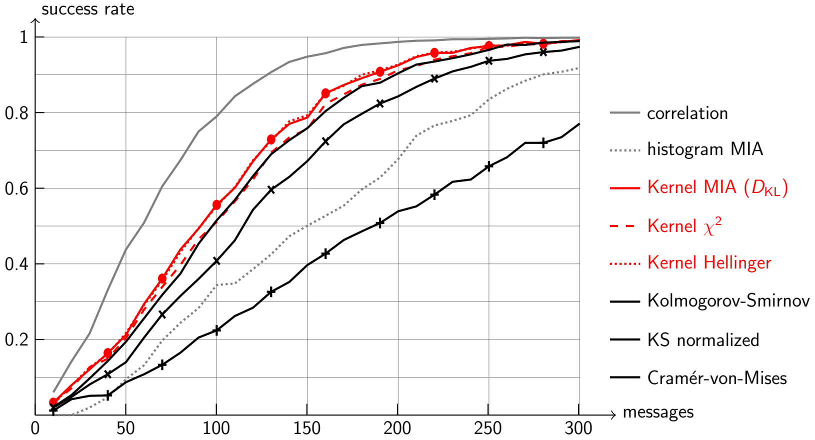
Success rate of different distinguishers

Experimental results



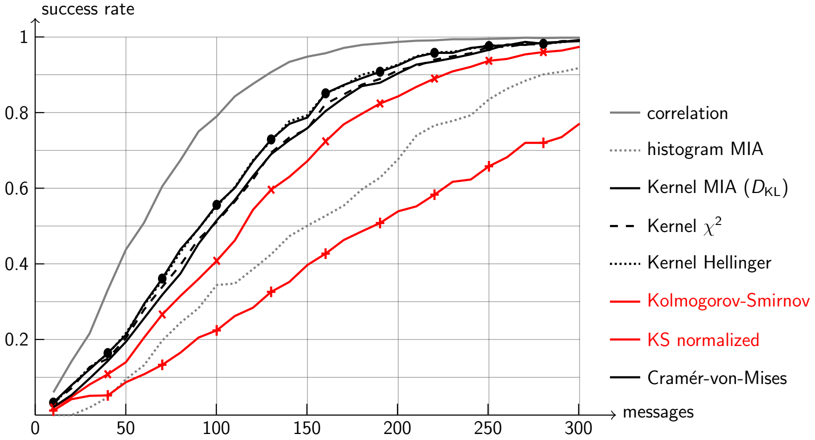
Success rate of different distinguishers

Experimental results



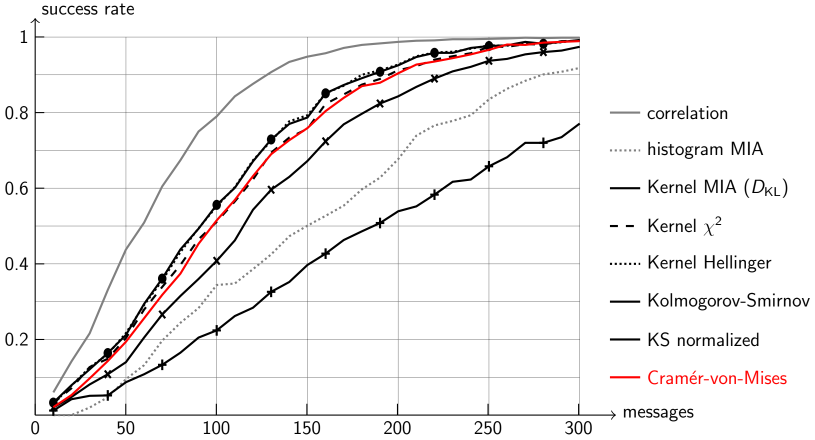
Success rate of different distinguishers

Experimental results



Success rate of different distinguishers

Experimental results

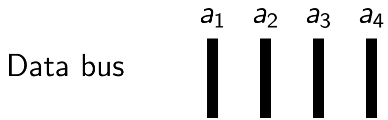


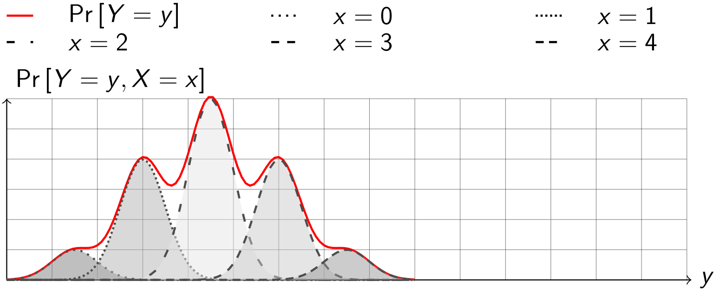
Success rate of different distinguishers

Outline

1. How to use MIA: the information theoretic toolbox
2. When to use it: MIA versus correlation
3. Why to use it: MIA as an evaluation metric

An example: leaky bit on a data bus

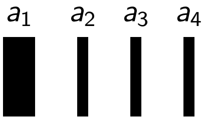


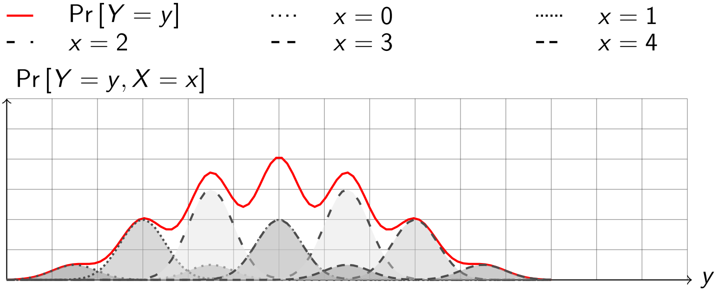


Effect of a leaky bit on the pdfs

An example: leaky bit on a data bus

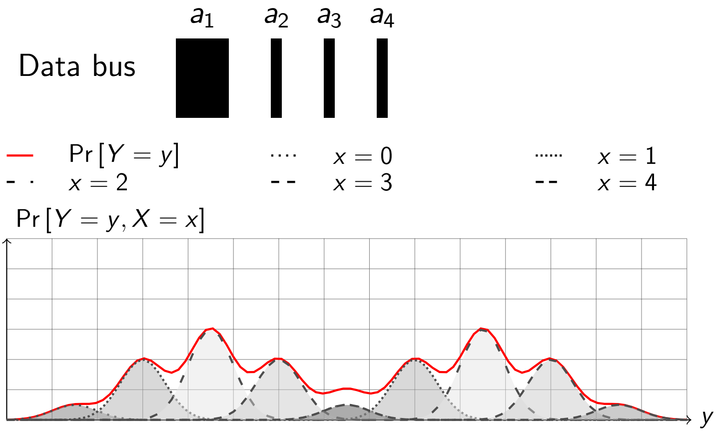
Data bus





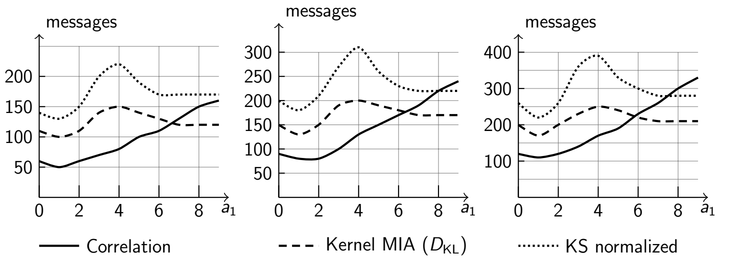
Effect of a leaky bit on the pdfs

An example: leaky bit on a data bus



Effect of a leaky bit on the pdfs

An example: leaky bit on a data bus



Weight of the first leaking bit vs number of messages for a success rate of 50% (left), 75% (middle) and 90% (right)

Limitations

MIA is not the only way to go here: DPA would work!

What about:

- protected logics
- masking scheme

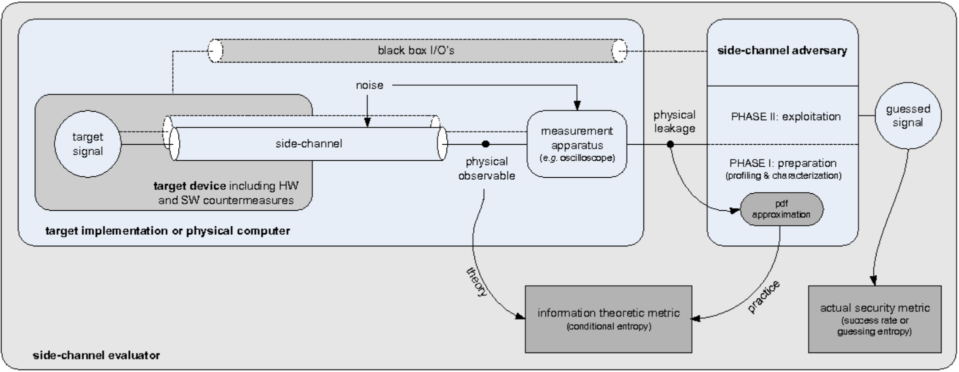
More resilient to erroneous leakage models
But not immune, requires $\mathrm{I}(X_g; Y) > \mathrm{I}(X_w; Y)$

Outline

1. How to use MIA: the information theoretic toolbox
2. When to use it: MIA versus correlation
3. Why to use it: MIA as an evaluation metric

MIA versus Mutual Information Metric

- Eurocrypt 2009:



MIA is not MIM

More precisely:

1. MIA: $\hat{\mathrm{I}}(X; Y)$ / MIM: $\mathrm{I}(K; Y)$
2. MIM directly targets the key dependencies
3. MIA requires an intermediate variable
4. MIM approximates $\mathrm{I}(K; Y)$ with “templates”
5. MIA estimates $\hat{\mathrm{I}}(X; Y)$ “on-the-fly”

$\longrightarrow$ If the leakage model used by the adversary is not perfect, MIA will underestimate the leakage:

$$\mathrm{I}(K; Y) > \hat{\mathrm{I}}(X; Y)$$

Summarizing

- MIA is a “toolbox”
- MIA is more resilient to erroneous leakage models
- MIA and MIM are two complementary tools with different purpose: generic adversary and generic evaluation tool

Any Questions?

T. W. Anderson.
On the distribution of the two-sample cramér-von mises criterion.
The Annals of Mathematical Statistics, 33 (3) : 1148–1159, 1962.

Sébastien Aumonier.
Generalized correlation power analysis.
In *Ecrypt Workshop on Tools For Cryptanalysis*. Krakòw, Poland, September 2007.

E. Brier, C. Clavier, F. Olivier. Correlation power analysis with a leakage model. In *CHES 2004*, LNCS, vol 3156, pp 16-29, Boston, MA, USA, August 2004.

T.M. Cover, J.A. Thomas.
Elements of Information Theory.
Wiley, 1991.

Imre Csiszár and Paul C. Shields.
Information theory and statistics: a tutorial.
Commun. Inf. Theory, vol 1, num 4, pp 417-528, 2004.

DPA Contest 2008/2009, http://www.dpacontest.org/

B. Gierlichs, L. Batina, P. Tuyls, B. Preneel.
Mutual information analysis.
In *CHES 2008*, LNCS, vol 5154, pp 426-442, Washington DC, USA, August 2008.

P. Kocher, J. Jaffe, B. Jun, Differential power analysis.
In *Crypto 1999*, LNCS, vol 1666, pp 398-412, Santa-Barbara, CA, USA, August 1999.

Emmanuel Prouff and Matthieu Rivain.
Theoretical and practical aspects of mutual information based side channel analysis.

To appear in *ACNS, Applied Cryptography and Network Security*, LNCS, Paris, June 2009.

Francois-Xavier Standaert, Tal G. Malkin, and Moti Yung.
A unified framework for the analysis of side-channel key recovery attacks (extended version).
Cryptology ePrint Archive, Report 2006/139, 2006.
http://eprint.iacr.org/.