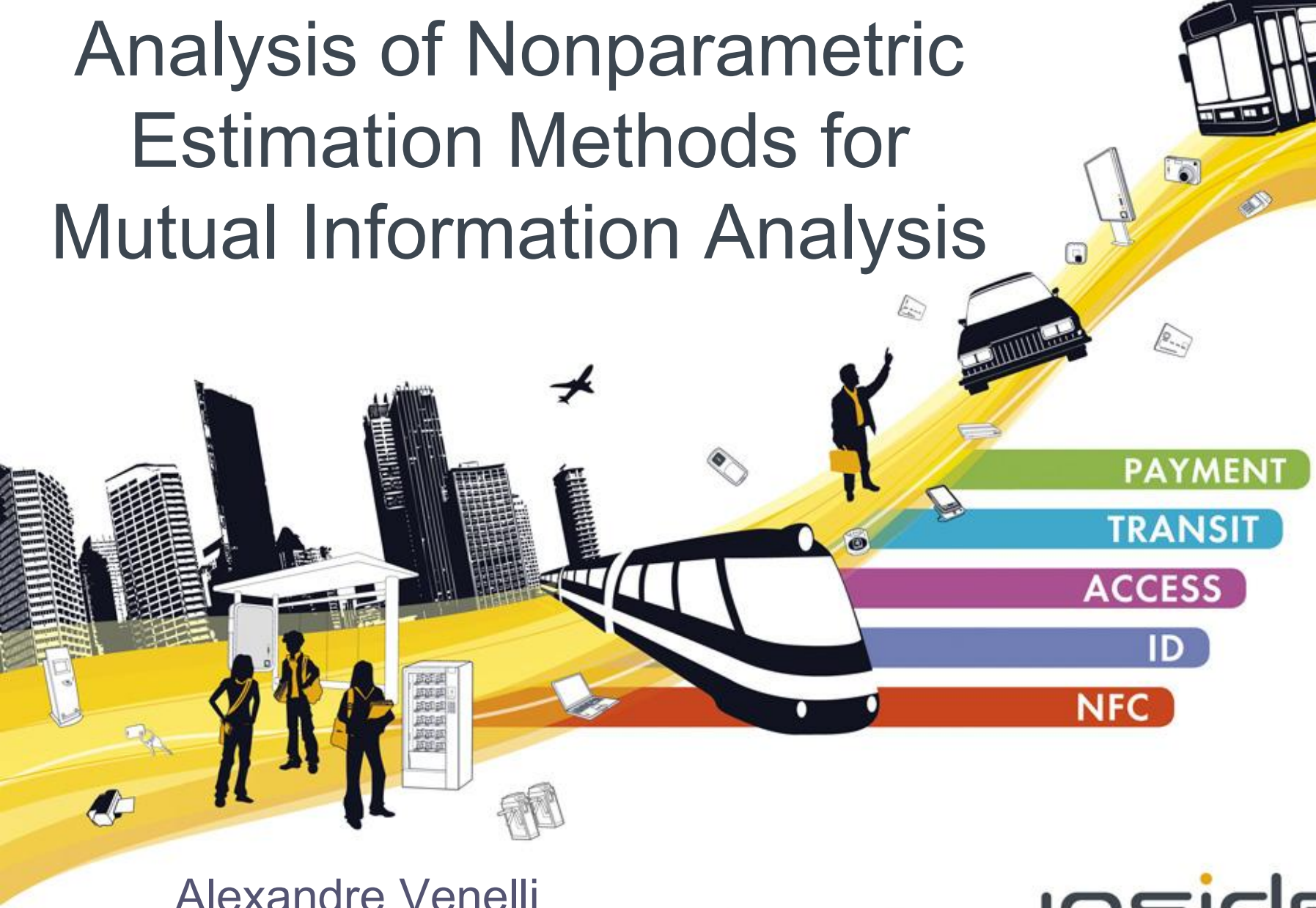
Analysis of Nonparametric Estimation Methods for Mutual Information Analysis



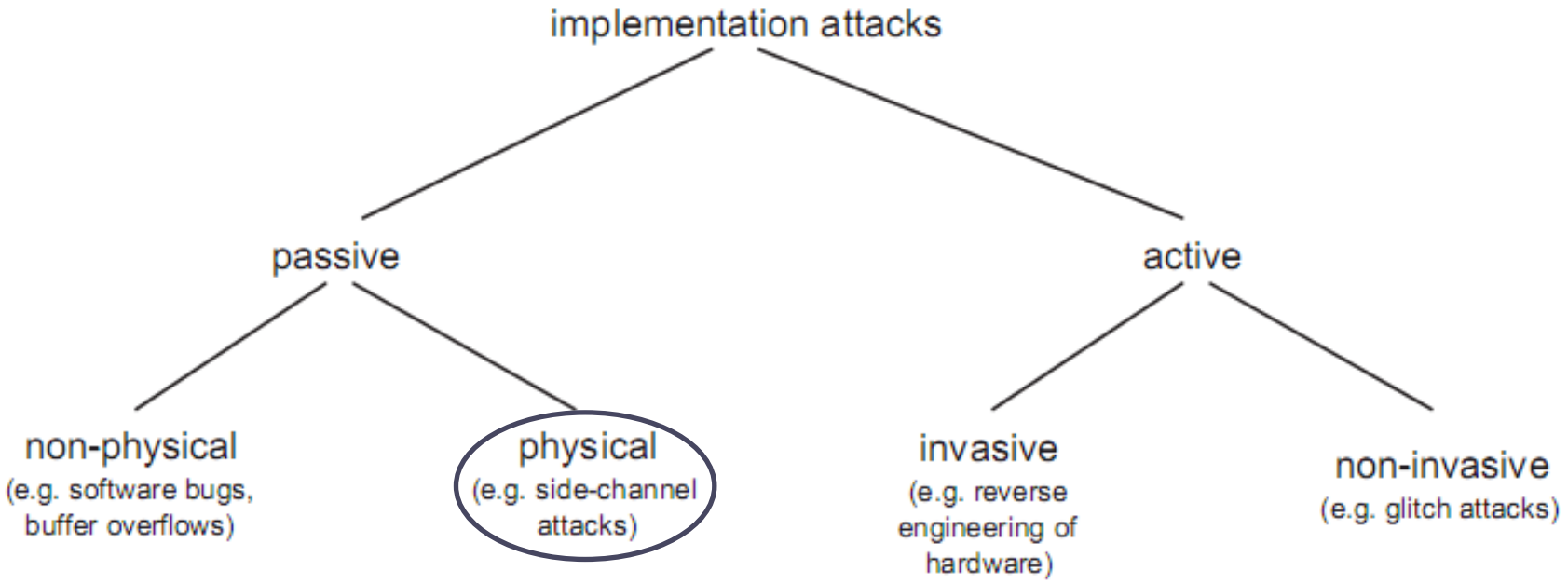
Alexandre Venelli

Outline

- Differential Side-Channel Analysis (DSCA)
- Mutual Information Analysis (MIA)
- Study of nonparametric PDF estimation methods for MIA
- Experimental results
- Conclusion

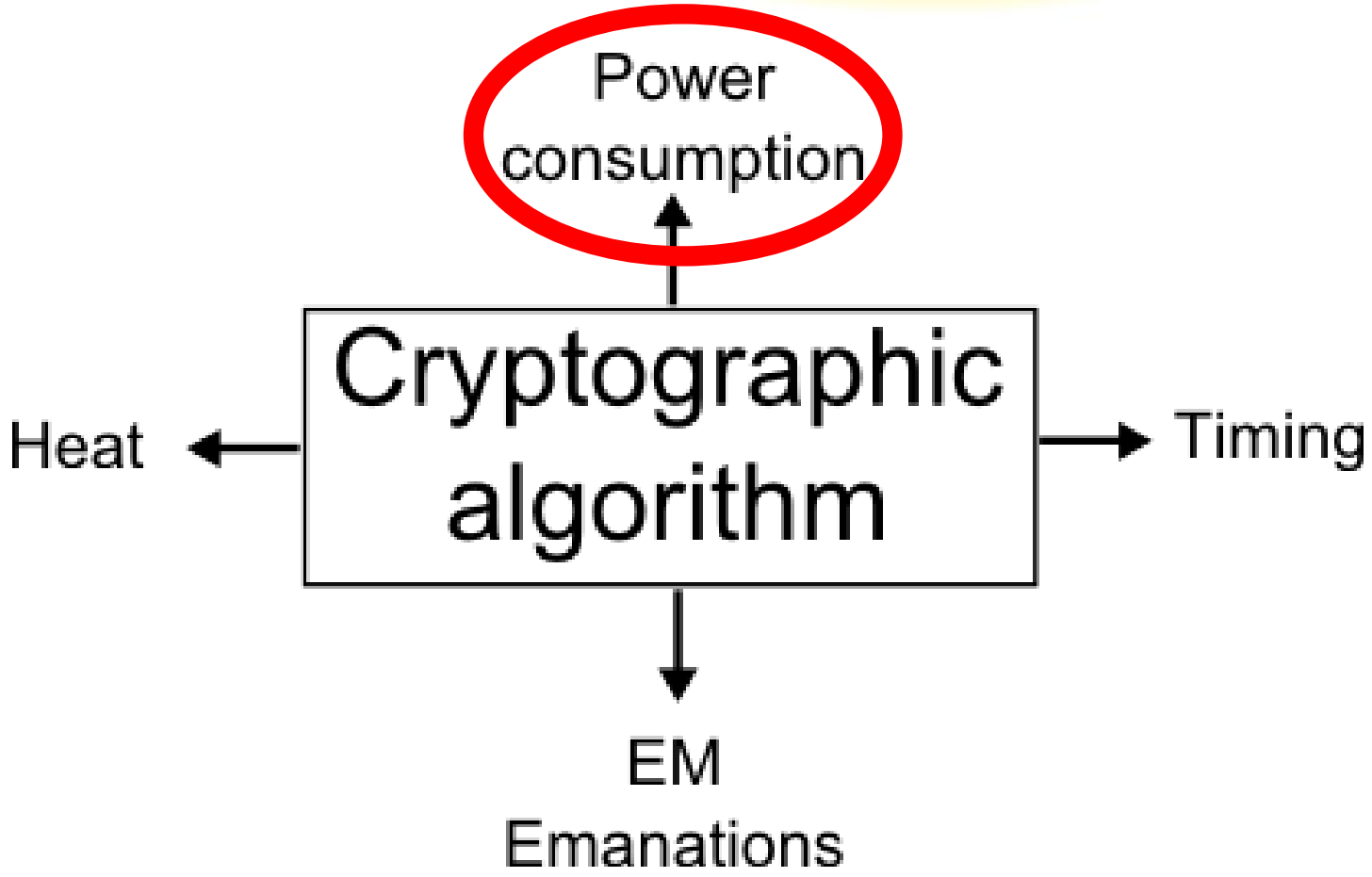
Attacks on cryptosystems

- **Mathematical attacks**
  - Cryptanalysis, brute force, …
- **Implementation attacks**

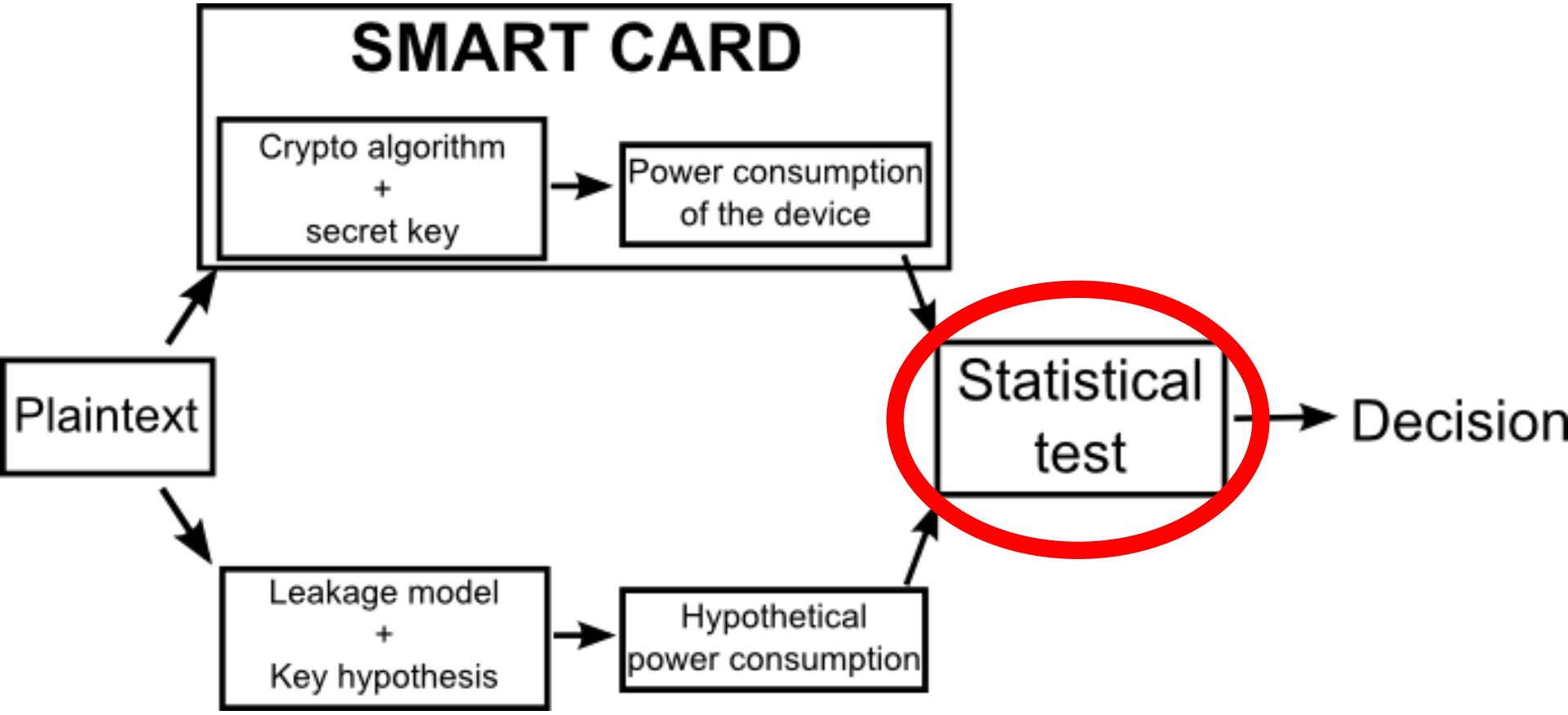


inside CONTACTLESS

Side-channel leakages



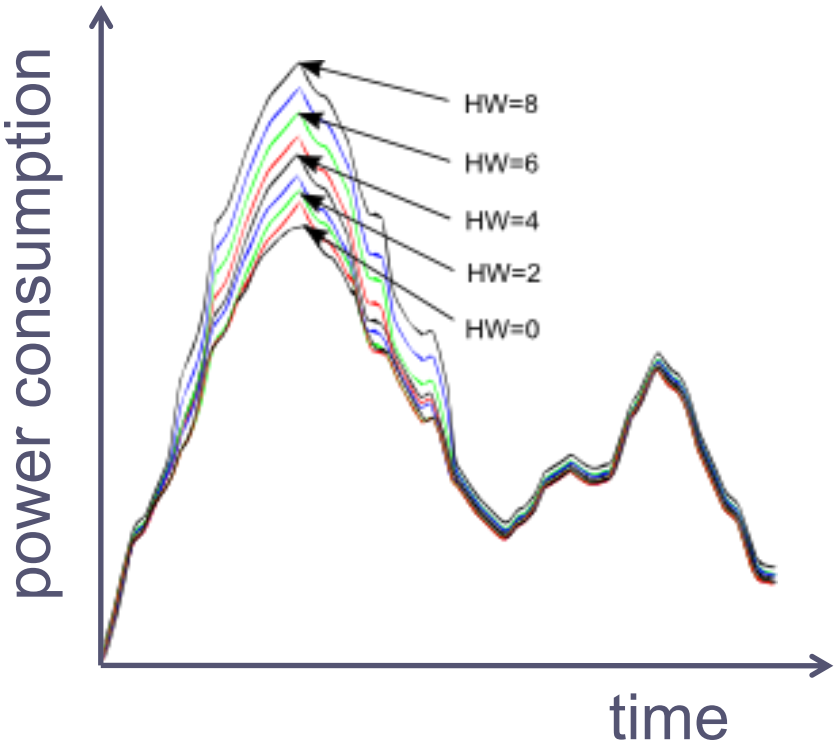
Differential side-channel analysis workflow



Power analysis and leakage model

- Messerges et al. 1999
- Brier et al. 2004

$$P(t) = a.HW(M) + b$$



inside CONTACTLESS

Brief history of statistical tests used in SCA (1)

• Kocher et al. 1999	Simplified T-test (DPA)
• Brier et al. 2004	Pearson correlation factor (CPA)
• Batina et al. 2008	Spearman factor (SPE)
• Batina et al. 2009	Differential Cluster Analysis (DCA)
• Veyrat-Charvillon et al. 2009	Cramér-von Mises test (CVM)

Brief history of statistical tests used in SCA (2)

• Gierlichs et al. 2008	Mutual Information (MIA)
• Prouff et al. 2009	MIA + finite mixtures
• Venelli 2010	MIA + B-spline estimation
• Thanh-Ha Le et al. 2010	MIA + Cumulant-based estimation

Remainder on information theory (1)

- Let X be a r.v. with n values $\{x_1, \ldots, x_n\}$
- Let f be the probability density function (PDF) of X
- Entropy of X

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

- Mutual Information (MI)

$$I(X;Y) = H(X) - H(X|Y)$$

$$I(X;Y) = H(X) + H(Y) - H(X,Y)$$

inside CONTACTLESS

Remainder on information theory (2)

- Rényi entropy

$$H_\alpha(X) = \begin{cases} \frac{1}{1-\alpha}\log\sum_x f(x)^\alpha & \text{for } \alpha \geq 0, \alpha \neq 1 \\ -\sum_x f(x)\log(f(x)) & \text{for } \alpha = 1 \end{cases}$$

- Generalized Mutual Information (GMIA), Pompe et al. 1993

$$I_2(X;Y) = H_2(X) + H_2(Y) - H_2(X,Y)$$

inside CONTACTLESS

Problem : estimate MI

- Mutual information
  - powerful
  - difficult to estimate
- Goal : estimate MI → Entropy → PDF given a small finite set of data
- Two main families of PDF estimation methods
  - parametric
  - nonparametric

Parametric estimation (1)

- Assumption : data sampled from a known family of distributions (Gaussian, exponential, …)
- Parameters are optimized by fitting the model to the data set
- Examples of estimators :
  - Maximum likelihood
  - Edgeworth
  - Least-square
  - Cumulants
  - …

Parametric Estimation (2)

Cumulant-based Estimation

- Thanh-Ha Le et al. 2010
  - Edgeworth expansion + cumulants

$$I(U) = I(U_1; U_2; ...; U_n) \approx \frac{1}{4} \sum_{ij \neq ii} (R_{ij}^U)^2 + \frac{1}{12} \sum_{ijk \neq iii} (T_{ijk}^U)^2 + \frac{1}{48} \sum_{ijkl \neq iiii} (Q_{ijkl}^U)^2$$

$$\begin{aligned} R_{ij}^X &= E(\overline{X}_i \overline{X}_j) \\ T_{ijk}^X &= E(\overline{X}_i \overline{X}_j \overline{X}_k) \\ Q_{ijkl}^X &= E(\overline{X}_i \overline{X}_j \overline{X}_k \overline{X}_l) - E(\overline{X}_i \overline{X}_j) E(\overline{X}_k \overline{X}_l) \\ &\quad - E(\overline{X}_i \overline{X}_k) E(\overline{X}_j \overline{X}_l) - E(\overline{X}_i \overline{X}_l) E(\overline{X}_j \overline{X}_k) \end{aligned}$$

- For 1rst order SCA, we only have U=[U1,U2]

Nonparametric estimation

- Assumption : none about the distribution of the population, « model-free » methods
- Parameters are often chosen more or less « blindly »
- Examples of estimators :
  - Histograms
  - Kernel Density Estimation
  - K-Nearest Neighbors
  - B-splines

Parametric vs. Nonparametric

- Why nonparametric statistics ?

- Nonparametric statistics enable us to process
  - Data of « low quality »,
  - From small samples,
  - On variables about which nothing is known (concerning their distribution)

- Often the case, in the context of DSCA

Histogram based Estimation (HE)

	
- Easy to calculate and understand	- Systematic errors due to the finite size of the dataset

inside CONTACTLESS

Kernel Density Estimation (KDE)

	
- Better convergence to the underlying distribution	- Slower to compute than HE

inside CONTACTLESS

K-Nearest Neighbors (KNN)

	
- Seems unbiased for independent X,Y - Smaller errors than KDE - Only decent method for high dimensional variables	- Medium slow to compute

inside CONTACTLESS

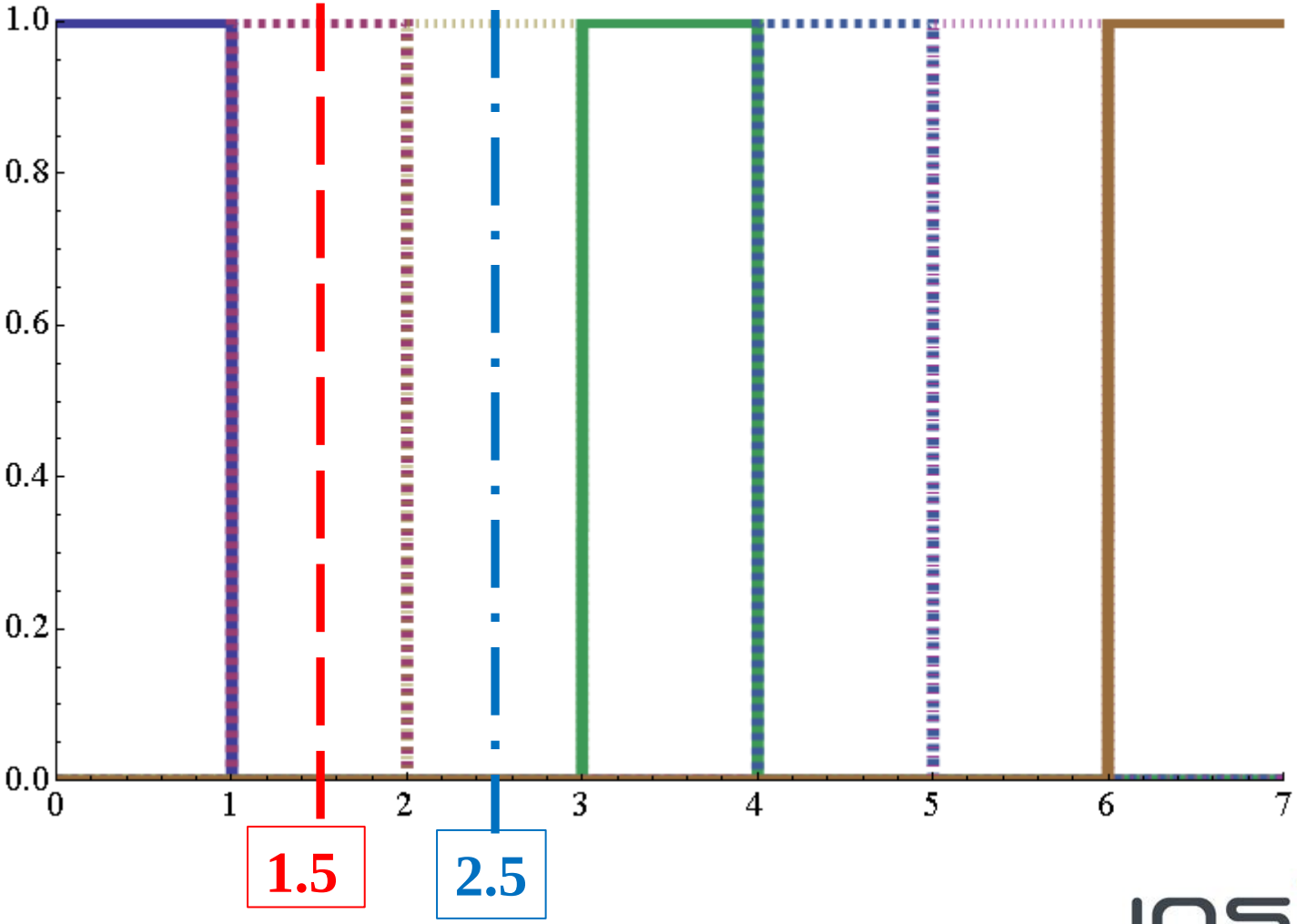
B-Splines Estimation (BSE) (1)

	
-Computationally faster than KDE and KNN - Interesting property in the side-channel context	- Slower than histograms

inside CONTACTLESS

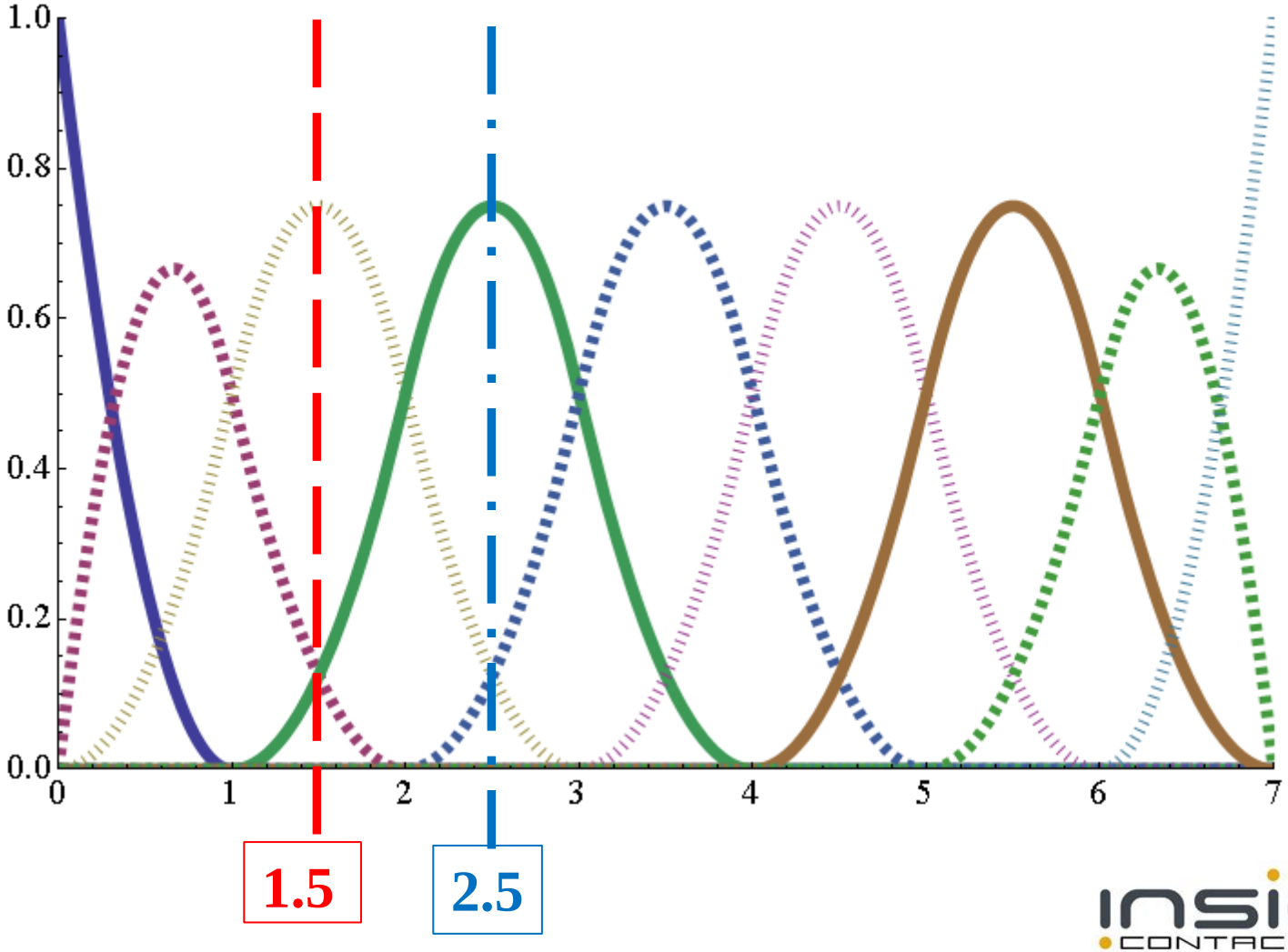
B-Splines Estimation (2)

Histograms



B-Splines Estimation (3)

Degree 2 B-Splines basis functions



Experimental results (1)

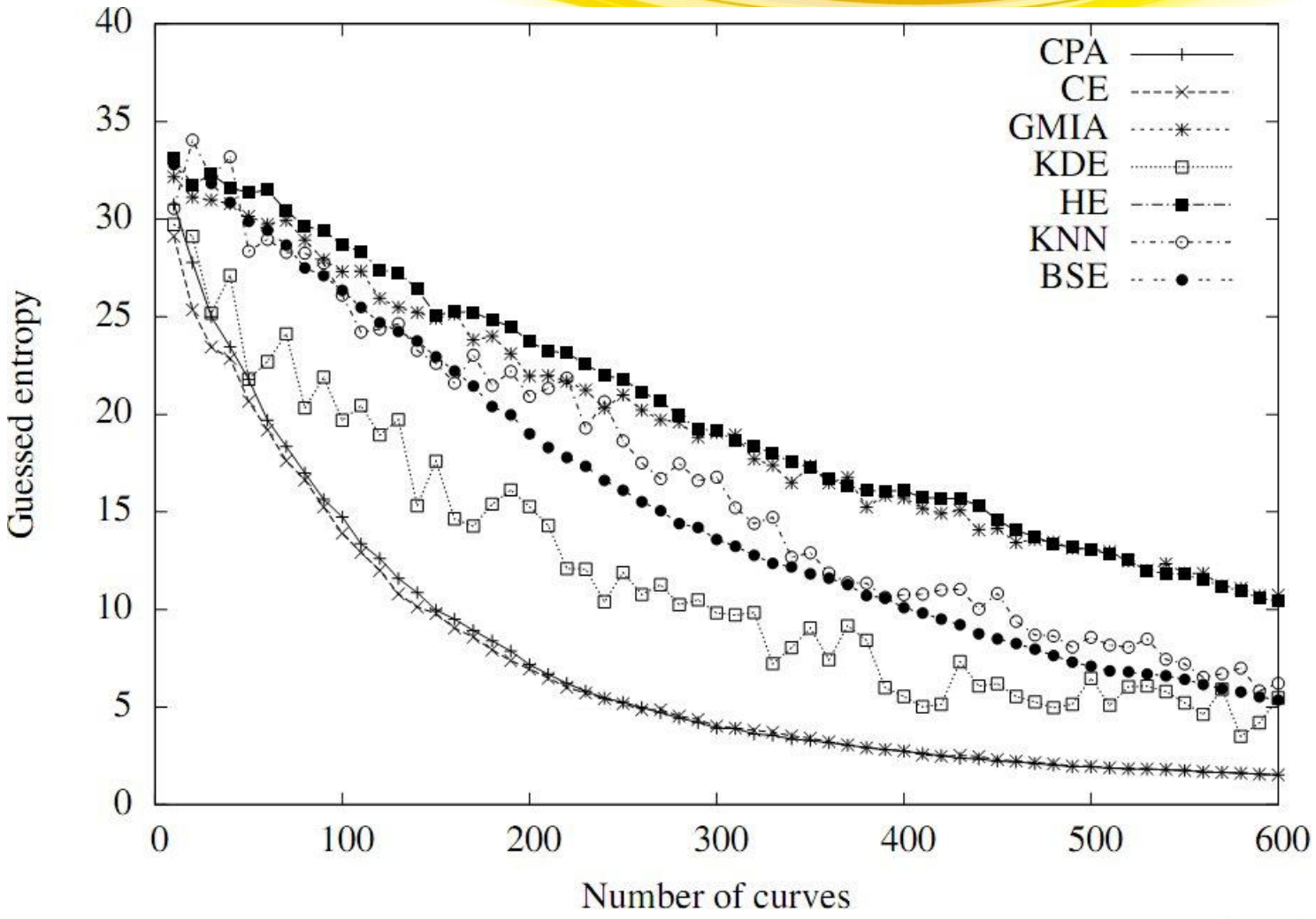
Metrics

- **Two metrics (Standaert et al. 2008) :**
  - First order success rate : given a number of traces, the probability that the correct hypothesis is the first best hypothesis of an attack
  - Guessed entropy : average position of the correct hypothesis in the sorted hypothesis vector of an attack

- **Attacks setups :**
  - DPA Contest v1 (http://www.dpacontest.org)
    - Hardware DES
    - Output of the Sbox at the last round
  - STK600+ATMega2561
    - Software multi-precision multiplication
    - Intermediate 8x8 multiplications

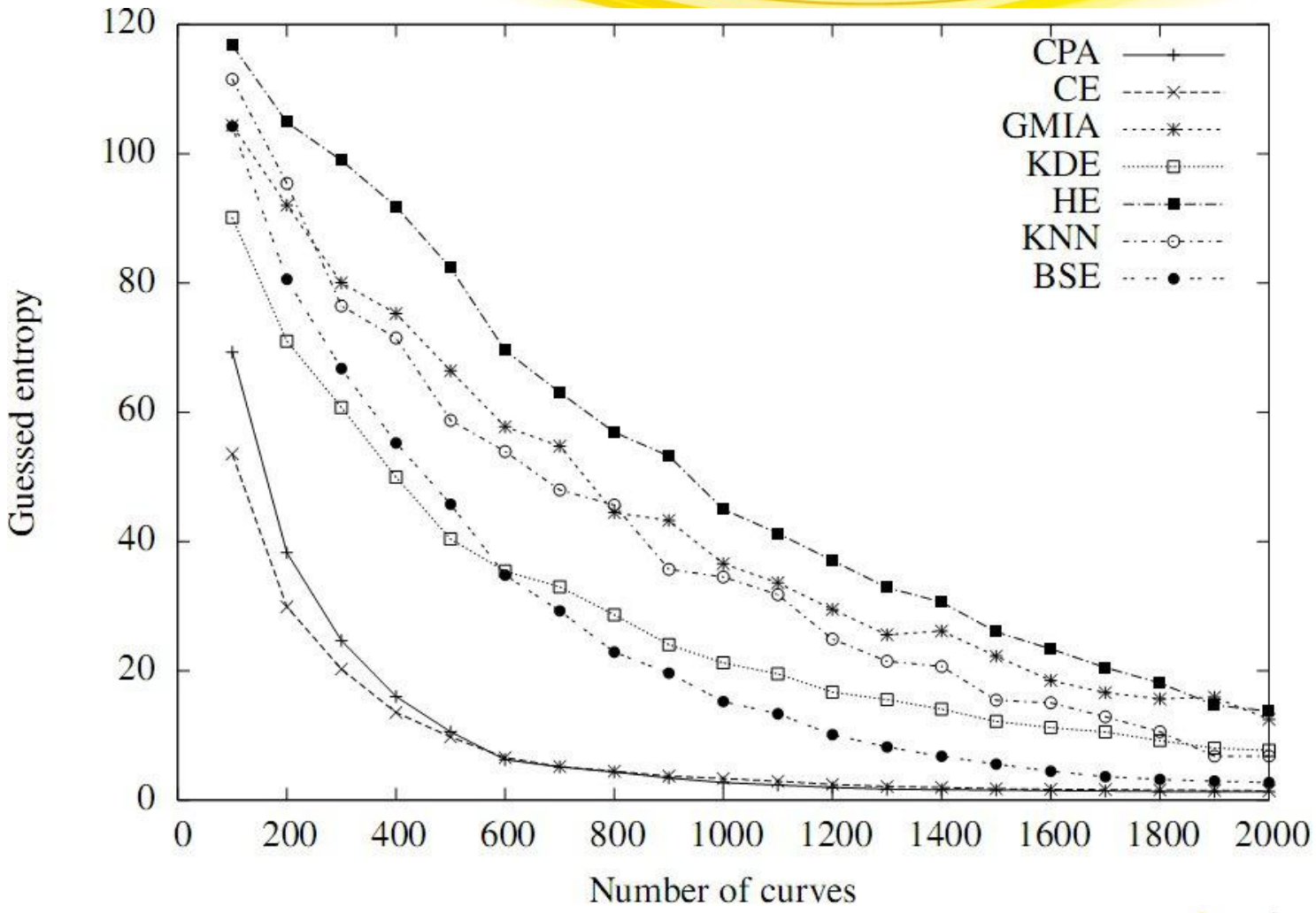
Experimental results (2)

DPA Contest v1 DES



Experimental results (3)

STK600/ATMega2561 multi-precision multiplication



Conclusion

- MIA + efficient PDF estimation
- Nonparametric estimation makes sense in the DSCA context
- However, the power consumption of CMOS devices seems highly linear in the Hamming weight of processed data
- Future of MIA
  - Higher order SCA
  - Devices using different logic

Thank you for your attention !

Contact : avenelli@insidefr.com