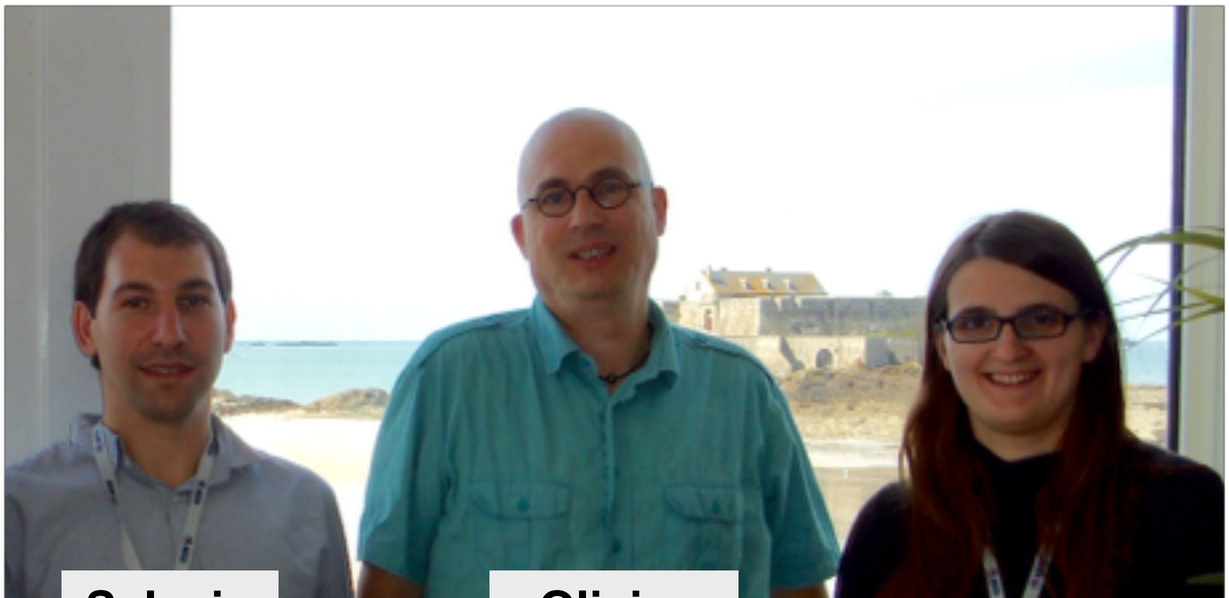
INSTITUT
Mines-Télécom

# “*Distinguishing Distinguishers*”
## A Theoretical Approach to Side-Channel Analysis

Annelie Heuser, Telecom ParisTech, LTCI, CNRS

# PhD

- **September 2012 - December 2015**
- **Supervisors:**



# Fellowship

- Google european doctoral fellowship in Privacy (3 years)
- 13 selected candidates throughout Europe in 2013
- Mentor at Google Zurich
- Interns, conferences, workshops, tech talks at Google

# Today’s World

- Growing number of embedded systems in our daily life
- Concerns Privacy, Safety, Security

# Today's World

- **Growing number of embedded systems in our daily life**
- **Concerns Privacy, Safety, Security**

# Side-Channel Attacks in a Nutshell

- **Cryptanalysis is "impossible"**

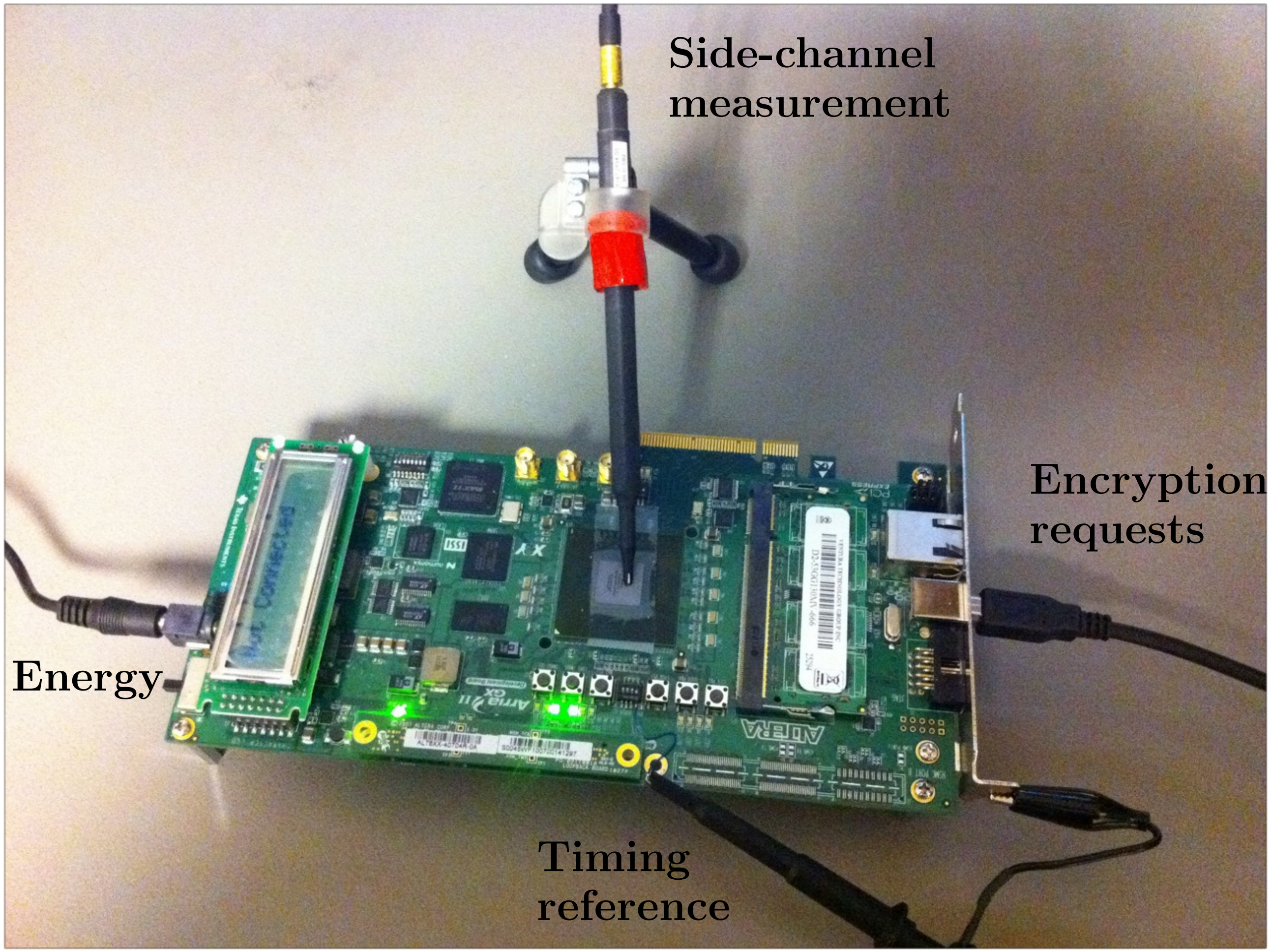
# Side-Channel Attacks in a Nutshell

- **Cryptanalysis is “impossible”**
- **Use an additional side-channel**

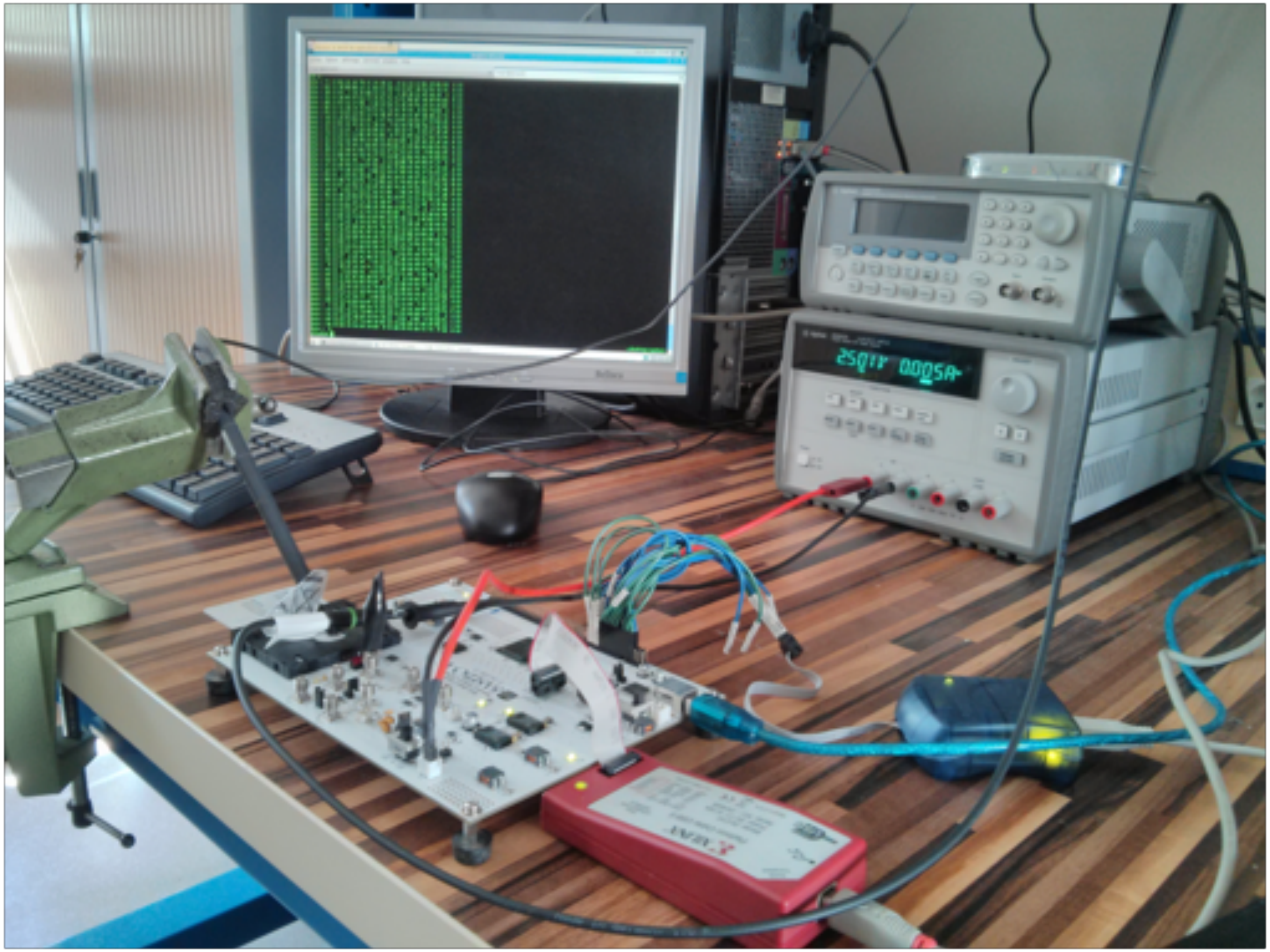


TELECOM ParisTech

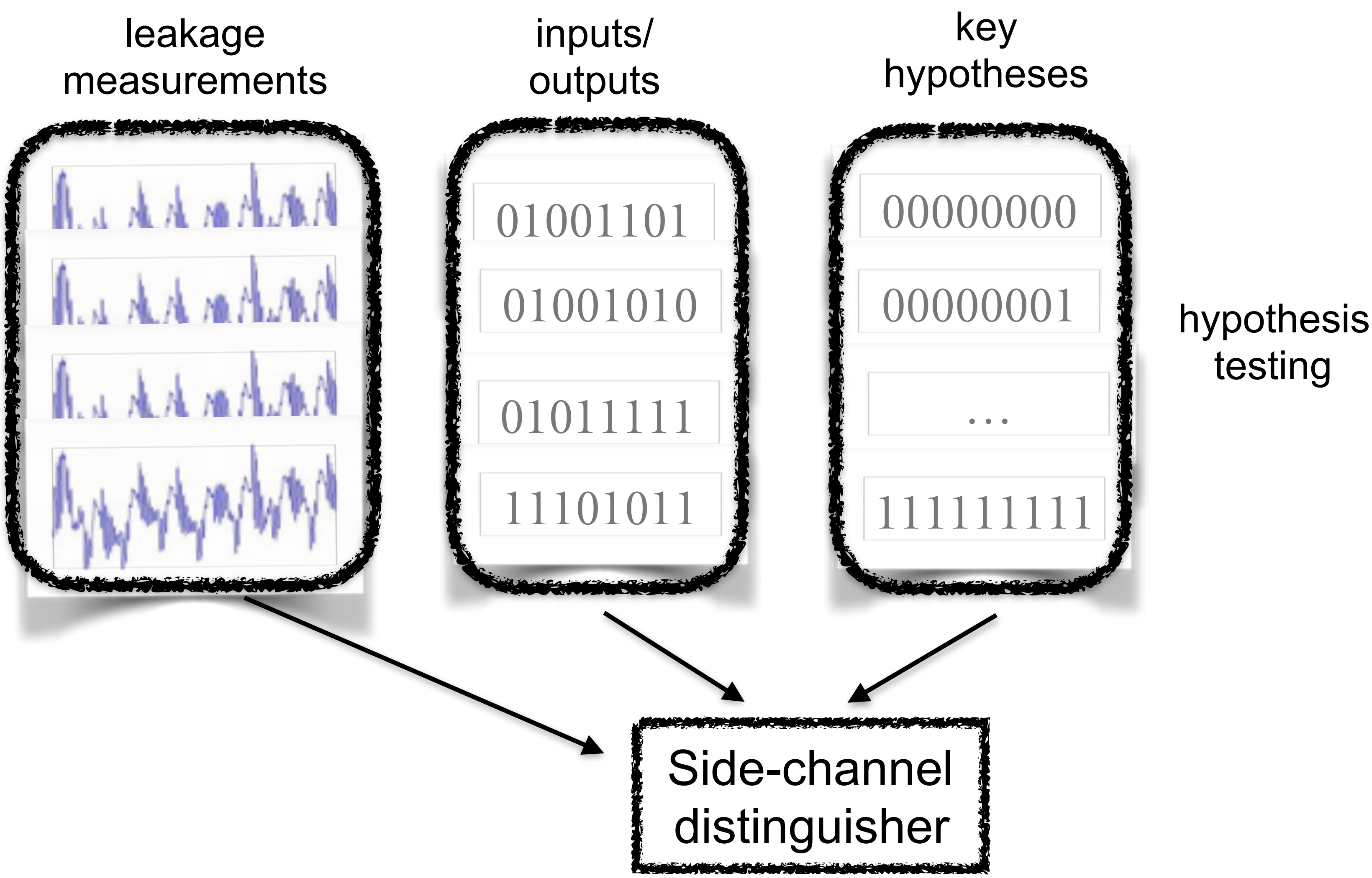
# Side-Channel Analysis



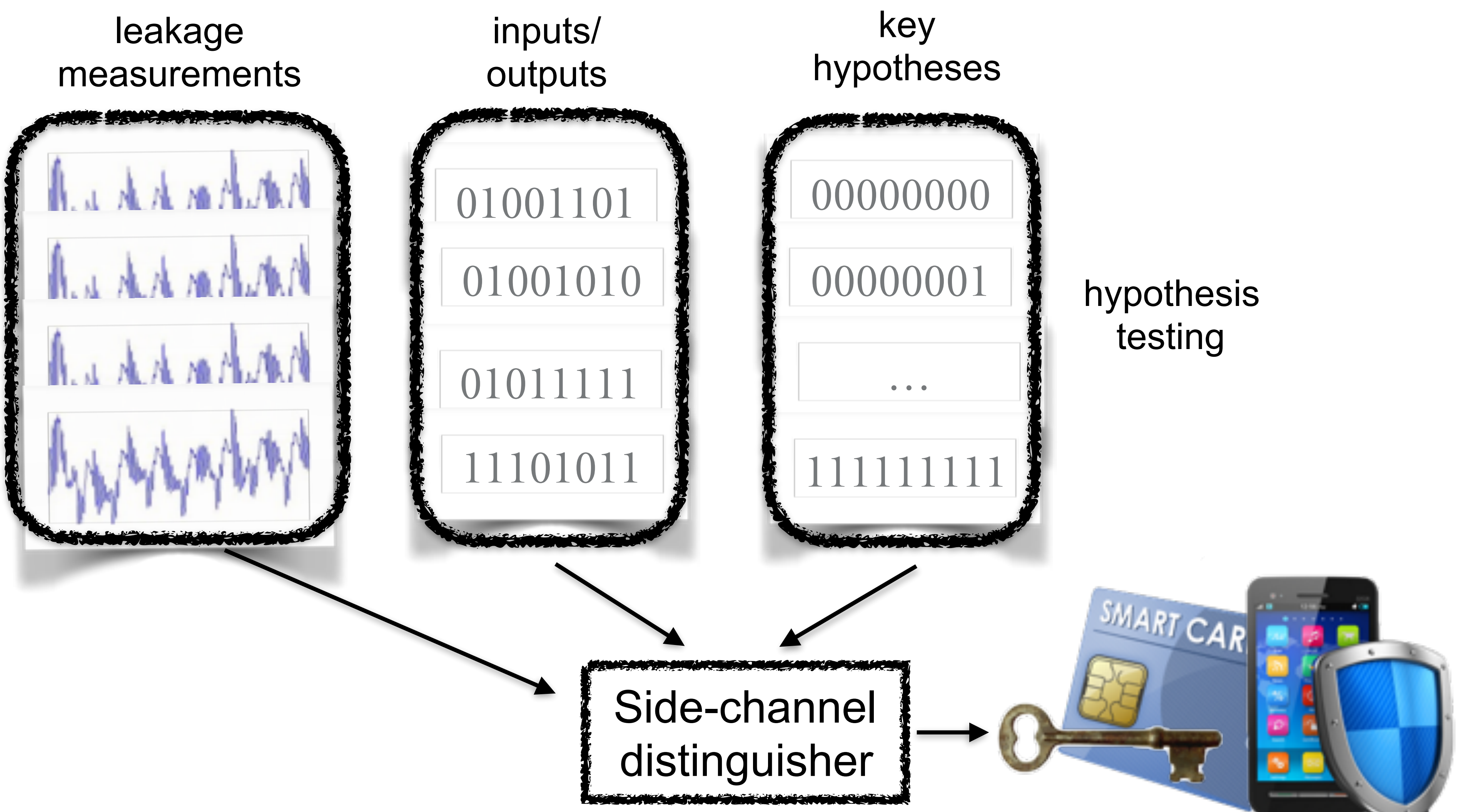
# Side-Channel Analysis



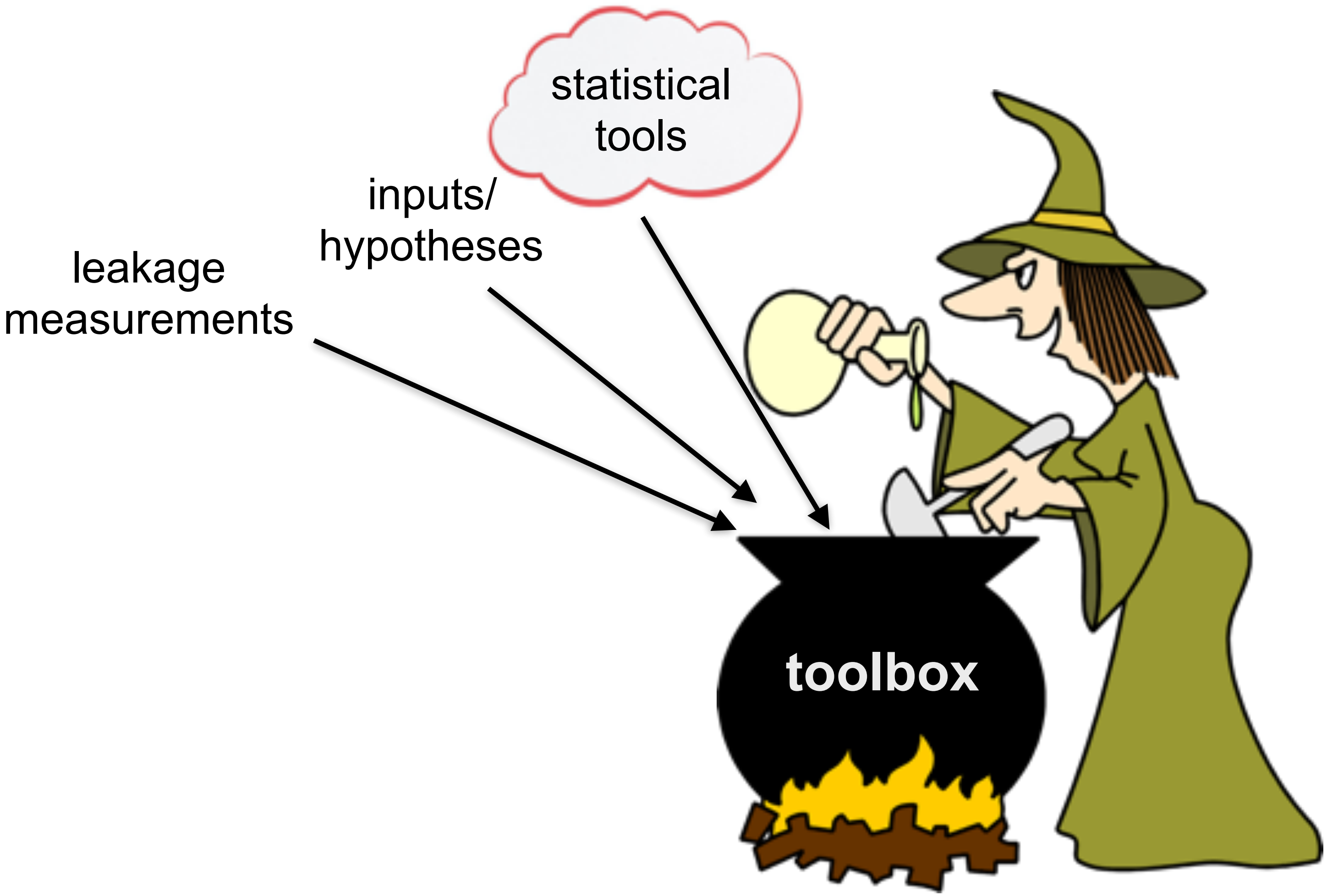
# Side-Channel Analysis



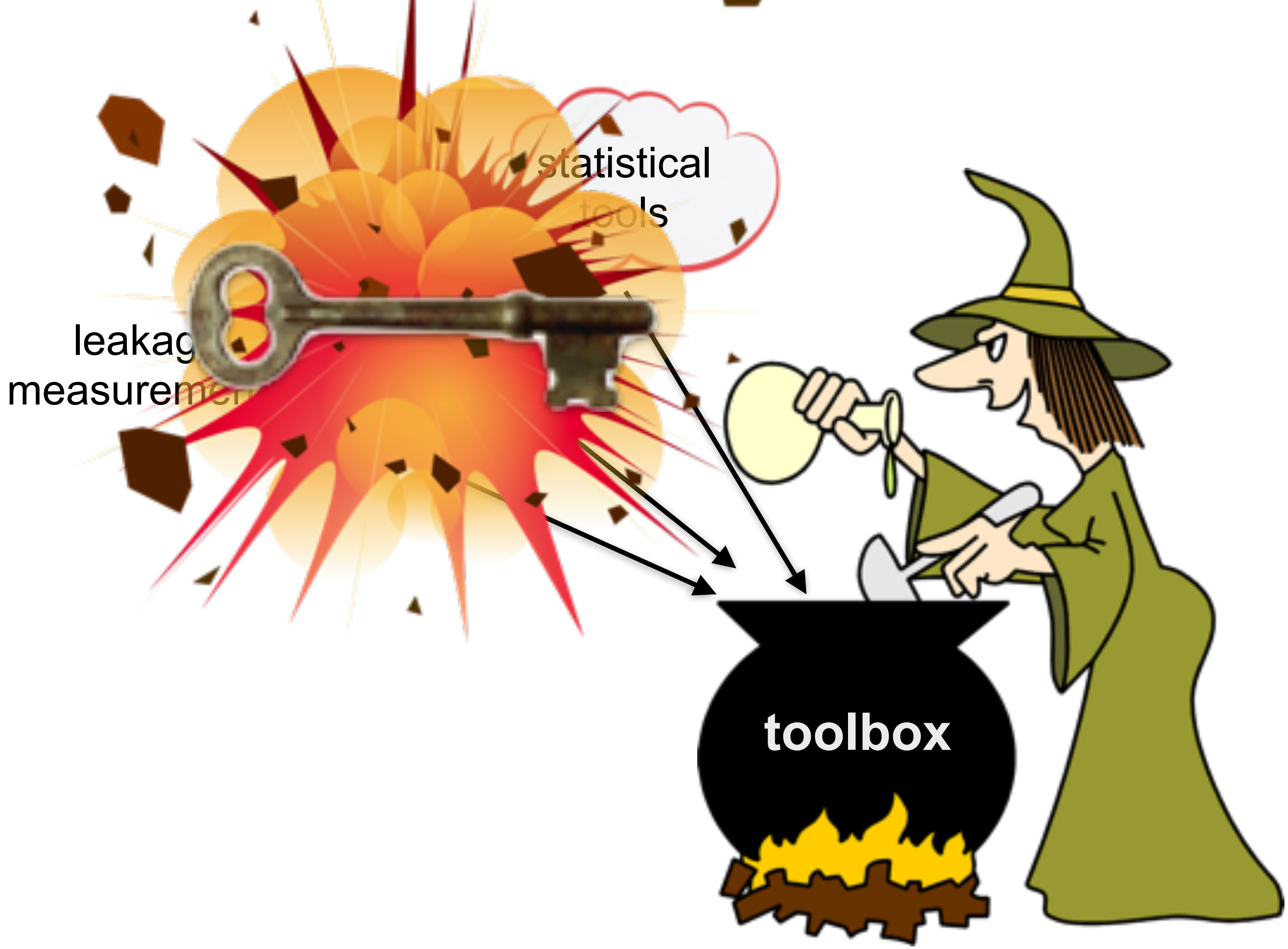
# Side-Channel Analysis



# State-of-the-Art

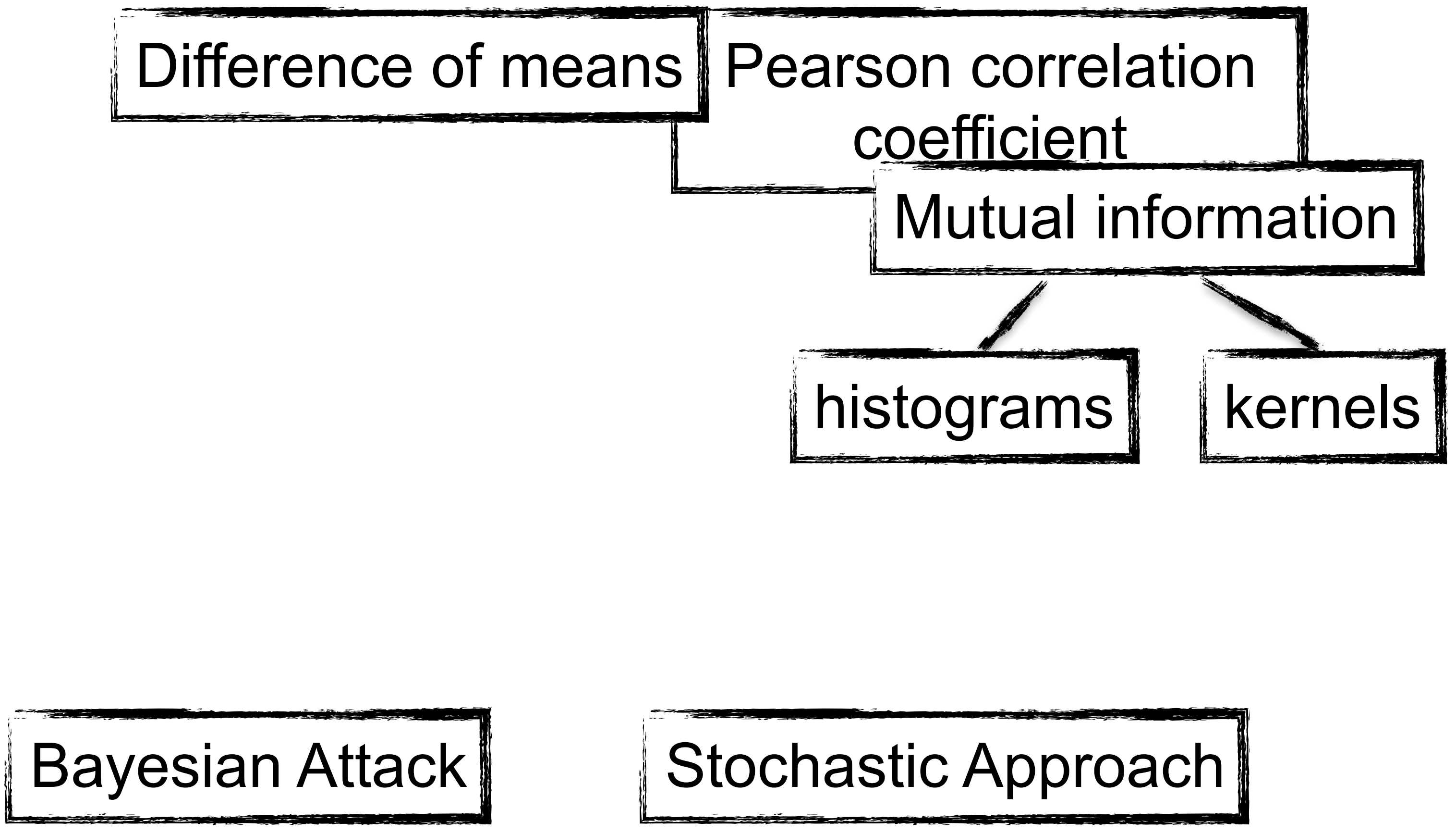


# State-of-the-Art



# State-of-the-Art

# State-of-the-Art



# State-of-the-Art

Difference of means

Pearson correlation coefficient

Linear regression

Mutual information

- histograms
- kernels

Kolmogorov-Smirnov

Copula distinguisher

Bayesian Attack

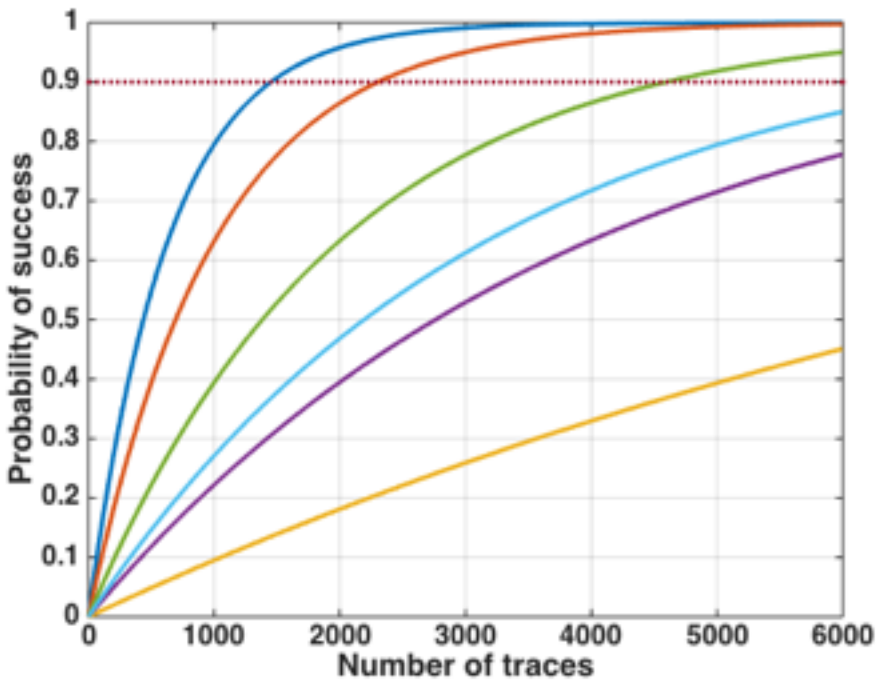
Stochastic Approach

# State-of-the-Art

- Difference of means
- Pearson correlation coefficient
- Linear regression
- Mutual information
  - histograms
  - kernels
- Kolmogorov-Smirnov
- Copula distinguisher
- Variance based distinguisher
- Cramér-von-Mises
- Template Attack
- SVM Attack
- Decision Trees
- Bayesian Attack
- Stochastic Approach
- ...

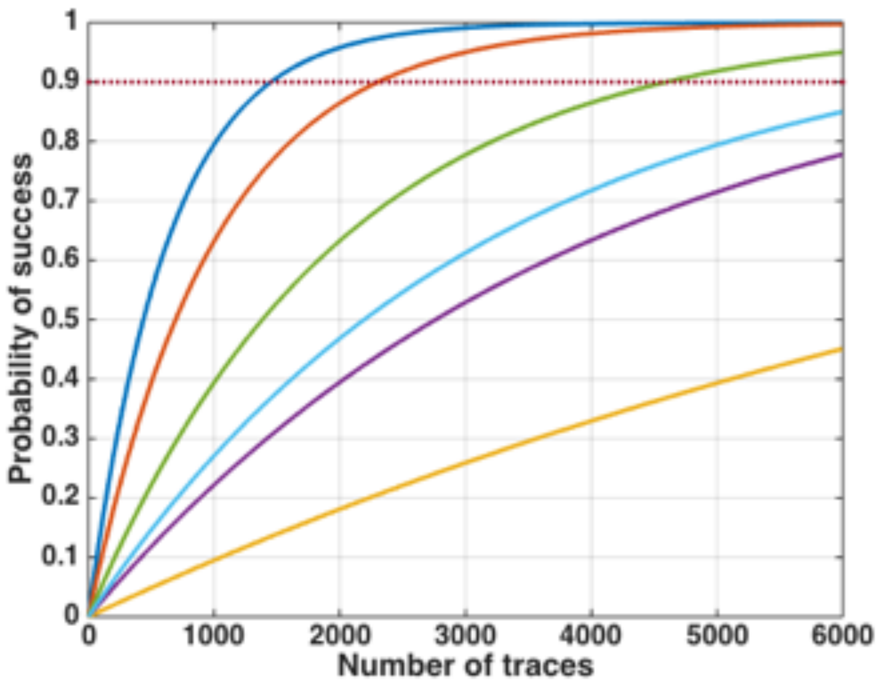
# Security Assessment

- Security evaluation
- False confidence



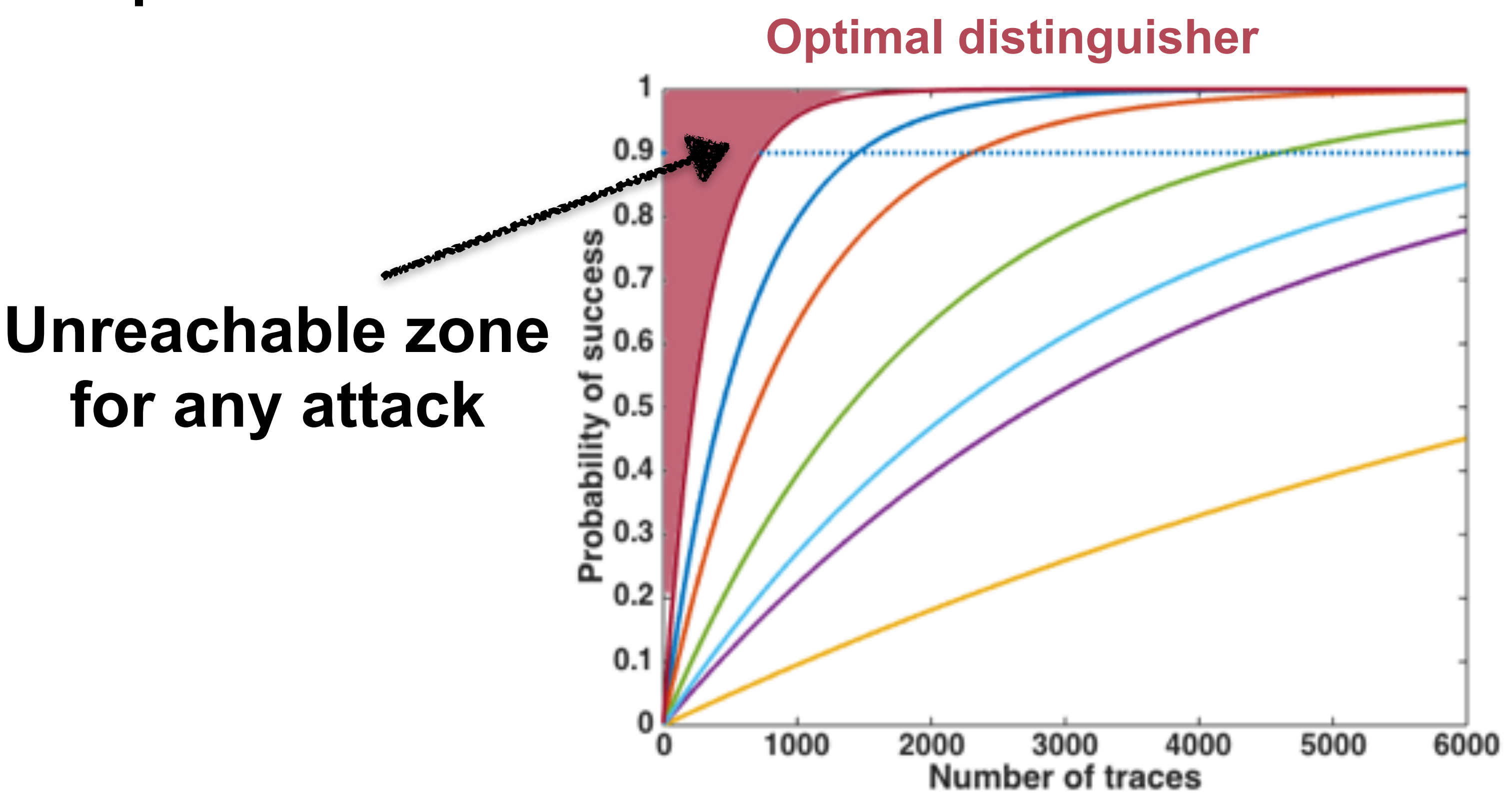

# Security Assessment

- Security evaluation
- False confidence

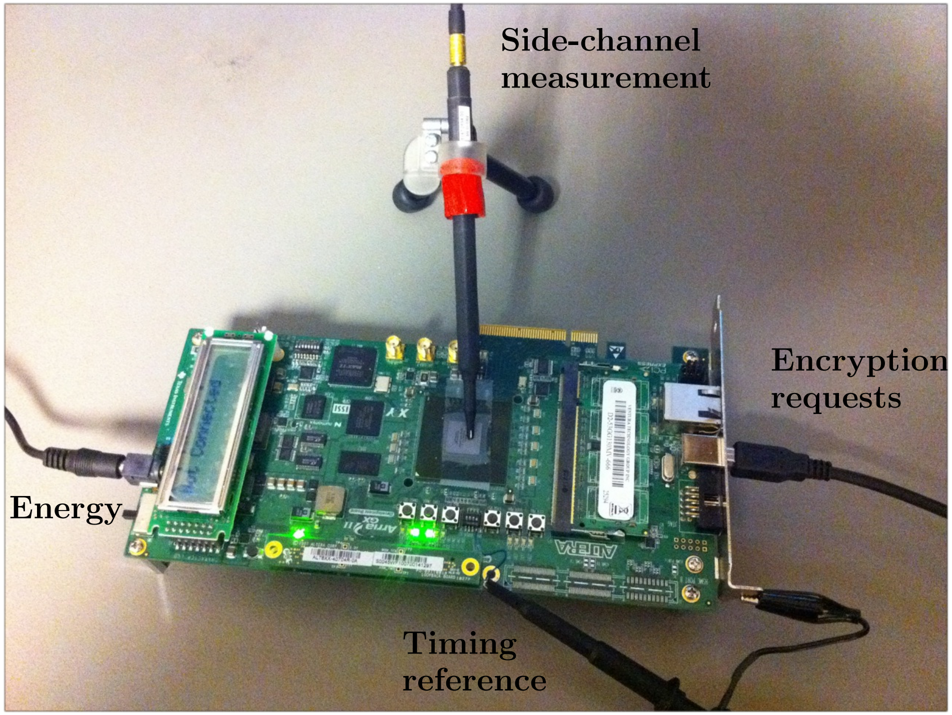



# Optimal Distinguisher

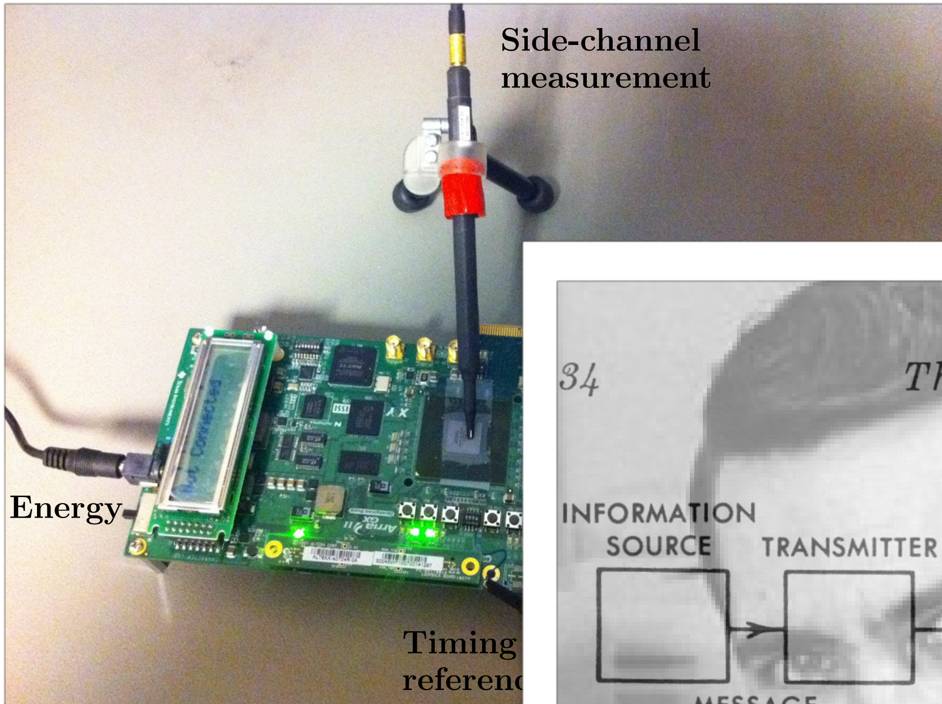
- In a given side-channel context
- What is the best possible distinguisher among all possible ones?

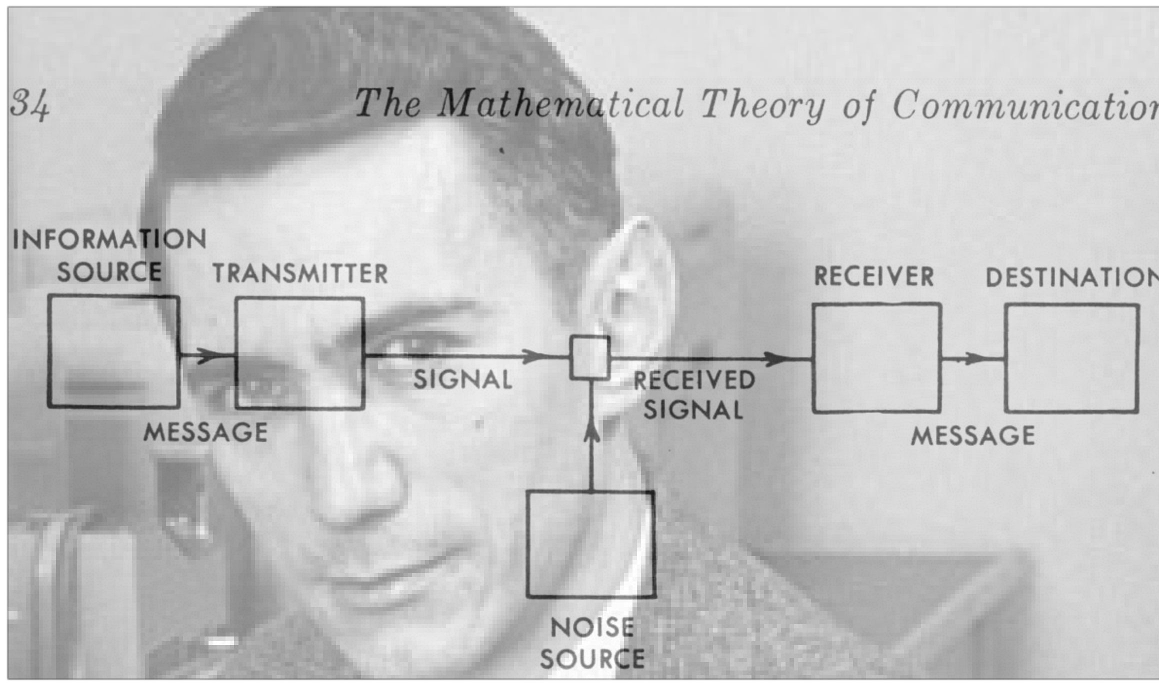


# Side-Channel Analysis As A Digital Communication Problem

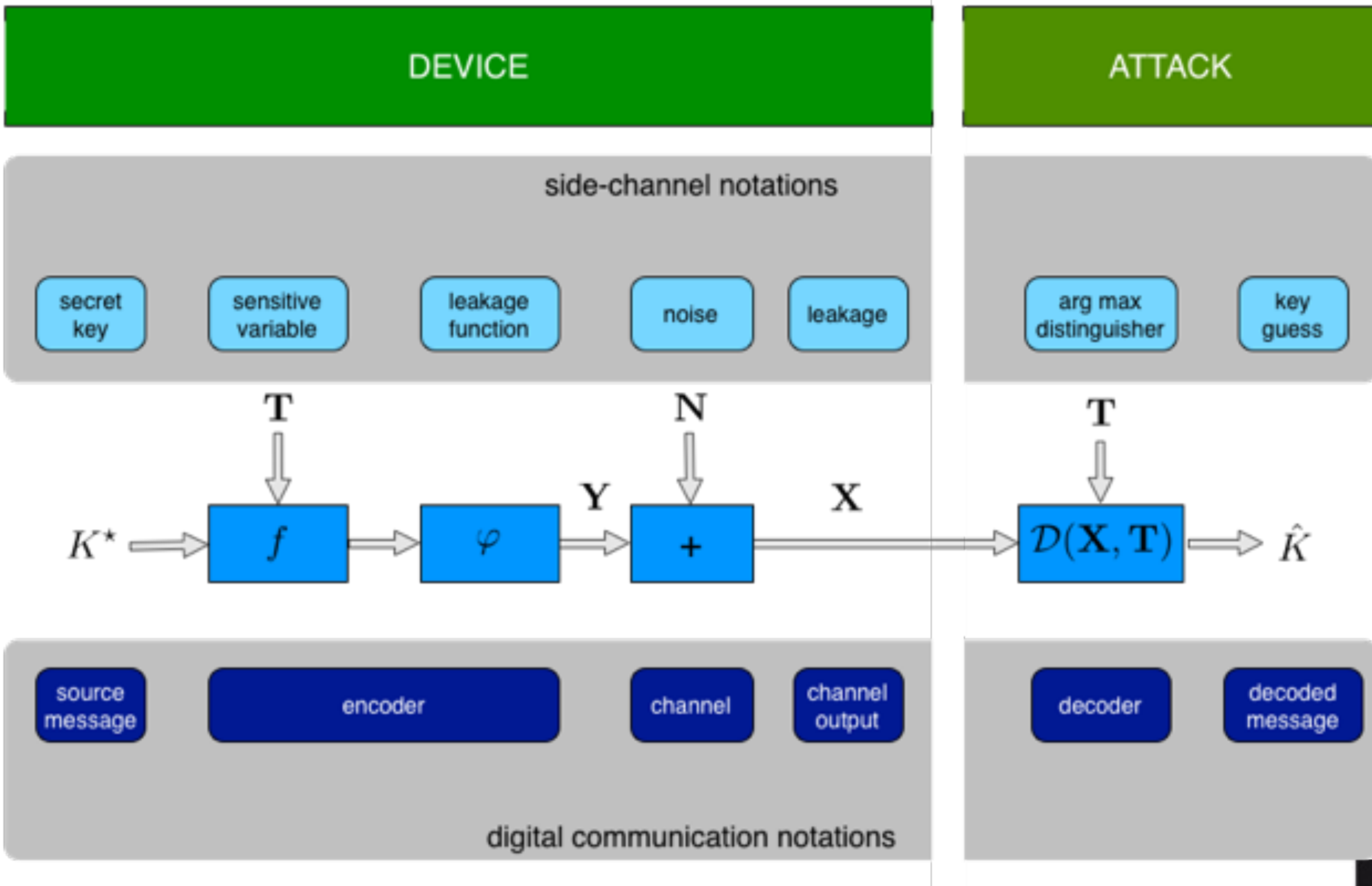


# Side-Channel Analysis As A Digital Communication Problem

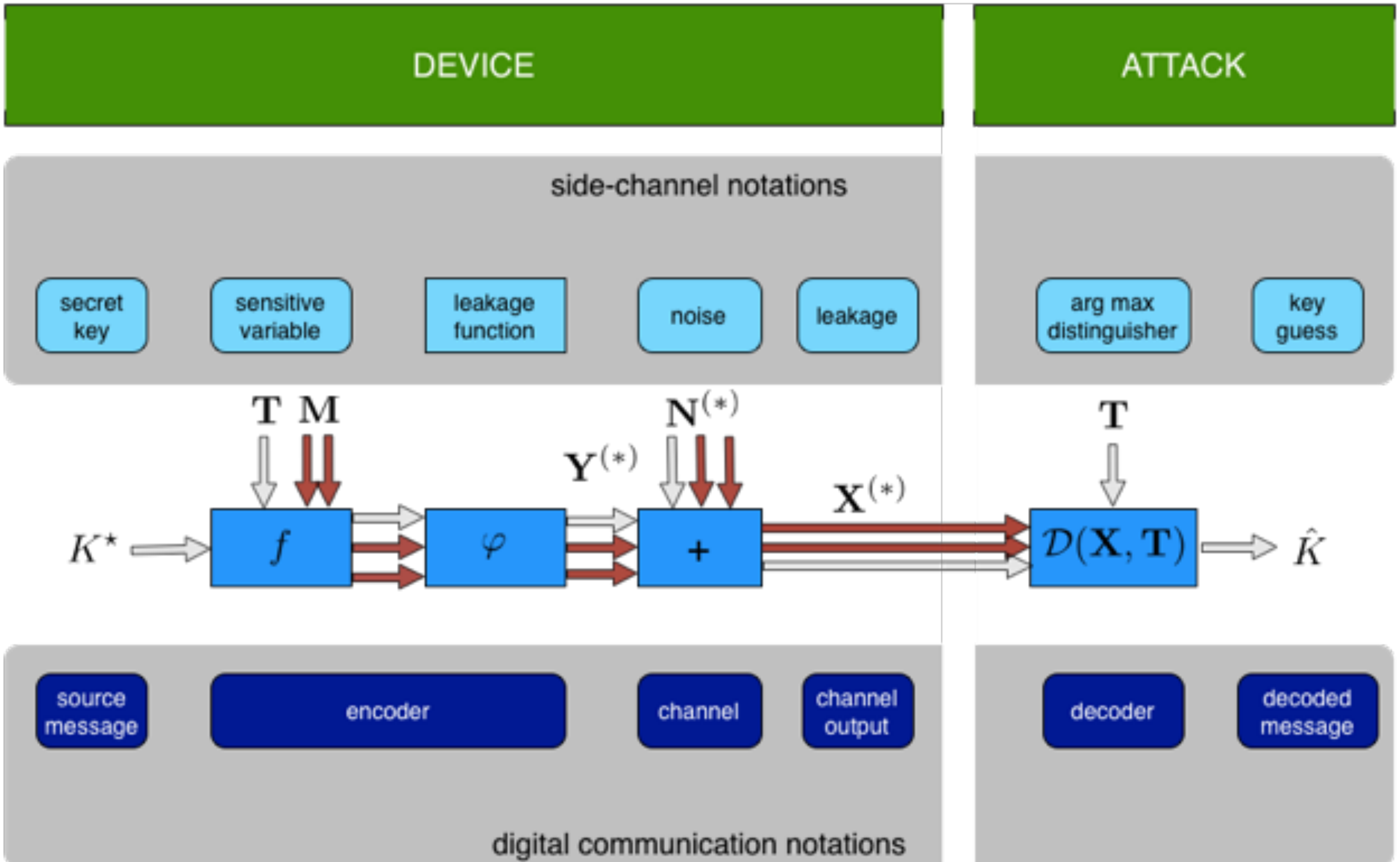


# Side-Channel Analysis As A Digital Communication Problem



# Side-Channel Analysis As A Digital Communication Problem



# Leave Prejudices & Misconceptions Behind…

- “Probability estimation is crucial”
- “Correlation Power Analysis is optimal”
- “Against first-order masked implementations, product combining is optimal”

# Leave Prejudices & Misconceptions Behind…

- “Probability estimation is crucial” **NO, we have proven it**
- “Correlation Power Analysis is optimal”
- “Against first-order masked implementations, product combining is optimal”

# Leave Prejudices & Misconceptions Behind…

- "Probability estimation is crucial" **NO, we have proven it**
- "Correlation Power Analysis is optimal" **Not always…**
- "Against first-order masked implementations, product combining is optimal"

# Leave Prejudices & Misconceptions Behind…

- "Probability estimation is crucial" — **NO, we have proven it**
- "Correlation Power Analysis is optimal" — **Not always…**
- "Against first-order masked implementations, product combining is optimal" — **NO, and we have proven it**

# Leave Prejudices & Misconceptions Behind…

- "Probability estimation is crucial" **NO, we have proven it**
- "Correlation Power Analysis is optimal" **Not always…**
- "Against first-order masked implementations, product combining is optimal" **NO, and we have proven it**

**Theorem** (Optimal expression for Gaussian noise). *When the noise is zero mean Gaussian, $N \sim \mathcal{N}(0, \sigma^2)$, the optimal distinguishing rule is*

$$\mathcal{D}_{opt}^{M,G}(\mathbf{x}, \mathbf{t}) = \arg\max_{k} \langle \mathbf{x} | \mathbf{y}(k) \rangle - \frac{1}{2} \|\mathbf{y}(k)\|_2^2.$$

# Leave Prejudices & Misconceptions Behind…

- "Probability estimation is crucial" **NO, we have proven it**
- "Correlation Power Analysis is optimal" **Not always…**
- "Against first-order masked implementations, product combining is optimal" **NO, and we have proven it**

**Theorem** (Optimal expression for Gaussian noise). *When the noise is zero mean Gaussian, $N \sim \mathcal{N}(0, \sigma^2)$, the optimal distinguishing rule is*

$$\mathcal{D}_{opt}^{M,G}(\mathbf{x}, \mathbf{t}) = \arg\max_k \langle \mathbf{x} | \mathbf{y}(k) \rangle - \frac{1}{2} \|\mathbf{y}(k)\|_2^2.$$

**Theorem** (Correlation power analysis). *When the leakage arises from $X = aY(K^\star) + b + N$ where $N$ is zero-mean Gaussian, the optimal distinguishing rule $\hat{k} = \arg\min_{k^\star} \min_{a,b} \|\mathbf{x} - a\mathbf{y}(k^\star) - b\|^2$ is equivalent to maximizing the absolute value of the empirical Pearson's coefficient:*

$$\hat{k} = \arg\max_{k^\star} |\hat{\rho}(k^\star)| = |\widehat{\mathrm{Cov}}(\mathbf{x}, \mathbf{y}(k^\star))| \Big/ \sqrt{\widehat{\mathrm{Var}}(\mathbf{x}) \cdot \widehat{\mathrm{Var}}(\mathbf{y}(k^\star))}.$$

# Mathematical Derivations

**Theorem** (Optimal expression for uniform and Laplacian noises). *When $f$ and $\varphi$ are known such that $Y(k) = \varphi(f(k,T))$, and the leakage arises from $X = Y(k^*) + N$ with $N \sim \mathcal{U}(0,\sigma^2)$ or $N \sim \mathcal{L}(0,\sigma^2)$, then the optimal distinguishing rule becomes*

- *Uniform noise distribution:* $\mathcal{D}_{opt}^{M,U}(\mathbf{x},\mathbf{t}) = \arg\max_k \; -\|\mathbf{x}-\mathbf{y}(k)\|_\infty$,
- *Laplace noise distribution:* $\mathcal{D}_{opt}^{M,L}(\mathbf{x},\mathbf{t}) = \arg\max_k \; -\|\mathbf{x}-\mathbf{y}(k)\|_1$.

**Theorem** (Optimal expression for unknown weights). *Let $\mathbf{Y}_{\boldsymbol{\alpha}}(k) = \sum_{j=1}^{n} \alpha_j [f(\mathbf{T},k)]_j$ and $\mathbf{Y}_j(k) = [f(\mathbf{T},k)]_j$, where the weights are independently deviating normally from the Hamming weight model, i.e., $\forall j \in [\![1,8, \alpha_j \sim \mathcal{N}(1,\sigma_\alpha^2)$. Then the optimal distinguishing rule is*

$$\mathcal{D}_{opt}^{\alpha,G}(\mathbf{x},\mathbf{t}) = \arg\max_k \, (\gamma\langle\mathbf{x}|\mathbf{y}(k)\rangle + \mathbf{1})^t \cdot (\gamma Z(k)+I)^{-1} \cdot (\gamma\langle\mathbf{x}|\mathbf{y}(k)\rangle + \mathbf{1}) - \sigma_\alpha^2 \ln\det(\gamma Z(k)+I), \quad (1)$$

*where $\gamma = \frac{\sigma_\alpha^2}{\sigma^2}$ is the epistemic to stochastic noise ratio (ESNR), $\langle\mathbf{x}|\mathbf{y}\rangle$ is the vector with elements $(\langle\mathbf{x}|\mathbf{y}(k)\rangle)_j = \langle\mathbf{x}|\mathbf{y}_j(k)\rangle$, $Z(k)$ is the $n\times n$ Gram matrix with entries $Z_{j,j'}(k) = \langle\mathbf{y}_j(k)|\mathbf{y}_{j'}(k)\rangle$, $\mathbf{1}$ is the all-one vector, and $I$ is the identity matrix.*

**Theorem** (Second-order HOOD). *If the model (i.e., $\varphi^{(s)}$) is known to the attacker for all $s$ in the direct scale, then the* second-order HOOD *is*

$$\mathcal{D}_{opt}^2(\mathbf{x}^{(\star)},\mathbf{t}^{(\star)}) = \arg\max_{k\in\mathcal{K}} p_k(\mathbf{x}^{(\star)}|\mathbf{t}^{(\star)}) = \arg\max_{k\in\mathcal{K}} \prod_{q=1}^{Q} \sum_{m^{(\star)}\in\mathcal{M}^{(\star)}} \mathbb{P}(m^{(\star)}) \prod_{s=0}^{1} p_k(x_q^{(s)}|t_q^{(s)}, m^{(s)}).$$

$$\mathcal{D}_{C\text{-CPA}}^{mt,\sigma\uparrow}(\mathbf{x}^{(\star)},\mathbf{t}) = \arg\max_{k\in\mathcal{K}} \sum_{s\in\mathbb{F}_2^n} \rho(c_X^{n\text{-}prod}(\mathbf{x}^{(s)},\mathbf{x}^{(2^n)}), c_Y^{\text{opt}}(\mathbf{y}^{(s)},\mathbf{y}^{(2^n)})) - \frac{1}{2}\rho(\mathbf{x}^{(s)}, c_Y^{\text{opt}}(\mathbf{y}^{(s)},\mathbf{y}^{(2^n)^2}))$$

**Proposition** (Second-order HOOD for Gaussian noise). *Assuming that $N^{(s)} \sim \mathcal{N}(0,{\sigma^{(s)}}^2)$ then the second-order optimal distinguisher in the direct scale becomes*

$$\mathcal{D}_{opt}^{2,G}(\mathbf{x}^{(0)},\mathbf{x}^{(1)},\mathbf{t}) = \arg\max_{k\in\mathcal{K}} \prod_{q=1}^{Q} \sum_{m\in\mathcal{M}} \exp\left\{-\frac{1}{2}\left(\frac{-2x_q^{(0)}y^{(0)}(t_q,k,m)+y^{(0)}(t_q,k,m)^2}{{\sigma^{(0)}}^2} + \frac{-2x_q^{(1)}y^{(1)}(t_q,k,m)+y^{(1)}(t_q,k,m)^2}{{\sigma^{(1)}}^2}\right)\right\}.$$

**Proposition** (Second-order HOOD for low Gaussian noise). *Assuming that both shares have the same low noise standard deviation $\sigma = \sigma^{(0)} = \sigma^{(1)}$ then the optimal distinguisher reduces at first order to*

$$\mathcal{D}_{opt}^{2,G,\sigma\downarrow}(\mathbf{x}^{(\star)},\mathbf{t}) = \arg\min_{k\in\mathcal{K}} \sum_{q=1}^{Q} \max_{m\in\mathbb{F}_2^n} (x_q^{(0)} - y^{(0)}(t_q,k,m))^2 + (x_q^{(1)} - y^{(1)}(t_q,k,m))^2$$

# Mathematical Derivations

**Theorem** (Optimal expression for uniform and Laplacian noises). *When $f$ and $\varphi$ are known such that $Y(k) = \varphi(f(k,T))$, and the leakage arises from $X = Y(k^*) + N$ with $N \sim \mathcal{U}(0,\sigma^2)$ or $N \sim \mathcal{L}(0,\sigma^2)$, then the optimal distinguishing rule becomes*

- *Uniform noise distribution:* $\mathcal{D}_{opt}^{M,U}(\mathbf{x},\mathbf{t}) = \arg\max_k \; -\|\mathbf{x}-\mathbf{y}(k)\|_\infty$,
- *Laplace noise distribution:* $\mathcal{D}_{opt}^{M,L}(\mathbf{x},\mathbf{t}) = \arg\max_k \; -\|\mathbf{x}-\mathbf{y}(k)\|_1$.

**Theorem** (Optimal expression for unknown weights). *Let $\mathbf{Y}_{\boldsymbol{\alpha}}(k) = \sum_{j=1}^{n} \alpha_j [f(\mathbf{T},k)]_j$ and $\mathbf{Y}_j(k) = [f(\mathbf{T},k)]_j$, where the weights are independently deviating normally from the Hamming weight model, i.e., $\forall j \in [\![1,8, \alpha_j \sim \mathcal{N}(1,\sigma_\alpha^2)$. Then the optimal distinguishing rule is*

$$\mathcal{D}_{opt}^{\alpha,G}(\mathbf{x},\mathbf{t}) = \arg\max_k \, (\gamma\langle\mathbf{x}|\mathbf{y}(k)\rangle + \mathbf{1})^t \cdot (\gamma Z(k) + I)^{-1} \cdot (\gamma\langle\mathbf{x}|\mathbf{y}(k)\rangle + \mathbf{1}) - \sigma_\alpha^2 \ln\det(\gamma Z(k) + I), \quad (1)$$

*where $\gamma = \frac{\sigma_\alpha^2}{\sigma^2}$ is the epistemic to stochastic noise ratio (ESNR), $\langle\mathbf{x}|\mathbf{y}\rangle$ is the vector with elements $(\langle\mathbf{x}|\mathbf{y}(k)\rangle)_j = \langle\mathbf{x}|\mathbf{y}_j(k)\rangle$, $Z(k)$ is the $n\times n$ Gram matrix with entries $Z_{j,j'}(k) = \langle\mathbf{y}_j(k)|\mathbf{y}_{j'}(k)\rangle$, $\mathbf{1}$ is the all-one vector, and $I$ is the identity matrix.*

**Theorem** (Second-order HOOD). *If the model (i.e., $\varphi^{(s)}$) is known to the attacker for all $s$ in the direct scale, then the* second-order HOOD *is*

$$\mathcal{D}_{opt}^{2}(\mathbf{x}^{(\star)},\mathbf{t}^{(\star)}) = \arg\max_{k\in\mathcal{K}} \; p_k(\mathbf{x}^{(\star)}|\mathbf{t}^{(\star)})$$

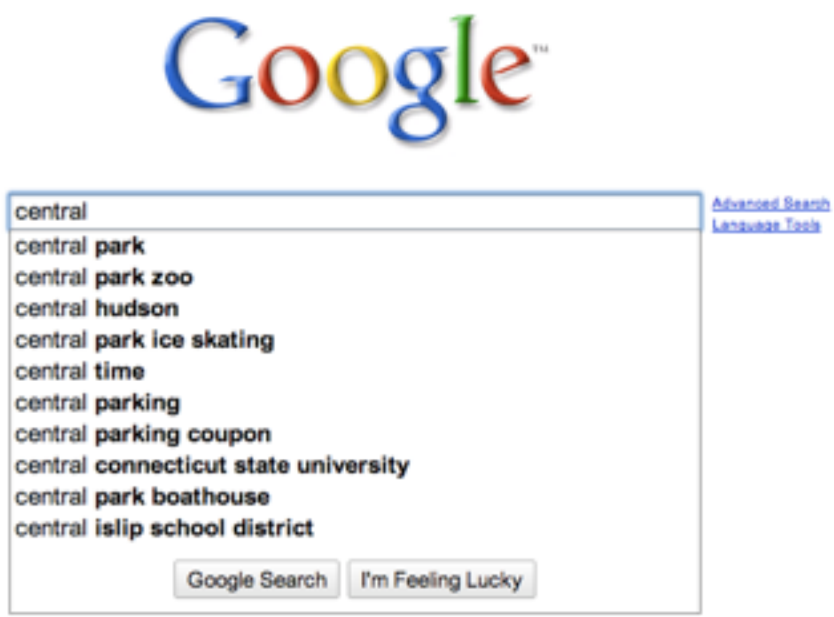
$\mathcal{D}_{C\text{-}CPA}^{mt,\sigma\uparrow}(\mathbf{x}^{(\star)},\mathbf{t})$

**Sylvain Guilley, Annelie Heuser, Olivier Rioul,**
***Methods for recovering secret data of a cryptographic device and for validating the security of such a device,***
**Brevet déposé par l'INSTITUT MINES-TELECOM, PCT N° PCT/IB2014/003248, 2014**

# Scientific Commitments/ Achievements

# Scientific Commitments/ Achievements

- **Reviewer:** COSADE 2013/2014, HOST 2013/2014, HASP 2013, SPACES 2013/2014, JCEN (several times), IEEE Trans. of Information Forensics & Security, IEEE Transactions on Computers
- **Conference organizer:** COSADE 2011-2014
- **Advisor:** 4 interns, students group (Polytechnique)



# Scientific Commitments/ Achievements

- **Reviewer:** COSADE 2013/2014, HOST 2013/2014, HASP 2013, SPACES 2013/2014, JCEN (several times), IEEE Trans. of Information Forensics & Security, IEEE Transactions on Computers
- **Conference organizer:** COSADE 2011-2014
- **Advisor:** 4 interns, students group (Polytechnique)
- **7 workshops:** Cryptarchi 2011, Cryptarchi 2013-2015, CrossFyre 2011-2013, Kryptotag 2010
- **2 poster presentations:** CHES 2013, 2014

# Scientific Commitments/ Achievements

- **Reviewer:** COSADE 2013/2014, HOST 2013/2014, HASP 2013, SPACES 2013/2014, JCEN (several times), IEEE Trans. of Information Forensics & Security, IEEE Transactions on Computers
- **Conference organizer:** COSADE 2011-2014
- **Advisor:** 4 interns, students group (Polytechnique)
- **7 workshops:** Cryptarchi 2011, Cryptarchi 2013-2015, CrossFyre 2011-2013, Kryptotag 2010
- **2 poster presentations:** CHES 2013, 2014
- **15 international conferences:** CHES 2015, INDOCRYPT 2015, ACNS 2014, ASIACRYPT 2014, CHES 2014, COSADE 2014, CRiSIS 2014, SPACE 2014, CARDIS 2013, COSADE 2012, CT-RSA 2012, DATE 2012, HOST 2012, IIH-MSP 2012, DSD 2011
- **3 journals:** JCEN 4(4), JCEN 3(4), JCEN 3(3)
- **1 book chapter** in Trusted Computing for Embedded Systems, Springer 2015

# Scientific Commitments/ Achievements

- **Reviewer:** COSADE 2013/2014, HOST 2013/2014, HASP 2013, SPACES 2013/2014, JCEN (several times), IEEE Trans. of Information Forensics & Security, IEEE Transactions on Computers
- **Conference organizer:** COSADE 2011-2014
- **Advisor:** 4 interns, students group (Polytechnique)
- **7 workshops:** Cryptarchi 2011, Cryptarchi 2013-2015, CrossFyre 2011-2013, Kryptotag 2010
- **2 poster presentations:** CHES 2013, 2014
- **15 international conferences:** CHES 2015, INDOCRYPT 2015, ACNS 2014, ASIACRYPT 2014, CHES 2014, COSADE 2014, CRiSIS 2014, SPACE 2014, CARDIS 2013, COSADE 2012, CT-RSA 2012, DATE 2012, HOST 2012, IIH-MSP 2012, DSD 2011
- **3 journals:** JCEN 4(4), JCEN 3(4), JCEN 3(3)
- **1 book chapter** in Trusted Computing for Embedded Systems, Springer 2015
- **1 Baby**

# Currently…

- Post-doctoral researcher
- European project SECODE (CHIST-ERA)
- Secure Codes to Thwart Cyber-Physical Attacks

# Thank you!


annelie.heuser@telecom-paristech.fr