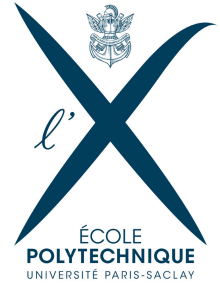
Institut
Mines-Telecom

Masks will Fall Off

Higher-Order Optimal Distinguishers

Nicolas Bruneau, Sylvain Guilley,
Annelie Heuser, Olivier Rioul

ASIACRYPT 2014, Kaohsiung, Taiwan

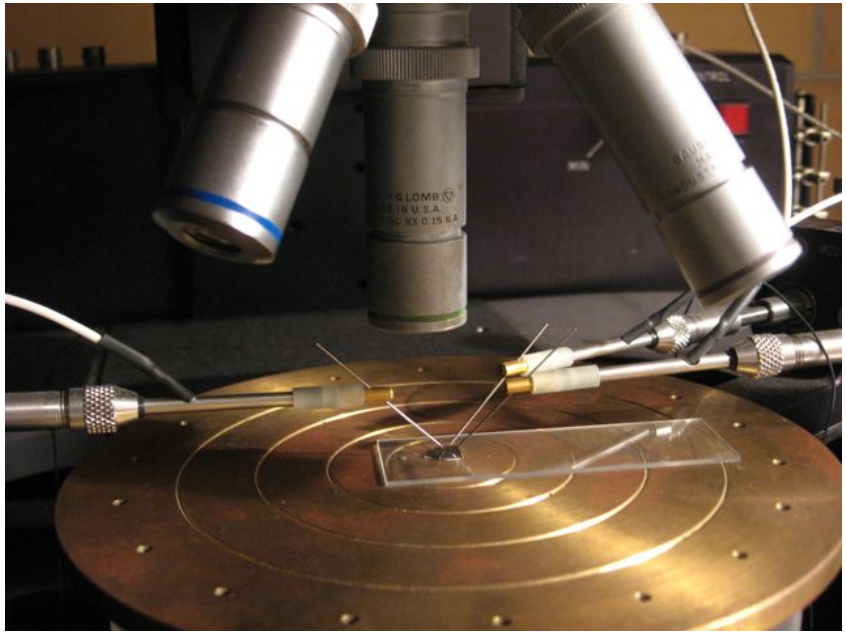
Nicolas BRUNEAU	Sylvain GUILLEY	Annelie HEUSER	Olivier RIOUL
is also with	*is also with*	*is PhD fellow at*	*is also Prof at*
			

Outline

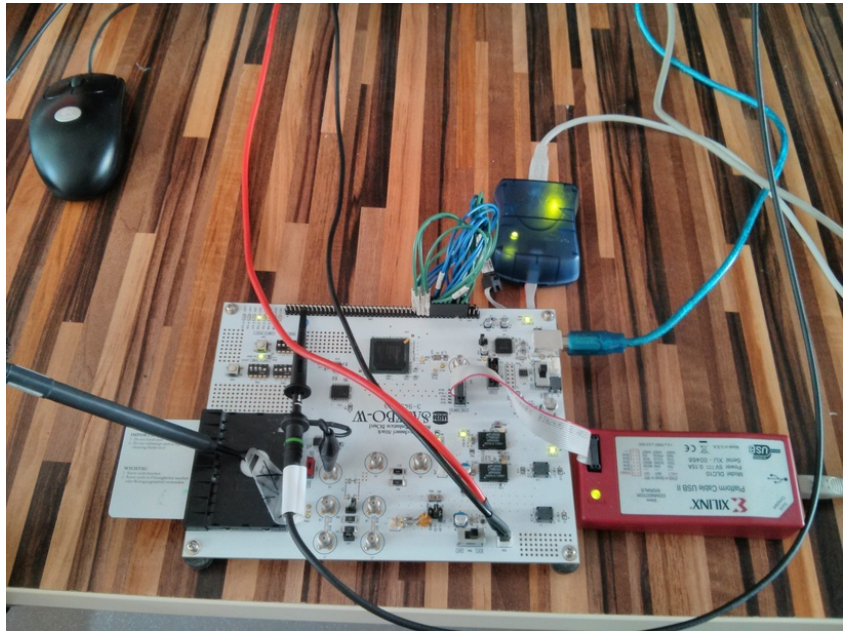
Outline

Higher-order Side-channel Analysis ($d = 2$)

Digital

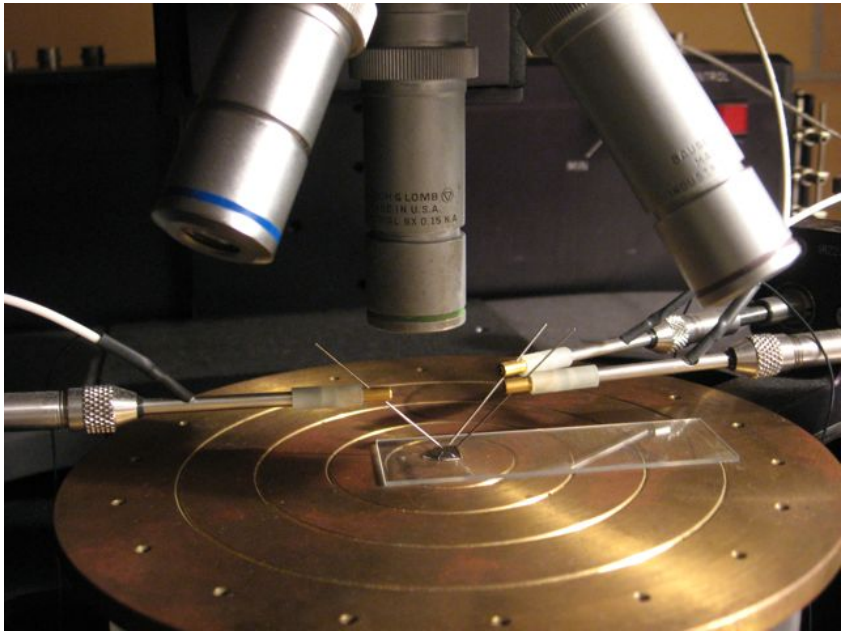


Analog

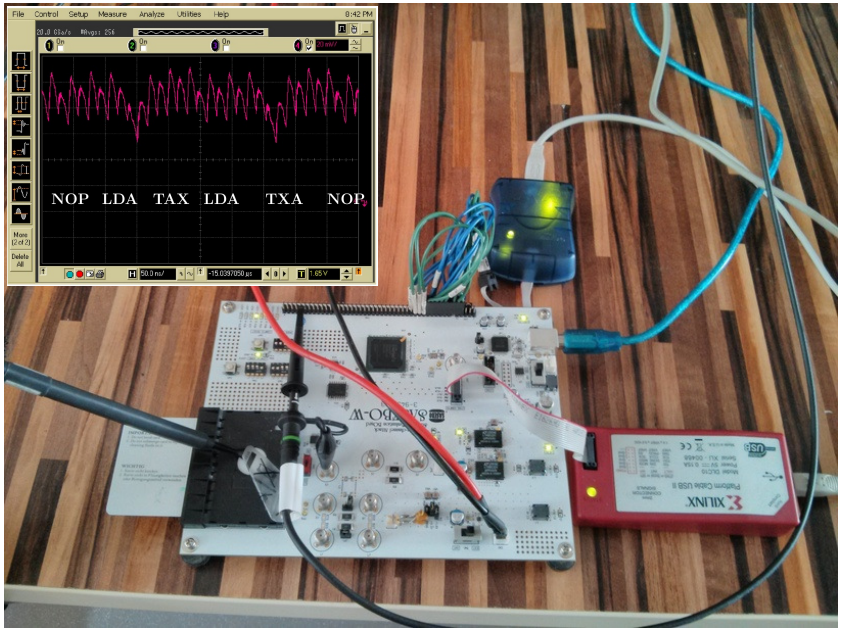


Higher-order Side-channel Analysis ($d = 2$)

Digital

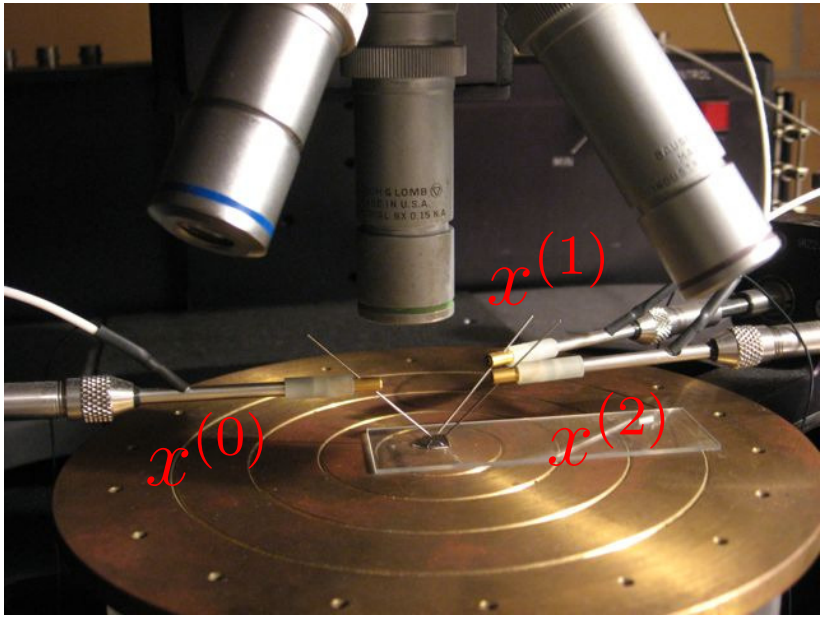


Analog

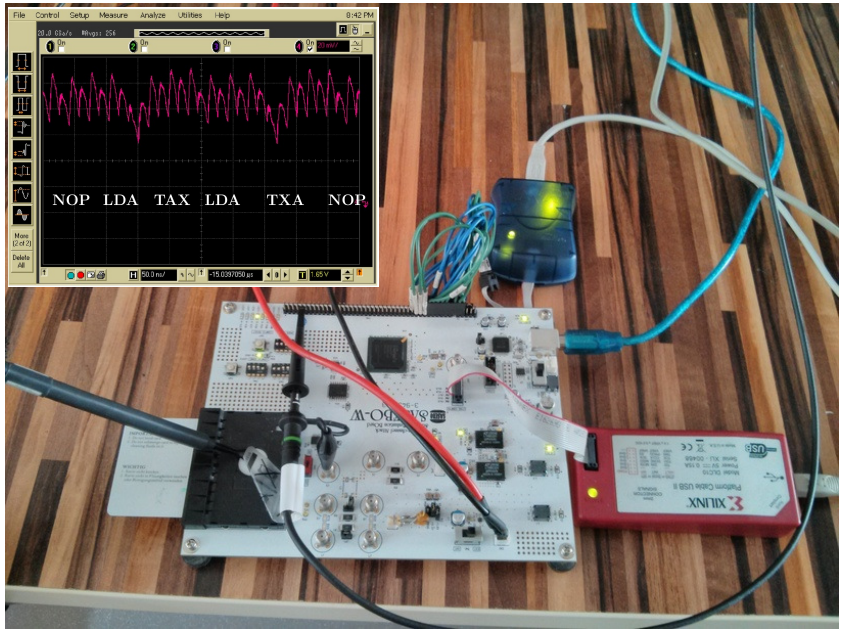


Higher-order Side-channel Analysis ($d = 2$)

Digital ..... $x^{(\omega)} \in \mathbb{F}_2$

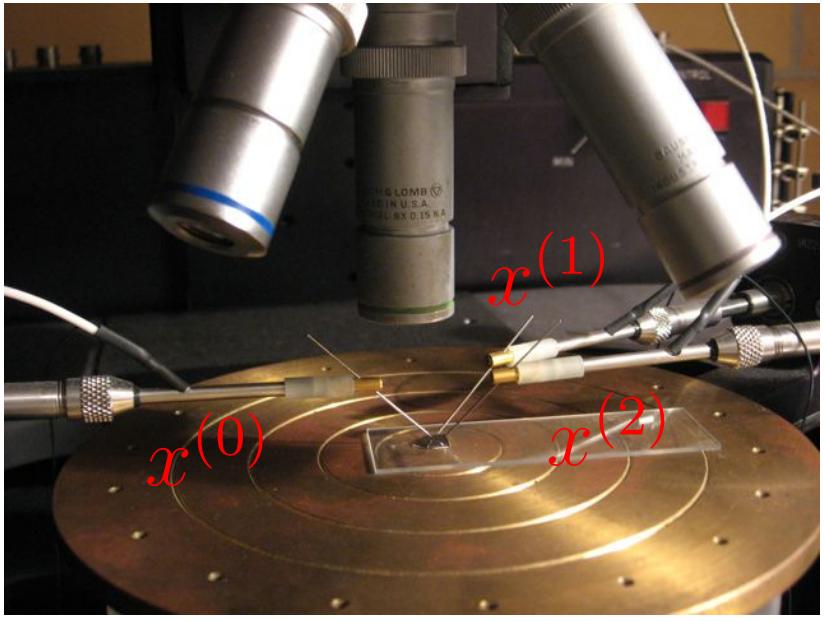


Analog

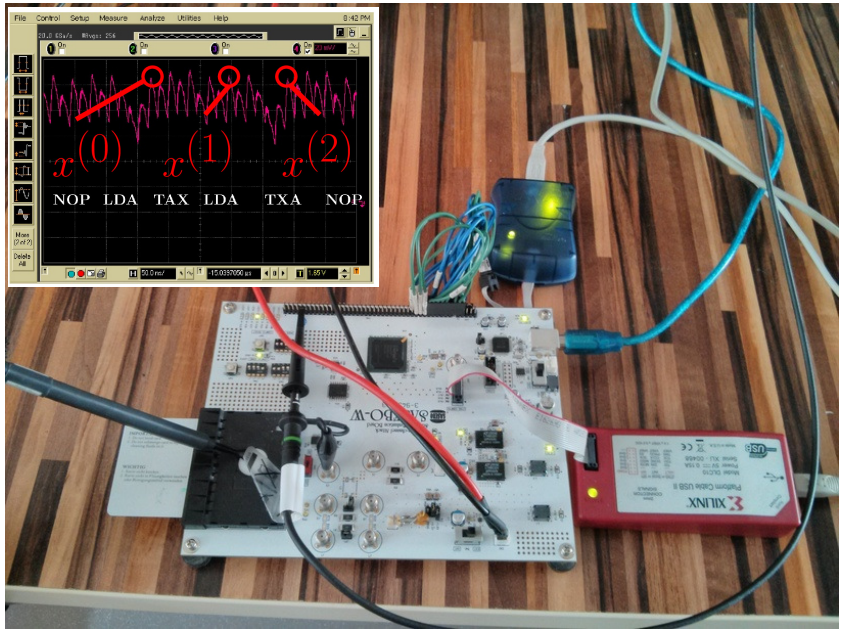


Higher-order Side-channel Analysis ($d = 2$)

Digital ..... $x^{(\omega)} \in \mathbb{F}_2$

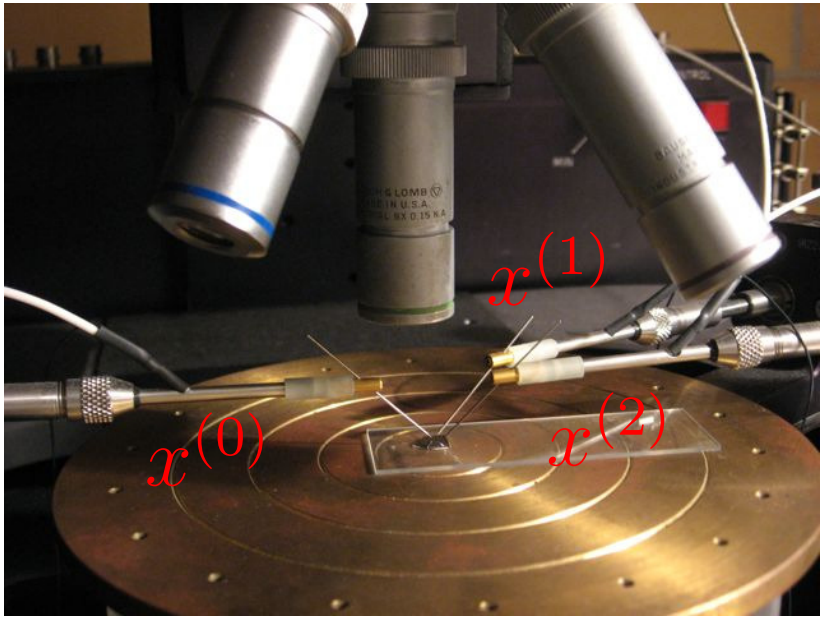


Analog ..... $x^{(\omega)} \in \mathbb{R}$



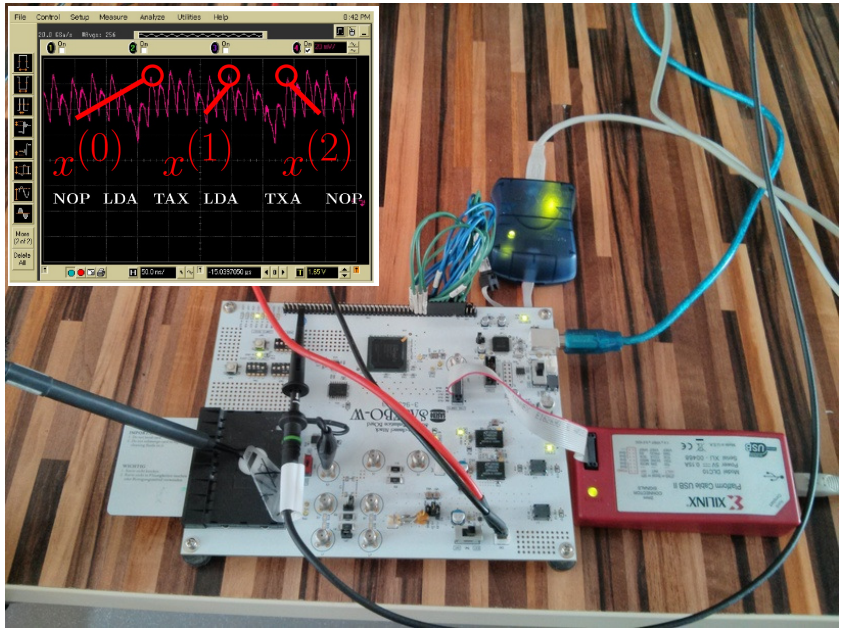
Higher-order Side-channel Analysis ($d = 2$)

Digital ..... $x^{(\omega)} \in \mathbb{F}_2$



Physical security: probing, etc. (Ishai, Sahai, Wagner – CRYPTO '03)

Analog ..... $x^{(\omega)} \in \mathbb{R}$

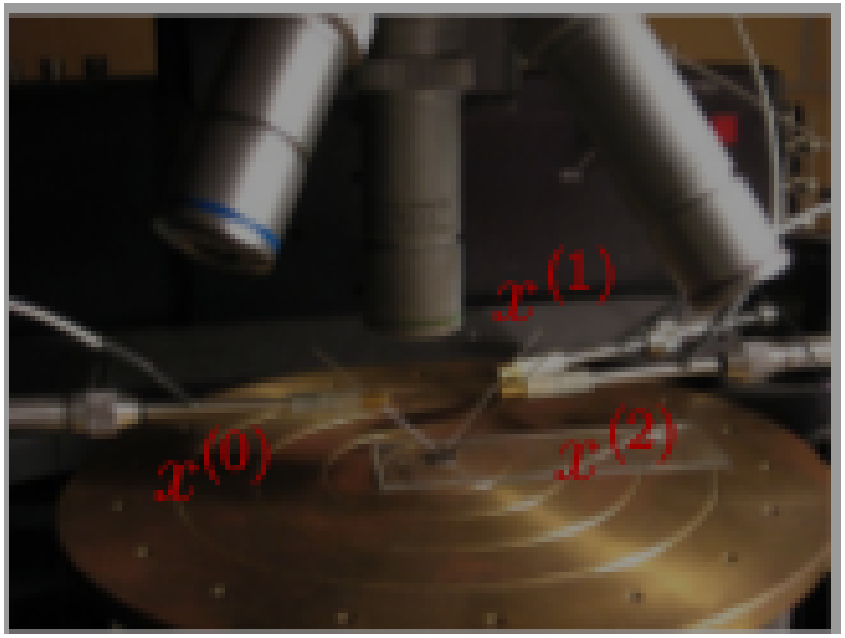


Physical security: EM attacks (Kocher, Jaffe, Jun – CRYPTO '99)

$\Longrightarrow$ **Our focus in this paper!**

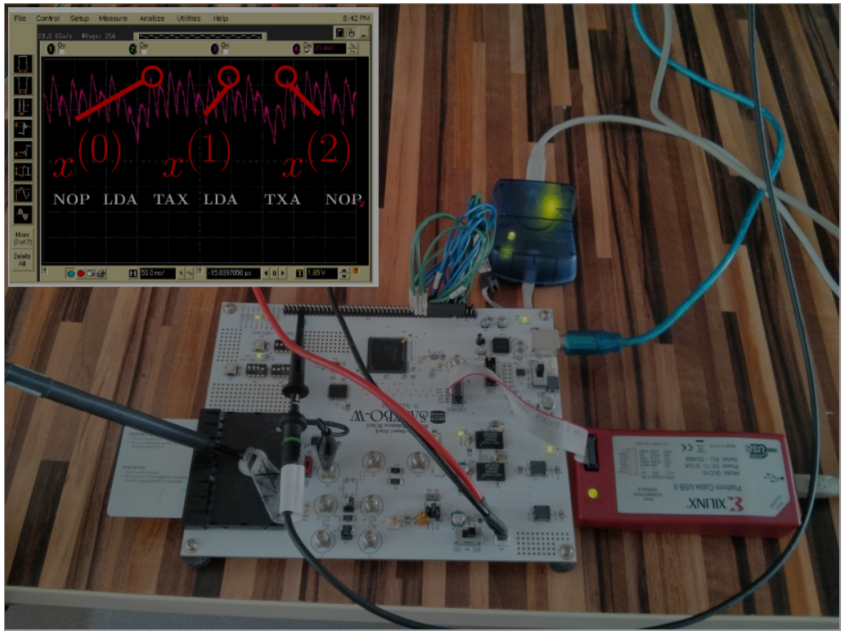
Higher-order Side-channel Analysis ($d = 2$)

Digital ..... $x^{(\omega)} \in \mathbb{F}_2$



Parallel with cyber security: debugger, heartbleed, etc.

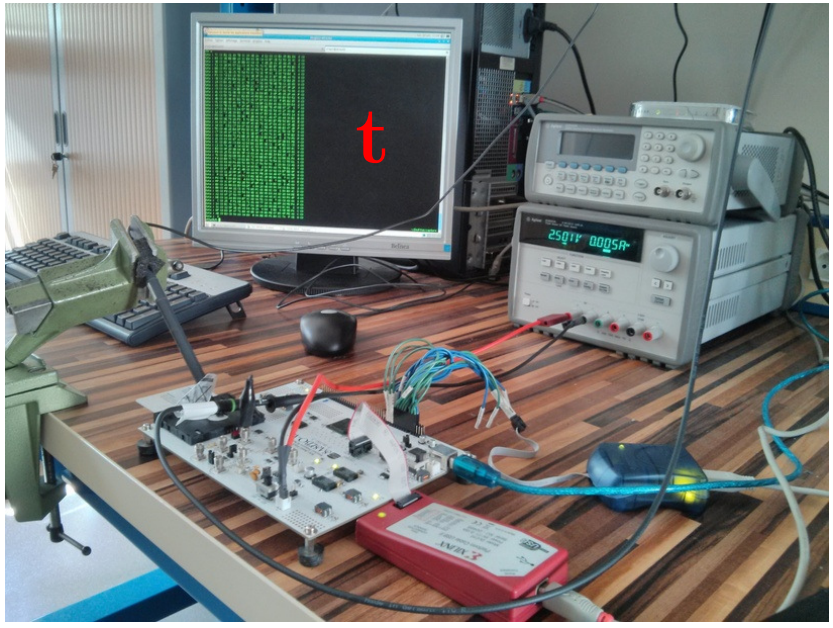
Analog ..... $x^{(\omega)} \in \mathbb{R}$

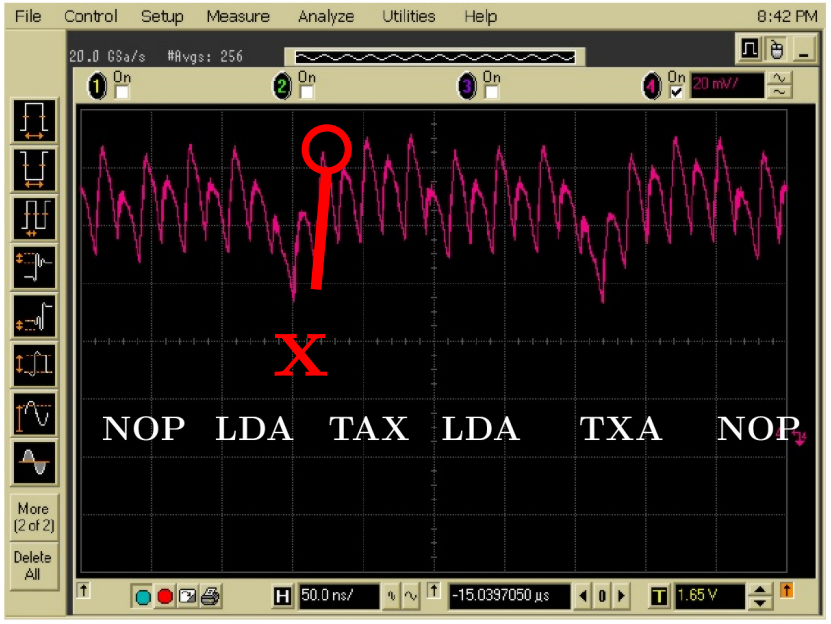


Parallel with cyber security: packet arrival timing, etc.

Outline

Side-Channel Analysis as a Digital Com. Problem (CHES '14 [HRG14])

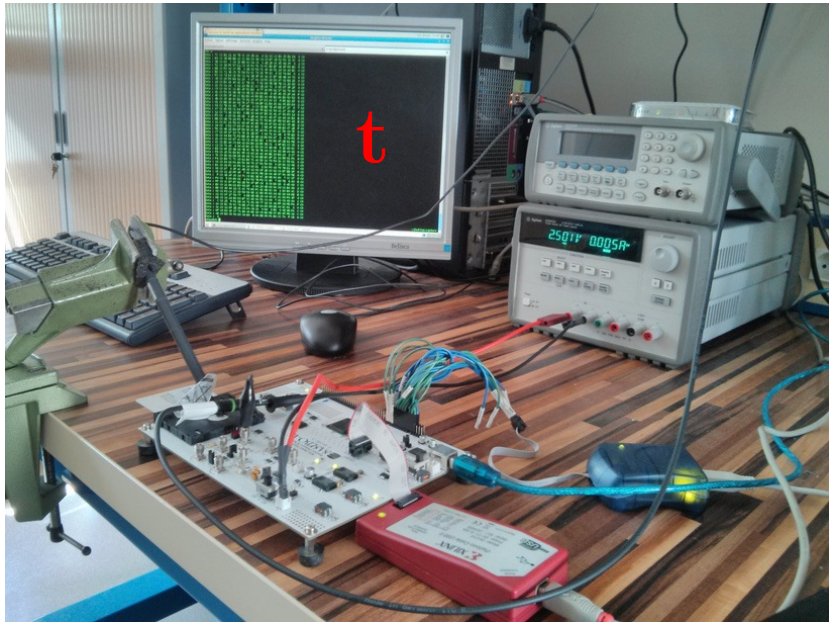


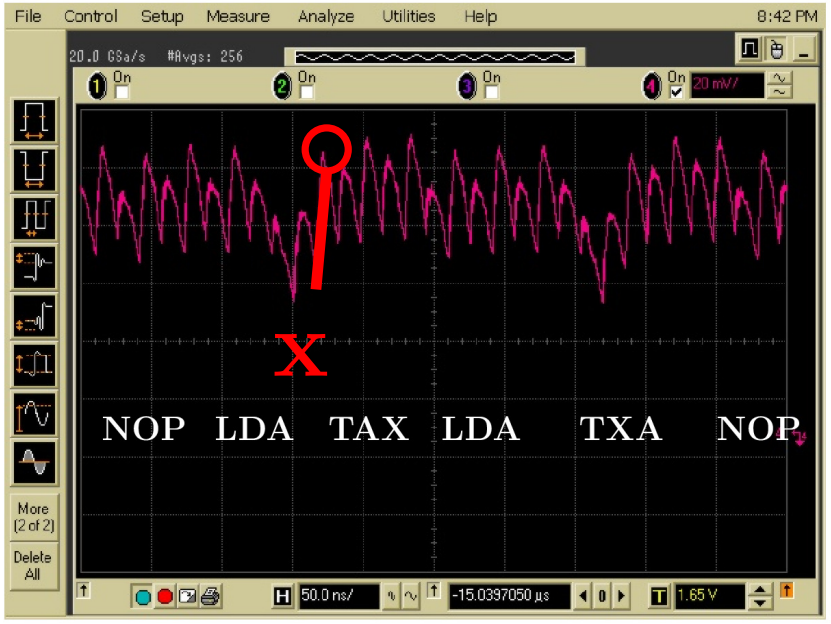


t: t_1 t_2 t_3 . . . t_q

x: x_1 x_2 x_3 . . . x_q

Side-Channel Analysis as a Digital Com. Problem (CHES '14 [HRG14])

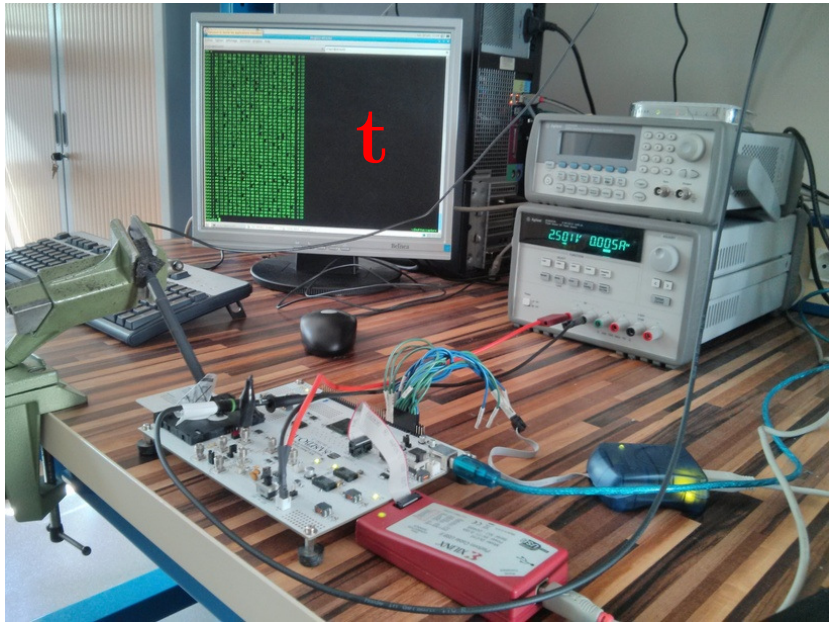


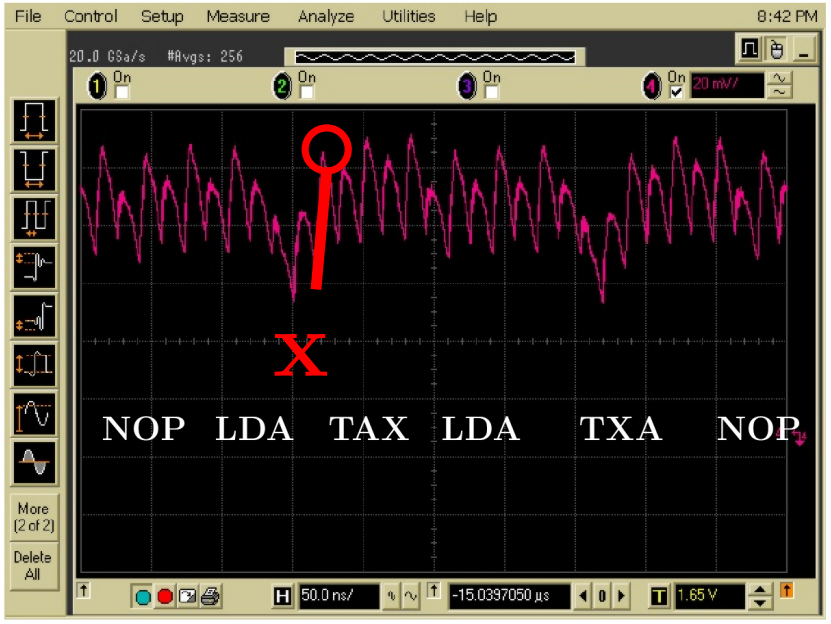


t: t_1 t_2 t_3 . . . t_q

x: x_1 x_2 x_3 . . . x_q

Side-Channel Analysis as a Digital Com. Problem (CHES '14 [HRG14])

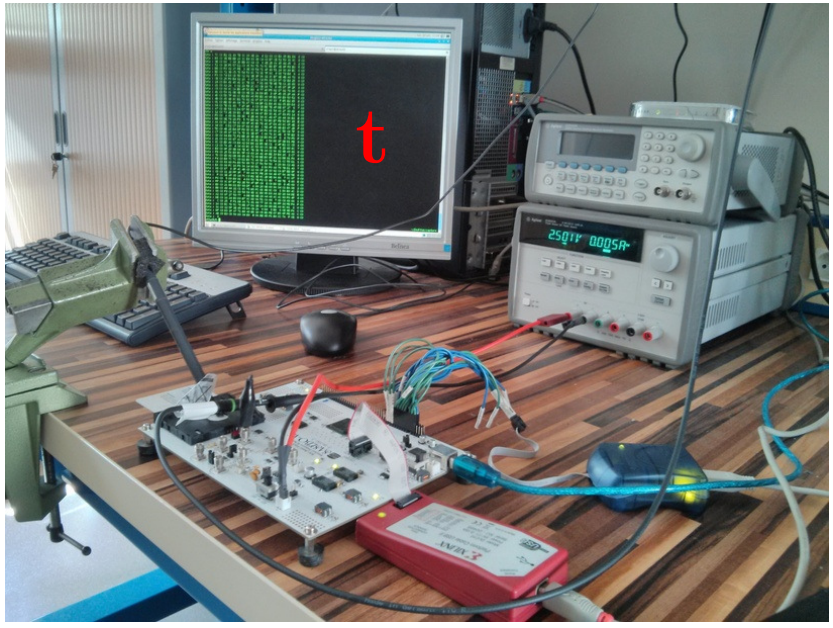


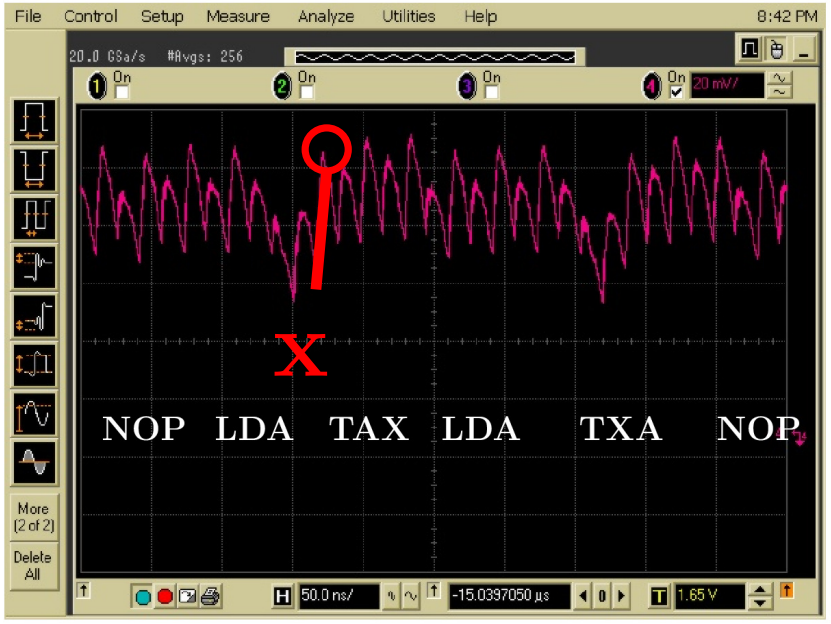


t: t_1 t_2 t_3 . . . t_q

x: x_1 x_2 x_3 . . . x_q

Side-Channel Analysis as a Digital Com. Problem (CHES '14 [HRG14])

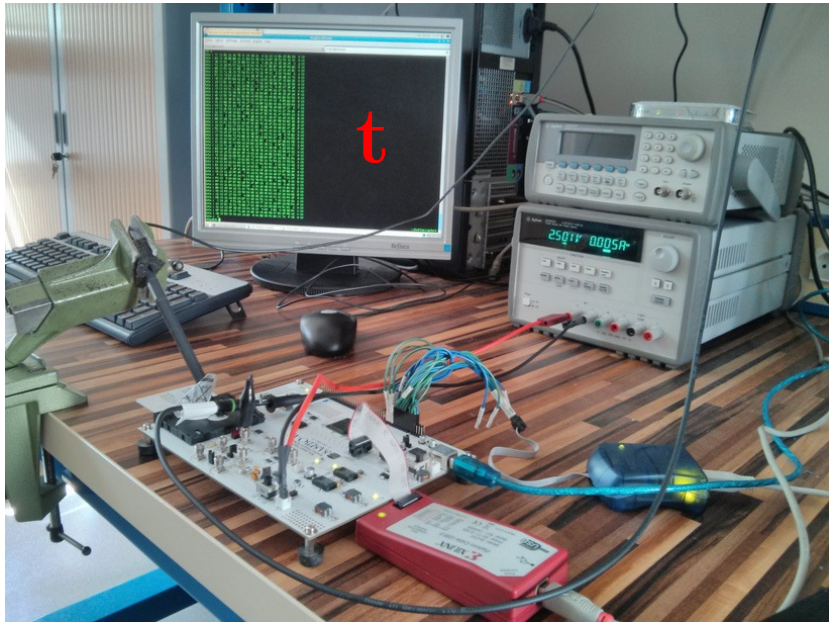


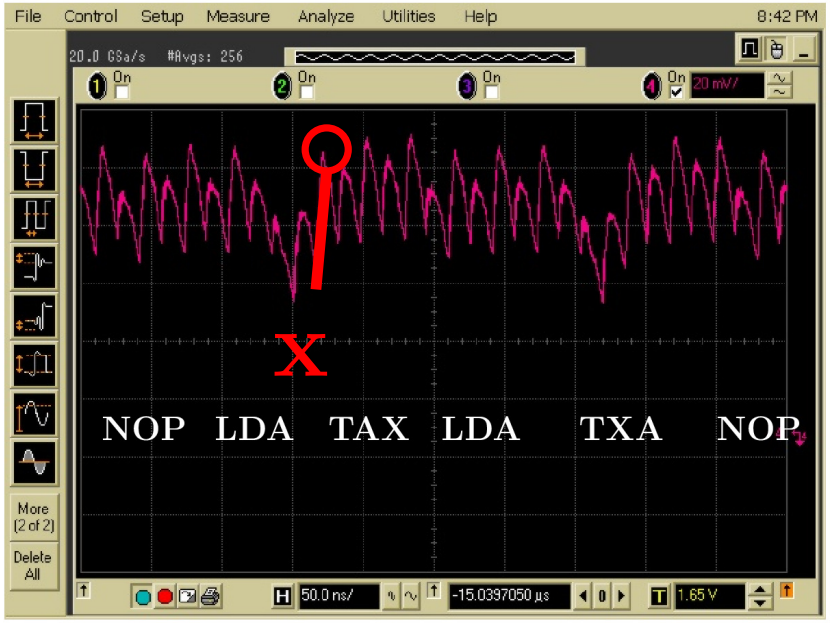


t: t_1 t_2 t_3 ... t_q

x: x_1 x_2 x_3 ... x_q

Side-Channel Analysis as a Digital Com. Problem (CHES '14 [HRG14])

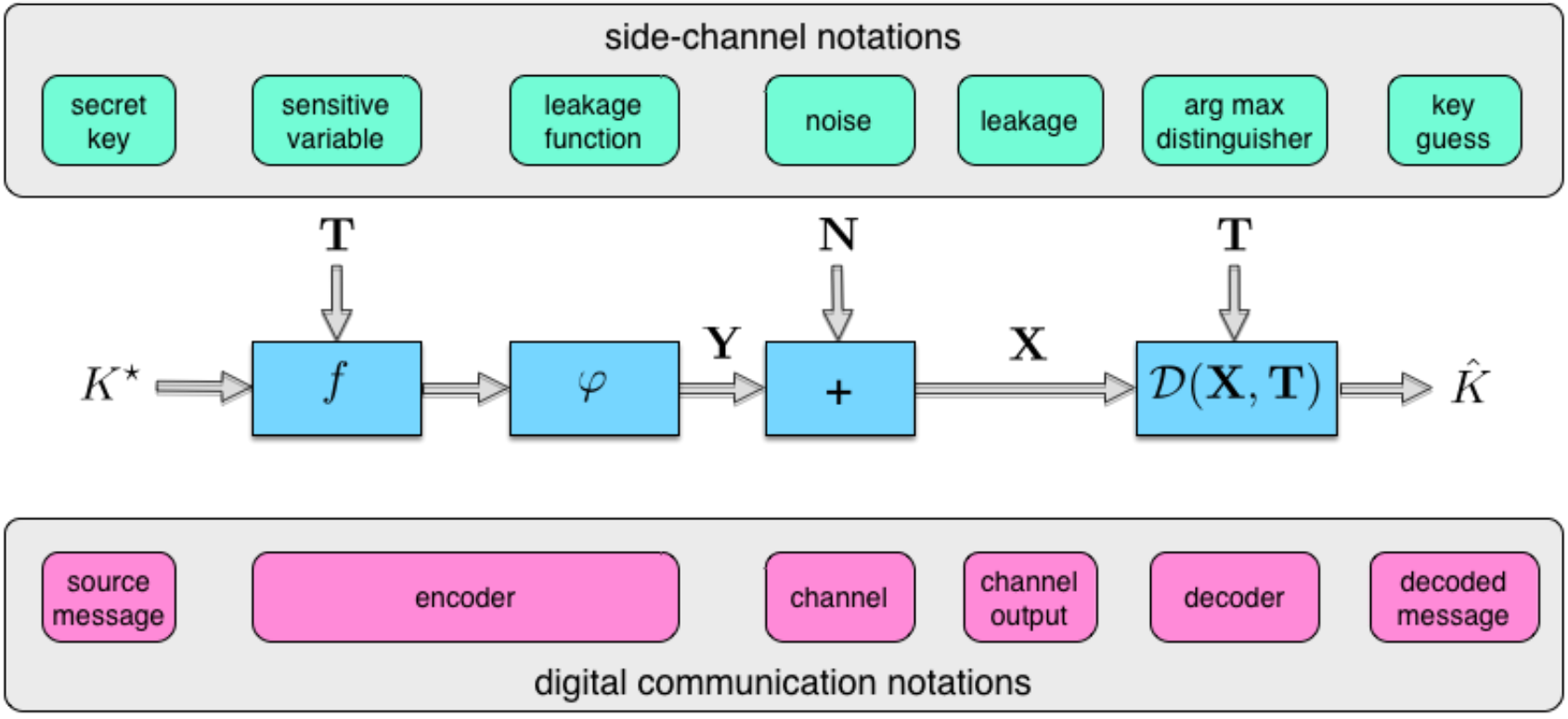




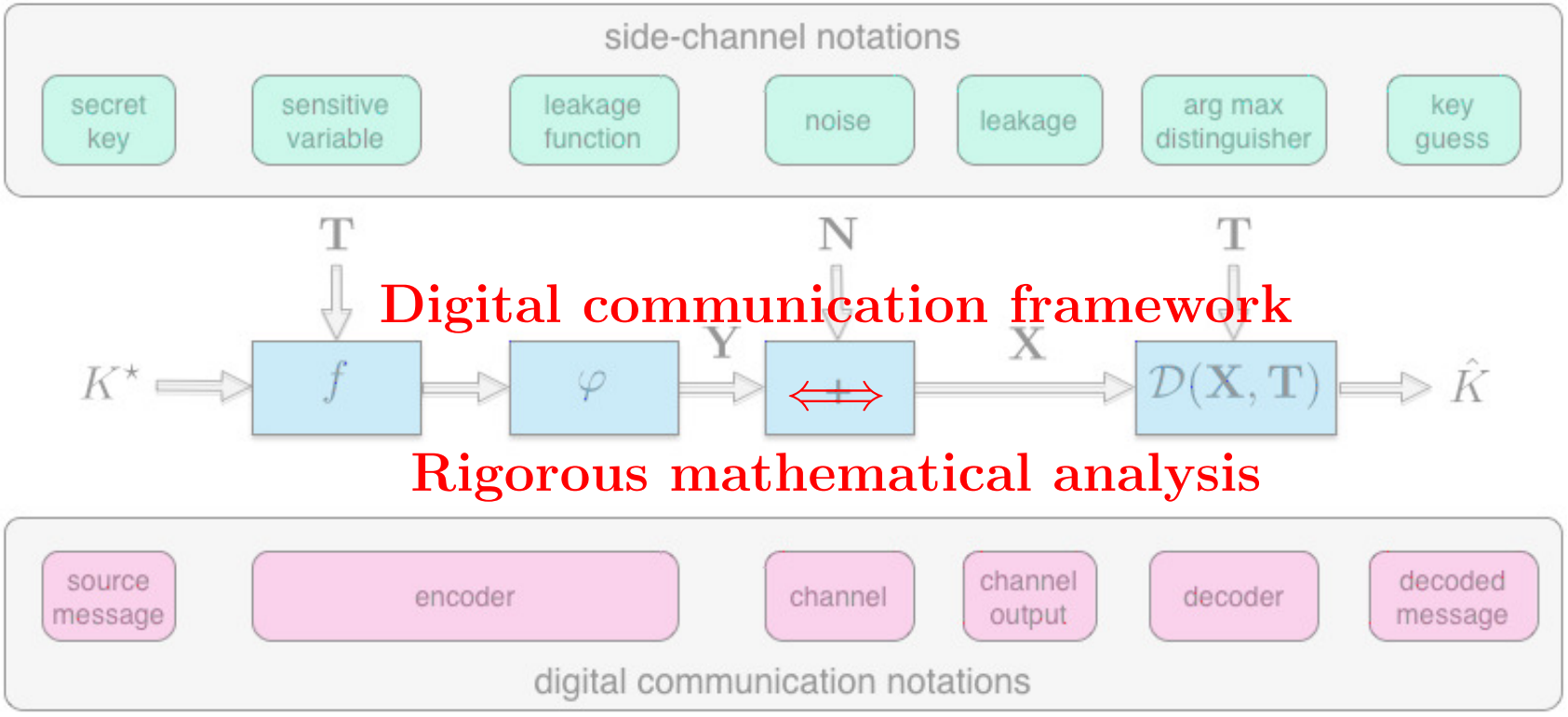
t: t_1 t_2 t_3 . . . t_q

x: x_1 x_2 x_3 . . . x_q

Side-Channel Analysis as a Digital Com. Problem (CHES '14 [HRG14])



Side-Channel Analysis as a Digital Com. Problem (CHES '14 [HRG14])



Optimal distinguisher [HRG14]

Ingredients

$\mathbf{t}$ and $\mathbf{x}$.

Objective

$$\text{maximize } \mathbb{P}_S, \text{ where } \mathbb{P}_S = \mathbb{P}(\hat{K} = K^\star) \ . \tag{1}$$

Optimal distinguisher [HRG14]

Solution

$$\max \mathbb{P}_S = \mathbb{P}(\hat{K} = K^\star) \iff$$

$$\mathcal{D}(\mathbf{x}, \mathbf{t}) = argmax_k \ \mathbb{P}(k)\ p(\mathbf{x}|\mathbf{t}, k) \qquad \text{[MAP]}$$

Also known as *template attack*
(Chari, Rao, Rohatgi – CHES '02 [CRR02]).

Optimal distinguisher [HRG14]

Solution

$$\max \mathbb{P}_S = \mathbb{P}(\hat{K} = K^\star) \iff$$
$$\mathcal{D}(\mathbf{x}, \mathbf{t}) = argmax_k \; p(\mathbf{x}|\mathbf{y}(\mathbf{t}, k)) \qquad \text{[ML]}$$

Also known as *template attack*
(Chari, Rao, Rohatgi – CHES '02 [CRR02]).

Derivation

We find optimal distinguishers that ...

- **are not** Difference of Means
- **are not** Pearson Correlation
- **are not** rank-based (Spearman, Kendall, Gini, ...) Correlation
- **are not** Mutual Information
- **are not** Kolmogorov-Smirnov Distance
- ...

- **are** actually nothing known so far...

Derivation

We find optimal distinguishers that ...

- **are not** Difference of Means
- **are not** Pearson Correlation
- **are not** rank-based (Spearman, Kendall, Gini, ...) Correlation
- **are not** Mutual Information
- **are not** Kolmogorov-Smirnov Distance
- ...

- **are** actually nothing known so far...

Derivation

We find optimal distinguishers that ...

- **are not** Difference of Means
- **are not** Pearson Correlation
- **are not** rank-based (Spearman, Kendall, Gini, ...) Correlation
- **are not** Mutual Information
- **are not** Kolmogorov-Smirnov Distance
- ...

- **are** actually nothing known so far...

Derivation

We find optimal distinguishers that ...

- **are not** Difference of Means
- **are not** Pearson Correlation
- **are not** rank-based (Spearman, Kendall, Gini, ...) Correlation
- **are not** Mutual Information
- **are not** Kolmogorov-Smirnov Distance
- ...

- **are** actually nothing known so far...

Derivation

We find optimal distinguishers that ...

- **are not** Difference of Means
- **are not** Pearson Correlation
- **are not** rank-based (Spearman, Kendall, Gini, ...) Correlation
- **are not** Mutual Information
- **are not** Kolmogorov-Smirnov Distance
- ...

- **are** actually nothing known so far...

Derivation

We find optimal distinguishers that ...

- **are not** Difference of Means
- **are not** Pearson Correlation
- **are not** rank-based (Spearman, Kendall, Gini, ...) Correlation
- **are not** Mutual Information
- **are not** Kolmogorov-Smirnov Distance
- ...

- **are** actually nothing known so far...

Derivation

We find optimal distinguishers that ...

- **are not** Difference of Means
- **are not** Pearson Correlation
- **are not** rank-based (Spearman, Kendall, Gini, ...) Correlation
- **are not** Mutual Information
- **are not** Kolmogorov-Smirnov Distance
- ...

- **are** actually nothing known so far...

Example: optimal distinguisher for one bit

Solution

- **is**:

$$\mathcal{D}^{M,G}_{opt(1\text{ bit})}(\mathbf{x},\mathbf{t}) = \arg\max_{k^\star} \sum_{i|y_i(k^\star)=1} x_i - \sum_{i|y_i(k^\star)=-1} x_i \ .$$

- but is **neither**:

$$\mathcal{D}^{M,G}_{\text{KJJ}}(\mathbf{x},\mathbf{t}) = \arg\max_{k^\star} \overline{\mathbf{x}_{+1}} - \overline{\mathbf{x}_{-1}}$$

(Kocher, Jaffe, Jun – CRYPTO '99)

nor:

$$\mathcal{D}^{M,G}_{\text{CKN}}(\mathbf{x},\mathbf{t}) = \arg\max_{k^\star} (\overline{\mathbf{x}_{+1}} - \overline{\mathbf{x}_{-1}}) \Big/ \sqrt{\frac{\sigma^2_{\mathbf{x}_{+1}}}{n_{+1}} + \frac{\sigma^2_{\mathbf{x}_{-1}}}{n_{-1}}} \ .$$

(Coron, Kocher, Naccache – FC '00)

Example: optimal distinguisher for one bit

Solution

- **is**:

$$\mathcal{D}^{M,G}_{opt(1\text{ bit})}(\mathbf{x},\mathbf{t}) = \arg\max_{k^\star} \sum_{i|y_i(k^\star)=1} x_i - \sum_{i|y_i(k^\star)=-1} x_i \ .$$

- but is **neither**:
  $\mathcal{D}^{M,G}_{\text{KJJ}}(\mathbf{x},\mathbf{t}) = \arg\max_{k^\star} \overline{\mathbf{x}_{+1}} - \overline{\mathbf{x}_{-1}}$
  (Kocher, Jaffe, Jun – CRYPTO '99)
  nor:
  $$\mathcal{D}^{M,G}_{\text{CKN}}(\mathbf{x},\mathbf{t}) = \arg\max_{k^\star} (\overline{\mathbf{x}_{+1}} - \overline{\mathbf{x}_{-1}}) \Big/ \sqrt{\frac{\sigma^2_{\mathbf{x}_{+1}}}{n_{+1}} + \frac{\sigma^2_{\mathbf{x}_{-1}}}{n_{-1}}} \ .$$
  (Coron, Kocher, Naccache – FC '00)

Example: optimal distinguisher in multi-bit

Solution ([HRG14], and also [MOS09, app. D])

- **is**:

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

- but is **neither** "covariance" **nor** "correlation",

- (but gets closer to CPA as SNR decreases).

Example: optimal distinguisher in multi-bit

Solution ([HRG14], and also [MOS09, app. D])

- **is**:

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

- but is **neither** "covariance" **nor** "correlation",

- (but gets closer to CPA as SNR decreases).

Example: optimal distinguisher in multi-bit

Solution ([HRG14], and also [MOS09, app. D])

- **is**:

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

- but is **neither** "covariance" **nor** "correlation",

- (but gets closer to CPA as SNR decreases).

Masking Countermeasure

Example (First-order software masking)

For example a first-order masking scheme ($d = 1$) might leak with

$$\begin{cases} X^{(0)} &= \mathrm{HW}[M] + N^{(0)}, \\ X^{(1)} &= \mathrm{HW}[\mathtt{Sbox}[T \oplus k^*] \oplus M] + N^{(1)}. \end{cases}$$

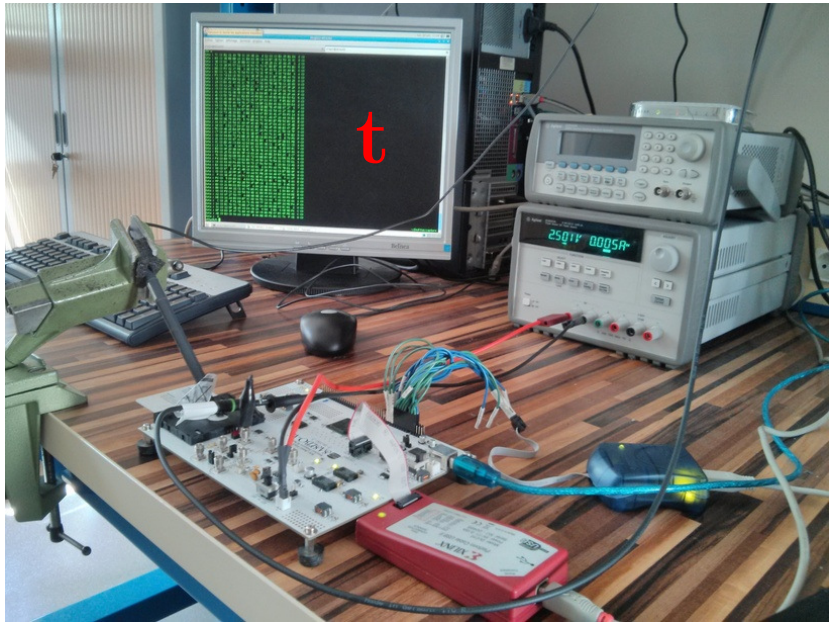
Masking Countermeasure

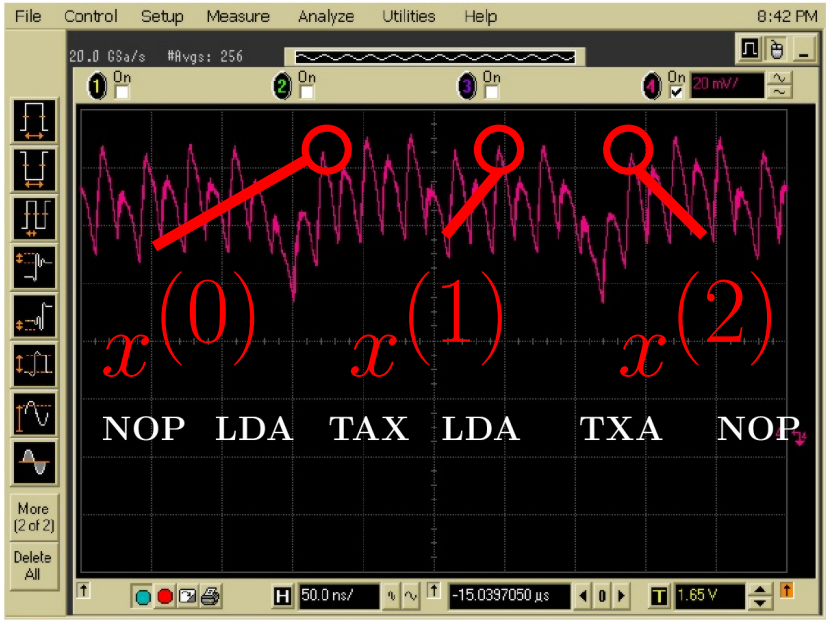
Example (Tables pre-computation)

Again when assuming a Hamming weight leakage model, a masking scheme using Sbox recomputation [KJJ99] might leak with

$$\begin{cases} X^{(\omega)} &= \mathrm{HW}[\omega \oplus M] + N^{(\omega)}, \qquad \forall \omega \in \{0, 1, \ldots, 2^n - 1\} \\ X^{(2^n)} &= \mathrm{HW}[T \oplus k^* \oplus M] + N^{(2^n)}. \end{cases}$$

Side-Channel Analysis as a Digital Com. Problem [HRG14] ($d = 2$)

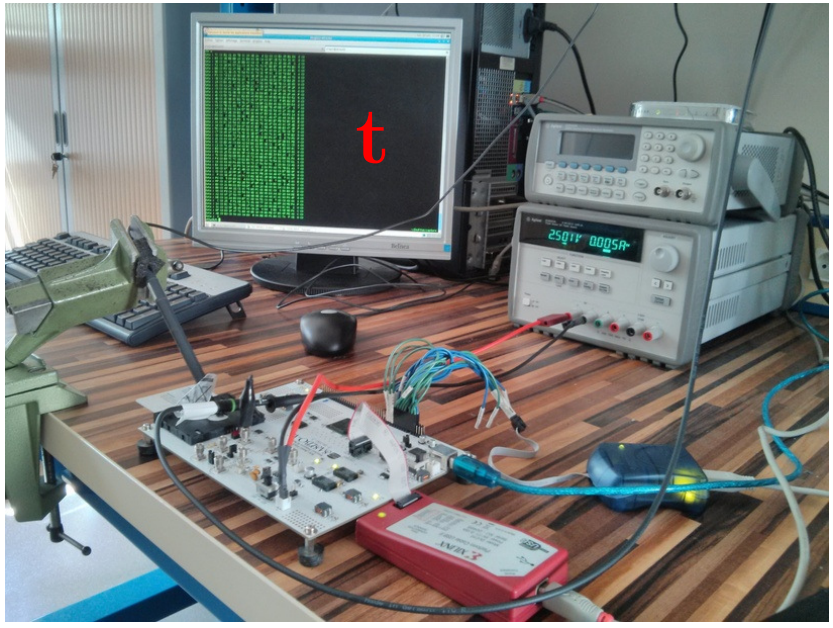


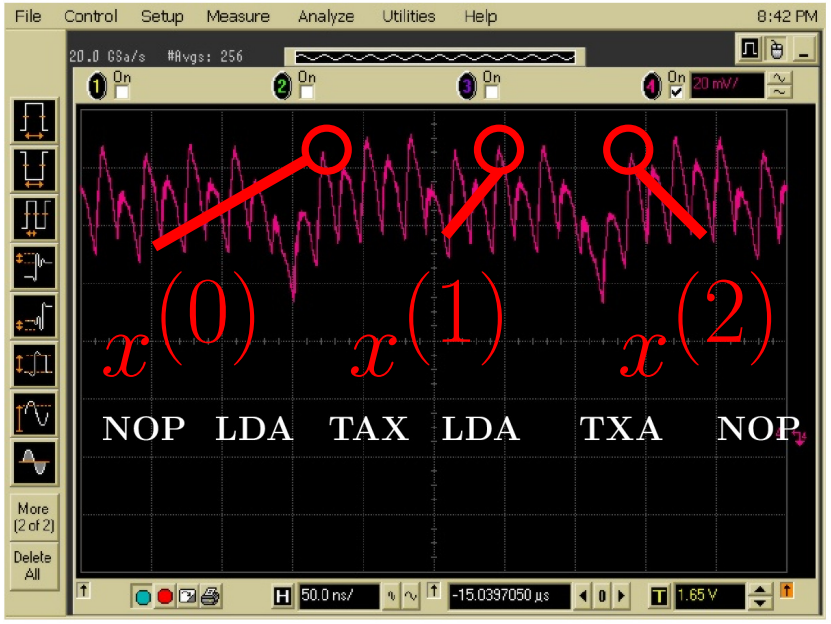


$\mathbf{t}$: t_1 t_2 t_3 $\ldots$ t_q

$\mathbf{x}^{(\star)}$: $x_1^{(\star)}$ $x_2^{(\star)}$ $x_3^{(\star)}$ $\ldots$ $x_q^{(\star)}$

Side-Channel Analysis as a Digital Com. Problem [HRG14] ($d = 2$)

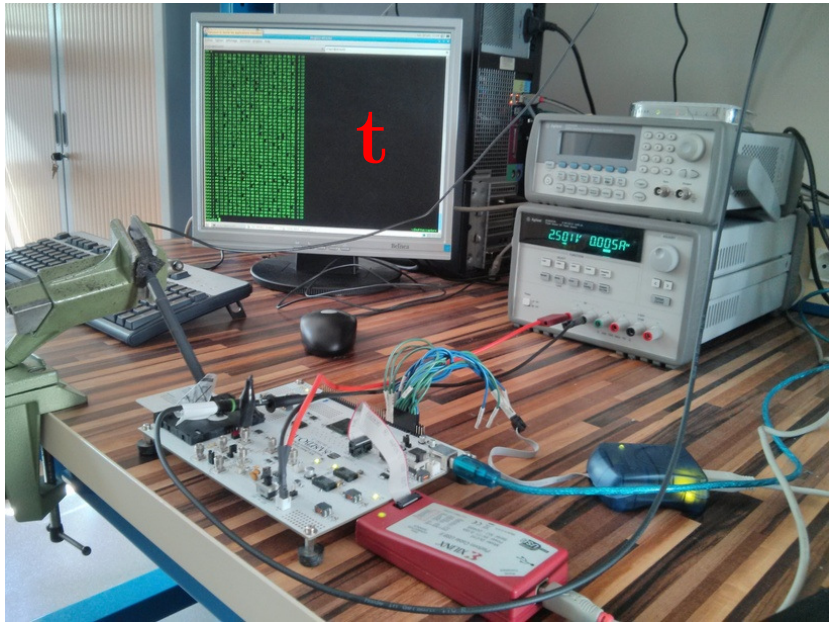


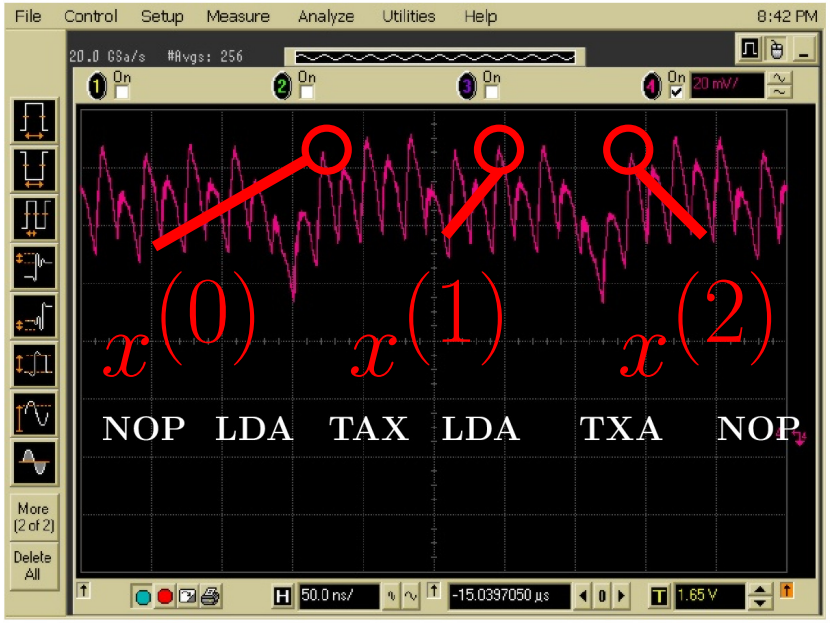


$\mathbf{t}$: t_1 t_2 t_3 $\ldots$ t_q

$\mathbf{x}^{(\star)}$: $x_1^{(\star)}$ $x_2^{(\star)}$ $x_3^{(\star)}$ $\ldots$ $x_q^{(\star)}$

Side-Channel Analysis as a Digital Com. Problem [HRG14] ($d = 2$)

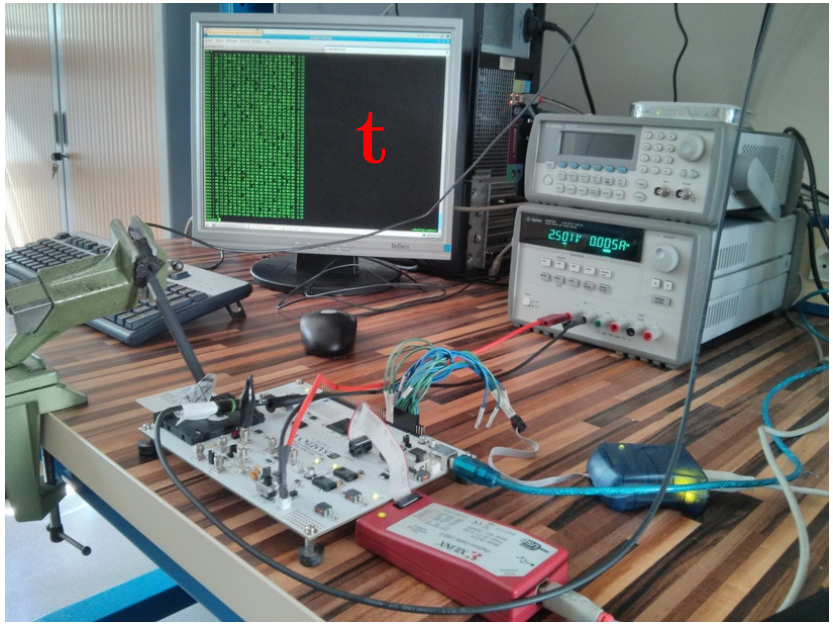


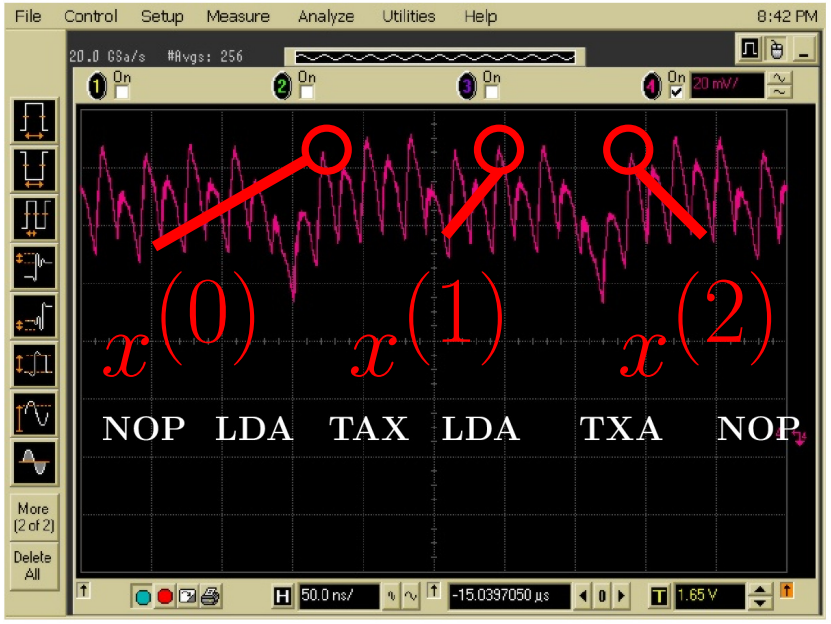


$\mathbf{t}$: t_1 t_2 t_3 ... t_q

$\mathbf{x}^{(\star)}$: $x_1^{(\star)}$ $x_2^{(\star)}$ $x_3^{(\star)}$ ... $x_q^{(\star)}$

Side-Channel Analysis as a Digital Com. Problem [HRG14] ($d = 2$)

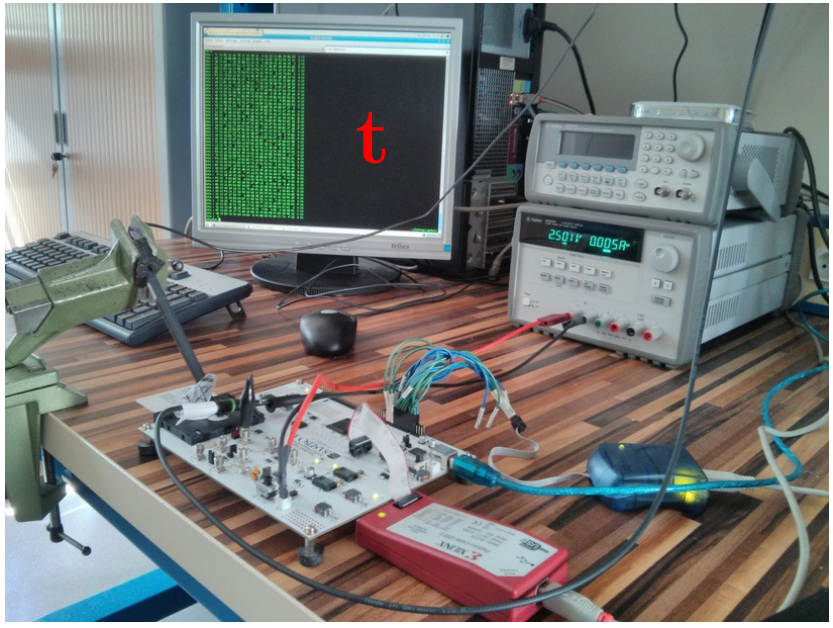


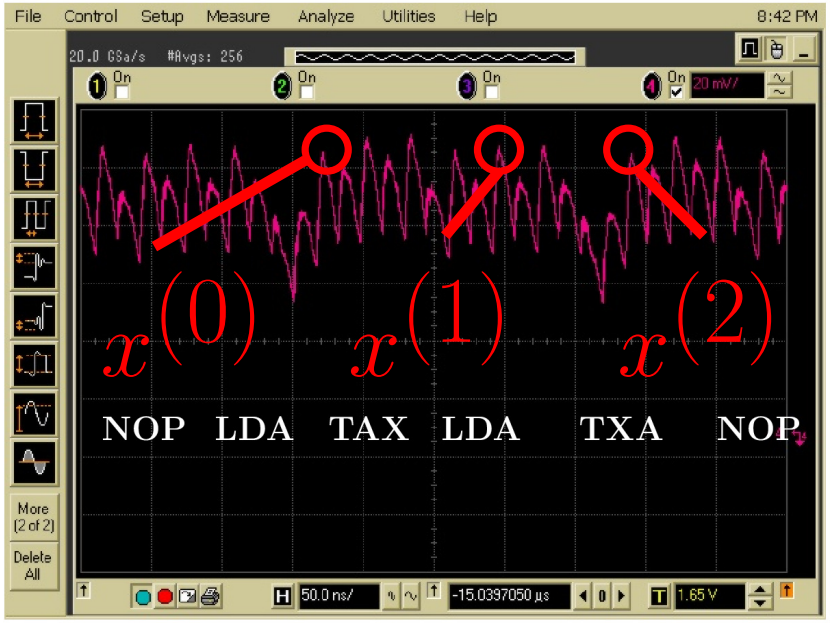


$\mathbf{t}$: t_1 t_2 t_3 ... t_q

$\mathbf{x}^{(\star)}$: $x_1^{(\star)}$ $x_2^{(\star)}$ $x_3^{(\star)}$ ... $x_q^{(\star)}$

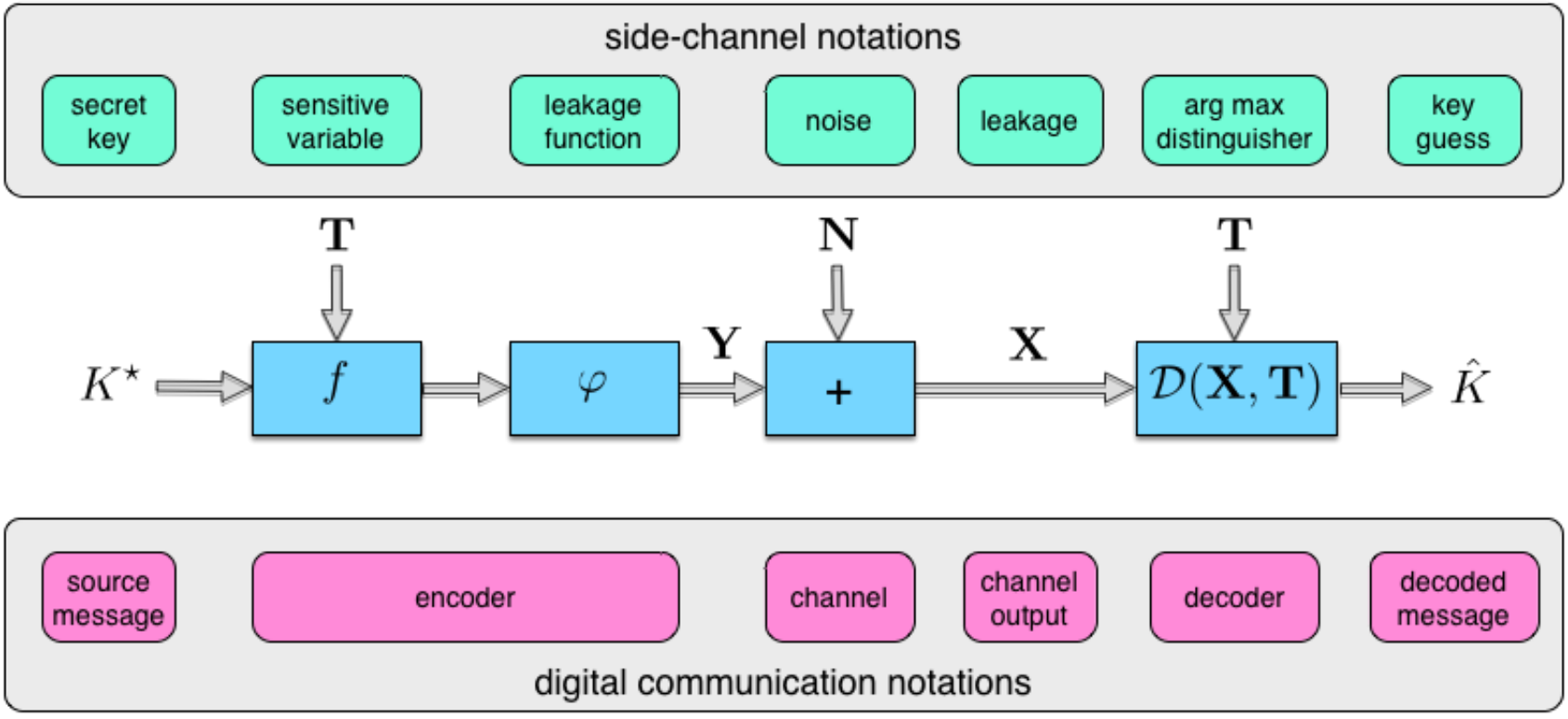
Side-Channel Analysis as a Digital Com. Problem [HRG14] ($d = 2$)



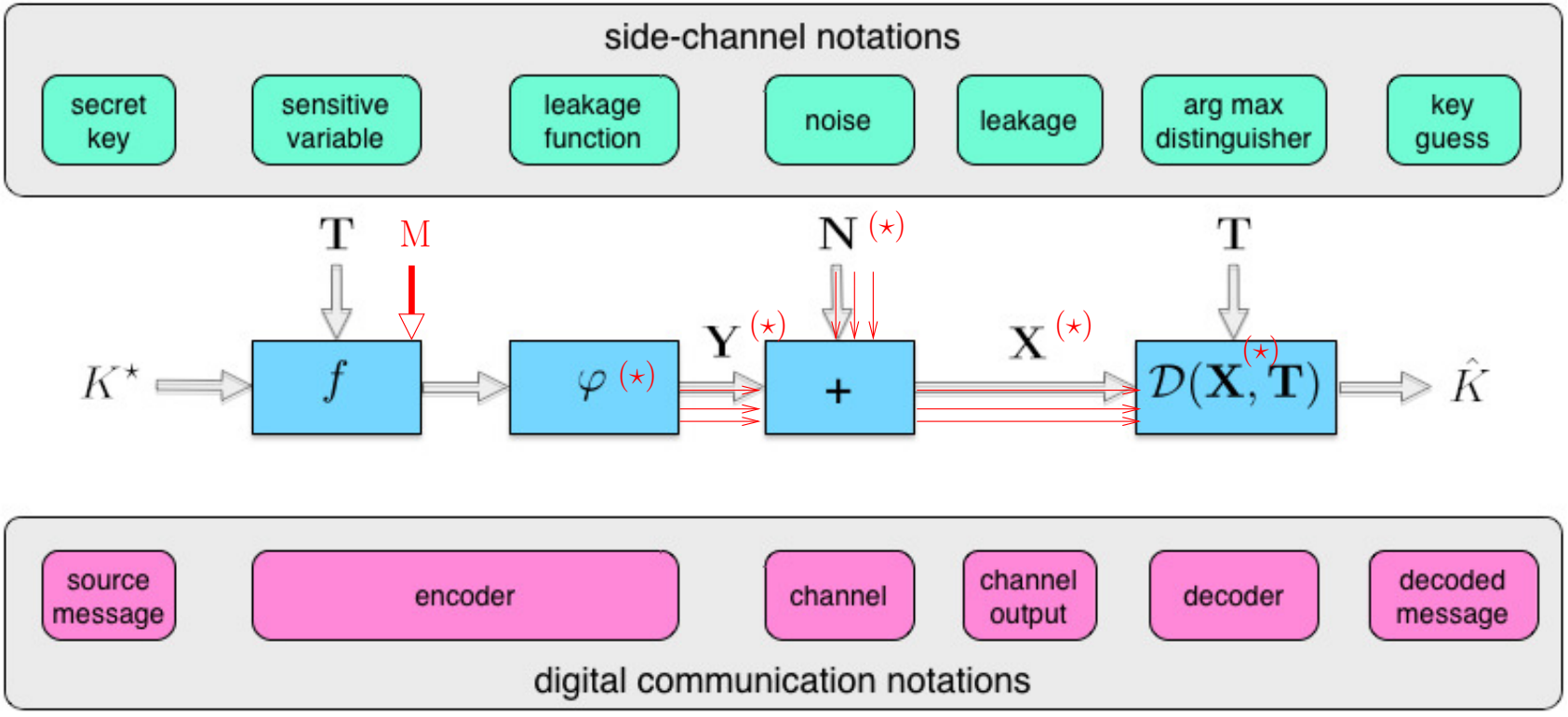


$\mathbf{t}$: $t_1\ t_2\ t_3\ \ldots\ t_q$ $\quad$ $\mathbf{x}^{(\star)}$: $x_1^{(\star)}\ x_2^{(\star)}\ x_3^{(\star)}\ \ldots\ x_q^{(\star)}$

Side-Channel Analysis as a Digital Com. Problem



Side-Channel Analysis as a Digital Com. Problem (with support of countermeasures)



Combination Functions for Higher-Order CPA

State-of-the-art

Leakage combination: $c_X : \mathcal{X}^{d+1} \to \mathbb{R}$
Model combination: $c_Y : \mathcal{T}^{d+1} \to \mathbb{R}$

Example

- Product combining ..... [CJRR99]
- Difference combining ..... [Mes00]
- Sinus combining ..... [OM07]
- Normalized product function ..... [PRB09]
  - maximizes absolute correlation

Motivation

Combination functions...

- are more inspired from an engineering perspective than a sound mathematical tool (necessary evil)
- go hand in hand with information loss (empirically [SVCO+10])
- **How to handle more than 2 leaks?**
- ↯ **product of two Gaussian noises is not Gaussian!**

Motivation

Combination functions...

- are more inspired from an engineering perspective than a sound mathematical tool (necessary evil)
- go hand in hand with information loss (empirically [SVCO+10])
- How to handle more than 2 leaks?
- ↯ product of two Gaussian noises is not Gaussian!

Motivation

Combination functions...

- are more inspired from an engineering perspective than a sound mathematical tool (necessary evil)
- go hand in hand with information loss (empirically [SVCO+10])
- **How to handle more than 2 leaks?**
- ↯ product of two Gaussian noises is not Gaussian!

Motivation

Combination functions...

- are more inspired from an engineering perspective than a sound mathematical tool (necessary evil)
- go hand in hand with information loss (empirically [SVCO+10])
- **How to handle more than 2 leaks?**
- ↯ **product of two Gaussian noises is not Gaussian!**

Outline

Optimal distinguisher (CHES '14 [HRG14])

Solution

$$\max \mathbb{P}_S = \mathbb{P}(\hat{K} = K^\star) \iff$$
$$\mathcal{D}(\mathbf{x}, \mathbf{t}) = argmax_k \; p(\mathbf{x}|\mathbf{t}, k)$$

Also known as *template attack*
(Chari, Rao, Rohatgi – CHES '02 [CRR02]).

Optimal distinguisher

Solution

$$\max \mathbb{P}_S = \mathbb{P}(\hat{K} = K^\star) \iff$$
$$\mathcal{D}(\mathbf{x}, \mathbf{t}) = argmax_k \sum_{\mathbf{m}} \mathbb{P}(\mathbf{m})\ p(\mathbf{x}|\mathbf{t}, k, \mathbf{m})$$

Also known as *template attack*
(Chari, Rao, Rohatgi – CHES '02 [CRR02]).

Optimal distinguisher for masking [This paper]

Solution for multi-dimensional leakage

$$\max \mathbb{P}_S = \mathbb{P}(\hat{K} = K^\star) \iff$$
$$\mathcal{D}(\mathbf{x}^{(\star)}, \mathbf{t}^{(\star)}) = argmax_k \; p(\mathbf{x}^{(\star)} | \mathbf{y}^{(\star)}(\mathbf{t}^{(\star)}, k)) \iff$$
$$\mathcal{D}(\mathbf{x}^{(\star)}, \mathbf{t}^{(\star)}) = argmax_k \sum_{\mathbf{m}^{(\star)}} \mathbb{P}(\mathbf{m}^{(\star)}) \; p(\mathbf{x}^{(\star)} | \mathbf{y}^{(\star)}(\mathbf{t}^{(\star)}, k, \mathbf{m}^{(\star)})) \; .$$

$\Longrightarrow$ in $p(\mathbf{x}^{(\star)} | \mathbf{t}^{(\star)}, k, \mathbf{m}^{(\star)})$, the only R.V. is the noise!

Optimal distinguisher for masking [This paper]

Solution for multi-dimensional leakage

$$\max \mathbb{P}_S = \mathbb{P}(\hat{K} = K^\star) \iff$$
$$\mathcal{D}(\mathbf{x}^{(\star)}, \mathbf{t}^{(\star)}) = argmax_k \; p(\mathbf{x}^{(\star)} | \mathbf{y}^{(\star)}(\mathbf{t}^{(\star)}, k)) \iff$$
$$\mathcal{D}(\mathbf{x}^{(\star)}, \mathbf{t}^{(\star)}) = argmax_k \sum_{\mathbf{m}^{(\star)}} \mathbb{P}(\mathbf{m}^{(\star)}) \; p(\mathbf{x}^{(\star)} | \mathbf{y}^{(\star)}(\mathbf{t}^{(\star)}, k, \mathbf{m}^{(\star)})) \; .$$

$\implies$ in $p(\mathbf{x}^{(\star)} | \mathbf{t}^{(\star)}, k, \mathbf{m}^{(\star)})$, the only R.V. is the noise!

Explicit Derivations

Theorem (Second-order HOOD)

If the model (i.e., $\varphi^{(\omega)}$) is known to the attacker for all ω, then the second-order HOOD is:

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

$$= \arg\max_{k \in \mathcal{K}} \prod_{i=1}^{q} \sum_{m^{(\star)} \in \mathcal{M}^{(\star)}} \mathbb{P}(m^{(\star)}) \prod_{\omega=0}^{1} p_k(x_i^{(\omega)} | t_i^{(\omega)}, m^{(\omega)}).$$

Explicit Derivations

Theorem (High-order HOOD)

If the model (i.e., $\varphi^{(\omega)}$) is known to the attacker for all ω, then the high-order HOOD is:

$$\mathcal{D}_{opt}^{d+1}(\mathbf{x}^{(\star)}, \mathbf{t}^{(\star)}) = \arg\max_{k \in \mathcal{K}} p_k(\mathbf{x}^{(\star)} | \mathbf{t}^{(\star)})$$

$$= \arg\max_{k \in \mathcal{K}} \prod_{i=1}^{q} \sum_{m^{(\star)} \in \mathcal{M}^{(\star)}} \mathbb{P}(m^{(\star)}) \prod_{\omega=0}^{d} p_k(x_i^{(\omega)} | t_i^{(\omega)}, m^{(\omega)}).$$

Explicit Derivations

Theorem (High-order HOOD — is *additive*)

If the model (i.e., $\varphi^{(\omega)}$) is known to the attacker for all ω, then the high-order HOOD is:

$$\mathcal{D}_{opt}^{d+1}(\mathbf{x}^{(\star)}, \mathbf{t}^{(\star)}) = \arg\max_{k \in \mathcal{K}} p_k(\mathbf{x}^{(\star)} | \mathbf{t}^{(\star)})$$

$$= \arg\max_{k \in \mathcal{K}} \sum_{i=1}^{q} \log \sum_{m^{(\star)} \in \mathcal{M}^{(\star)}} \mathbb{P}(m^{(\star)}) \prod_{\omega=0}^{d} p_k(x_i^{(\omega)} | t_i^{(\omega)}, m^{(\omega)}).$$

Gaussian Noise

Second-order HOOD for Gaussian noise

Assuming that $N^{(\omega)} \sim \mathcal{N}(O, {\sigma^{(\omega)}}^2)$ then the second-order optimal distinguisher becomes

$$\mathcal{D}_{opt}^{2,G}(\mathbf{x}^{(0)}, \mathbf{x}^{(1)}, \mathbf{t}) =$$

$$\arg\max_{k \in \mathcal{K}} \prod_{i=1}^{q} \sum_{m \in \mathcal{M}} \exp\left\{ -\frac{1}{2} \left(\frac{-2x_i^{(0)} y^{(0)}(t_i, k, m) + y^{(0)}(t_i, k, m)^2}{{\sigma^{(0)}}^2} \right.\right.$$

$$\left.\left. + \frac{-2x_i^{(1)} y^{(1)}(t_i, k, m) + y^{(1)}(t_i, k, m)^2}{{\sigma^{(1)}}^2} \right) \right\}.$$

High Gaussian Noise

Second-order HOOD for *high* Gaussian noise

$$\mathcal{D}_{opt}^{2,G,\sigma\uparrow}(\mathbf{x}^{(\star)},\mathbf{t}) =$$

$$\arg\max_{k\in\mathcal{K}} \prod_{i=1}^{q} \sum_{m\in\mathcal{M}} \exp\left\{ \frac{x_i^{(0)} y^{(0)}(t_i,k,m)}{\sigma^{(0)^2}} + \frac{x_i^{(1)} y^{(1)}(t_i,k,m)}{\sigma^{(1)^2}} \right\}.$$

High Gaussian Noise

Second-order HOOD for *high* Gaussian noise

$$\mathcal{D}_{opt}^{2,G,\sigma\uparrow}(\mathbf{x}^{(\star)},\mathbf{t}) =$$

$$\arg\max_{k\in\mathcal{K}} \sum_{i=1}^{q} \log \sum_{m\in\mathcal{M}} \exp\left\{ \frac{x_i^{(0)} y^{(0)}(t_i,k,m)}{\sigma^{(0)^2}} + \frac{x_i^{(1)} y^{(1)}(t_i,k,m)}{\sigma^{(1)^2}} \right\}.$$

Comparison with Second-Order CPA

Second-order HOOD for high noise $\iff$ second-order CPA

The second-order HOOD for high noise can be approximated as

$$\mathcal{D}_{opt}^{2,G,\sigma\uparrow} \approx \arg\max_k \left\langle \mathbf{x}^{(0)} \cdot \mathbf{x}^{(1)} \middle| \sum_{m\in\mathcal{M}} y^{(0)}(\mathbf{t},k,m) \cdot y^{(1)}(\mathbf{t},k,m) \right\rangle,$$

Accordingly,...

- ... the normalized product function is *optimal*
- ... *direct* scale $\implies$ *proportional* scale
  (but the "sign" has to be known)

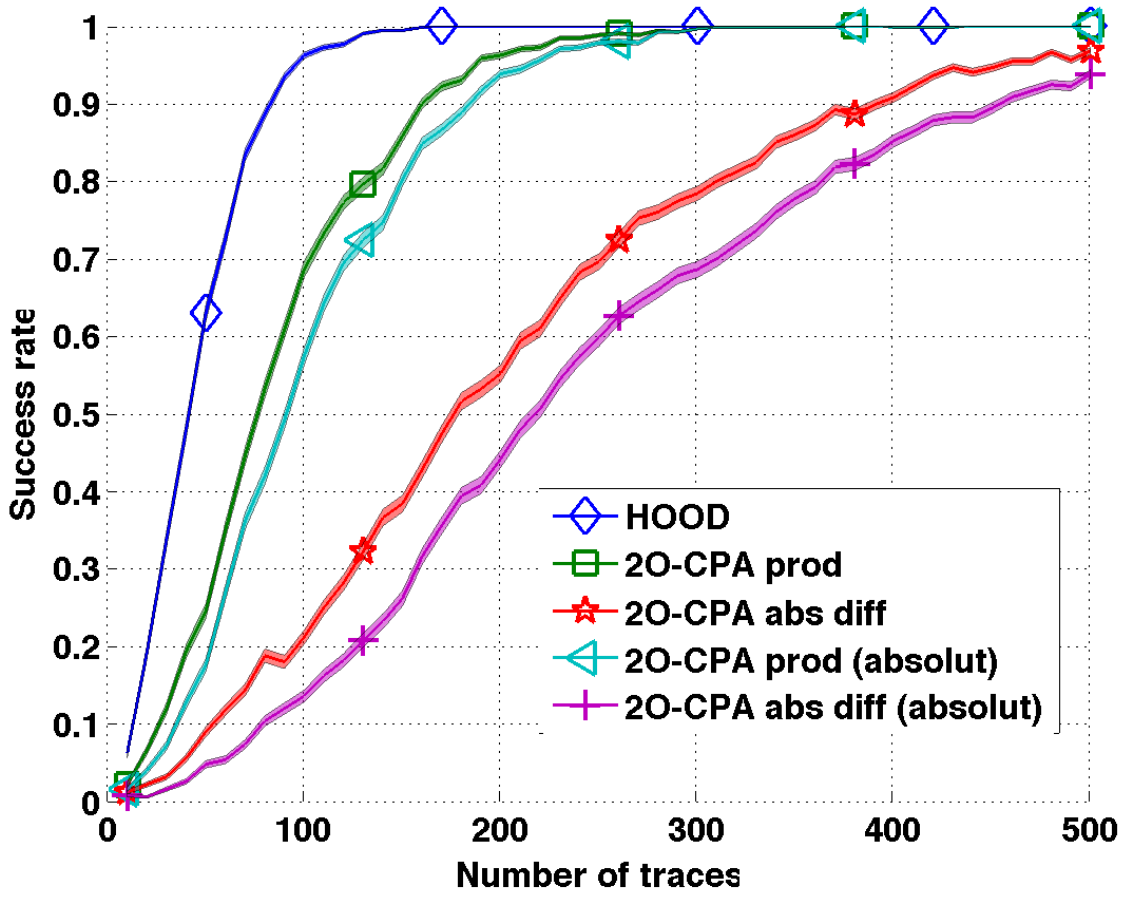
Low Gaussian Noise

Second-order HOOD for *low* Gaussian noise

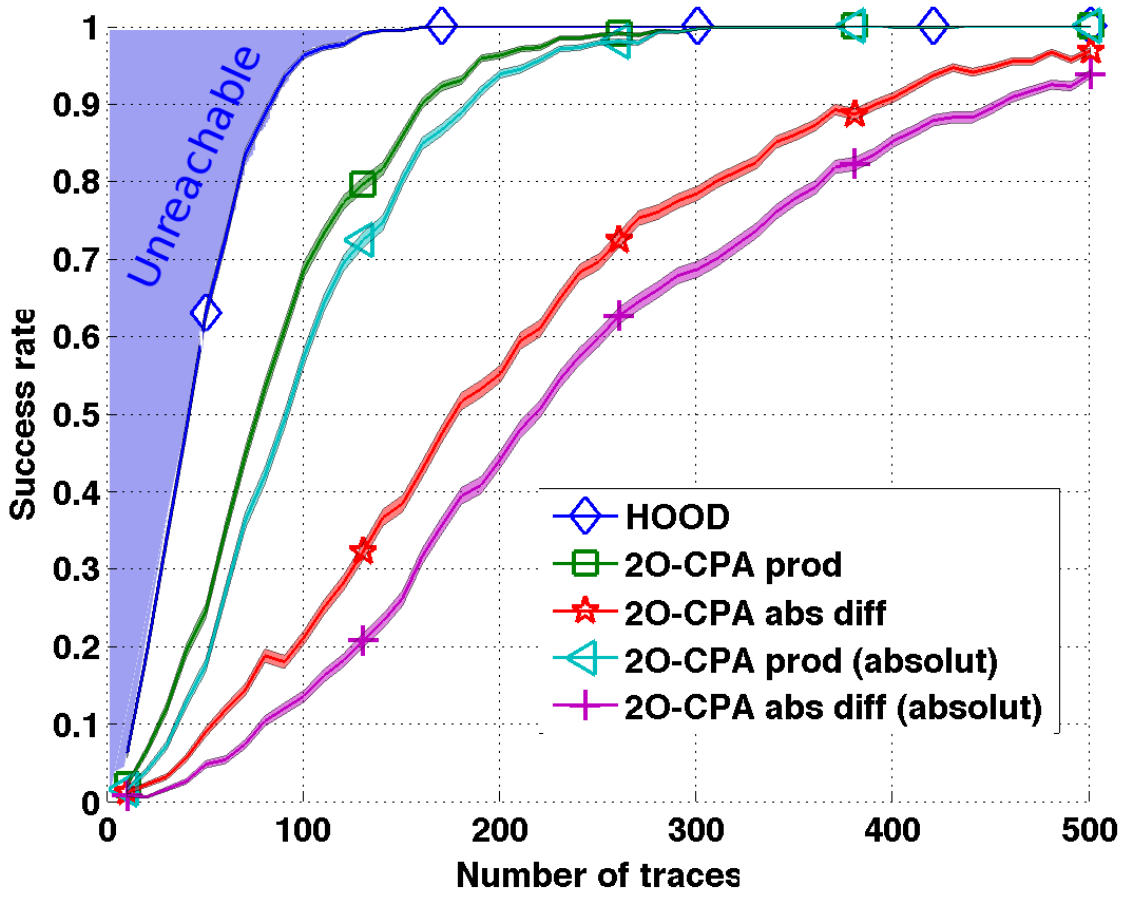
$$\mathcal{D}_{opt}^{2,G,\sigma\downarrow}(\mathbf{x}^{(\star)}, \mathbf{t}) =$$

$$\arg\min_{k\in\mathcal{K}} \sum_{i=1}^{q} \max_{m\in\mathbb{F}_2^n} (x_i^{(0)} - y^{(0)}(t_i, k, m))^2 + (x_i^{(1)} - y^{(1)}(t_i, k, m))^2.$$

Experiments



Experiments



Outline

Algorithm (used in smartcards)

input : t, one byte of plaintext, and k, one byte of key
output: The application of AddRoundKey and SubBytes on t, i.e., $S(t \oplus k)$

1 $m \leftarrow_{\mathcal{R}} \mathbb{F}_2^n$, $m' \leftarrow_{\mathcal{R}} \mathbb{F}_2^n$ // Draw of random input and output masks ;
2 **for** $\omega \in \{0, 1, \ldots, 2^n - 1\}$ **do** // Sbox masking
3 $\quad z \leftarrow \omega \oplus m$ // Masked input ;
4 $\quad z' \leftarrow S[\omega] \oplus m'$ // Masked output ;
5 $\quad S'[z] \leftarrow z'$ // Creating the masked Sbox entry ;
6 **end**
7 $t \leftarrow t \oplus m$ // Plaintext masking ;
8 $t \leftarrow t \oplus k$ // Masked AddRoundKey ;
9 $t \leftarrow S'[t]$ // Masked SubBytes ;
10 $t \leftarrow t \oplus m'$ // Demasking ;
11 **return** t

Algorithm (used in smartcards)

input : t, one byte of plaintext, and k, one byte of key
output: The application of AddRoundKey and SubBytes on t, i.e., $S(t \oplus k)$

1 $m \leftarrow_{\mathcal{R}} \mathbb{F}_2^n$, $m' \leftarrow_{\mathcal{R}} \mathbb{F}_2^n$ // Draw of random input and output masks ;
2 **for** $\omega \in \{0, 1, \ldots, 2^n - 1\}$ **do** // Sbox masking
3 $\quad z \leftarrow \omega \oplus m$ // Masked input ;
4 $\quad z' \leftarrow S[\omega] \oplus m'$ // Masked output ;
5 $\quad S'[z] \leftarrow z'$ // Creating the masked Sbox entry ;
6 **end**
7 $t \leftarrow t \oplus m$ // Plaintext masking ;
8 $t \leftarrow t \oplus k$ // Masked AddRoundKey ;
9 $t \leftarrow S'[t]$ // Masked SubBytes ;
10 $t \leftarrow t \oplus m'$ // Demasking ;
11 **return** t

Usual 2-variate 2nd-order attack

Algorithm (used in smartcards)

input : t, one byte of plaintext, and k, one byte of key
output: The application of AddRoundKey and SubBytes on t, i.e., $S(t \oplus k)$

1 $m \leftarrow_{\mathcal{R}} \mathbb{F}_2^n$, $m' \leftarrow_{\mathcal{R}} \mathbb{F}_2^n$ // Draw of random input and output masks ;
2 **for** $\omega \in \{0, 1, \ldots, 2^n - 1\}$ **do** // Sbox masking
3 $z \leftarrow \omega \oplus m$ // Masked input ;
4 $z' \leftarrow S[\omega] \oplus m'$ // Masked output ;
5 $S'[z] \leftarrow z'$ // Creating the masked Sbox entry ;
6 **end**
7 $t \leftarrow t \oplus m$ // Plaintext masking ;
8 $t \leftarrow t \oplus k$ // Masked AddRoundKey ;
9 $t \leftarrow S'[t]$ // Masked SubBytes ;
10 $t \leftarrow t \oplus m'$ // Demasking ;
11 **return** t

Improved $(2^n + 1)$-variate 2nd-order attack

Algorithm (used in smartcards)

input : t, one byte of plaintext, and k, one byte of key
output: The application of AddRoundKey and SubBytes on t, i.e., $S(t \oplus k)$

1 $m \leftarrow_{\mathcal{R}} \mathbb{F}_2^n$, $m' \leftarrow_{\mathcal{R}} \mathbb{F}_2^n$ // Draw of random input and output masks ;
2 **for** $\omega \in \{0, 1, \ldots, 2^n - 1\}$ **do** // Sbox masking
3 $z \leftarrow \omega \oplus m$ // Masked input ;
4 $z' \leftarrow S[\omega] \oplus m'$ // Masked output ;
5 $S'[z] \leftarrow z'$ // Creating the masked Sbox entry ;
6 **end**
7 $t \leftarrow t \oplus m$ // Plaintext masking ;
8 $t \leftarrow t \oplus k$ // Masked AddRoundKey ;
9 $t \leftarrow S'[t]$ // Masked SubBytes ;
10 $t \leftarrow t \oplus m'$ // Demasking ;
11 **return** t

Improved 257-variate 2nd-order attack

Previous Attacks

- Second-order attacks, collision attacks, ...
- However: a more powerful attack would consist in using all the leakages from the Sbox recomputation

2-stage CPA attack [TWO13]

$$2{\times}\mathrm{CPA}^{mt}(\mathbf{x},\mathbf{t}) = \arg\max_{k\in\mathcal{K}} \rho(\mathbf{x}^{(2^n)}, y^{(2^n)}(\mathbf{t},k,\hat{\mathbf{m}})),$$

where $\forall i\ \hat{m}_i$ is the mask that maximizes the correlation between $x_i^{(\omega)}$ and $y_i^{(\omega)} = \omega \oplus m_i$ for $\omega \in \{0,\ldots,2^n-1\}$.

Previous Attacks

- Second-order attacks, collision attacks, ...
- However: a more powerful attack would consist in using all the leakages from the Sbox recomputation

2-stage CPA attack [TWO13]

$$2{\times}\mathrm{CPA}^{mt}(\mathbf{x},\mathbf{t}) = \arg\max_{k\in\mathcal{K}} \rho(\mathbf{x}^{(2^n)}, y^{(2^n)}(\mathbf{t},k,\hat{\mathbf{m}})),$$

where $\forall i\ \hat{m}_i$ is the mask that maximizes the correlation between $x_i^{(\omega)}$ and $y_i^{(\omega)} = \omega \oplus m_i$ for $\omega \in \{0,\ldots,2^n-1\}$.

Optimal?

Previous Attacks

- Second-order attacks, collision attacks, ...
- However: a more powerful attack would consist in using all the leakages from the Sbox recomputation

2-stage CPA attack [TWO13]

$$2{\times}\mathrm{CPA}^{mt}(\mathbf{x},\mathbf{t}) = \arg\max_{k\in\mathcal{K}} \rho(\mathbf{x}^{(2^n)}, y^{(2^n)}(\mathbf{t},k,\hat{\mathbf{m}})),$$

where $\forall i\ \hat{m}_i$ is the mask that maximizes the correlation between $x_i^{(\omega)}$ and $y_i^{(\omega)} = \omega \oplus m_i$ for $\omega \in \{0,\ldots,2^n-1\}$.

Optimal? **No!**

Higher-order optimal distinguisher

Theorem (HOOD for masking tables)

$$\mathcal{D}_{opt}^{mt,G}(\mathbf{x}^{(\star)},\mathbf{t}) =$$

$$\arg\max_{k\in\mathcal{K}} \sum_{i=1}^{q} \log\left\{\sum_{m\in\mathbb{F}_2^n} \exp\left\{\sum_{\omega\in\mathbb{F}_2^n} \frac{1}{\sigma^{(\omega)^2}}\left(x_i^{(\omega)}\varphi(\omega\oplus m)-\frac{1}{2}\varphi^2(\omega\oplus m)\right)\right.\right.$$

$$\left.\left.+\frac{1}{\sigma^{(2^n)^2}}\left(x_i^{(2^n)}\varphi(t_i\oplus m\oplus k)-\frac{1}{2}\varphi^2(t_i\oplus m\oplus k)\right)\right\}\right\}.$$

Higher-order optimal distinguisher

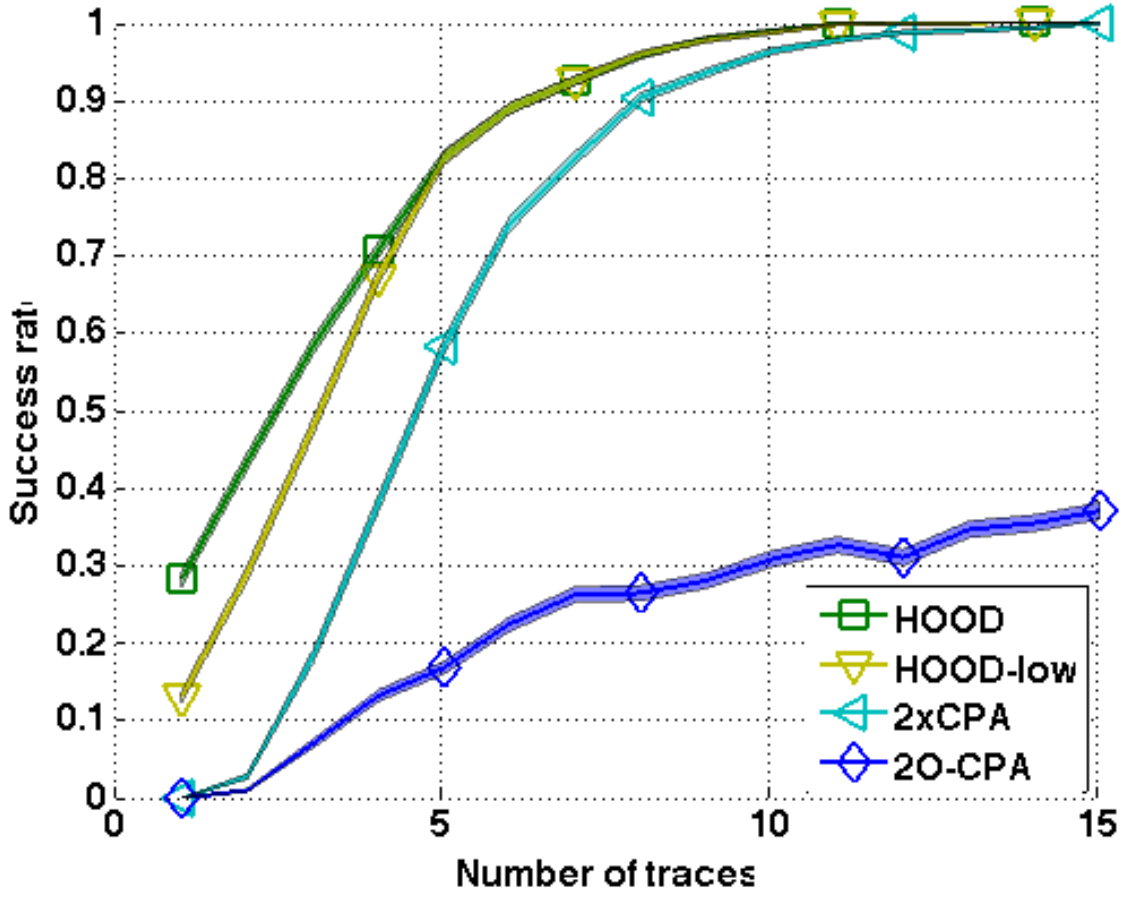
HOOD for masking tables for low SNR

For large Gaussian noise (or low SNR) the distinguisher becomes

$$\mathcal{D}_{opt}^{mt,G,\sigma\uparrow}(\mathbf{x}^{(\star)},\mathbf{t}) =$$

$$\arg\max_{k\in\mathcal{K}} \sum_{\omega\in\mathbb{F}_2^n} \frac{1}{\sigma^{(\omega)^2}} \sum_{i=1}^{q} \left(\begin{array}{l} x_i^{(\omega)} x_i^{(2^n)} \sum_m \varphi(\omega\oplus m)\varphi(t_i\oplus k\oplus m) \\ -\frac{1}{2} x_i^{(2^n)} \sum_m \varphi(t_i\oplus k\oplus m)\varphi(\omega\oplus m)^2 \\ -\frac{1}{2} x_i^{(\omega)} \sum_m \varphi(\omega\oplus m)\varphi(t_i\oplus k\oplus m)^2 \\ +\frac{1}{4} \sum_m \varphi(\omega\oplus m)^2\varphi(t_i\oplus k\oplus m)^2 \end{array} \right).$$

Experimental Validation



Outline

Conclusion

- Methodology of attack...
- ... that works even with protections!
- To our surprize, optimal distinguishers **confirm** the state-of-the-art:
  - **only** under specific conditions (e.g., high Gaussian noise)
- **but not in general**:
  - Especially if SNR is high
  - Or if the leakage is highly multi-variate

Conclusion

- Methodology of attack...
- ... that works even with protections!
- To our surprize, optimal distinguishers **confirm** the state-of-the-art:
  - **only** under specific conditions (e.g., high Gaussian noise)
- **but not in general**:
  - Especially if SNR is high
  - Or if the leakage is highly multi-variate

Conclusion

- Methodology of attack...
- ... that works even with protections!
- To our surprize, optimal distinguishers **confirm** the state-of-the-art:
  - **only** under specific conditions (e.g., high Gaussian noise)
- **but not in general**:
  - Especially if SNR is high
  - Or if the leakage is highly multi-variate

Conclusion

- Methodology of attack...
- ... that works even with protections!
- To our surprize, optimal distinguishers **confirm** the state-of-the-art:
  - **only** under specific conditions (e.g., high Gaussian noise)
- **but not in general**:
  - Especially if SNR is high
  - Or if the leakage is highly multi-variate

Perspectives

- Integrate the *uncertainty* about the *model* (epistemic noise)
- Give us an implementation ...
- ... and we derive the optimal distinguisher!

Questions?

Perspectives

- Integrate the *uncertainty* about the *model* (epistemic noise)
- Give us an implementation ...
- ... and we derive the optimal distinguisher!

Questions?

Perspectives

- Integrate the *uncertainty* about the *model* (epistemic noise)
- Give us an implementation ...
- ... and we derive the optimal distinguisher!

Questions?

Perspectives

- Integrate the *uncertainty* about the *model* (epistemic noise)
- Give us an implementation ...
- ... and we derive the optimal distinguisher!

Questions?

Institut
Mines-Telecom

Masks will Fall Off

Higher-Order Optimal Distinguishers

Nicolas Bruneau, Sylvain Guilley,
Annelie Heuser, Olivier Rioul

ASIACRYPT 2014, Kaohsiung, Taiwan

[CJRR99] Suresh Chari, Charanjit S. Jutla, Josyula R. Rao, and Pankaj Rohatgi.
Towards Sound Approaches to Counteract Power-Analysis Attacks.

In *CRYPTO*, volume 1666 of *LNCS*. Springer, August 15-19 1999.
Santa Barbara, CA, USA. ISBN: 3-540-66347-9.

[CRR02] 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.

[HRG14] Annelie Heuser, Olivier Rioul, and Sylvain Guilley.
Good Is Not Good Enough - Deriving Optimal Distinguishers from Communication Theory.
In Lejla Batina and Matthew Robshaw, editors, *Cryptographic Hardware and Embedded Systems - CHES 2014 - 16th International Workshop, Busan, South Korea, September 23-26, 2014. Proceedings*, volume 8731 of *Lecture Notes in Computer Science*, pages 55–74. Springer, 2014.

[KJJ99] Paul C. Kocher, Joshua Jaffe, and Benjamin Jun.
Differential Power Analysis.
In *Proceedings of CRYPTO'99*, volume 1666 of *LNCS*, pages 388–397. Springer-Verlag, 1999.

[Mes00] Thomas S. Messerges.
Using Second-Order Power Analysis to Attack DPA Resistant Software.
In *CHES*, volume 1965 of *LNCS*, pages 238–251. Springer-Verlag, August 17-18 2000.
Worcester, MA, USA.

[MOS09] Stefan Mangard, Elisabeth Oswald, and François-Xavier Standaert.
One for All - All for One: Unifying Standard DPA Attacks.
Cryptology ePrint Archive, Report 2009/449, 2009.

[OM07] Elisabeth Oswald and Stefan Mangard.
Template Attacks on Masking — Resistance Is Futile.
In Masayuki Abe, editor, *CT-RSA*, volume 4377 of *Lecture Notes in Computer Science*, pages 243–256. Springer, 2007.

[PRB09] Emmanuel Prouff, Matthieu Rivain, and Régis Bevan.
Statistical Analysis of Second Order Differential Power Analysis.
IEEE Trans. Computers, 58(6):799–811, 2009.

[SVCO+10] François-Xavier Standaert, Nicolas Veyrat-Charvillon, Elisabeth Oswald, Benedikt Gierlichs, Marcel Medwed, Markus Kasper, and Stefan Mangard.
The World is Not Enough: Another Look on Second-Order DPA.
In *ASIACRYPT*, volume 6477 of *LNCS*, pages 112–129. Springer, December 5-9 2010.
Singapore.
http://www.dice.ucl.ac.be/~fstandae/PUBLIS/88.pdf.

[TWO13] Michael Tunstall, Carolyn Whitnall, and Elisabeth Oswald.
Masking Tables - An Underestimated Security Risk.
In Shiho Moriai, editor, *FSE*, volume 8424 of *Lecture Notes in Computer Science*, pages 425–444. Springer, 2013.

HOOD $\Longleftrightarrow$ Higher-order CPA

Relationship between HOOD and CPA for masking tables

When all noise variances are equal, i.e., $\sigma = \sigma^{(\omega)}\ \forall \omega$ we can further simplifies to

$$\mathcal{D}_{opt}^{mt,G,\sigma\uparrow}(\mathbf{x}^{(\star)},\mathbf{t}) = \arg\max_{k\in\mathcal{K}} \sum_{\omega\in\mathbb{F}_2^n}\sum_{i=1}^{q}\Bigg(x_i^{(\omega)}x_i^{(2^n)}\sum_{m\in\mathbb{F}_2^n}\varphi(\omega\oplus m)\varphi(t_i\oplus k\oplus m) - \frac{1}{2}x_i^{(\omega)}\sum_{m\in\mathbb{F}_2^n}\varphi(\omega\oplus m)\varphi^2(t_i\oplus k\oplus m)\Bigg),$$

which becomes close to a combination of higher-order CPAs, i.e.,

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