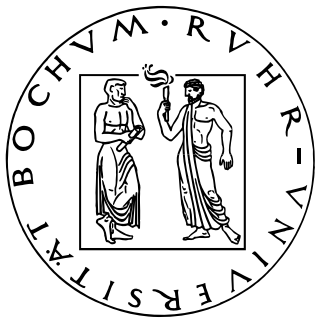
Gaussian Mixture Models for Higher-Order Side Channel Analysis

Kerstin Lemke-Rust and Christof Paar

Ruhr University Bochum, Germany

September 11, 2007



Outline

Presentation Outline

- Introduction
- Our Model
  - Masked Implementation
  - Side Channel Adversary
- Gaussian Mixture Models
  - Expectation-Maximization (EM) Algorithm
- Experimental Case Study
- Further Directions
- Conclusion

First Order Side Channel Analysis and Masking

Side Channel Leakage

Measurable observables (power consumption, EM emanation) depend on (key-dependent) internal states of a cryptographic implementation.

First Order Side Channel Analysis

First-order side channel analysis applies statistical tests using

- measurement data and
- one key-dependent internal state of the cryptographic implementation.

Masking

Masking hides key-dependent internal states by adding random numbers. First-order side channel analysis can be prevented.

Higher-Order Side Channel Analysis and Masking

Higher-Order Side Channel Analysis

Higher-order side channel analysis applies statistical tests using

- measurement data and
- multiple internal states of a cryptographic implementation.

Higher-Order Side Channel Analysis and Masking

Higher-order side channel analysis is essential if

- the cryptographic implementation applies an effective masking scheme.

Example

Second-order side channel analysis considers two internal states (e.g., the mask and the masked key-dependent internal state) and can defeat first-order masking.

This Contribution

Previous Contributions

- Univariate statistics (First-Order DPA, Second-Order DPA,...)
- Multivariate statistics (Templates, Stochastic Methods,...) with complete knowledge of the adversary at profiling.

This Contribution

- Multivariate statistics with incomplete knowledge of the adversary at profiling, i.e., the adversary does not know random numbers used for masking.
- Use of Gaussian Mixture Models for estimating multivariate probability density functions (p.d.f.s) for each key dependency at profiling.

Our Model: Masked Implementation

Concrete Settings

- First order boolean masking scheme: data $x \in \{0,1\}^d$, key $k \in \{0,1\}^d$, and mask $y \in \{0,1\}^d$
- Two internal states for side channel analysis:
  - Differential analysis: y and $y \oplus k \oplus x$
- m-dimensional side channel observable $\vec{l}(x,k,y) = (l_1, \ldots, l_m)^T$ with $\vec{i}(x,k,y) = (i_1, \ldots, i_m)^T$ representing one measurement vector.
- For each pair (x,k,y) $\vec{z} := \vec{i}(x,k,y) \in \mathbb{R}^m$ is distributed according to an m-variate Gaussian density

$$\mathcal{N}(\vec{z}, \vec{\mu}, \mathbf{\Sigma}) = \frac{1}{\sqrt{(2\pi)^m |\mathbf{\Sigma}|}} \exp\left[-\frac{1}{2}(\vec{z} - \vec{\mu})^T \mathbf{\Sigma}^{-1} (\vec{z} - \vec{\mu})\right]$$

Our Model: Side Channel Adversary $\mathcal{A}$

Profiling Phase

Training Device

Input:

- N vectorial measurement samples $\vec{i}(x, k, y)$
- Known data $x \in \{0,1\}^d$
- Known key $k \in \{0,1\}^d$

Output:

- Multivariate p.d.f. $f^{(x,k)}$ of the side channel leakage for each pair of (x, k)

Key Recovery Phase

Target Device

Input:

- N° vectorial measurement samples $\vec{i}(x, k^\circ, y)$
- Known data $x \in \{0,1\}^d$
- Multivariate p.d.f. $f^{(x,k)}$ of the side channel leakage for each pair of (x, k)

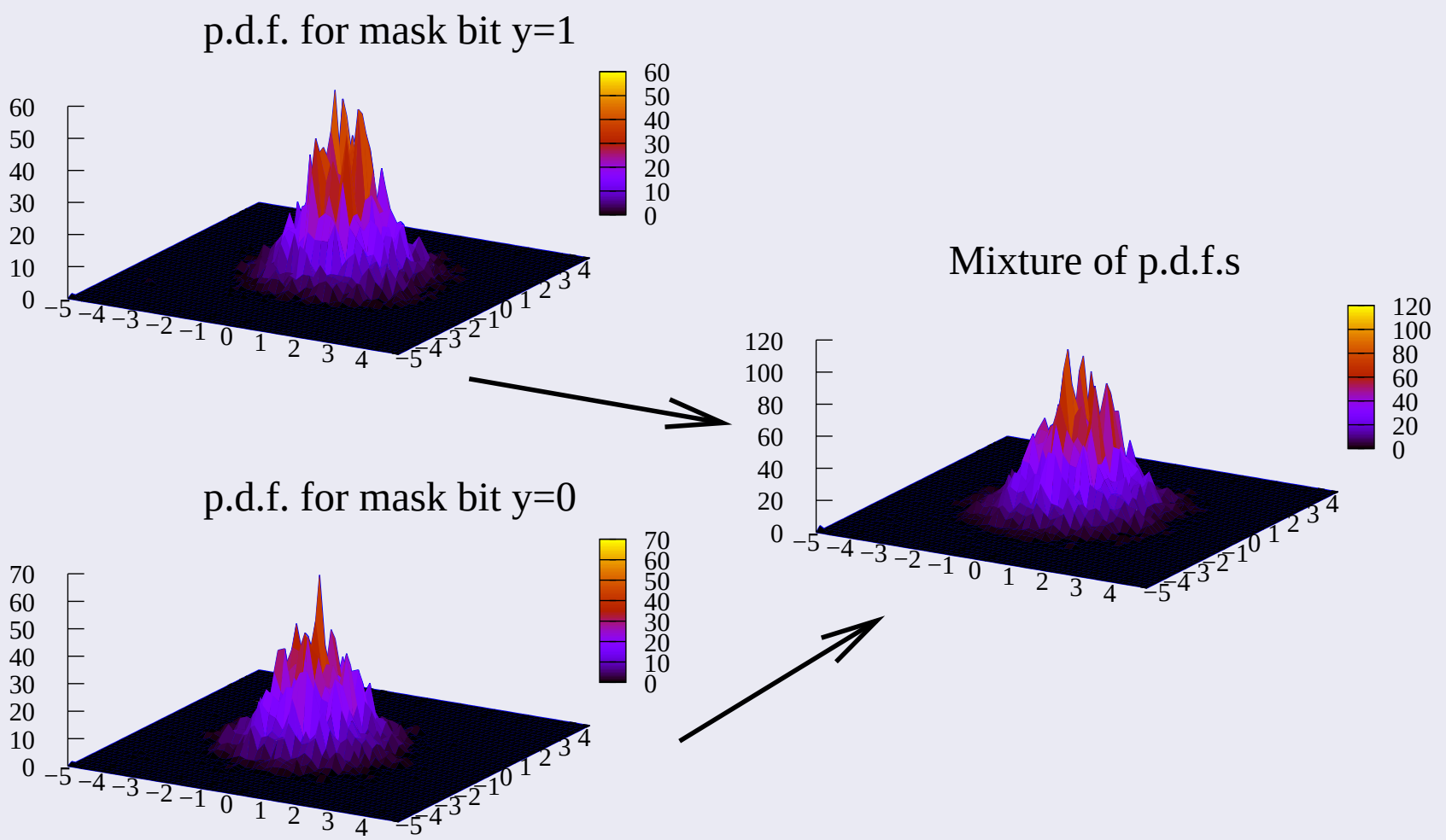
Output:

- Key guess $k^* \in \{0,1\}^d$

Adversary Success

Adversary $\mathcal{A}$ is successful if $k^* = k^\circ$.

Experimental Mixture of p.d.f.s



Gaussian Mixture Models for Profiling

Mixture p.d.f.s and Component p.d.f.s

For each (x, k) $\mathcal{A}$ observes a *mixture p.d.f.*

$$f(\vec{z}, \theta^{(x,k)}) = \sum_{j=0}^{2^d-1} \alpha_j^{(x,k)} \mathcal{N}(\vec{z}, \vec{\mu}_j^{(x,k)}, \mathbf{\Sigma}_j^{(x,k)}) \tag{1}$$

that consists of 2^d m-variate Gaussian *component p.d.f.s* $\mathcal{N}(\vec{z}, \vec{\mu}_j^{(x,k)}, \mathbf{\Sigma}_j^{(x,k)})$ for each mask j.

The $\alpha_j^{(x,k)}$ satisfy

$$\alpha_j^{(x,k)} \geq 0, j = 0, \ldots, 2^d - 1 \text{ and } \sum_{j=0}^{2^d-1} \alpha_j^{(x,k)} = 1.$$

Problem Statement

Problem statement for $\mathcal{A}$ at profiling

Given a mixture $f(\vec{z}, \theta^{(x,k)})$ in (1) estimate the parameters

$$\theta^{(x,k)} = \left(\alpha_0^{(x,k)}, \vec{\mu}_0^{(x,k)}, \mathbf{\Sigma}_0^{(x,k)}, \ldots, \alpha_{2^d-1}^{(x,k)}, \vec{\mu}_{2^d-1}^{(x,k)}, \mathbf{\Sigma}_{2^d-1}^{(x,k)}\right) .$$

Side information:

- The number of component p.d.f.s is known to be 2^d.
- The component p.d.f.s are uniformly distributed in an effective masking scheme:

$$\alpha_j^{(x,k)} \approx 2^{-d}$$

- For key recovery, the labels of the component p.d.f.s (i.e., the masks) are not needed to be identified.

Variants for Parameter Estimation

Variants for Parameter Estimation

Table: Number of free parameters in the Gaussian mixture model.

Variant	$\alpha_j^{(x,k)}$	$\vec{\mu}_j^{(x,k)}$	$\boldsymbol{\Sigma}_j^{(x,k)}$ or $\boldsymbol{\Sigma}^{(x,k)}$	Total
1	$\times$	$2^d m$	$\times$	$2^d m$
2	$2^d - 1$	$2^d m$	$\times$	$2^d(1+m) - 1$
3	$2^d - 1$	$2^d m$	$(m^2+m)/2$	$2^d(1+m) + (m+m^2)/2 - 1$
4	$2^d - 1$	$2^d m$	$2^d(m^2+m)/2$	$2^d(1+3m/2+m^2/2) - 1$

Example

If $d = 1$ and $m = 2$ the number of free parameters is 4 for Variant 1, 5 for Variant 2, 8 for Variant 3, and 11 for Variant 4.

Expectation-Maximization (EM) Algorithm in Profiling Phase

Expectation-Maximization (EM) Algorithm

- Maximizes the likelihood function

$$f(\vec{z}_1, \theta^{(x,k)}) \cdot f(\vec{z}_2, \theta^{(x,k)}) \cdot \ldots \cdot f(\vec{z}_{N^{(x,k)}}, \theta^{(x,k)}). \quad (2)$$

- Number of samples for each (x, k): $N^{(x,k)} \approx \frac{N}{2^{2d}}$
- Iterative algorithm that requires initial values for the set of parameters $\alpha_j^{(x,k)}$, $\vec{\mu}_j^{(x,k)}$ and $\mathbf{\Sigma}_j^{(x,k)}$.
- After each iteration (shown on next slide), compute (2) and check for convergence.
- Repeat the EM Algorithm with other (e.g., randomized) initial values.

Expectation-Maximization (EM) Algorithm for Variant 4

Expectation Step (E-Step):

$$\alpha_{jn} := \frac{\hat{\alpha}_j^{(x,k)} \, \mathcal{N}(\vec{z}_n, \hat{\vec{\mu}}_j^{(x,k)}, \hat{\mathbf{\Sigma}}_j^{(x,k)})}{\sum_{i=0}^{2^d-1} \hat{\alpha}_i^{(x,k)} \, \mathcal{N}(\vec{z}_n, \hat{\vec{\mu}}_i^{(x,k)}, \hat{\mathbf{\Sigma}}_i^{(x,k)})}$$

Maximization Step (M-Step):

$$\hat{\alpha}_j^{(x,k)} = \frac{1}{N^{(x,k)}} \sum_{n=1}^{N^{(x,k)}} \alpha_{jn} \quad , \quad \hat{\vec{\mu}}_j^{(x,k)} = \frac{1}{\sum_{n=1}^{N^{(x,k)}} \alpha_{jn}} \sum_{n=1}^{N^{(x,k)}} \alpha_{jn} \vec{z}_n$$

$$\hat{\mathbf{\Sigma}}_j^{(x,k)} = \frac{1}{\sum_{n=1}^{N^{(x,k)}} \alpha_{jn}} \sum_{n=1}^{N^{(x,k)}} \alpha_{jn} \left(\vec{z}_n - \hat{\vec{\mu}}_j^{(x,k)}\right) \left(\vec{z}_n - \hat{\vec{\mu}}_j^{(x,k)}\right)^T$$

See paper for the details of the 3 variants.

Key Recovery Phase

Decision Strategy

Adversary $\mathcal{A}$ computes

$$\mathcal{L}_k := \sum_{i=1}^{N^\circ} \ln f(\vec{z}_i | k, x_i) = \sum_{i=1}^{N^\circ} \ln \left(\sum_{j=0}^{2^d-1} \alpha_j^{(x_i,k)} \mathcal{N}(\vec{z}_i, \vec{\mu}_j^{(x_i,k)}, \mathbf{\Sigma}_j^{(x_i,k)}) \right)$$

for each of the 2^d key hypotheses k using known $x_i \in \{0,1\}^d$ and decides in favour of that key hypothesis k^* that leads to the maximum likelihood:

$$k^* := \arg\max_k \mathcal{L}_k \ .$$

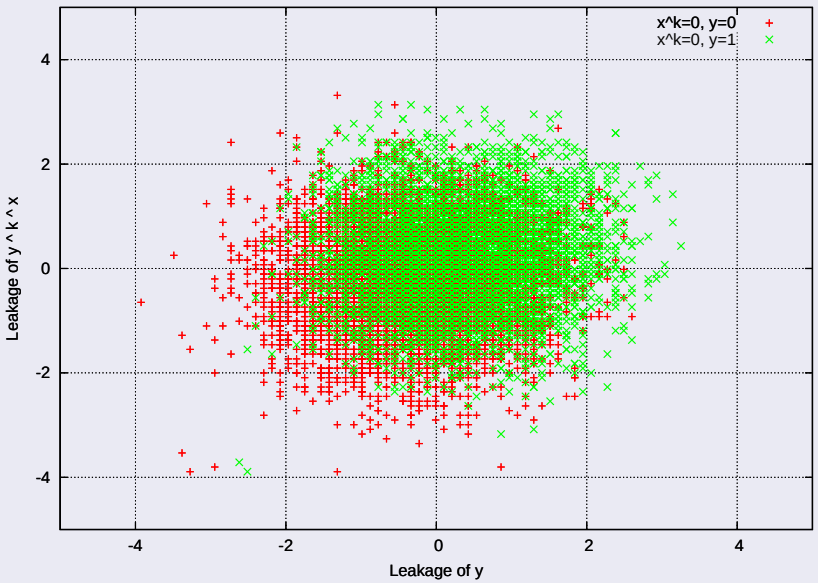
Experimental Case Study

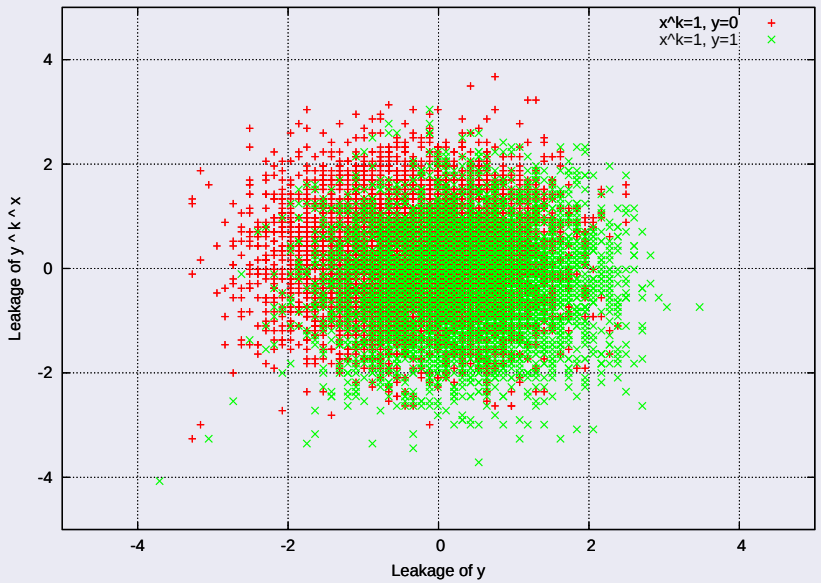
Settings

- Power consumption measurements of an 8-bit microprocessor AT90S8515 running a boolean masking scheme.
- $(d = 1,\ m = 2)$ setting: $y \in \{0, 1\}$ and $y \oplus k \oplus x \in \{0, 1\}$.
- Assumption: two p.d.f.s $f^{(x \oplus k)}$ are sufficient for the characterization problem (instead of four p.d.f.s $f^{(x,k)}$).
- Comparison
  - EM Estimates vs. Templates (Adversary with complete knowledge at profiling).
  - EM Estimates vs. Second-Order DPA (Adversary without profiling stage, as proposed by Messerges).

Empirical component p.d.f.s (Supervised Learning)

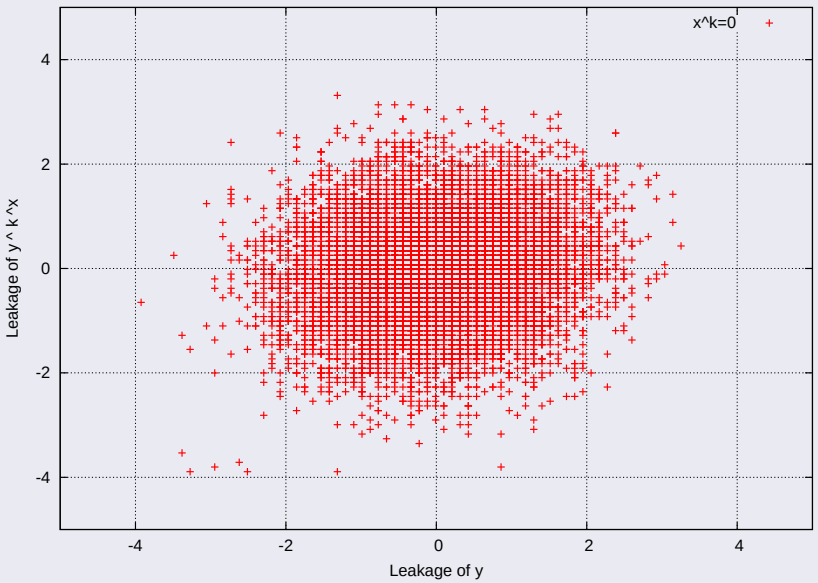
Left plot: component p.d.f.s for $(x \oplus k = 0)$
Right plot: component p.d.f.s for $(x \oplus k = 1)$

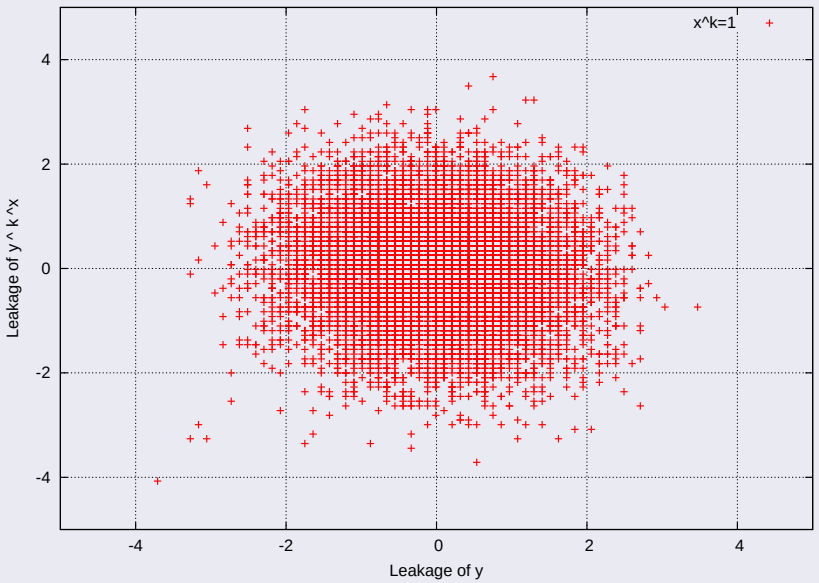




Empirical mixture p.d.f.s (Unsupervised Learning)

Left plot: mixed p.d.f. for $(x \oplus k = 0)$
Right plot: mixed p.d.f. for $(x \oplus k = 1)$





Estimated Parameters for the Gaussian component p.d.f.s

$x \oplus k$	y	μ_1	μ_2	σ_{11}	σ_{22}	$\sigma_{12} = \sigma_{21}$
Templates						
0	0	-0.343609	-0.264896	0.890693	0.929354	0.027368
0	1	0.363384	0.258210	0.849087	0.890358	0.046014
1	0	-0.353654	0.255177	0.885363	0.943963	0.042504
1	1	0.349743	-0.267222	0.877618	0.965020	0.062675
EM Algorithm, Variant 1						
$x \oplus k$	j	μ_1	μ_2	σ_{11}	σ_{22}	$\sigma_{12} = \sigma_{21}$
0	0	-0.228378	-0.222345	1.0	1.0	0.0
0	1	0.252548	0.218852	1.0	1.0	0.0
1	0	0.152021	-0.158530	1.0	1.0	0.0
1	1	-0.173202	0.166899	1.0	1.0	0.0
EM Algorithm, Variant 4						
$x \oplus k$	j	μ_1	μ_2	σ_{11}	σ_{22}	$\sigma_{12} = \sigma_{21}$
0	0	0.625019	-0.019519	0.636527	0.926563	0.143675
0	1	-0.520327	0.013980	0.695991	1.022322	0.134543
1	0	0.610178	-0.093003	0.610554	0.937405	-0.025781
1	1	-0.549292	0.088531	0.695000	1.024076	0.006803

Experimental Success Rates at Key Recovery

N°	Templates	EM Algorithm		Second-Order DPA
		Variant 4	Variant 1	
10	58.17 %	58.49 %	58.77 %	54.84 %
20	62.82 %	62.26 %	61.63 %	56.74 %
50	68.43 %	68.26 %	67.90 %	61.67 %
100	75.33 %	74.52 %	74.59 %	67.46 %
200	83.85 %	83.13 %	81.22 %	73.93 %
400	91.59 %	91.05 %	89.52 %	81.89 %
1000	98.88 %	98.68 %	98.09 %	92.77 %
2000	99.94 %	99.95 %	99.91 %	98.44 %

Further Directions for Higher-Order Analysis

Further Directions

1. Increase of m (number of instants):
   - should clearly improve success rates for key recovery.
2. Increase of d (number of bits):
   - results in two drawbacks:
     - Number of free parameters increases exponentially.
     - Number of measurements that are available for estimation decreases exponentially.
   - The benefit of an improved signal-to-noise ratio (due to more chosen bits) may be thwarted.
3. Location of relevant instants without knowing the masks:
   - check for all possible combinations (for $m = 2$),
   - second-order DPA (for $m = 2$),
   - combinations of $m = 2$ candidates, or
   - principal component analysis.

Conclusion

- Gaussian mixture models and the EM algorithm can be used for profiling of a masked implementation with incomplete knowledge.
- For a single-bit second-order setting:
  - Experimental key recovery efficiency is clearly better than second-order DPA and close to templates.
- Masking may not be sufficient to secure cryptographic implementations.