Leakage Detection with the χ^2-Test

Amir Moradi[1], Bastian Richter[1],
Tobias Schneider[2] and François-Xavier Standaert[2]

[1] Horst Görtz Institute for IT Security, Ruhr-Universität Bochum, Germany
[2] ICTEAM/ELEN/Crypto Group, Université catholique de Louvain, Belgium

CHES 2018, Amsterdam 10.09.2018

Motivation

Security Evaluation

- How to assure that a device does not leak sensitive values during execution of a cryptographic operation?

- Often performed based on attacks (e.g. in Common Criteria)
  - High complexity
  - Every attack has to be optimized
  - Easy to miss an attack vector

Motivation
Leakage Detection

- General approach to detect leakage independent of models or attack methods

- Reduction to general statistical assumptions without a specific model for the implementation (black box)

TVLA based on Welch‘s t-test [1]:

1. Reduction to two classes (e.g. fixed-vs.-random)
2. Simple statistical treatment (estimation of statistical moments)

[1] Goodwill et al., A testing methodology for side-channel resistance validation. NIST non-invasive attack testing workshop, 2011.

Motivation

t-Test Problems

- The two properties can lead to problems

1. **Reduction to two classes**
   - **False negative** because of leakage which is too similar in two classes but would be detectable with more classes

Motivation

t-Test Problems

- The two properties can lead to problems

1. **Reduction to two classes**
   - **False negative** because of leakage which is too similar in two classes but would be detectable with more classes

2. **Estimation and comparison of separate moments**
   - **False negative** because of leakage distributed over multiple moments

Motivation
t-Test Problems

- The two properties can lead to problems

1. **Reduction to two classes**
   - **False negative** because of leakage which is too similar in two classes but would be detectable with more classes

χ^2-Test works with multiple classes

2. **Estimation and comparison of separate moments**
   - **False negative** because of leakage distributed over multiple moments

Motivation
t-Test Problems

- The two properties can lead to problems

1. **Reduction to two classes**
   – **False negative** because of leakage which is too similar in two classes but would be detectable with more classes

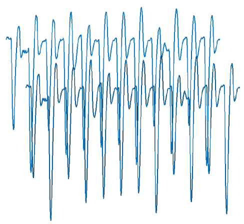
$\boxed{\chi^2\text{-Test works with multiple classes}}$

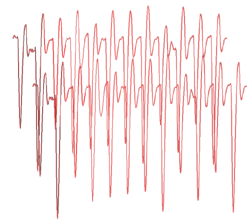
2. **Estimation and comparison of separate moments**
   – **False negative** because of leakage distributed over multiple moments

$\boxed{\chi^2\text{-Test is based on the whole distribution}}$

χ^2-Test Methodology

Fixed vs. Random

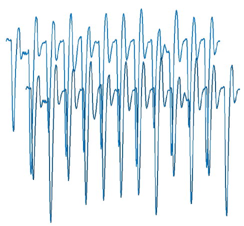


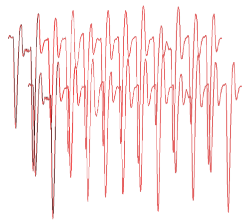


1. Measure traces for random or fixed input in random order

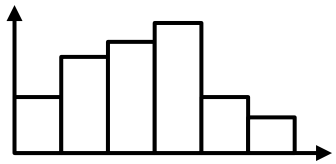
χ^2-Test Methodology

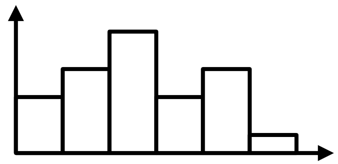
Fixed vs. Random





1. Measure traces for random or fixed input in random order

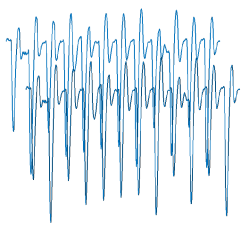


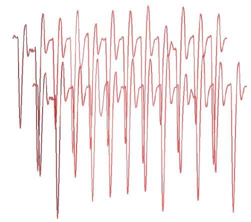


2. Compute histograms for each point of classes

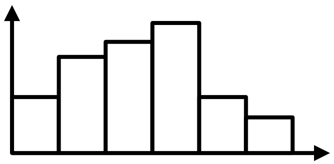
χ^2-Test Methodology

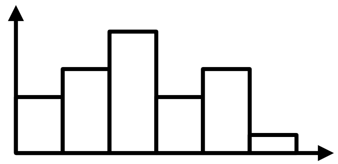
Fixed vs. Random





1. Measure traces for random or fixed input in random order





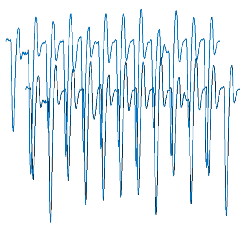
2. Compute histograms for each point of classes

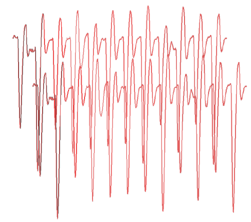
		Bin					
	$F_{i,j}$	0	1	2	3	4	5
Class	0	3	4	6	3	4	1
	1	2	3	7	5	3	2

3. Compute contingency table $F_{i,j}$ from histograms

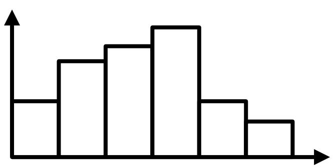
χ^2-Test Methodology

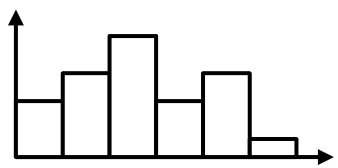
Fixed vs. Random





1. Measure traces for random or fixed input in random order





2. Compute histograms for each point of classes

Bin

Class	$F_{i,j}$	0	1	2	3	4	5
	0	3	4	6	3	4	1
	1	2	3	7	5	3	2

$$x = \sum_{i=0}^{r-1} \sum_{j=0}^{c-1} \frac{(F_{i,j} - E_{i,j})^2}{E_{i,j}},$$

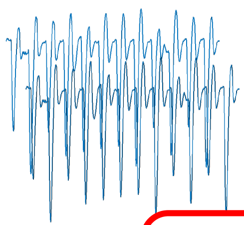
$$p = \int_{x}^{\infty} \mathsf{f}(x, v)\, dx,$$

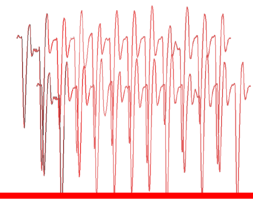
3. Compute contingency table $F_{i,j}$ from histograms

4. Compute x, v, and p from table $F_{i,j}$

χ^2-Test Methodology

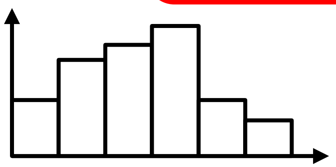
Fixed vs. Random

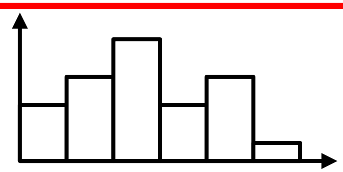




1. Measure traces for random or fixed input in random order

Same procedure as for t-test [1]





2. Compute histograms for each point of classes

Class \ Bin: $F_{i,j}$	0	1	2	3	4	5
0	3	4	6	3	4	1
1	2	3	7	5	3	2

$$x = \sum_{i=0}^{r-1} \sum_{j=0}^{c-1} \frac{(F_{i,j} - E_{i,j})^2}{E_{i,j}},$$

$$p = \int_{x}^{\infty} \mathrm{f}(x, v)\, dx,$$

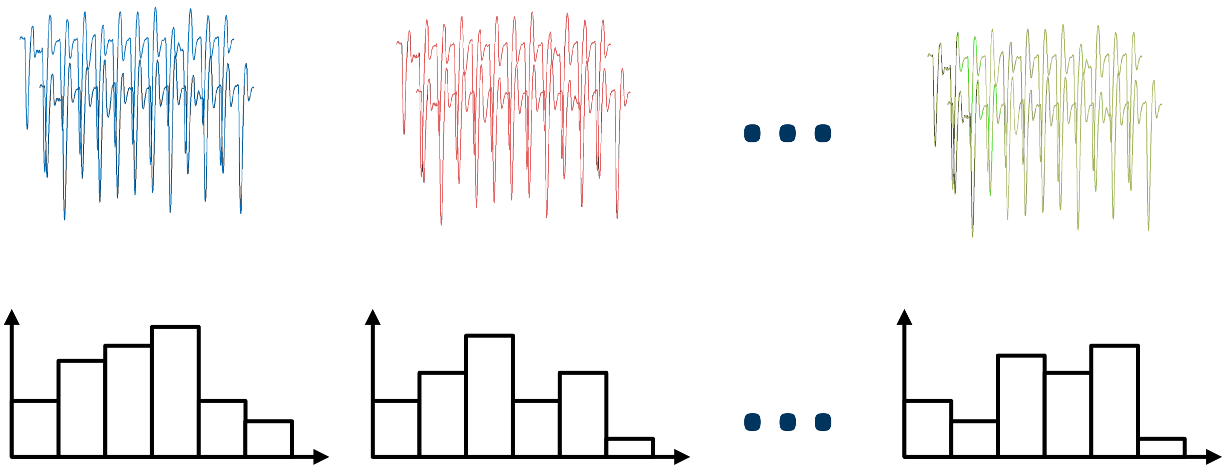
3. Compute contingency table $F_{i,j}$ from histograms
4. Compute x, v, and p from table $F_{i,j}$

[1] Reparaz et al., Fast Leakage Assessment, CHES 2017

χ^2-Test Methodology

Fixed vs. Fixed

RUB



1. Measure traces for r different inputs in random order
2. Compute histograms for each point of classes
3. Compute contingency table $F_{i,j}$ from histograms
4. Compute x, v, and p from table $F_{i,j}$

Bin

Class

$F_{i,j}$	0	1	2	3	4	5
0	3	4	6	3	4	1
1	2	3	7	5	3	2
$\vdots$			$\vdots$			
r-1	1	4	6	2	3	5

$$x = \sum_{i=0}^{r-1} \sum_{j=0}^{c-1} \frac{(F_{i,j} - E_{i,j})^2}{E_{i,j}},$$

$$p = \int_x^{\infty} \mathsf{f}(x, v)\, dx,$$

χ^2-Test

Pearson's χ^2-Test of Independence

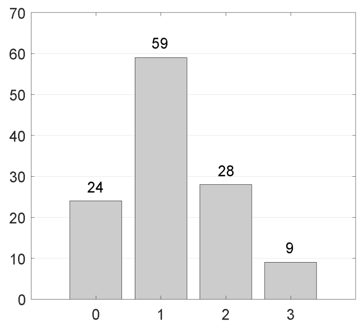
RUB

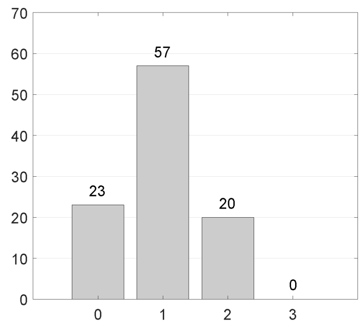
- **Null hypothesis:** The occurrences of the observations are independent.
- If the test concludes that the null hypothesis is rejected, the leakage is assumed to be informative.
- Evaluation of independence based on contingency table of frequencies.
- In contrast to the t-test we have to calculate the p-values for the χ^2-test as the degree-of-freedom does not converge.
  - ➔ We chose $\mathrm{p} = 10^{-5}$ as threshold (equivalent to $t = 4.5$).

χ^2-Test

Pearson's χ^2-Test of Independence

1. Build contingency table $F_{i,j}$ from histograms

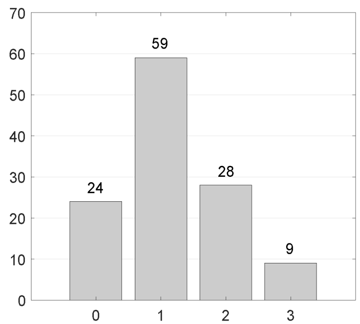


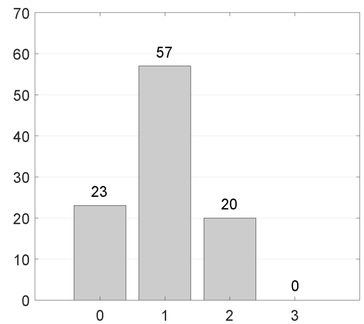


$F_{i,j}$	$j = 0$	$j = 1$	$j = 2$	$j = 3$	total
$i = 0$	24	59	28	9	120
$i = 1$	23	57	20	0	100
total	47	116	48	9	220

Pearson's χ^2-Test of Independence

1. Build contingency table $F_{i,j}$ from histograms





$F_{i,j}$	$j = 0$	$j = 1$	$j = 2$	$j = 3$	total
$i = 0$	24	59	28	9	120
$i = 1$	23	57	20	0	100
total	47	116	48	9	220

2. Calculate expected values $E_{i,j}$ for each cell

$$E_{0,0} = \frac{(24 + 59 + 28 + 9) \cdot (24 + 25)}{220} = \frac{120 \cdot 47}{220} \approx 25.64$$

$$E_{i,j} = \frac{\left(\sum_{k=0}^{c-1} F_{i,k}\right) \cdot \left(\sum_{k=0}^{r-1} F_{k,j}\right)}{N}$$

$E_{i,j}$	$j = 0$	$j = 1$	$j = 2$	$j = 3$
$i = 0$	25.64	63.18	26.18	4.91
$i = 1$	21.36	52.73	21.82	4.09

RUB

3. Calculate χ^2-test statistic x and degree-of-freedom v

$$x = \sum_{i=0}^{r-1} \sum_{j=0}^{c-1} \frac{(F_{i,j} - E_{i,j})^2}{E_{i,j}}$$

$$v = (2-1) \cdot (4-1) = 3$$

For cell (0,0): $\frac{(24 - 25.64)^2}{25.64} \approx 0.10$

$$x = 0.10 + 0.29 + 0.13 + 3.41 + 0.13 \\ +0.35 + 0.15 + 4.09 = 8.64$$

χ^2-Test

Pearson's χ^2-Test of Independence

RUB

3. Calculate χ^2-test statistic x and degree-of-freedom v

$$x = \sum_{i=0}^{r-1} \sum_{j=0}^{c-1} \frac{(F_{i,j} - E_{i,j})^2}{E_{i,j}}$$

$$v = (2-1) \cdot (4-1) = 3$$

For cell (0,0): $\frac{(24 - 25.64)^2}{25.64} \approx 0.10$

$$x = 0.10 + 0.29 + 0.13 + 3.41 + 0.13 \\ +0.35 + 0.15 + 4.09 = 8.64$$

4. Derive p value using the χ^2 probability density function

$$p = \int_x^\infty f(x, v)\, dx$$

$$f(x, v) = \begin{cases} \frac{x^{\frac{v}{2}-1} e^{-\frac{x}{2}}}{2^{\frac{v}{2}} \Gamma(\frac{v}{2})}, & x > 0 \\ 0, & otherwise \end{cases}$$

$$p \approx 0.0345$$

Simulated Experiments

Univariate

RUB

- Simulation of masked hardware design with parallel processing of d shares
- Secret value X is split in to d Boolean shares X_i

$$X = X_0 \oplus X_1 \oplus \ldots \oplus X_{d-1}$$

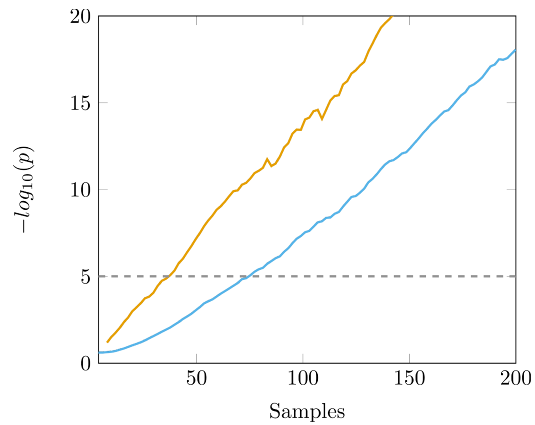
- Leakage is combined with Hamming Weight leakage function with additive Gaussian noise for three different SNRs

$$L = \sum_{i=0}^{d-1} HW(X_i) + \mathcal{N}_{0,\sigma}$$

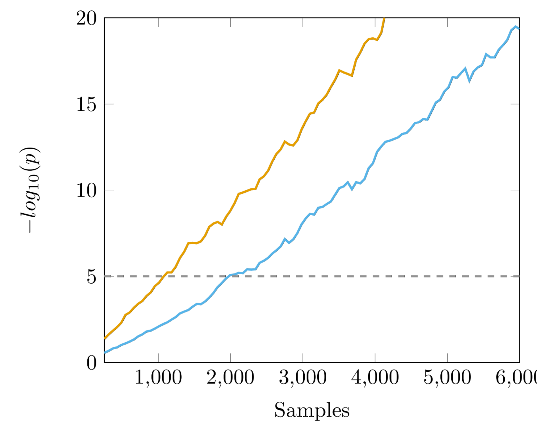
- Traces are generated for Fixed vs. Random test

Simulated Experiments
Univariate Results - Orders

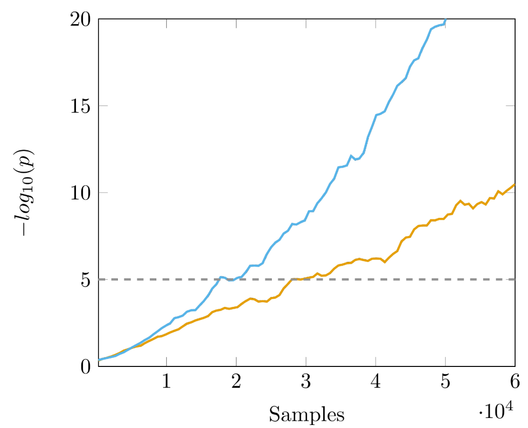
- t-test significantly outperforms χ^2-test for lower orders ($d = 1, 2$)
- χ^2-test improves with higher orders and is significantly better in order $d = 4$
- Advantage is expected to increase with higher orders



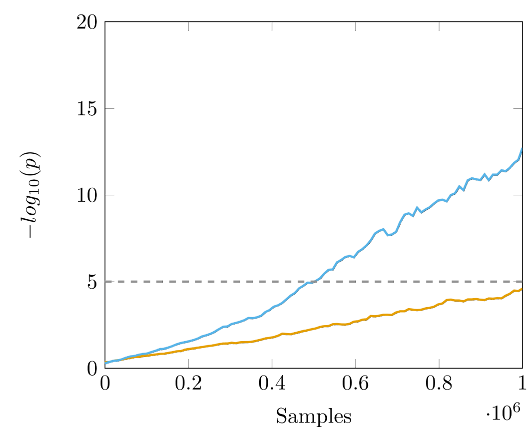
$d = 1$, $SNR_2 = 1.0$



$d = 2$, $SNR_2 = 1.0$



$d = 3$, $SNR_2 = 1.0$



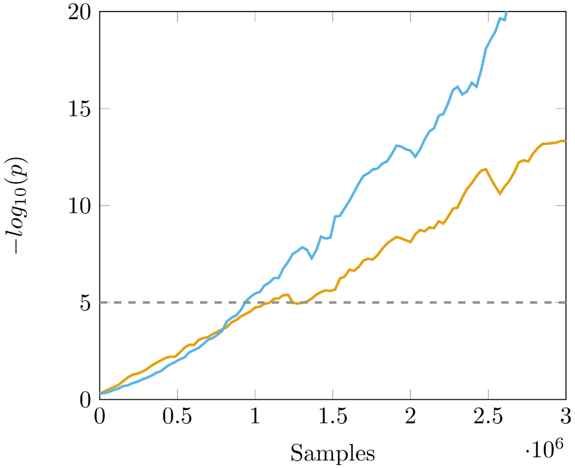
$d = 4$, $SNR_2 = 1.0$

χ^2-Test t-Test

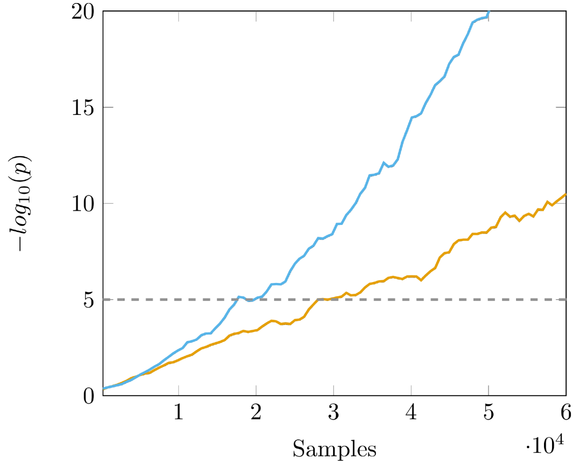
Simulated Experiments

Univariate Results - SNR

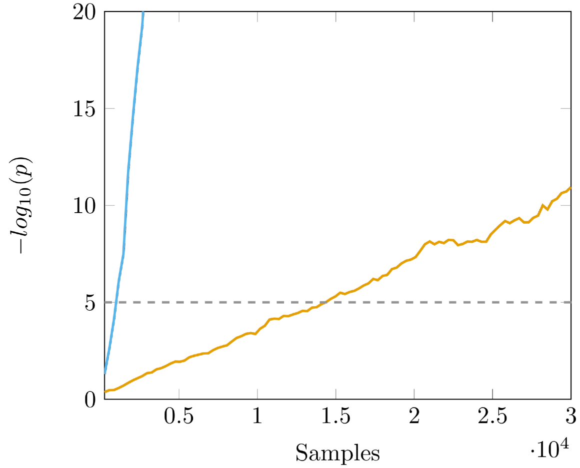
- Results of χ^2-test are highly affected by the SNR
- For low SNRs the advantage in higher orders disappears
- A high SNR significantly improves the detection of leakage
- Effect of SNR on t-test is much lower



$d = 3, SNR_1 = 0.1$



$d = 3, SNR_2 = 1.0$



$d = 3, SNR_3 = 10$

χ^2-Test t-Test

Simulated Experiments
Multivariate

- Simulation of software or serialized hardware masking

- Leakage of different shares at separate points in time

$$L_{t_i} = HW(X_i) + \mathcal{N}_{0,\sigma}, \qquad 0 \leq i < d$$

- We evaluated three different options to combine leakage

- **Normalized Product:** (χ^2- and t-test)

$$L' = \prod_{i=0}^{d-1} (L_{t_i} - \mu_{t_i})$$

Simulated Experiments

Multivariate

- **Sum Combining:**
  Possible because whole distribution and not only the means are compared

$$L' = \sum_{i=0}^{d-1} L_{t_i}$$

+ Noise Terms are not multiplied

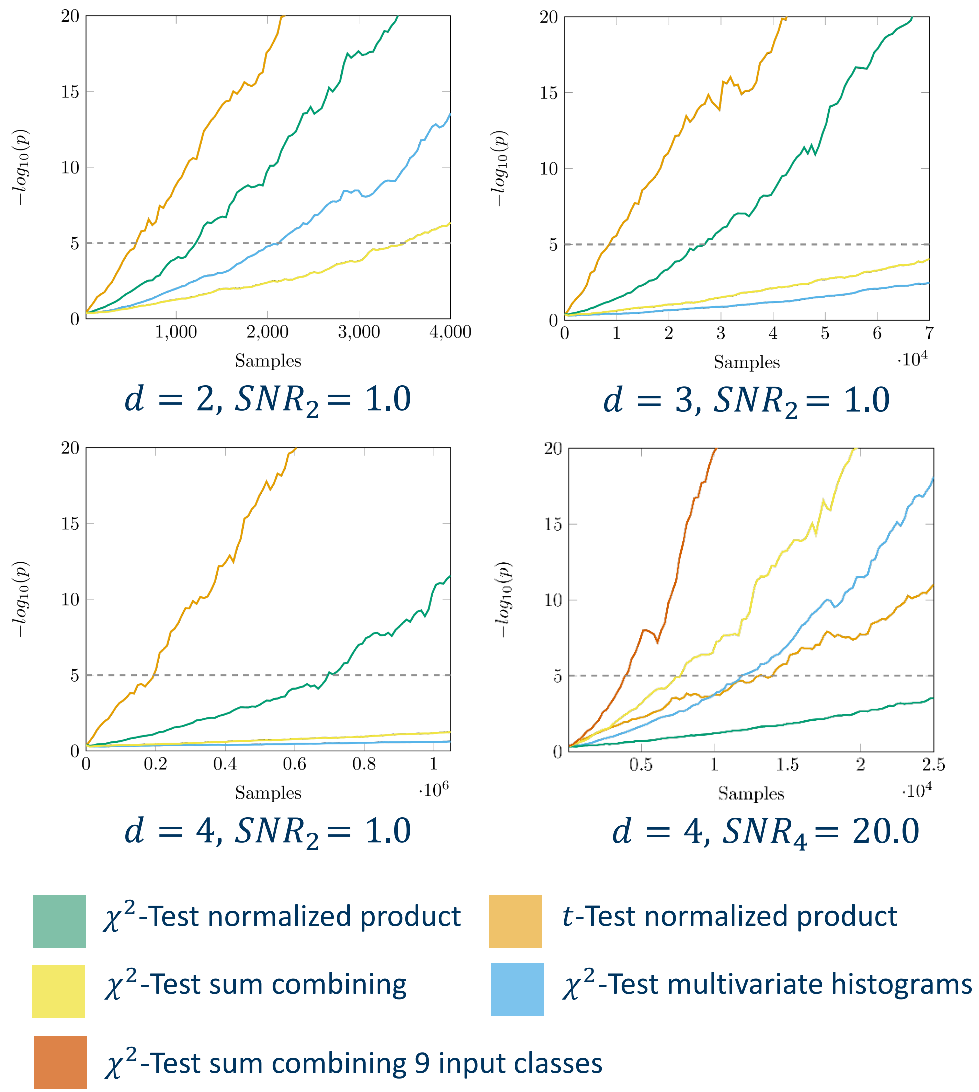
- **Multivariate Histograms:**

$$L' = (L_{t_0}, L_{t_1}, \ldots, L_{t_{d-1}})$$

Simulated Experiments

Multivariate Results

- Unless for very high SNRs t-test works better than the χ^2-test
- The normalized product works best for all orders with non-negligible noise
- Sum combining and multivariate histograms only improve the results for very low noise

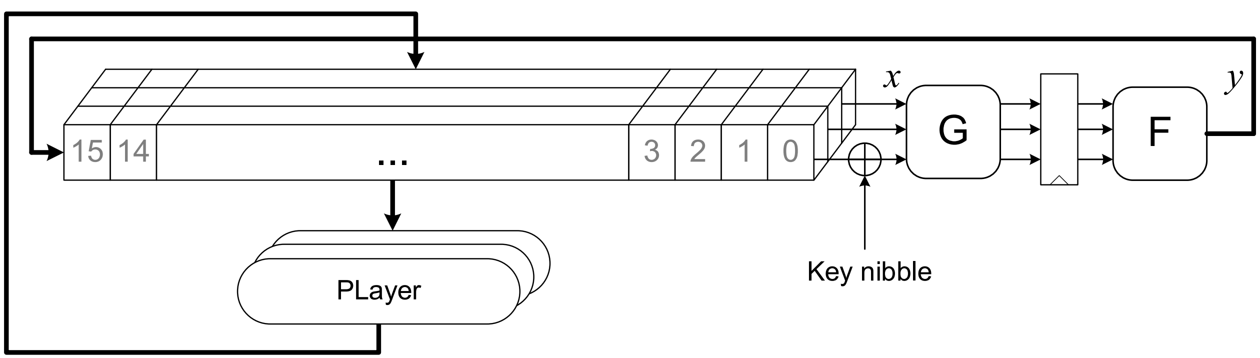


Experiments

Target

- Threshold Implementation of PRESENT with 3 shares
- S-box split up into two functions G and F
- Byte-serial implementation with shift register for state

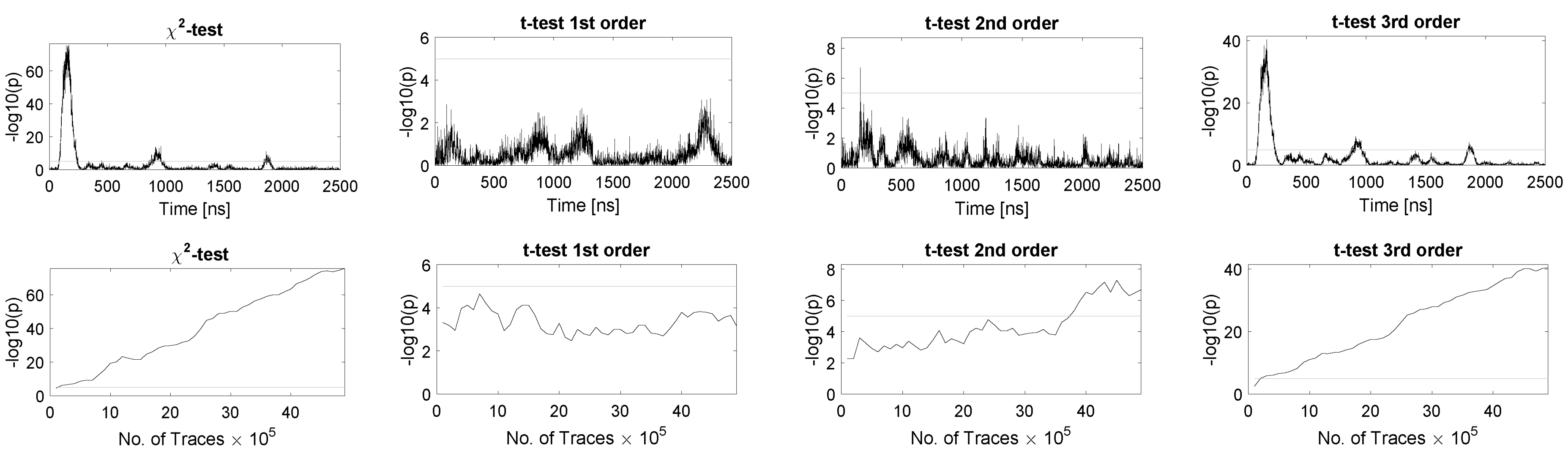
- Implemented on Spartan-6 (SAKURA-G)
- Running at 160 MHz and measured at 1 GS/s



Experiments

Fixed vs. Random

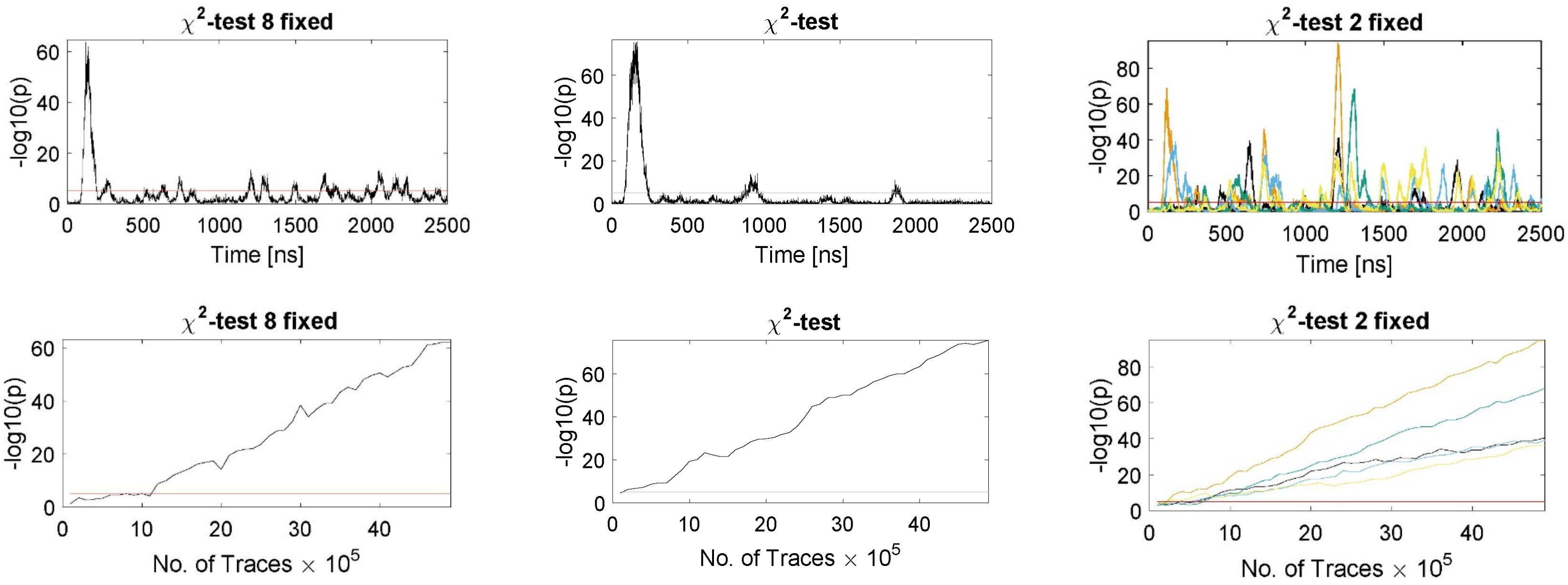
- As expected leakage detected in orders $d \geq 2$ with main leakage in third order
- χ^2-test shows similar shape as 3rd-order t-test
- Confidence is significantly higher for χ^2-test



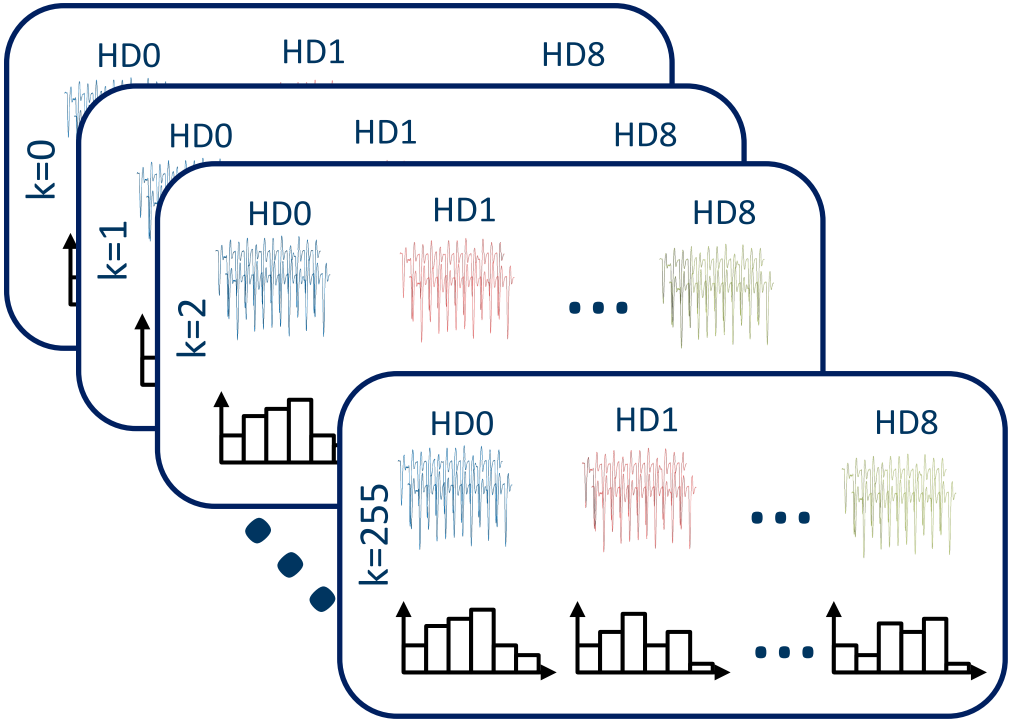
Experiments

Fixed vs. Fixed

- Traces recorded for eight different fixed plaintexts
- χ^2-test can process the plaintexts as eight classes
- Main leakage at beginning similar to fixed vs. random
- Detects leakage at late times with lower confidence as for the pairs of plaintexts



Experiments
Distinguisher



- Use multi-class capability as distinguisher with model
- For each key candidate k
  1. Sort traces into classes by model (e.g. HD)
  2. Calculate histograms for the classes
  3. Calculate x, v and p
- Rank key candiates by p-value

Experiments

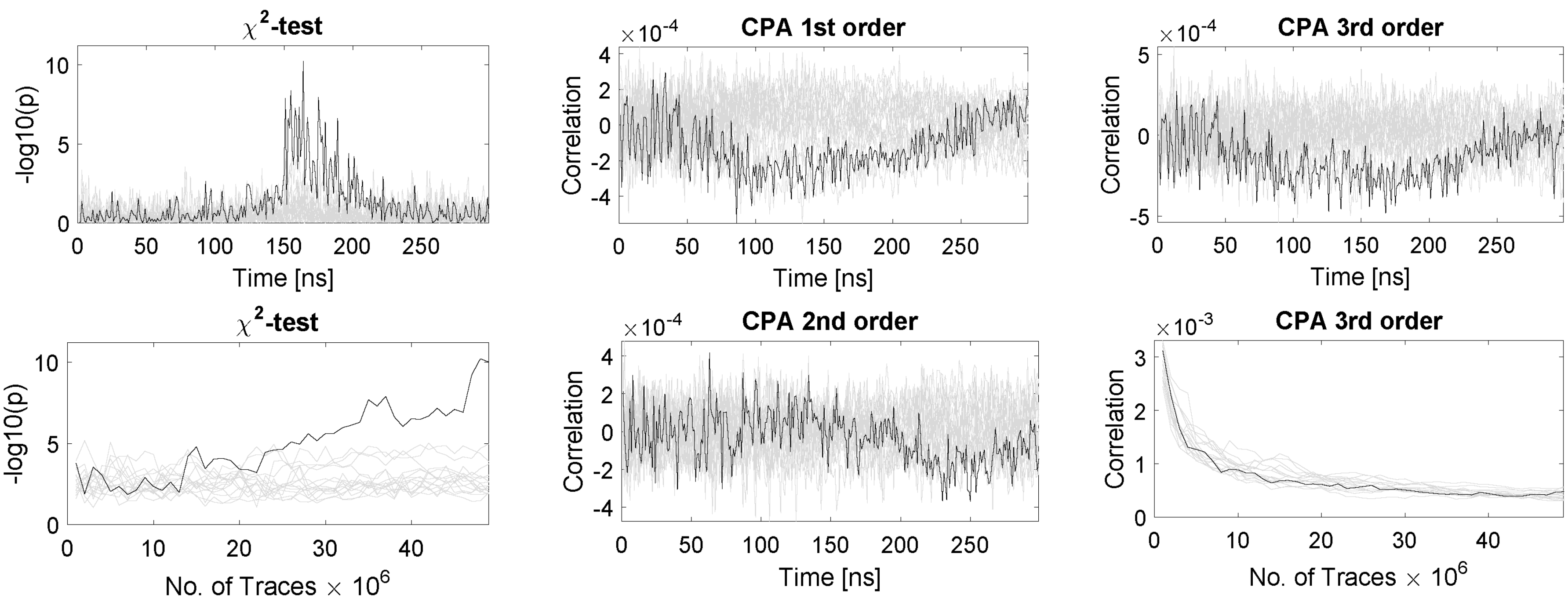
Distinguisher

RUB

- Utilizes leakage in the whole distribution and not only a single moment
- Similar to Mutual Information Analysis (MIA) but provides a confidence level for each key candidate
- Number of classes has to be lower than number of key candidates (same for MIA)

Experiments
Distinguisher

- CPA and χ^2-test with HD-Model of consecutive S-boxes
- None of the higher-order CPAs is successful (with 50M traces)
- χ^2-test is successfull after 28 million traces



Conclusion

RUB

- We presented the χ^2-test as a complement to the t-test
- It is able to outperform the t-test if:
  - The noise level is not sufficient
  - The leakage is distributed over multiple statistical moments
- It should only be used together with the t-test, since there are cases which are not detected.
- Use t-test to evaluate the security order and χ^2-test to evaluate the noise level.

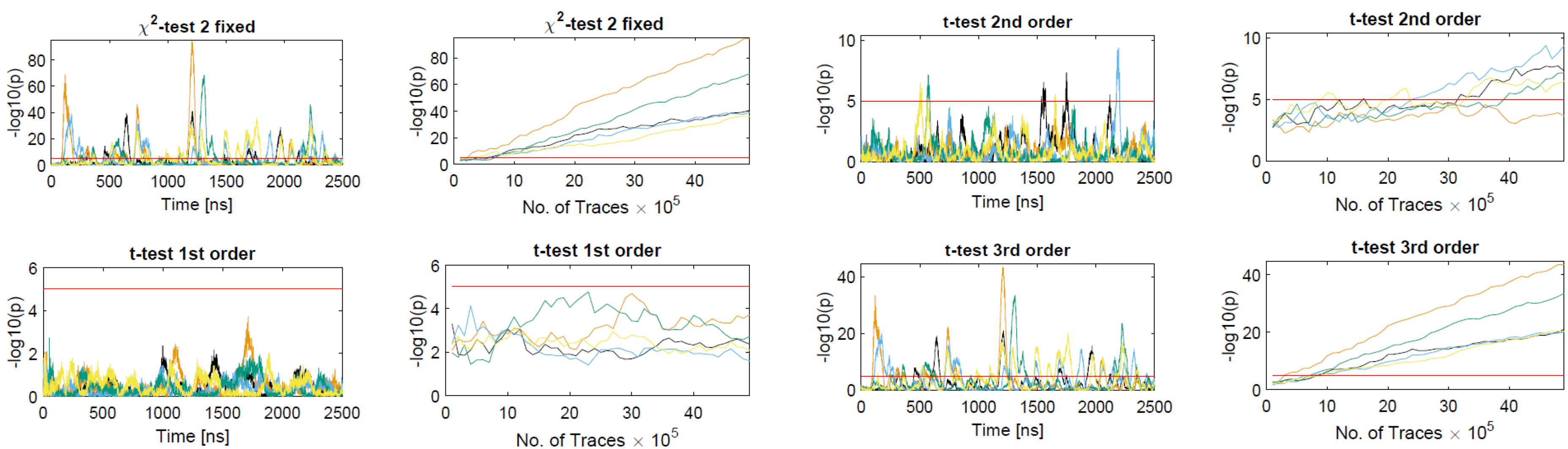
Thank You For Your Attention!
Any Questions?

Experiments

Fixed vs. Fixed

RUB

- Traces recorded for eight different fixed plaintexts
- Five combinations of two plaintexts plotted
- Different combinations detect leakage at different times
- Third order t-test and χ^2-test again similar
- χ^2-test again gives higher confidence



Experiments

Performance

- t-Test and χ^2-Test implemented in C++
- Based on histograms computed before for both tests
- Both tests need approx. 2.8 µs per point
  on an Intel i7-6600U @2.6GHz
- Calculation of t-Test only speeds up by 0.4 µs when omitting the calcualtion of p value and degree of freedom