A New Difference Method for Side-Channel Analysis with High-Dimensional Leakage Models

Annelie Heuser, Michael Kasper, Werner Schindler, Marc Stöttinger

Technische Universität Darmstadt (TUD)

Fraunhofer Institute for Secure Information Technology (SIT)

Bundesamt für Sicherheit in der Informationstechnik (BSI)

Center for Advanced Security Research Darmstadt (CASED)

CASCADE

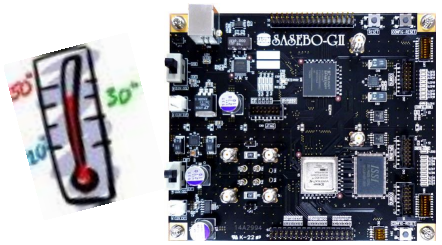
CONSTRUCTIVE ATTACKS | SIDE CHANNEL ANALYSIS | SECURE DESIGN

Overview

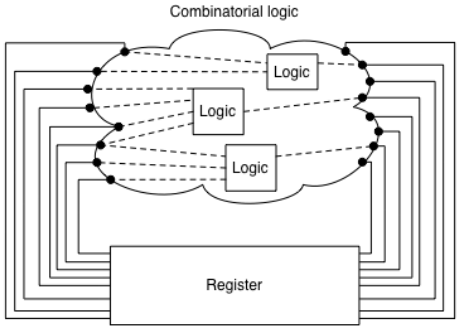
CASED

A New Difference Method for Side-Channel Analysis with High-Dimensional Leakage Models

New Difference Method



Leakage Models



illustrated by

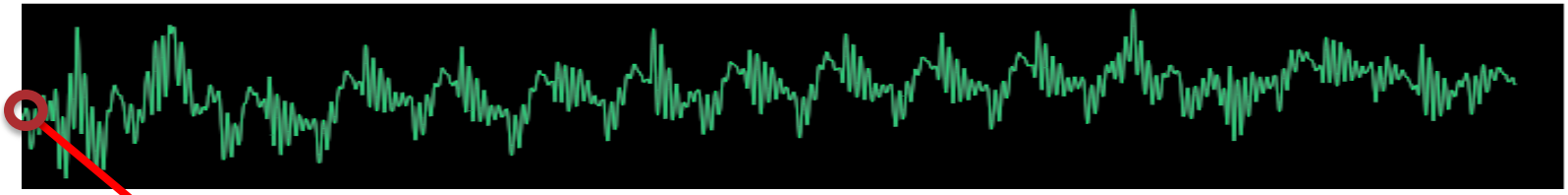
Dpacontest v2

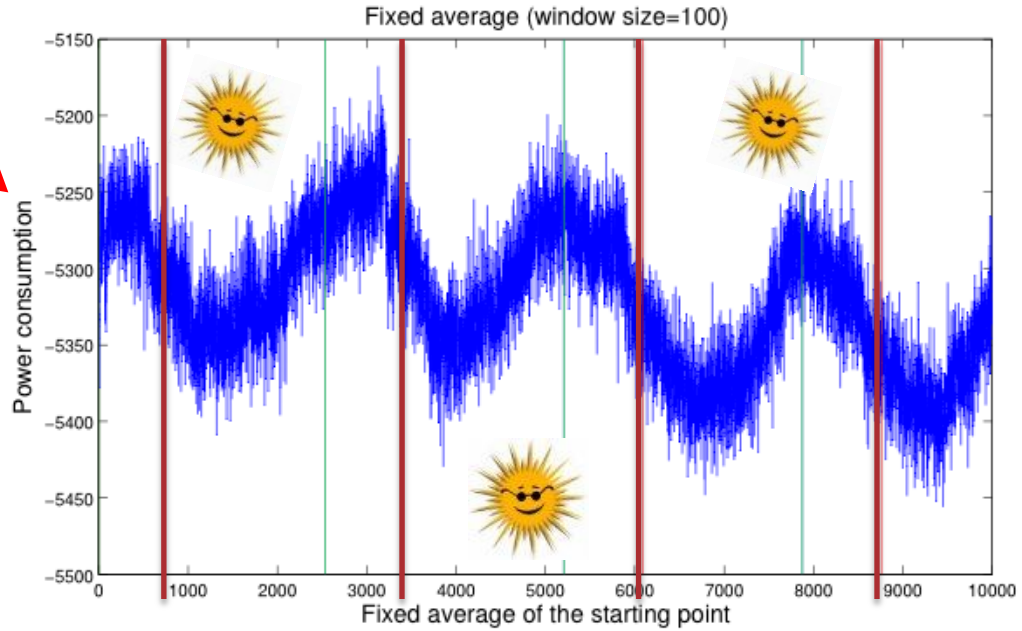
Dpacontest v2

- AES- 128 hardware implementation
- Public Access
  - Template Base
    - 1.000.000 measurements
    - Random keys / inputs
  - Public Base
    - 32 different fixed keys each 20.000 random inputs
- Organizers
  - Evaluation on a Private Base
  - Criteria: Partial Success Rate, Global Success Rate, Guessing Entropy

Dpacontest v2

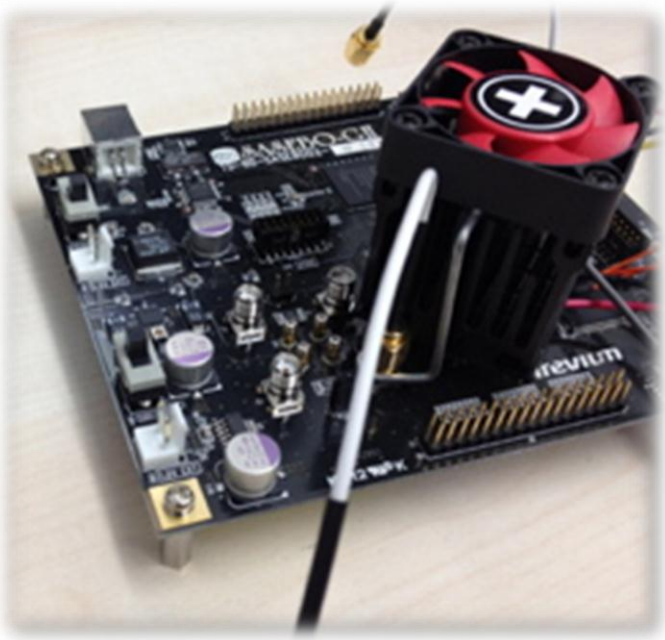
Template Base with 1.000.000 measurements, recorded during 3-4 days

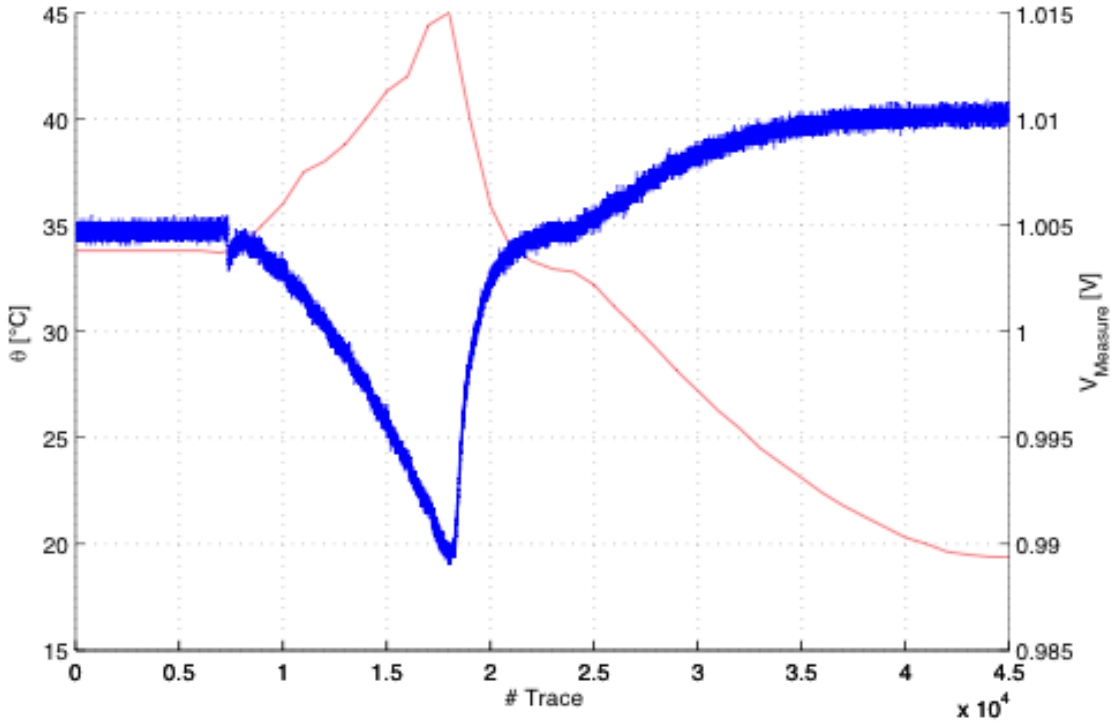




Environmental Influences

CASED





Temperature

Power consumption

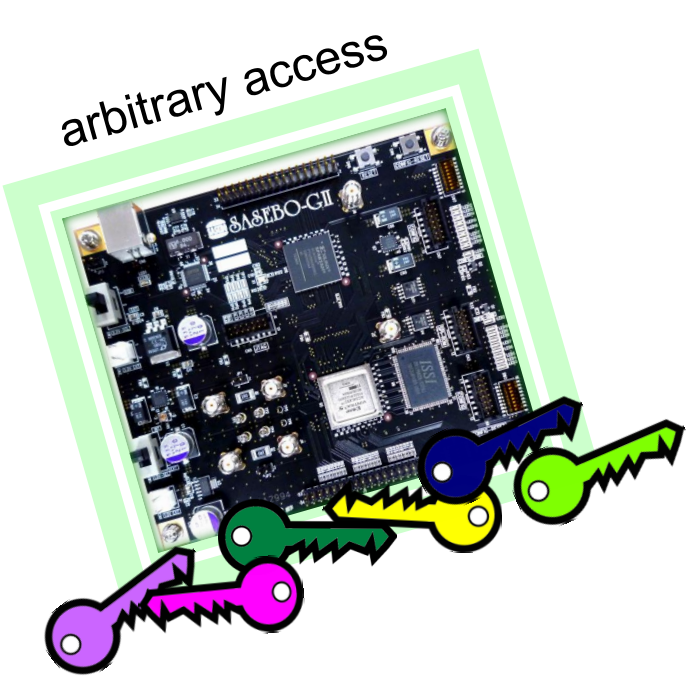
TECHNISCHE UNIVERSITÄT DARMSTADT

Fraunhofer SIT

h.da HOCHSCHULE DARMSTADT UNIVERSITY OF APPLIED SCIENCES

LOEWE – Landes-Offensive zur Entwicklung Wissenschaftlich-ökonomischer Exzellenz

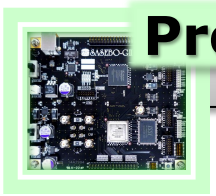
Profiled Side-Channel Attacks



- Classification: Template Attack [Chari03], + Model [AGH07]
- Model + Regression: Stochastic Approach [SLP05]

Stochastic Approach in a Nutshell

CASED



Profiling Phase

Stochastic Model

$$I_t(x,k) = \boxed{\begin{array}{c}\text{Leakage Model}\\ + \\ \text{Regression}\end{array}} + R_t$$

Power consumption at time t

Deterministic part

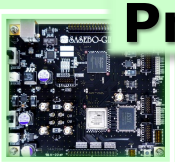
Noise

TECHNISCHE UNIVERSITÄT DARMSTADT Fraunhofer SIT h.da HOCHSCHULE DARMSTADT UNIVERSITY OF APPLIED SCIENCES LOEWE – Landes-Offensive zur Entwicklung Wissenschaftlich-ökonomischer Exzellenz

Stochastic Approach in a Nutshell

CASED

Profiling Phase



$$I_t(x,k) = h_t(x,k) + R_t$$

$$\begin{pmatrix} i_t(x_1,k) \\ i_t(x_2,k) \\ i_t(x_3,k) \\ \vdots \\ i_t(x_{N_1},k) \end{pmatrix} = \begin{pmatrix} 1 & g_{1,t;k} & \cdots & g_{u-1,t;k} \\ 1 & g_{1,t;k} & \cdots & g_{u-1,t;k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & g_{1,t;k} & \cdots & g_{u-1,t;k} \end{pmatrix} \beta^* + \epsilon$$

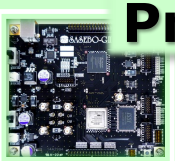
Leakage Model

Least Square Estimation $\widetilde{\beta^*} = (A^T A)^{-1} A^T \mathbf{i_t}$

Estimated Approximator $\widetilde{h}^*_{t;k}(\cdot,k) = \sum_{j=0}^{u-1} \widetilde{\beta}^*_{j,t;k} \cdot g_{j,t;k}(\cdot,k)$

Stochastic Approach in a Nutshell

CASED



Profiling Phase

$$I_t(x,k) = h_t(x,k) + R_t$$

$$R_{\mathbf{t}} = I_{\mathbf{t};k}(X,k) - h_{\mathbf{t};k}(X,k)$$

$$\mathbf{R_t} \sim \mathcal{N}(0,C)$$

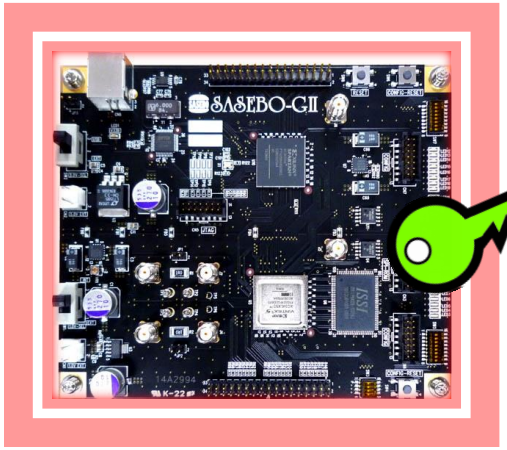
$$f_C\colon \mathbb{R}^m \to \mathbb{R} \qquad f_C(\mathbf{z}) = \frac{1}{\sqrt{(2\pi)^m \det C}} e^{-\frac{1}{2}\mathbf{z}^T C^{-1}\mathbf{z}}$$

TECHNISCHE UNIVERSITÄT DARMSTADT Fraunhofer SIT h_da HOCHSCHULE DARMSTADT UNIVERSITY OF APPLIED SCIENCES LOEWE – Landes-Offensive zur Entwicklung Wissenschaftlich-ökonomischer Exzellenz

Stochastic Approach in a Nutshell

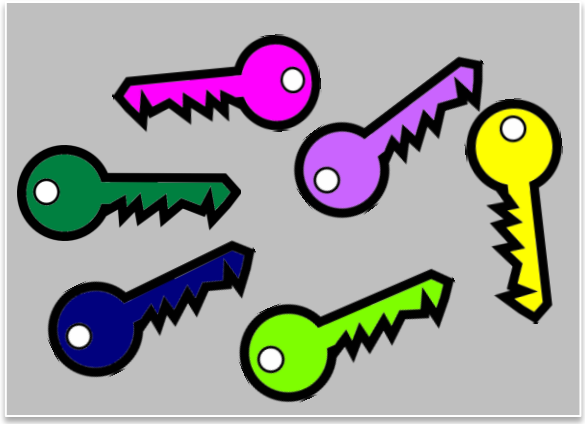
CASED

Attack Phase



Attacker gains N_3 measurements

with a secret key $k^\dagger$

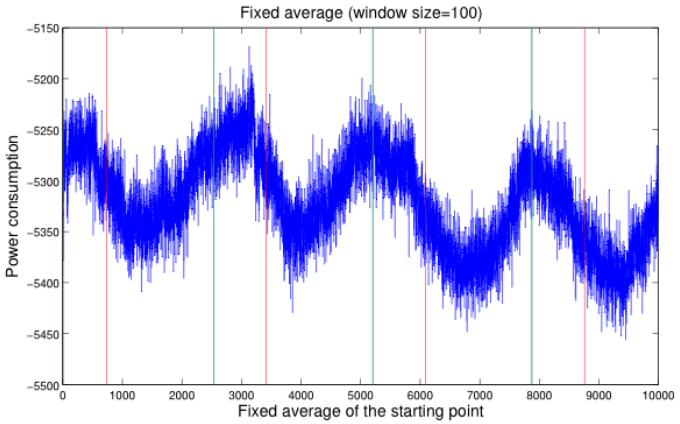


Maximum Likelihood

$$\prod_{l=1}^{N_3} f_C\left(\mathbf{i_t}(x_l, k^\dagger) - \widetilde{\mathbf{h}}^*_{\mathbf{t};k}(x_l, k)\right)$$

TECHNISCHE UNIVERSITÄT DARMSTADT Fraunhofer SIT h.da HOCHSCHULE DARMSTADT UNIVERSITY OF APPLIED SCIENCES LOEWE – Landes-Offensive zur Entwicklung Wissenschaftlich-ökonomischer Exzellenz

Drifting Offset



$$I_t(x,k) = h_t(x,k) + R_t$$

'normal' SA

OTM

$$I_t(x_l,k) = h_t(x_l,k) + \tau_{t;l} + R_t$$

$$\tau_{t;l} - \tau_{t;l+1} \approx 0$$

Consider differences of consecutive power consumptions!

Stochastic Approach + OTM

Consider differences of consecutive power consumptions!

OTM

Profiling Phase I

$$\widetilde{h}^*_{t;k}(\cdot,k) = \widetilde{\beta}^*_{0,t,k} + \sum_{j=1}^{u-1} \widetilde{\beta}^*_{j,t,k} g_{j,t;k}(\cdot,k)$$

Stochastic Approach + OTM

CASED

Consider differences of consecutive power consumptions!

OTM

Profiling Phase II

‘normal’ SA

$$R_{\mathbf{t}} \sim \mathcal{N}(0, C) \qquad R_{\mathbf{t}} = I_{\mathbf{t};k}(X, k) - h_{\mathbf{t};k}(X, k)$$

SA - OTM

$$\mathbf{I_t}(x_l, k) - \mathbf{h_t}(x_l, k) - \tau_{\mathbf{t};l}$$
$$-(\mathbf{I_t}(x_{l+1}, k) - \mathbf{h_t}(x_{l+1}, k) - \tau_{\mathbf{t};l+1}) \sim N(0, 2C)$$

Stochastic Approach + OTM

CASED

Consider differences of consecutive power consumptions!

OTM

Attack Phase

Maximum Likelihood

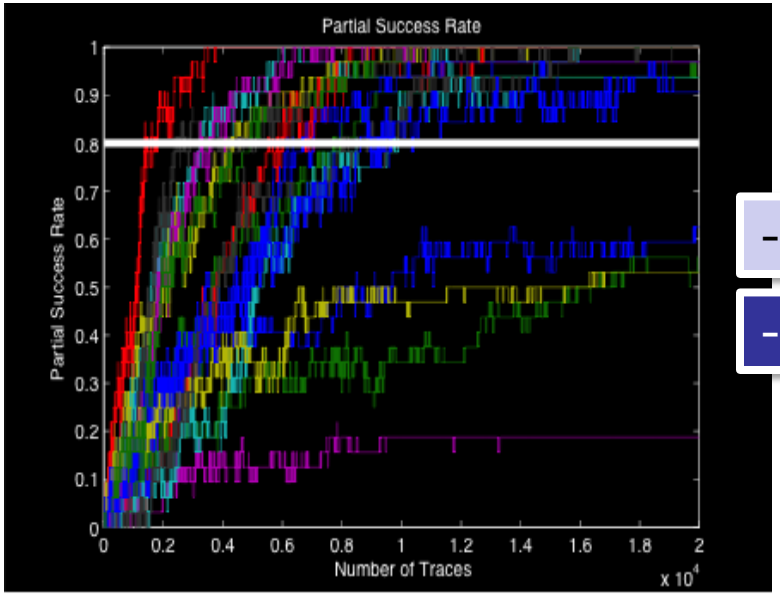
$$\prod_{l=1}^{N_3} f_C\left(\mathbf{i_t}(x_l, k^\dagger) - \widetilde{\mathbf{h}}^*_{\mathbf{t};k}(x_l, k)\right)$$

Independence → additional mathematical difficulties (details: paper)

Results: SA vs. SA-OTM

CASED

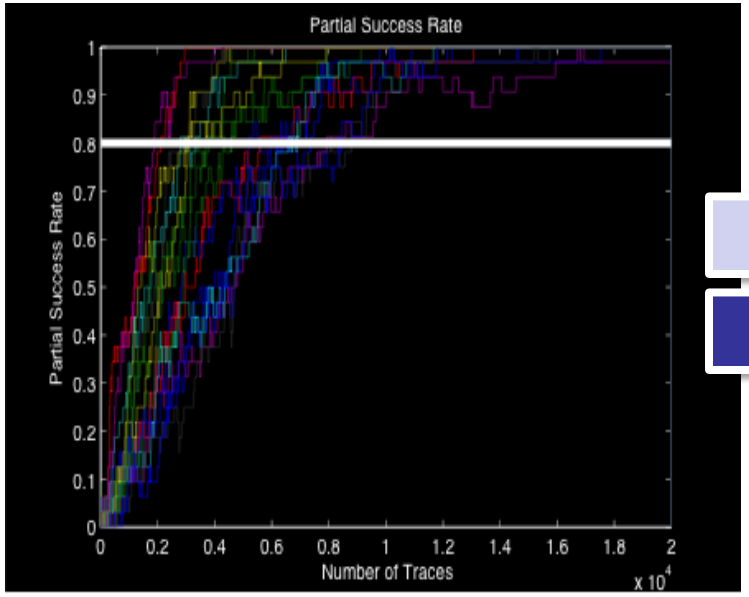
Stochastic Approach



--

--

Stochastic Approach + OTM



8781

4638

PSR = average % of obtaining the correct key byte

PGE = average ranking of the correct key byte

min stable PSR > 80%

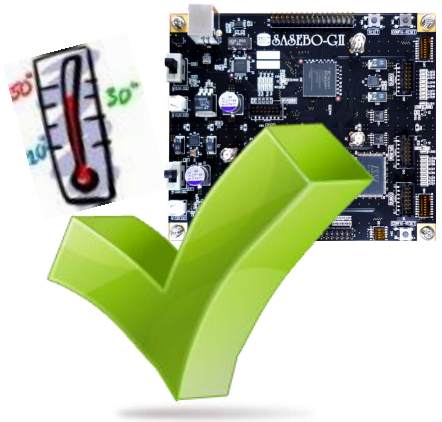
max PGE stable < 10

Overview

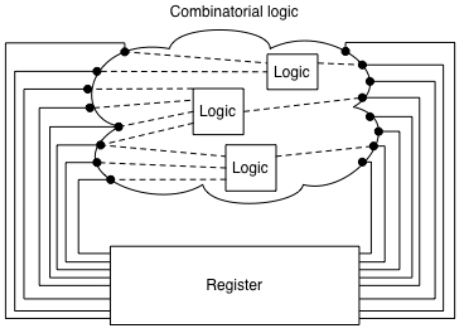
CASED

A New Difference Method for Side-Channel Analysis with High-Dimensional Leakage Models

New Difference Method



Leakage Models



TECHNISCHE UNIVERSITÄT DARMSTADT Fraunhofer SIT h.da HOCHSCHULE DARMSTADT UNIVERSITY OF APPLIED SCIENCES LOEWE – Landes-Offensive zur Entwicklung Wissenschaftlich-ökonomischer Exzellenz

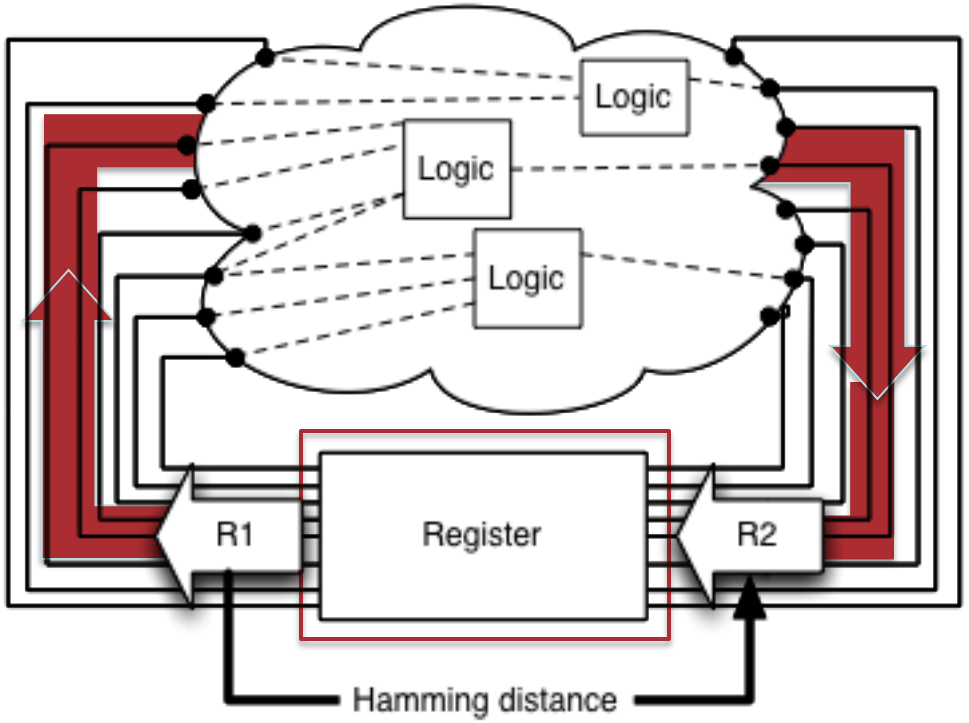
Leakage Models

Common Models:

- HD Register transitions
- bitwise HD transitions [KSS09]

High-dimensional Models:

- bitwise HD transitions + interactions between bit lines



Results: Stochastic Approach + OTM

Public Base

Dim	Interactions	PSR > 80%	PGE < 10
8	1	6781	4637
36	1,2	5876	2308
92	1,2,3	5195	2139
162	1,2,3,4	4353	1690
218	1,2,3,4,5	3552	1504
246	1,2,3,4,5,6	3769	1477
254	1,2,3,4,5,6,7	3720	1476
255	1,2,3,4,5,6,7,8	3718	1479

-13 % / -50 %

-23 % / -53 %

-35 % / -63 %

-47 % / -67 %

-44 % / -68 %

-45 % / -68 %

-45 % / -68 %

Results: Private Base

Dim	Interactions	PSR > 80%	PGE < 10
92	1,2,3	4358	1894
255+alignment	1,2,3,4,5,6,7,8	2748	1356

‘official winner’	5890	2767

unknown type attack

[LNOS11]	2155	3181

clockwise collision attack
additionally considers information from Round 9

Conclusion

CASED

A New Difference Method for Side-Channel Analysis with High-Dimensional Leakage Models

New Difference Method

- ❑ Environmental influences
- ❑ Tolerates drifting offsets

Leakage Models

- ❑ More precise models
- ❑ Captures also interactions between circuit elements

References

[Chari03] Chari, S., Rao, J.R., Rohatgi, P.: Template Attacks.
In: Kaliski, B., C ̧etin Kaya Koc ̧, Paar, C. (eds.) CHES 2003.
Lecture Notes in Computer Science, vol. 2523, pp. 13–28. Springer (2002)

[AGH07] EL AABID, M., GUILLEY, S, HOOGVORST, P. :Template Attacks with a Power Model
Cryptology ePrint Archive, Report 2007/443

[SLP05] Schindler,W.,Lemke,K.,Paar,C.:AStochasticModelforDifferentialSideChannelCrypt-
analysis. In: Rao, J.R., Sunar, B. (eds.) CHES 2005. Lecture Notes in Computer Science, vol.
3659, pp. 30–46. Springer (2005)

[MOP07] Stefan Mangard, Elisabeth Oswald, Thomas Popp:
Power analysis attacks - revealing the secrets of smart cards. Springer 2007: I-XXIII, 1-337

[KSS09] Kasper, M., Schindler, W., Stöttinger, M.: A Stochastic Method for Security Evaluation of
Cryptographic FPGA Implementations. In: FPT 2010. pp. 146–154. IEEE Press (2010)

[LNOS11] Li, Y. , Nakatsu, D., Li, Q., Ohta, K., Sakiyama, K.,
Clockwise Collision Analysis -- Overlooked Side-Channel Leakage Inside Your Measurements
Cryptology ePrint Archive, Report 2011/579

Getting more from PCA: First results of using Principal Component Analysis for extensive power analysis

Lejla Batina[1, 2], Jip Hogenboom[3] & Jasper G. J. van Woudenberg[4]

[1] Radboud University Nijmegen, The Netherlands

[2] KU Leuven, Belgium

[3] KPMG Advisory

[4] Riscure

Session ID: CRYP-402

Session Classification: Side Channel Attacks III

RSA CONFERENCE 2012

Outline

- Introduction
- Workings of PCA
- Previous works
- PCA for side-channel analysis
  - Motivating examples
- Experiments
  - Noise reduction
  - PCA transformation
  - PCA on misaligned traces
- Conclusions and Future work

Introduction and Motivation

- We know a lot about side-channel analysis, but...
- Many open problems and research directions
- Similar research questions as in other communities i.e. privacy, machine learning, image processing, etc.
- Various countermeasures makes the key-recovery ever challenging
- PCA considered very powerful tool for data reduction/approximation

Principal Component Analysis (PCA)

- Multivariate technique, known since ~1900
- Finds major patterns in data variability
- Used to reduce the noise or the dimensionality in a data set, while retaining the most variance
- Given data points in n-dimens. space, project into a subspace while preserving max info
- Transforms data to a new set of Principal Components (PCs) by means of eigenvectors
- Used e.g. in gene analysis and face recognition

PCA in side-channel terminology

- Useful for template attacks i.e. interesting points selection
- Used for new distinguishers (variance dependency)
- Reducing the dimensionality of data
- Learning about leakage model

PCA: finding projections

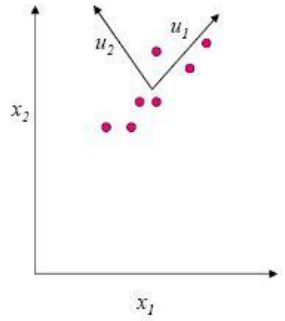
- Assume data is set of n-dimensional vectors, where m-th vector is: $x^m = (x_1^m, \cdots, x_n^m)$.
- We can represent these in any n-dim orthogonal basis vectors

$$x^m = \sum_{i=1}^{n} \alpha_i^m u_i, \quad u_i^T u_j = \delta_{ij}.$$

- PCA: for given $d < n$, find $(u_1, \ldots, u_d)$ that minimizes

$$E_d = \sum_{m=1}^{N} \left\| x^m - \hat{x}^m \right\|^2$$

$$\hat{x}^m = \frac{\sum_{i=1}^{N} x^m}{N} + \sum_{i=1}^{d} \alpha_i^m u_i.$$



PCA: finding projections to minimize the error

- Then:

$$E_d = \sum_{i=d+1}^{n} \sum_{m=1}^{N} \left[u_i^T \left(x^m - \bar{x} \right) \right]^2 = \sum_{i=d+1}^{n} u_i^T \sum u_i = \sum_{i=d+1}^{n} \lambda_i$$

$$\left(\hat{x}^m = \bar{x} + \sum_{i=1}^{d} \alpha_i^m u_i \right)$$

- E_d -> min for u_i eigen vectors of covariance matrix Σ, where:

$$\Sigma = (x_m - \bar{x})(x_m - \bar{x})^T$$

$$\Sigma u_i = \lambda_i u_i$$

PCA algorithm

1. Create a matrix of data Nxn (each row is one vector)
2. Subtract mean from all data points: $x_m - \bar{x}$
3. Compute covariance matrix Σ
4. Find eigenvectors and eigenvalues of Σ
5. Principal components are n eigenvectors with largest eigenvalues

$$\Sigma = U \Lambda U^{-1}$$

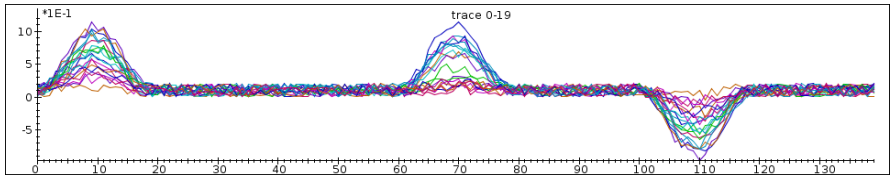
Previous works

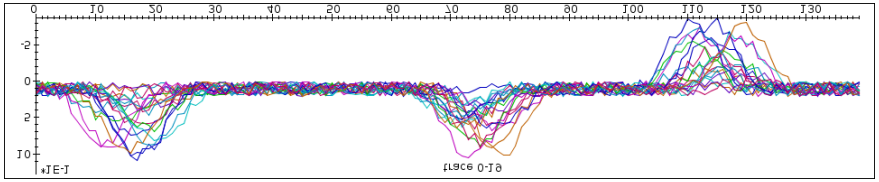
- [BN+03] first investigation
- [AP+06], [HG+11] PCA for templates
- [SN+10] PCA for key recovery
  - First Principal component as the side-channel distinguisher

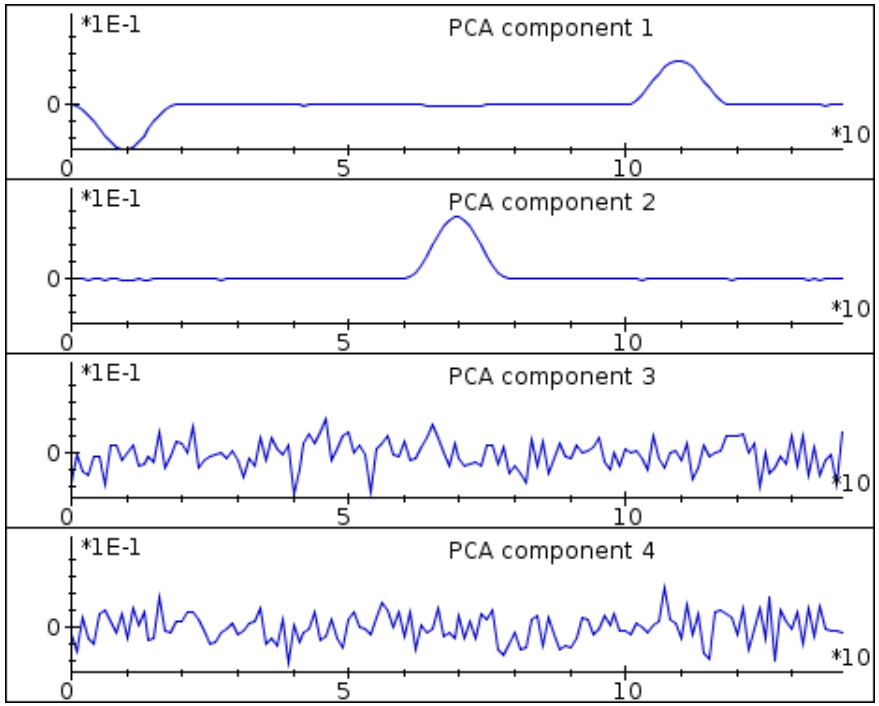
Multiple leakage points and PCA

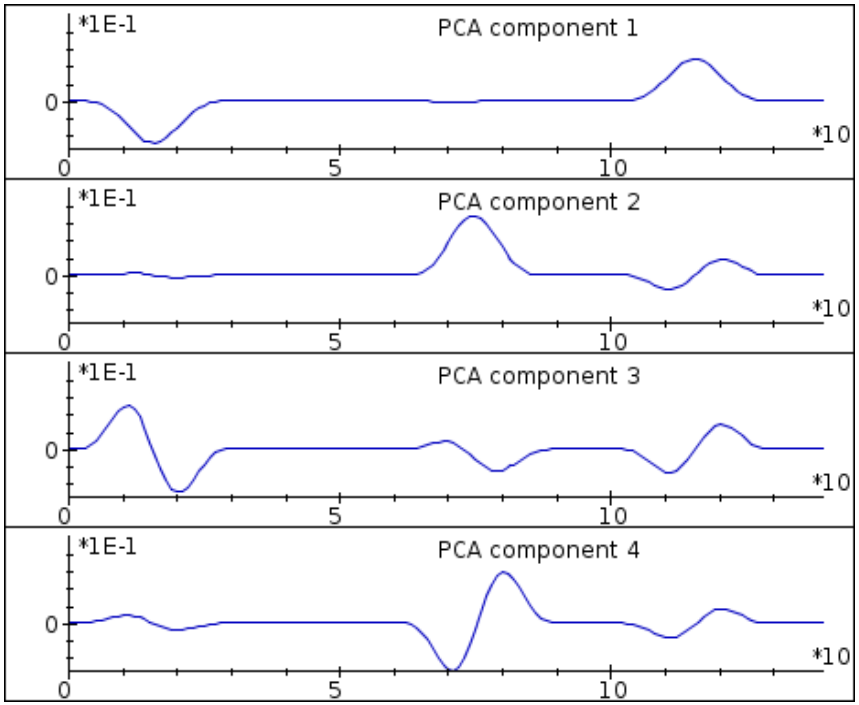
aligned

misaligned









PCA and noise reduction

- U is the feature vector

$$Y = U^T * X^T = (X * U)^T$$

- The PCA approximation with only p components

$$\hat{x} = \sum_{j=1}^{p} (u_j^T * x) * u_j$$

- The squared error

$$\sum_{i=p+1}^{n} \lambda_i$$

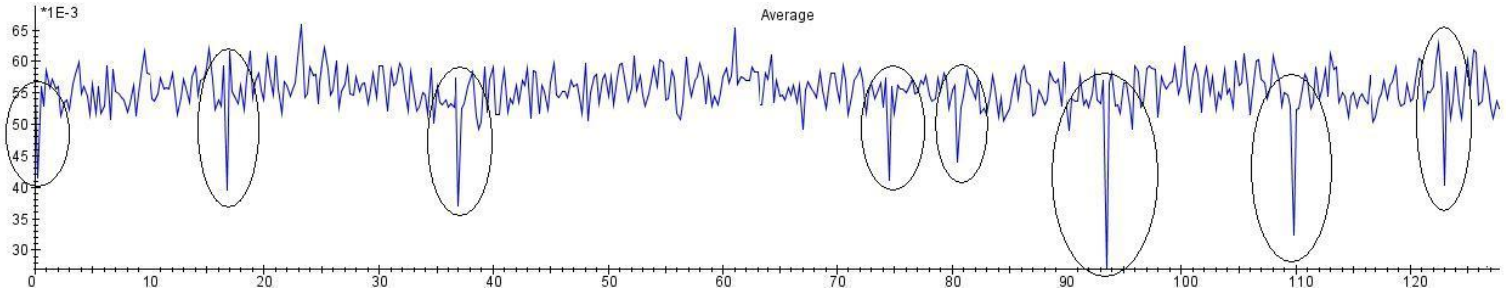
Experiments

Experiment 1: Noise reduction

- Key-related information shifts for various implementations and platforms
- Best strategy: PCA-> remove some PCs-> transform the data set back
- For a software implementation of DES
  - Different component keep different information related to specific S-boxes
  - In general, largest components contained a lot of (useless) noise
  - Templates could be made for a specific implementation/platform – which PCs should be kept

Experiment 2: PCA transformation

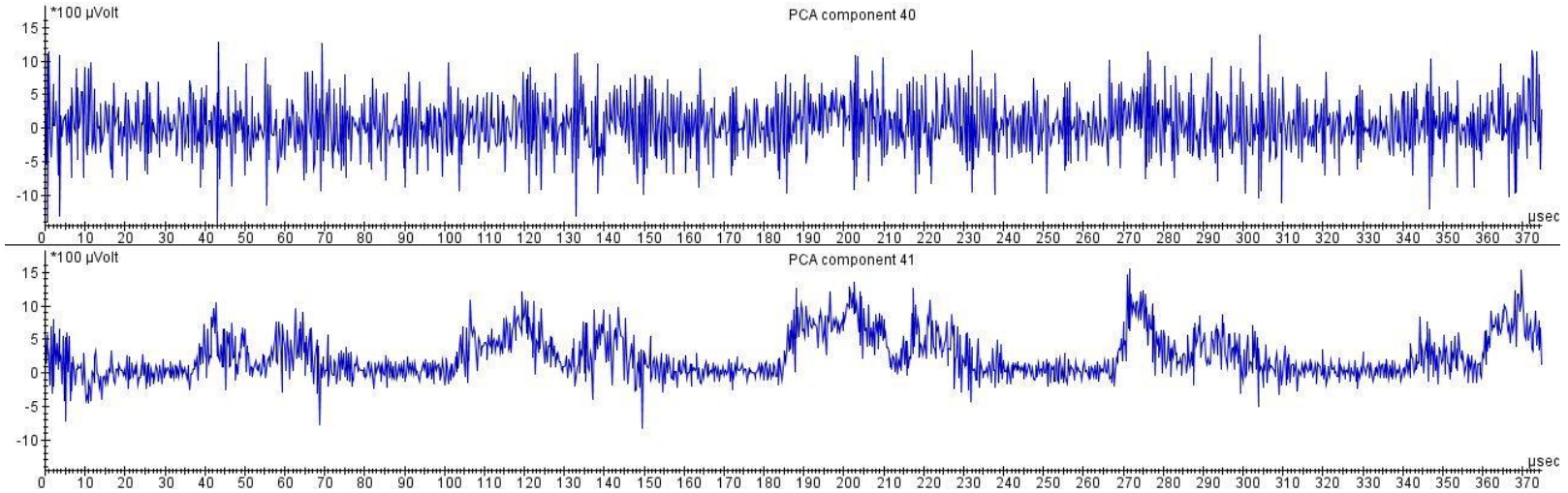
- CPA absolute average distinguisher – y-axis



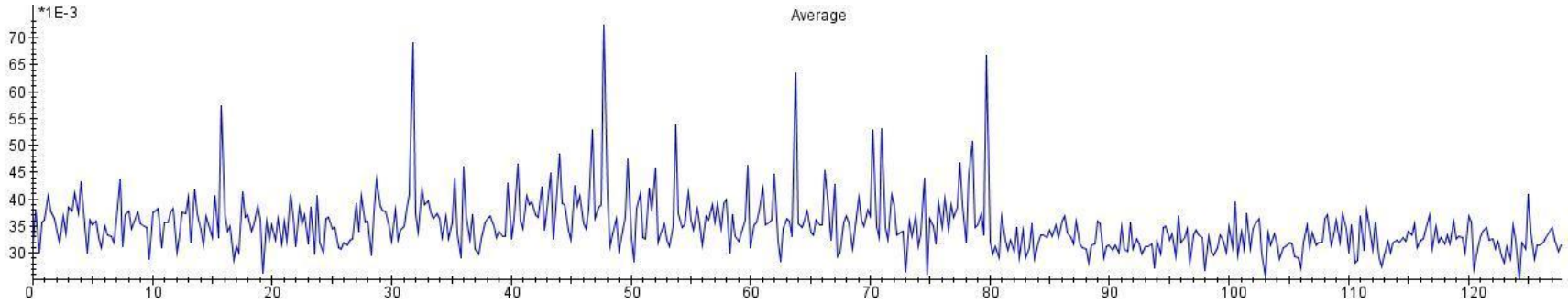
Key guesses for all S-boxes

Experiment 3: PCA and misalignment

- Specific PCs are more sensitive to exploitable leakage (e.g. 41-57)
- Absolute average distinguisher



Experiment 3: PCA and misalignment



Correct key guesses for 5 S-boxes

Comparison to other distinguishers and alignment techniques

- Computational issues:
  - Covariance matrix is typically very large (nxn)
  - Singular Value Decomposition (SVD)
- Comparison to static alignment
  - Due to the specifics of the PCA-based distinguisher, hard to compare with common methods

Conclusions

- PCA can be used for pre-processing
  - De-noising the data
  - Handling the countermeasures causing misalignment
- PCA for noise reduction: after removing noise-containing components, the key recovery improved
- PCA for key recovery: DPA on PCA-transformed data
- Our findings are confirmed on various platforms and algorithms

Some more reasoning + Future directions

- The linearity of PCA results in preserving leakage although spread over many time instances
- Kernel PCA: extracting also non-linear features
- Connection to clustering

Thanks

References:

- T. Mitchell: Machine learning, McGraw-Hill, 2003
- [BN+03] L. Bohy, M. Neve, D. Samyde, and J.-J. Quisquater. *Principal and independent component analysis for crypto-systems with hardware unmasked units*. In Proceedings of e-Smart 2003, 2003.
- [AP+06] C. Archambeau, E. Peeters, F.-X. Standaert, and J.-J. Quisquater, *Template attacks in principal subspaces*, In Cryptographic Hardware and Embedded Systems - CHES 2006, volume 4249 of *Lecture Notes in Computer Science*, pages 1–14, Springer, 2006
- [HG+11] G. Hospodar, B. Gierlichs, E. De Mulder, I. Verbauwhede, J. Vandewalle, *Machine learning in side-channel analysis: a first study*. J. Cryptographic Engineering (JCE) 1(4):293-302 (2011)
- [SN+10] Y. Souissi, M. Nassar, S. Guilley, J.-L. Danger, and F. Flament, *First principal components analysis: A new side channel distinguisher*, In K.-H. Rhee and D. Nyang, editors, Proceedings of ICISC'10, volume 6829 of LNCS, pages 407–419. Springer, 2010.