Understanding the Reasons for the Side-Channel Leakage is Indispensable for Secure Design

Werner Schindler

Federal Office for Information Security (BSI),
Bonn, Germany

Leuven, September 13, 2012

Outline

- Introduction and motivation
- Goals of a security evaluation
- The Stochastic Approach
  - basics in a nutshell
- How to obtain relevant design information
- Conclusions

- Side-channel analysis has been a hot topic in academia and industry for the last 15 years.
- In the early years the applied mathematical methods often wasted a lot of information.
- In the meanwhile the mathematical methods have become much more efficient.
- The time has been ripe for systematic methods!

How I came in touch with side-channel analysis (I)

- In 1999 I gave a course “Selected Topics in Modern Cryptography” at Darmstadt Technical University.
- I had to bridge a “gap” of one and a half 90 minute lectures. I remembered a timing attack from Jean-Jacques Quisquater and his research group (CARDIS 1998).
- I studied the paper and was quickly convinced that the attack could be improved significantly.

How I came in touch with side-channel analysis (II)

- I contacted Jean-Jacques and proposed a new decision strategy.
- For the same hardware the number of traces per attack dropped down from 200000 – 300000 to 5000, which is an increase of efficiency by factor ≈ 50 (Schindler, Koeune, Quisquater, 2001).
- New stochastic methods made this improvement possible.
- I thought it might be a good idea to write one paper on this topic…

Security evaluations (I)

- The resistance of smart cards, or more generally, of security implementations, against power attacks has been an important aspect of many security evaluations.
- It is very important for evaluators and designers to know the strongest attacks.
- Usually several side-channel attacks are applied (e.g. different DPA or CPA attacks). The target device is considered secure if it withstands all these attacks.

Federal Office for Information Security

Security evaluations (II)

- A successful attack shows that the device is vulnerable.
- But …
  - What are the consequences (countermeasures, limitation of the number of operations, re-design)?
- What is the conclusion if all attacks have been ineffective? Do stronger attacks exist?

Security evaluations (III)

- It is clearly desirable
  - to have reliable security evaluations
  - to get more than a one-bit information (successful attack is known / is not known).
- Reliable and trustworthy evaluation methods are needed!
- Ideally, a security evaluation should disclose potential weaknesses, allowing target-oriented re-design if necessary (constructive side-channel analysis).

DPA / CPA

- DPA and CPA are the „classics“ in power analysis.
- DPA and CPA are correlation attacks
  - + easy to apply, no profiling
  - - exploit only a fraction of the available information

Template attacks

- exploit power information from several time instants $t_1 < \ldots < t_m$
- electrical current vectors are interpreted as realizations of m-dimensional random vectors with unknown probability distribution.
- These random vector may depend on
  - (x,k): part of the plaintext / ciphertext x, subkey k
  - (x,z,k): part of the plaintext / ciphertext x, masking value z, and subkey k
  - f(x,k): e.g., $f(x,k) := ham(x \oplus k)$ (model-based templates)

Template attacks (II)

- profiling phase (training device):
  - estimation of a probability density for each (x,k), resp. for each (x,z,k), resp. for each f(x,k) (templates)
- attack (target device)
  - substitution of the measured current values into the templates (→ maximum likelihood principle)

A successful template attack shows that the target implementation is vulnerable but it does not explain how to fix the problem.

The stochastic approach

- target: block cipher
- exploits power measurements at several time instants $t_1 < t_2 < \ldots < t_m$
- The measurement values are interpreted as values that are assumed by random variables.
- The stochastic approach combines engineers' expertise with efficient stochastic methods from multivariate statistics.

Literature

- Pioneer work:

  Schindler, Lemke, Paar (2005),
- Theoretical foundations and attack efficiency:

  Schindler, Lemke, Paar (2005), Lemke, Gierlichs, Paar (2006), Lemke-Rust, Paar (2007), Schindler (2008), Standaert, Koeune, Schindler (2009), Heuser, Kasper, Schindler, Stöttinger (2012)
- Design aspects:

  Kasper, Schindler, Stöttinger (2010),
  Heuser, Kasper, Schindler, Stöttinger (2011 + 2012)

The stochastic model (basic variant)

target algorithm: block cipher (e.g., AES; no masking)

$x \in \{0,1\}^p$ (known) part of the plaintext or ciphertext

$k \in \{0,1\}^s$ subkey **[AES: (typically) s = 8]**

t time instant

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

$I_t(x,k)$: random variable (depends on x and k)

quantifies the random-ness of the side-channel signal at time t

$h_t(x,k)$: deterministic part = leakage function (depends on x and k)

R_t: random variable $E(R_t) = 0$

noise (centered)

The stochastic model (masking)

$x \in \{0,1\}^p$ (known) part of the plaintext or ciphertext

$z \in M$ masking value

$k \in \{0,1\}^s$ subkey **[AES: (typically) s = 8]**

$t \in \{t_1, t_2, \ldots, t_m\}$ time instant

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

random variable
(depends on x,z,k)

deterministic part
= leakage function
(depends on x,z,k)

random variable
$E(R_t) = 0$

quantifies the random-
ness of the side-channel
signal at time t

noise (centered)

Note

- The *leakage functions*
  $\mathbf{h_{t1}(\cdot\, , \cdot, \cdot,), h_{t2}(\cdot\, , \cdot\, , \cdot,), \ldots, h_{tm}(\cdot\, , \cdot, \cdot)}$

  and

- the probability distribution of the random vector
  $\mathbf{(R_{t1}, R_{t2}, \ldots, R_{tm})}$ („noise vector“)

  are <u>unknown</u> and have to be estimated with a training device.

Federal Office for Information Security

Profiling, Step 1 (I)

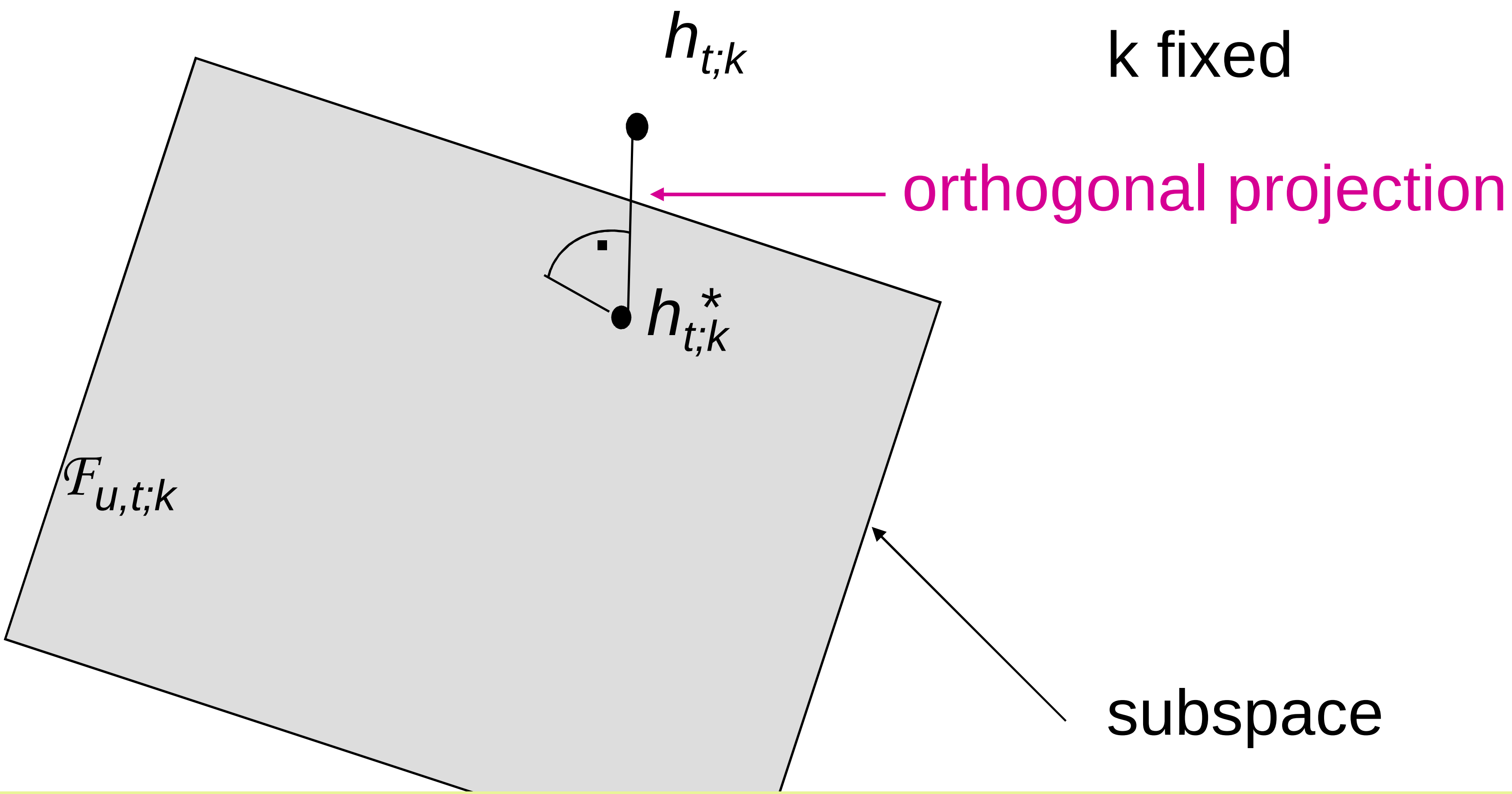
- Fix a subkey $k \in \{0,1\}^s$.
- The unknown function

  $$h_{t;k}: \in \{0,1\}^p \times M \times \{k\} \to R,\ h_{t;k}(x,z;k) := h_t(x,z;k)$$

  is interpreted as an element of a high-dimensional real vector space $\mathcal{F}_k$. In particular, $\dim(\mathcal{F}_k) = 2^p\,|M|$.
- <u>Goal:</u> Approximate $h_{t;k}$ by its image $h^*_{t;k}$ under the orthogonal projection onto a suitably selected low-dimensional vector subspace $\mathcal{F}_{u,t;k}$

Federal Office for Information Security

Geometric illustration



The image $h^*_{t,k}$ is the best approximator of $h_{t;k}$ in $\mathcal{F}_{u,t;k}$

Profiling, Step 1 (II)

$$\mathcal{F}_{u,t;k} := \{h' : \{0,1\}^p \times M \times \{\mathrm{k}\} \to R \mid \sum_{j=0}^{u-1} \beta'_{j,t;k}\, g_{j,t;k} \text{ with } \beta'_{j,t;k} \in \mathrm{R}\}$$

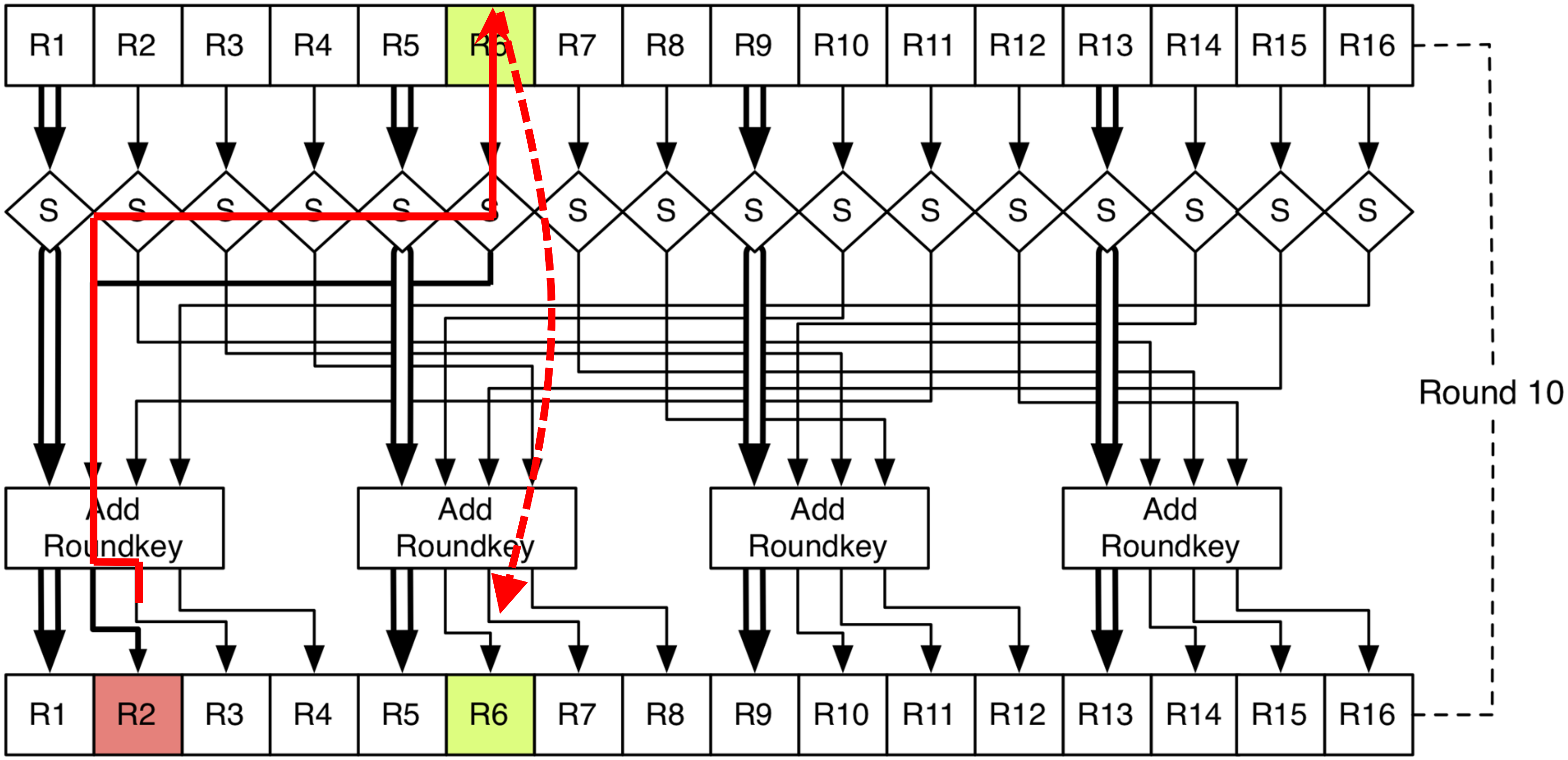
(masking case)

with basis functions $g_{j,t;k} : \{0,1\}^p \times M \times \{k\} \to R$

The basis $g_{0,t;k},\ldots,g_{u-1,t;k}$ shall be selected under consideration of the attacked device.

The estimation of $h^*_{t,k}$ can completely be moved to the low-dimensional subspace $\mathcal{F}_{u,t;k}$, which reduces the number of measurements to a small fraction.

Example: AES implementation on an FPGA (final round)



„Difference“ in register R6: R6 (new) ⊕ R6 (old)

AES implementation on an FPGA (I)

Target: Key byte $k_{(2)} \in \{0,1\}^8$ in round 10

$R_{(x)}$ value of register x after round 10

9-dimensional subspace:

$$g_{0,t;k(2)}((R_{(2)},R_{(6)}),k_{(2)}) = 1$$

$$g_{j,t;k(2)}((R_{(2)},R_{(6)}),k_{(2)}) = (R_{(6)} \oplus S^{-1}(R_{(2)} \oplus k_{(2)}))_j$$

for $1 \leq j \leq 8$

AES implementation on an FPGA (II)

Target: Key byte $k_{(2)} \in \{0,1\}^8$ in round 10

$R_{(x)}$ value of register x after round 10

2-dimensional subspace:

$g_{0,t;k(2)}((R_{(2)},R_{(6)}),k_{(2)}) = 1$

$g'_{1,t;k(2)}((R_{(2)},R_{(6)}),k_{(2)}) = ham(R_{(6)} \oplus S^{-1}(R_{(2)} \oplus k_{(2)}))$

This 2-dimensional subspace potentially contains less leakage information than the 9-dimensional subspace defined on the previous slide.

Profiling, Step 1 (I)

$$h^*_{t;k} = \sum_{j=0}^{u-1} \beta^*_{j,t;k} g_{j,t;k}$$ (best approximator of $h_{t;k}$ in $\mathcal{F}_{u,t;k}$)

- Task: Estimate the unknown coefficients $\beta^*_{0,t;k}, \ldots, \beta^*_{(u-1),t;k}$
- N_1 measurement values from the training device $i_t(x_1, z_1, k), \ldots i_t(x_{N\_1}, z_{N\_1}, k)$
- Least-square estimation:

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

Profiling, Step 2 (only relevant for attacks)

$(I_{t\_1}(x,z,k) - h^*_{t\_1;k}(x,z,k), \ldots , I_{t\_m}(x,z,k) - h^*_{t\_m}(x,z,k)) \approx$

$(I_{t\_1}(x,z,k) - h_{t\_1}(x,z,k), \ldots , I_{t\_m}(x,z,k) - h_{t\_m}(x,z,k)) =$

$(R_{t\_1}, \ldots , R_{t\_m}) \sim N(0,C)$

- Estimate the covariance matrix C (multivariate normal distribution), possibly with PCA

- $\rightarrow$ prob. density $f_{x,z;k}(\cdot)$ for $I_t(x,z,k)$

Federal Office for Information Security

Attack phase
(only relevant for attacks)

- Perform N_3 measurements on the target device
- Apply the maximum likelihood principle
  (analogous to template attacks)

NOTE: The random vector $I_t(x,Z,k)$
(unknown masking value) has density

$$\sum_{z' \in M} \mathrm{Prob}(Z = z')\, f_{x,z';k}(\cdot)$$

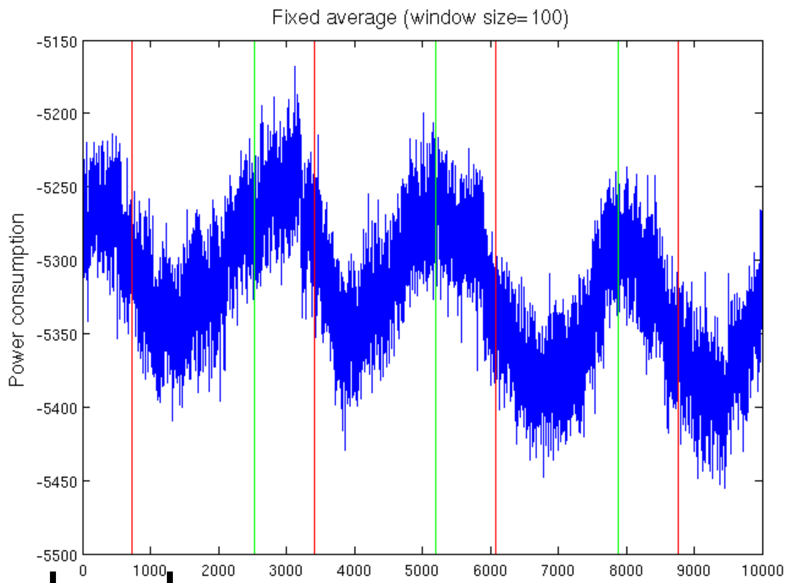
Be careful !

- Within long measurement series the environmental conditions might change, influencing the power consumption and thereby violating the (silent) assumption of having identical conditions all the time.

- Example:
  dpa-v2 power traces

0:00 am

+24h



(time-local average power consumption)

Drifting offset

- The average electrical current shows a periodic drift (← variation of the temperature in the lab).
- This drift in particular influences the data-independent coefficient.
- All profiling-based attacks suffer from this problem.

Stochastic approach – the OTM method

- exhanced stochastic model

  $I_t(x_v,k) = h_t(x_v,k) + \theta_v + R_t$

  drifting offset

- <u>Observation:</u> $\theta_{v+1} - \theta_v \approx 0$

- <u>Solution:</u> Consider overlapping differences
  $I_t(x_{v+1},k) - I_t(x_v,k) \approx N(h_{t;k}(x_{v+1},k) - h_{t;k}(x_v,k), 2C)$
  - use subspaces $\mathcal{F}^\circ_{u,t;k}$ <u>without</u> $g_{0,t;k} = 1$
  - additional mathematical problems
    but clear increase of efficiency

Stochastic approach: profiling workload

- Phase 1: 2^s (= # subkeys) measurement series; may reduce to 1 measurement series in case of symmetry (→ later)
- Phase 2: 1 measurement serie
- no additional steps in case of masking

Stochastic approach: attack efficiency

- The attack efficiency depends on the choice of the subspace.
- For suitable subspaces the attack efficiency should be close to (full) template attacks
- more efficient than DPA and CPA

Representation of the leakage

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

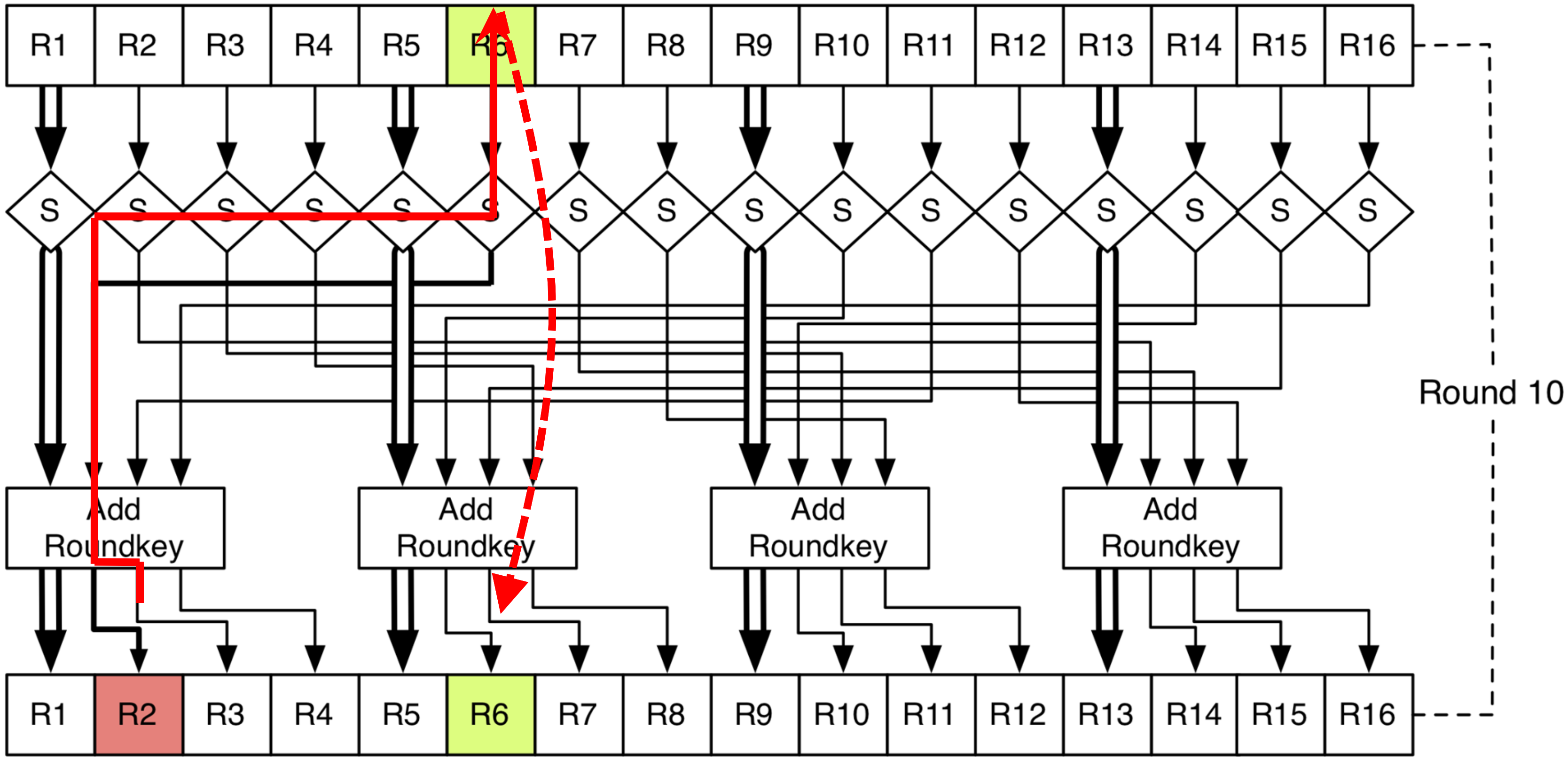
If $|\beta^{*}_{j,t;k}|$ is ‘large’ the ‘direction’ of the basis vector $g_{j,t;k}$ has significant impact on the data-dependent part of the leakage $h_{t;k}$.

Federal Office
for Information Security

Note

- To obtain design information only the first profiling phase is relevant (estimation of $h^*_{t,k}(\cdot,\cdot)$).
- These following results were obtained together with Annelie Heuser, Michael Kasper and Marc Stöttinger from my research group CASCADE at CASED (within the research project RESIST).
- For our experiments we used the SASEBO G-I evaluation board (with Virtex-II pro FPGA) and the SASEBO G-II evaluation board (with Spartan V FPGA).

Example: AES implementation on an FPGA (final round)



„Difference“ in register R6: R6 (new) ⊕ R6 (old)

Reminder: AES implementation on an FPGA

Target: Key byte $k_{(2)} \in \{0,1\}^8$ in round 10

$R_{(x)}$ value of register x after round 10

9-dimensional subspace:

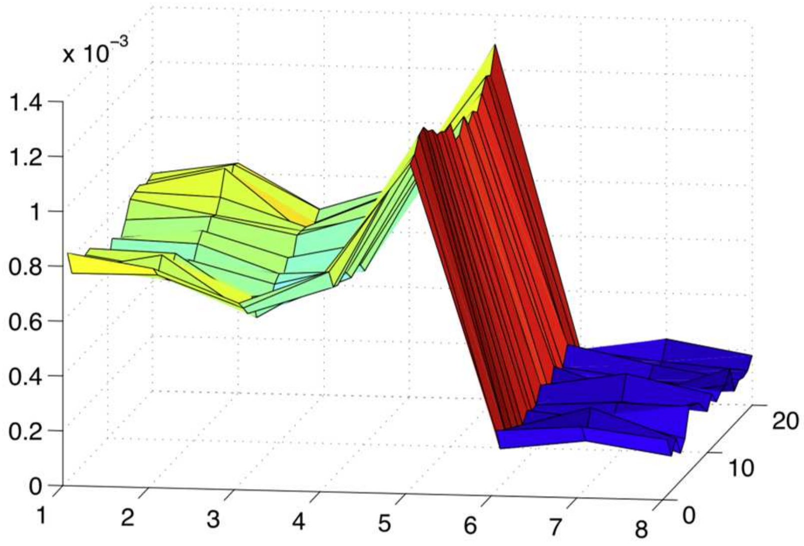
$$g_{0,t;k(2)}\ ((R_{(2)},R_{(6)}),k_{(2)}) = 1$$

$$g_{j,t;k(2)}\ ((R_{(2)},R_{(6)}),k_{(2)}) = (R_{(6)} \oplus S^{-1}(R_{(2)} \oplus k_{(2)}))_j - 0.5$$

for $1 \leq j \leq 8$

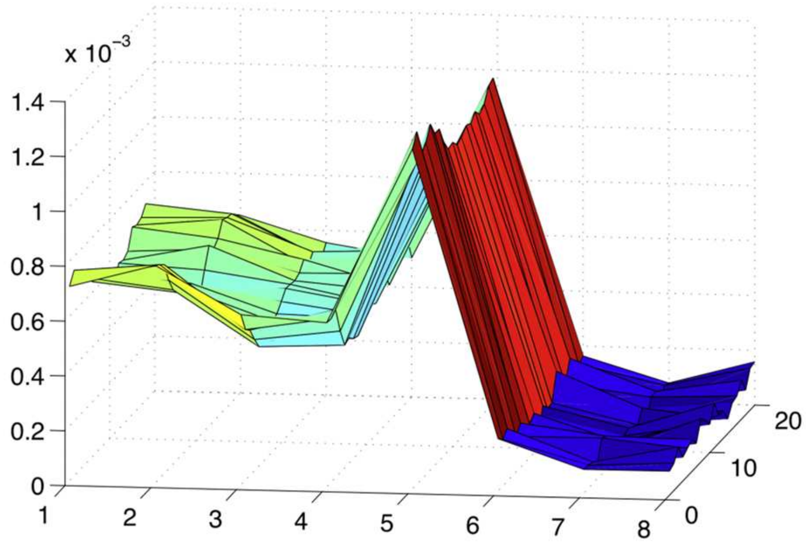
The term ‘ – 0.5 ‘ ensures that the basis vectors are centered (i.e. $E(g_{j,t;k(2)}) = 0$) for j>0, and $\beta_{0,t;k} = E(I_t(\cdot))$

β-Characteristic for an S-Box Design (FPGA, TBL)



AES TBL, $k_{(1)} = 19$:

$|\beta_1|,\ldots,|\beta_8|$ für $t_1,\ldots,t_{20}$

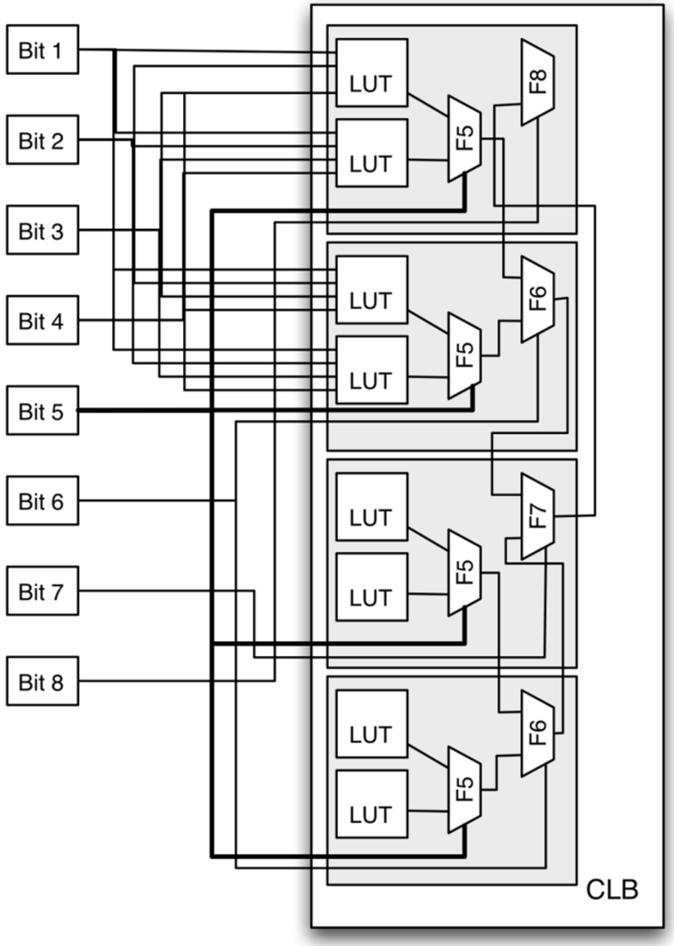


AES TBL, $k_{(1)} = 209$:

$|\beta_1|,\ldots,|\beta_8|$ für $t_1,\ldots,t_{20}$

$|\beta_5|$ is exceptionally large! Why?

A closer look at the implementation



- Part of the SBox after the synthesis process and the place & route process (Virtex-II pro family)
- The first layer of the multiplexer network is switched by the 5th bit
- Different propagation delays caused by LUT to the multiplexer produces data-dependent glitches.
- This implies bit-specific higher power consumption.

High-dimensional subspaces

Example: Attack on the key byte $k_{(2)}$

$\mathcal{B}_0 := \{g_{0,t;k(2)} = 1\}$

$g'_{j,t;k(2)}((R_{(2)},R_{(6)}),k_{(2)}) := (R_{(6)} \oplus S^{-1}(R_{(2)} \oplus k_{(2)}))_j$ for $1 \leq j \leq 8$

$\mathcal{B}_1 := \{g'_{j,t;k(2)} - 0.5 \mid 1 \leq j \leq 8\}$

High-dimensional subspaces

$$\mathcal{B}_i := \{g'_{j\_1,t;k(2)} \cdots g'_{j\_i,t;k(2)} - (0.5)^i \mid 1 \leq j_1 < \ldots < j_i \leq 8\}$$

Unordered i-fold products
(catches the interaction between up to i bit lines)

<u>Example:</u> $g'_{3,t;k(2)} \cdot g'_{7,t;k(2)} - 0.25 \in \mathcal{B}_2$

(catches the interaction between the bit lines 3 and 7)

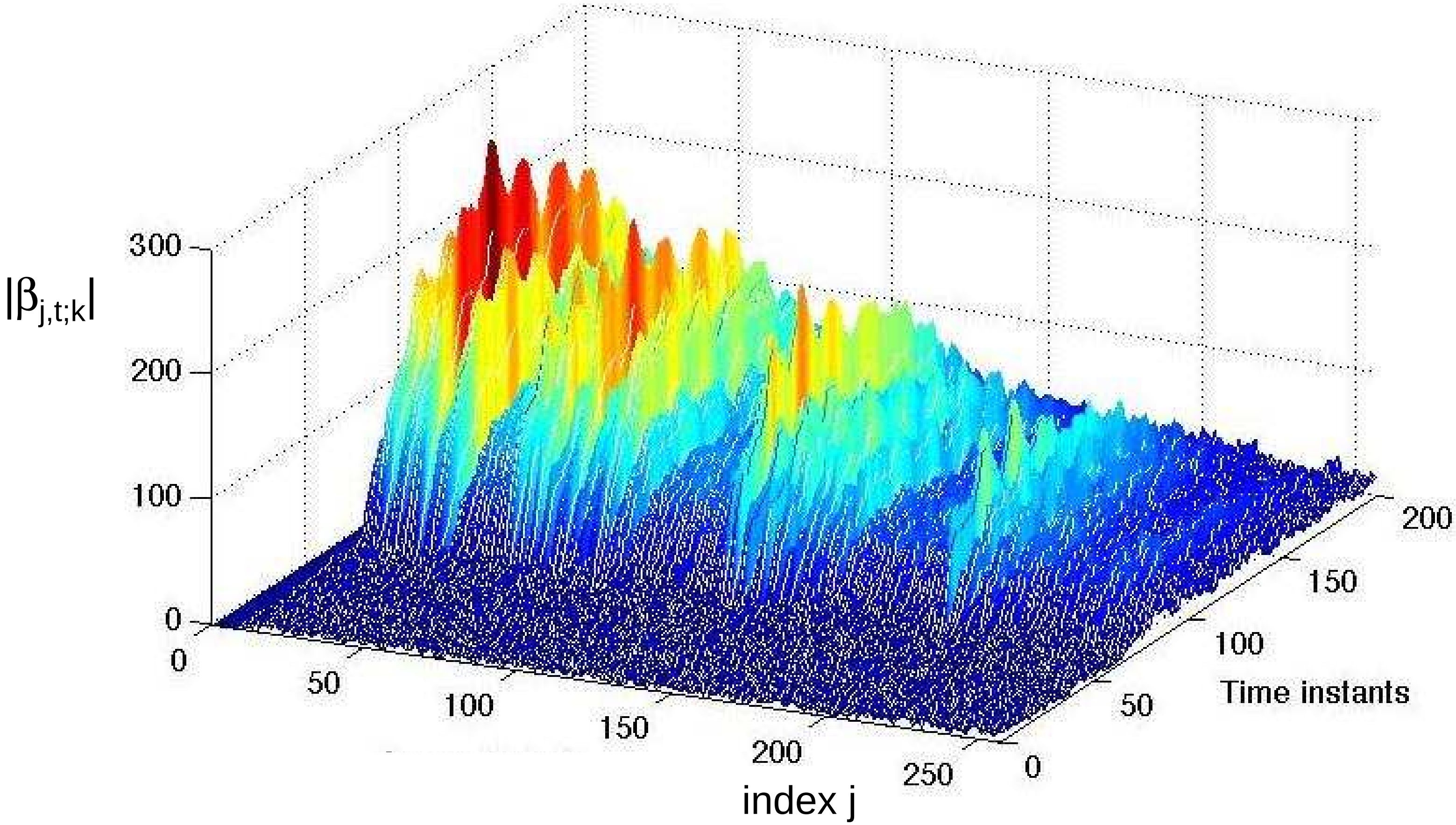
High-dimensional subspaces (OTM)

- The subspaces $\mathcal{F}^{\circ}_{u,t;k}$ are spanned by the following basis vectors
  - $\mathcal{B}_1$ (dim = 8)
  - $\mathcal{B}_1 \cup \mathcal{B}_2$ (dim = 36)
  - $\mathcal{B}_1 \cup \mathcal{B}_2 \cup \mathcal{B}_3$ (dim = 92)
  - $\mathcal{B}_1 \cup \mathcal{B}_2 \cup \mathcal{B}_3 \cup \mathcal{B}_4$ (dim = 162)
  - $\mathcal{B}_1 \cup \mathcal{B}_2 \cup \mathcal{B}_3 \cup \mathcal{B}_4 \cup \mathcal{B}_5$ (dim = 218)
  - $\mathcal{B}_1 \cup \mathcal{B}_2 \cup \mathcal{B}_3 \cup \mathcal{B}_4 \cup \mathcal{B}_5 \cup \mathcal{B}_6$ (dim = 246)
  - $\mathcal{B}_1 \cup \mathcal{B}_2 \cup \mathcal{B}_3 \cup \mathcal{B}_4 \cup \mathcal{B}_5 \cup \mathcal{B}_6 \cup \mathcal{B}_7$ (dim = 254)
  - $\mathcal{B}_1 \cup \mathcal{B}_2 \cup \mathcal{B}_3 \cup \mathcal{B}_4 \cup \mathcal{B}_5 \cup \mathcal{B}_6 \cup \mathcal{B}_7 \cup \mathcal{B}_8$ (dim = 255)

For the ‘standard method’ ‘$\mathcal{B}_0$’ is added to these bases, which increases the dimension by 1.

β- coefficients (256-dimensional subspace)

AES, last round, S-Box, COMP



Impact on the attack efficiency

- DPA contest v2: also SASEBO-G-II board with Spartan V - FPGA, S-box design: COMP

Federal Office
for Information Security

DPA-contest v2 / OTM method / public base

dim ($\mathcal{F}^{\circ}_{u,t;k}$)	PSR > 80 %	GSR > 80 %
8	8781	13020
36	5876	7533
92	5159	6734
162	4353	6144
218 (up to 5-fold products)	3552	4564
246	3769	4691
254	3720	4740
255	3718	4748
255 (with vertical trace alignment)	2682	3836

Research group

CASCADE

Observation

- Even some 5-fold products have significant contribution to the leakage.
- Crossover effects between neighboured bit lines cannot be the (only) reason.
- What is the reason for this behaviour? Glitches due to different time delays? (open question)
- Do other designs of the S-Box show qualitatively different results (maybe only significant contributions up to 3-fold products exist)? (open question)

Suitability of the leakage model

- High-dimensional subspaces $\mathcal{F}_{u,t;k}$ may provide more precise leakage models.
- An important question remains: Is the choice of the basis vectors appropriate?

Symmetries (I)

The basis vectors from our example

$$g_{j,t;k(2)}((R_{(2)},R_{(6)}),k_{(2)}) = (R_{(6)} \oplus S^{-1}(R_{(2)} \oplus k_{(2)}))_j - 0.5$$

depend only on

$$\varphi(R(2),R(6),k(2)) := R(6) \oplus S^{-1}(R(2) \oplus k(2))$$

('symmetry')

Symmetries (II)

- This reduces the argument of the leakage function from 24 bit to 8 bit …
- … and the dimension of the relevant (large) vector space from 2^{24} to 2^8.
- If the symmetry assumption (expressed by φ) is valid then for each j

$$\beta^*_{j,t;k'} = \beta^*_{j,t;k''} \qquad \text{for all } k',k'' \in \{0,1\}^s$$

Consequences

- In case of a (perfect) symmetry φ it suffices to estimate $h^*_{t,k}$ for any single subkey k.
- Any power curve related to some subkey k‘ can be ‚converted‘ into a power curve related to k
  → all power traces can be used for a single estimation process

Verification of a symmetry assumption (I)

- Any symmetry assumption influences the choice of the basis vectors.
- The suitability of the basis is very important for both attack and for getting useful design information.
- How can a symmetry assumption be verified?

Verification of a symmetry assumption (II)

- Crucial property: If the symmetry assumption is valid
  $\beta^*_{j,t;k'} = \beta^*_{j,t;k''}$ for all $k',k'' \in \{0,1\}^s$

- 1^{st} approach:
  Estimate the β- coefficients for several subkeys $k_1,k_2,..,k_v$
  - If the β- estimates are 'almost' equal:
    → confirmation of the symmetry assumption
  - If the β- estimates are very unequal:
    → rejection of the symmetry assumption

Federal Office for Information Security

Symmetry distance

For subkeys k' and k'' the ratio

$$\frac{2\sqrt{\sum_{j>0}\left(\beta_{j,t;k'}-\beta_{j,t;k''}\right)^2}}{\sqrt{\sum_{j>0}\beta_{j,t;k'}{}^2}+\sqrt{\sum_{j>0}\beta_{j,t;k''}{}^2}} \qquad (**)$$

quantifies the distance of their β-coefficients.

If the symmetry assumption is valid this term equals 0.

Symmetry distance (II)

This symmetry metric is invariant

- under the multiplication of the leakage function by positive scalars
- under all orthonormal bases of $\mathcal{F}_{u,t;k}$ with $g_{0,t;k}=1$

Action: Use a orthonormal basis and substitute the β-estimates into formula (**)

Leakage model $\mathcal{B}$ (distance model)

9-dimensional vector space (orthonormal basis)

$g_{0,t;k(2)}\ ((R_{(2)},R_{(6)}),k_{(2)}) = 1$

$g_{j,t;k(2)}\ ((R_{(2)},R_{(6)}),k_{(2)}) = 2((R_{(6)} \oplus S^{-1}(R_{(2)} \oplus k_{(2)}))_j - 0.5)$

for $1 \leq j \leq 8$

Here: $\varphi((R_{(2)},R_{(6)}),k_{(2)}) := R_{(6)} \oplus S^{-1}(R_{(2)} \oplus k_{(2)})$

(symmetry assumption $\mathcal{B}$)

This symmetry property transfers to

$h^*_{t,k(2)}\ ((R_{(2)},R_{(6)}),k_{(2)})$ and $\tilde{h}^*_{t,k(2)}\ ((R_{(2)},R_{(6)}),k_{(2)})$

Alternate leakage model $\mathcal{A}$ (weight model)

The basis vectors

9-dimensional vector space (orthonormal basis)

$g'_{0,t;k(2)}((R_{(2)},R_{(6)}),k_{(2)}) = 1$

$g'_{j,t;k(2)}((R_{(2)},R_{(6)}),k_{(2)}) = 2\,((S^{-1}(R_{(2)} \oplus k_{(2)}))_j - 0.5)$

for $1 \leq j \leq 8$

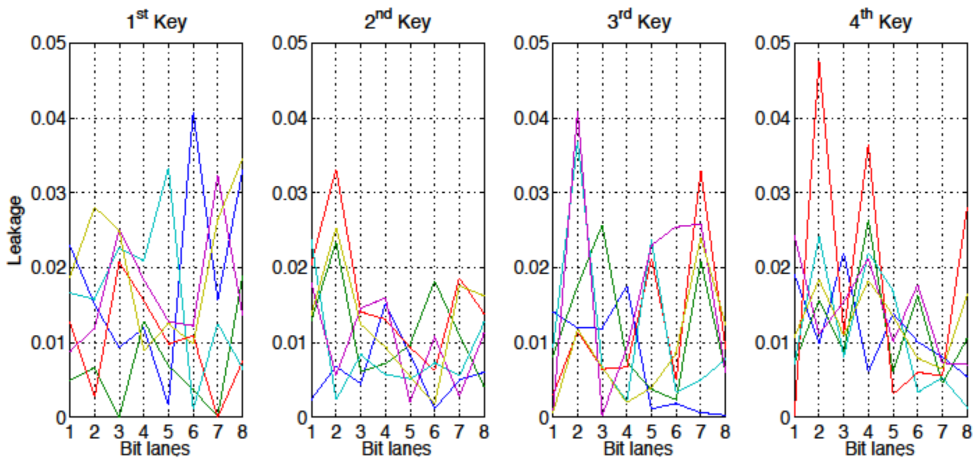
depend on $((R_{(2)},R_{(6)}),k_{(2)})$ only through

$\varphi_A((R_{(2)},R_{(6)}),k_{(2)}) := S^{-1}(R_{(2)} \oplus k_{(2)})$

(alternate symmetry assumption $\mathcal{A}$)

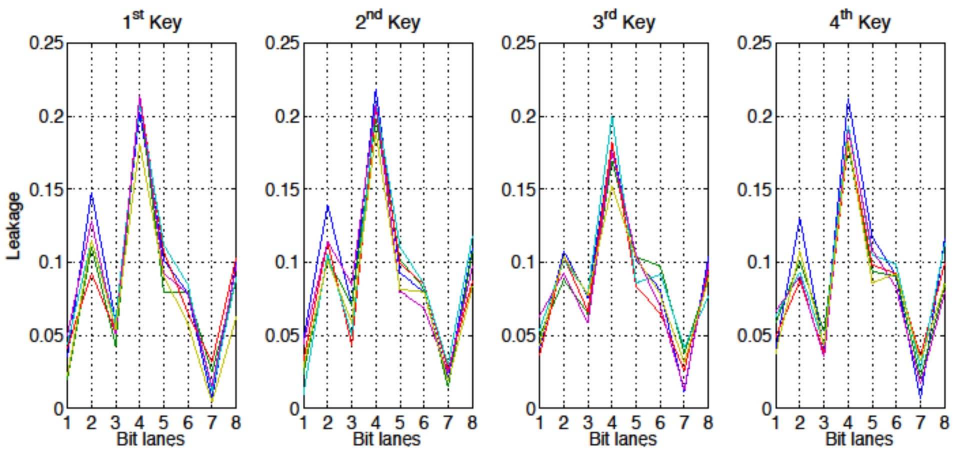
Federal Office
for Information Security

Comparison of β-coefficients



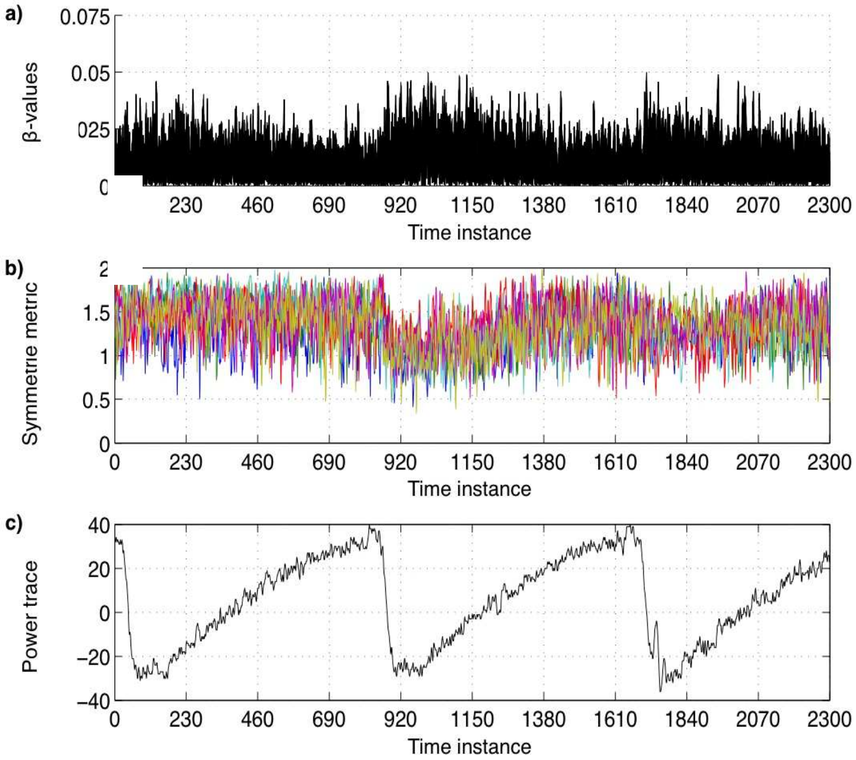
Leakage model $\mathcal{A}$

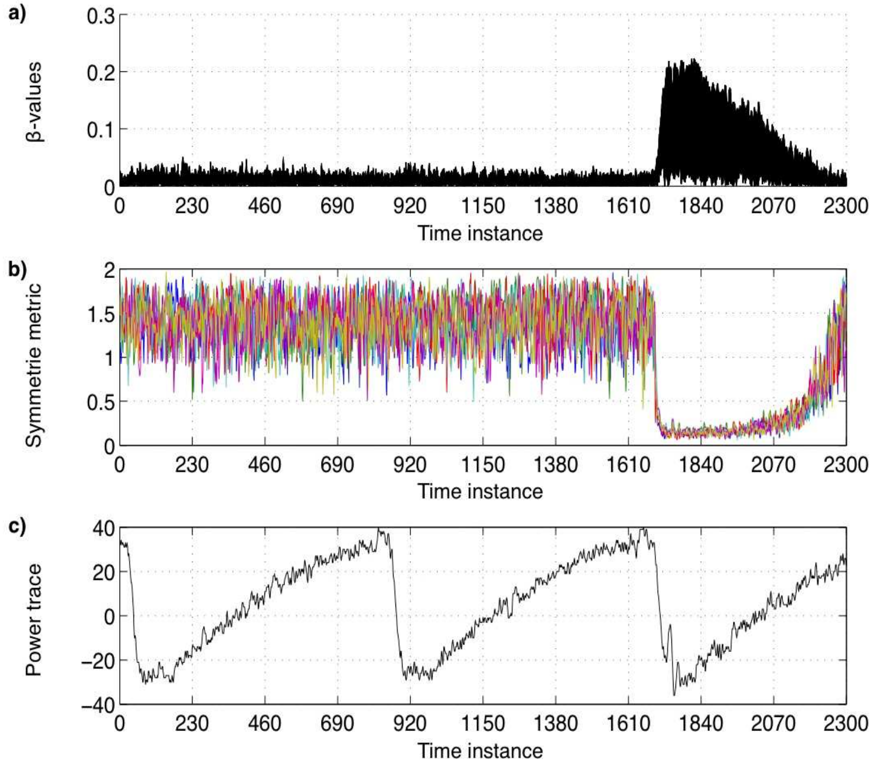
Equal colours refer to identical time instants



Leakage model $\mathcal{B}$

Experimental Results





Round 10

leakage model $\mathcal{A}$ leakage model $\mathcal{B}$

Further aspects

The stochastic approach can also be used to estimate

- $E_X((h_{t;k}(X,k) – h^*_{t;k}(X,k))^2)$,
  (This L^2-distance quantifies the approximation error of $h^*_{t;k}(\cdot,k)$.)
- the signal-to-noise ratio

Details: Heuser, Schindler, Stöttinger (DATE 2012)

Masking

- Masked implementations can be handled similarly if the masking values are known. (Profiling with unknown masking values is also possible but less efficient.)
- Additionally, it might be necessary to rate the effect of masking (e.g. by the estimation of L^1-distances of probability distributions).

Conclusion

The stochastic approach

- is an efficient attack tool
- provides a representation of the leakage with regard to a vector basis

The stochastic approach can also be used to

- identify and quantify properties / weaknesses, which (might) be relevant for the leakage
- to verify or falsify leakage models (within the limits of statistics)
- to support target-oriented (re-)design (constructive side-channel analysis)

Contact

Federal Office for Information Security (BSI)

Werner Schindler
Godesberger Allee 185-189
53175 Bonn, Germany

Tel: +49 (0)228-9582-5652
Fax: +49 (0)228-10-9582-5652

Werner.Schindler@bsi.bund.de
www.bsi.bund.de
www.bsi-fuer-buerger.de