Federal Office
for Information Security

Fraunhofer
SIT

Optimal CPA Attack

Michael Kasper[1], Werner Schindler[2]

1: Fraunhofer SIT, Darmstadt, Germany

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

Annecy, June 30, 2014

Outline of the talk

- Motivation
- Relation between CPA and the stochastic approach
- The optimal CPA function
- Experimental results
- Conclusion

Motivation

- Dozens of papers on CPA attacks have been published. Typically, CPA attacks focus on a single bit or apply a Hamming weight (distance) model.
- Natural questions:

  For a given implementation:
  - What is the most efficient CPA function?
  - What is the (theoretical) limit of a univariate CPA attack?

What's new?

- This talk considers univariate CPA attacks on unprotected block ciphers.
- We develop an explicit expression for the optimal CPA function (in dependency of the leakage distribution) for leakage models with given complexity.

Reference:

[1] Michael Kasper, Werner Schindler: Towards Optimal Correlation Analysis Attacks by High-Dimensional Stochastic Models. Status: Submitted.

Definition

$x \in \{0,1\}^p$ set of admissible (known) parts of the plaintext or ciphertext
[AES: typically, p = 8 or p = 16]

$k' \in \{0,1\}^s$ set of admissible subkeys
[AES: typically, s = 8]

k correct subkey

$f_t : \{0,1\}^p \times \{0,1\}^s \rightarrow R$ CPA function at time t;
[e.g., Hamming weight of an Sbox output]

$$\rho_N(i_t, f_t, k') = \frac{\frac{1}{N}\sum_j i_t(x_j, k) f_t(x_j, k') - mean(i_t)mean(f_t)}{\sqrt{empVar(i_t)}\sqrt{empVar(f_t)}}$$

CPA

- We interpret $x_1, x_2, \ldots, x_N$ as realizations of iid (independent and identically distributed) random variables $X_1, X_2, \ldots, X_N$.
- As N tends to infinity the term $\rho_N(i_t, f_t, k')$ converges to the correlation coefficient

$$corr(I_t(X,k), f_t(X,k')) = \frac{cov(I_t(X,k), f_t(X,k'))}{\sqrt{Var(I_t(X,k))}\sqrt{Var_X(f_t(X,k'))}}$$

- The CPA function $f_{t,2}$ is viewed better than $f_{t,1}$, if the absolute value $|corr\left(I_t(X,k), f_{t,2}(X,k)\right)|$ is larger than $|corr\left(I_t(X,k), f_{t,1}(X,k)\right)|$.

The stochastic model

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

Random variable
(depends on x and k)

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

deterministic part
= leakage function
(depends on x and k)

Random variable

$E(R_t) = 0$

Noise (centered)

Stochastic approach: Profiling, Step 1

- For fixed subkey $k \in \{0,1\}^s$ the unknown function

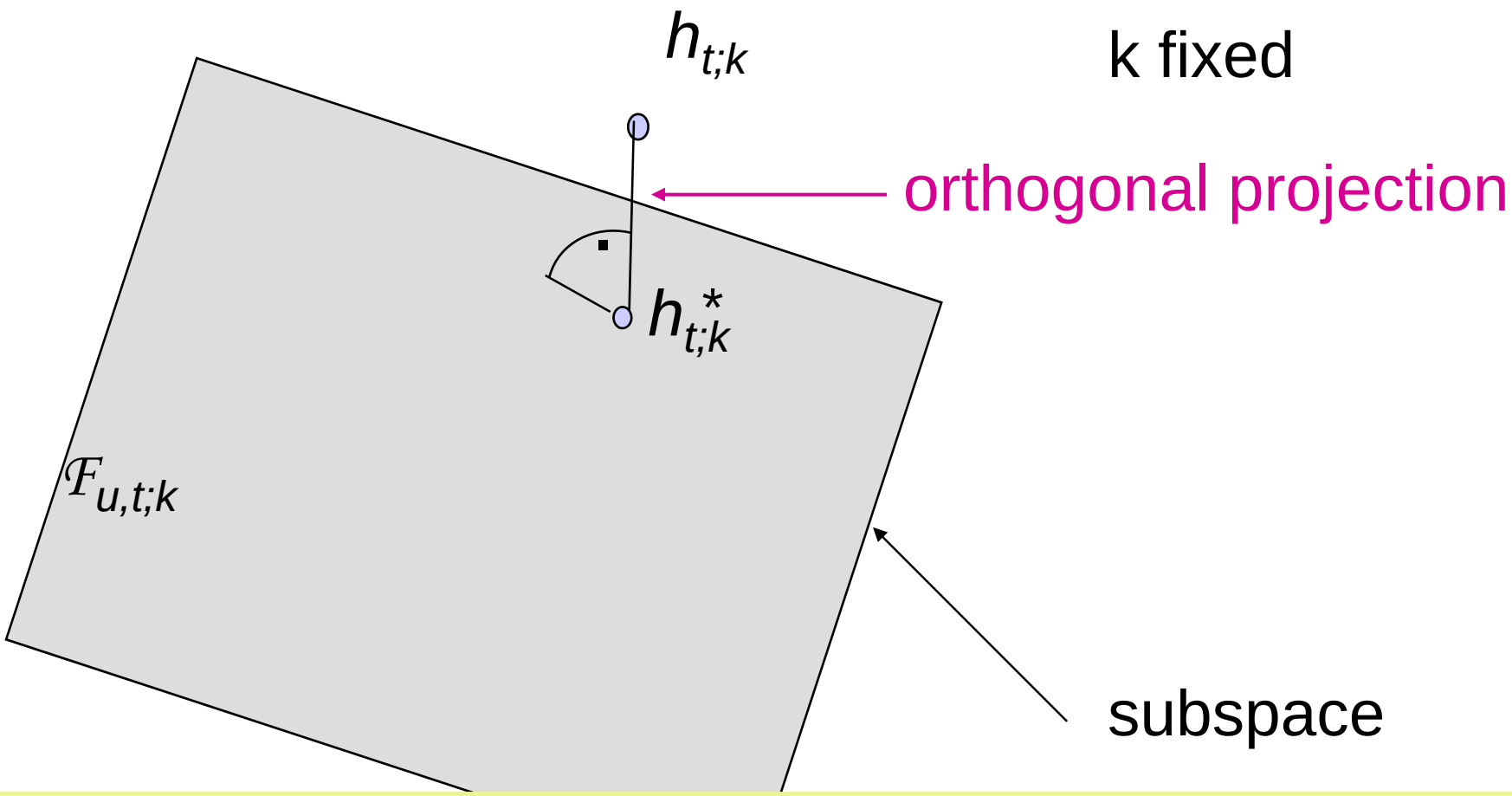
$$h_{t;k}: \in \{0,1\}^p \times \{k\} \to R, \quad h_{t;k}(x;k) := h_t(x;k)$$

is interpreted as an element of the real vector space

$$\mathcal{F}_k := \{h': \{0,1\}^p \times \{k\} \to R\} \qquad [\dim(\mathcal{F}_k) = 2^p]$$

- Stochastic approach: Approximate the leakage function $h_{t;k}$ by its image $h^*_{t;k}$ under the orthogonal projection onto a suitably selected low-dimensional vector subspace $\mathcal{F}_{u,t;k}$.

Geometric illustration



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

A scalar product on $\mathcal{F}_k$

- Assume that Prob(X=x) > 0 for all $\{0,1\}^p$. (Otherwise cancel those elements that occur with probability 0.)
- Then

  $(f_1,f_2) := \sum_{x\in\{0,1\}^p} Prob(X=x) f_1(x,k) f_2(x,k) = E(f_1 f_2)$

  defines a scalar product on $\mathcal{F}_k$
- Here and in the following $g_{0,t;k} = 1$, $g_{1,t;k},\ldots,g_{2^p-1,t;k}$ denotes an orthonormal basis of $\mathcal{F}_k$ w.r.t $(\cdot,\cdot)$.
- In particular,

$$(h_{t;k}, g_{i,t;k}) = \left(\sum_{j=0}^{2^p-1} \beta_{j,t;k} g_{j,t;k}, g_{i,t;k}\right) = \beta_{i,t;k}$$

CPA (from a more theoretical standpoint)

- Assume that $E(f_t(X,k)) = 0$. Then

$$\mathrm{corr}(l_t(X,k),f_t(X,k)) = \ldots = \mathrm{const}\cdot(h_{t;k}(X,k), \frac{f_t(X,k)}{\|f_t(X,k)\|}),$$

- CPA with the function $f_t(x,k)$ basically corresponds to the stochastic approach with the 1-dimensional subspace $< f_t(\cdot,k) >$

Optimal CPA

Theorem: Assume that $g_{0,t;k} = 1, g_{1,t;k}, \ldots, g_{u-1,t;k}$ is an orthonormal basis of $\mathcal{F}_{u,t;k}$, and $\sigma^2 := \mathrm{Var}(R_t)$.

Then $f_{opt,t} := h_{t;k}^* = \sum_{j=0}^{u-1} \beta_{j,t;k}\, g_{j,t;k}$ is the optimal CPA function, which is contained in the subspace $\mathcal{F}_{u,t;k}$.

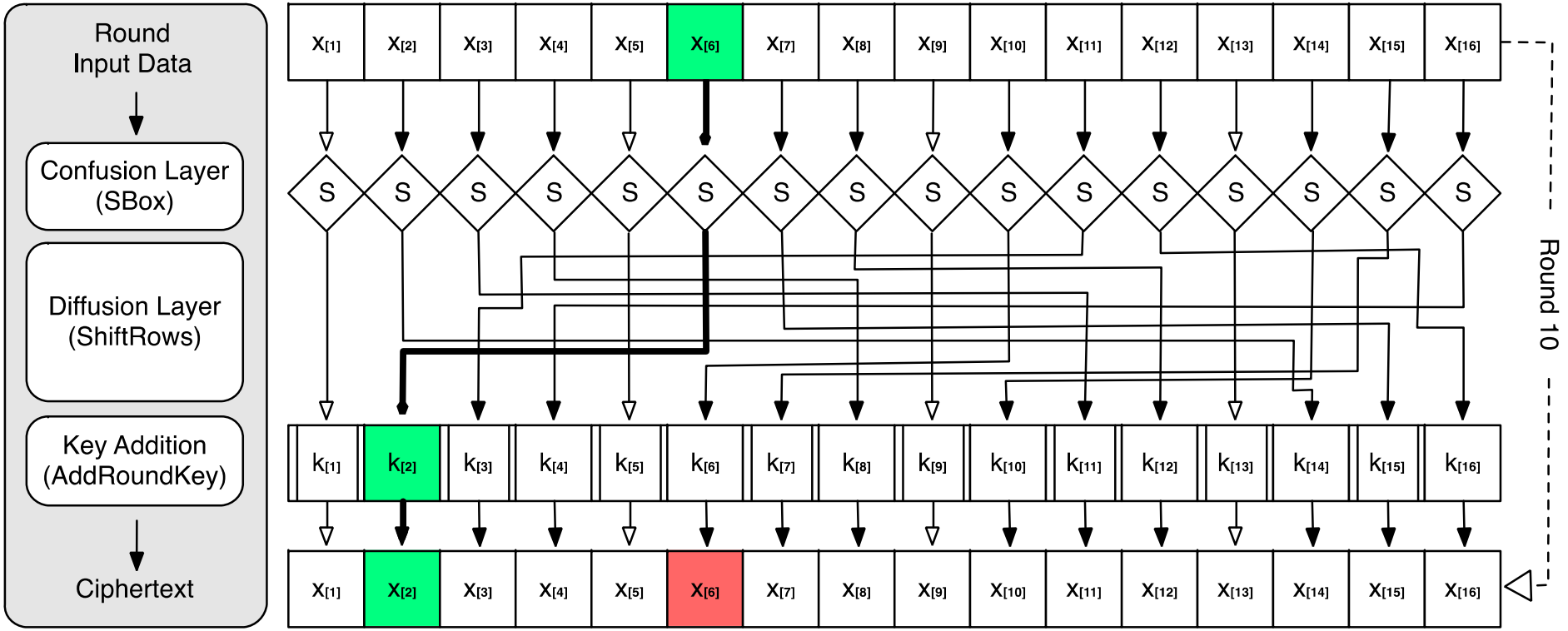
In particular,

$$\mathrm{corr}(I_t(X,k), f_{opt,t}(X,k)) = \frac{\sqrt{\sum_{j=1}^{u-1} \beta_{j,t;k}^2}}{\sqrt{\sum_{j=1}^{2\hat{}p-1} \beta_{j,t;k}^2 + \sigma^2}} = \frac{\sqrt{Var_X(h_{t;k}^*(X,k))}}{\sqrt{Var(I_t(X,k))}}$$

Practical relevance?

- We have determined an explicit formula for the optimal CPA function in different subspaces.
- How large is the information gain compared to ‘standard’ CPA functions?

AES: Final round



Leakage model

9-dimensional subspace

$g_{0,t;k}(x,k) := 1$

$g_{j,t;k}(x,k) := (x_{[b]} \oplus S^{-1}(x_{[a]} \oplus k_{[a]}))_j - 0.5$

for $1 \leq j \leq 8$

Here $x = (x_{[a]},x_{[b]})$ and $k = k_{[a]}$, e.g. (a,b) = (2,6).
Moreover,

$\hat{g}_{j,t,k}(x,k) := gj_{,t,k}(x,k) + 0.5$ for $1 \leq j \leq 8$

$$\mathcal{B}_i := \{\hat{g}_{j_1,t;k} \cdots \hat{g}_{j_i,t;k} - 2^{-i} \mid 1 \leq j_1 < \ldots < j_i \leq 8\}$$

$$\mathcal{B}_{1/2} := \{g_{1,t;k} + \cdots + g_{8,t;k}\}$$

High dimensional leakage models

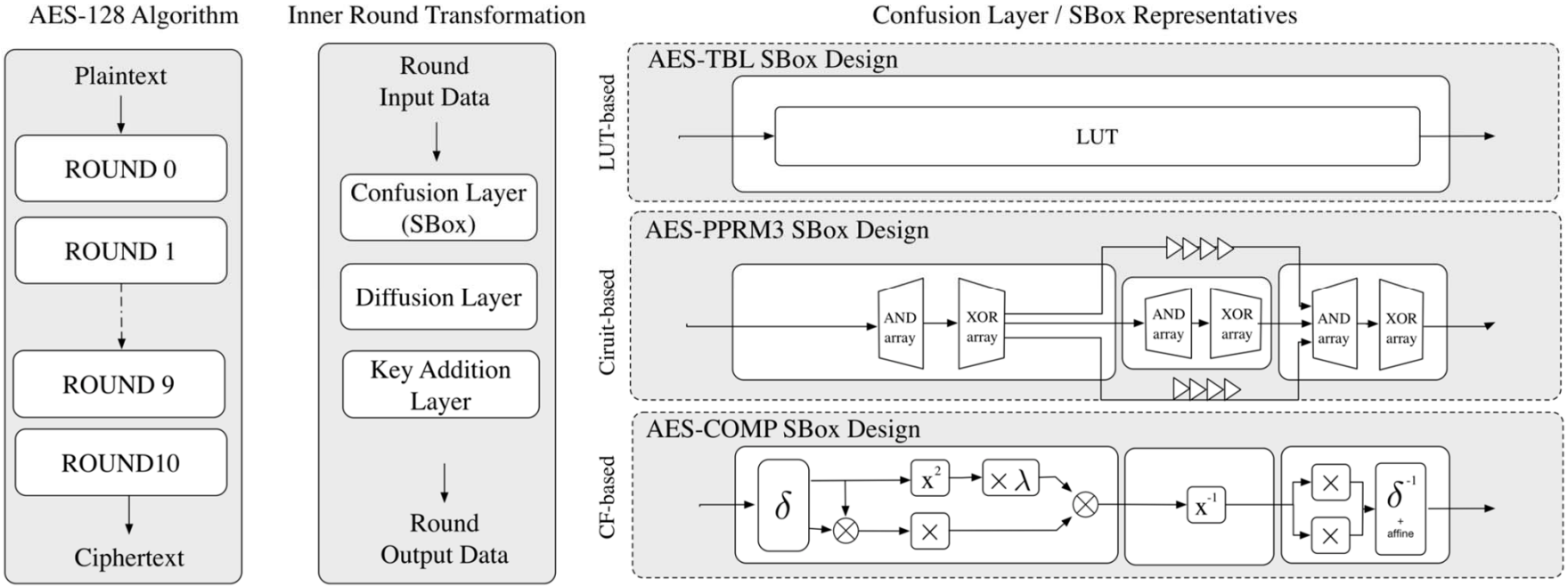
$\dim(\mathcal{F}_{u,t;k}) = u$	Set of basis functions
2	$\mathcal{B}_0 \cup \mathcal{B}_{1/2}$
9	$\mathcal{B}_0 \cup \mathcal{B}_1$
37	$\mathcal{B}_0 \cup \mathcal{B}_1 \cup \mathcal{B}_2$
93	$\mathcal{B}_0 \cup \mathcal{B}_1 \cup \mathcal{B}_2 \cup \mathcal{B}_3$
163	$\mathcal{B}_0 \cup \mathcal{B}_1 \cup \mathcal{B}_2 \cup \mathcal{B}_3 \cup \mathcal{B}_4$
219	$\mathcal{B}_0 \cup \mathcal{B}_1 \cup \mathcal{B}_2 \cup \mathcal{B}_3 \cup \mathcal{B}_4 \cup \mathcal{B}_5$
247	$\mathcal{B}_0 \cup \mathcal{B}_1 \cup \mathcal{B}_2 \cup \mathcal{B}_3 \cup \mathcal{B}_4 \cup \mathcal{B}_5 \cup \mathcal{B}_6$
255	$\mathcal{B}_0 \cup \mathcal{B}_1 \cup \mathcal{B}_2 \cup \mathcal{B}_3 \cup \mathcal{B}_4 \cup \mathcal{B}_5 \cup \mathcal{B}_6 \cup \mathcal{B}_7$
256	$\mathcal{B}_0 \cup \mathcal{B}_1 \cup \mathcal{B}_2 \cup \mathcal{B}_3 \cup \mathcal{B}_4 \cup \mathcal{B}_5 \cup \mathcal{B}_6 \cup \mathcal{B}_7 \cup \mathcal{B}_8$

High-dimensional leakage models also capture interactions between bit lines, in particular propagation glitches or cross-talk phenomena.

AES Sbox Design

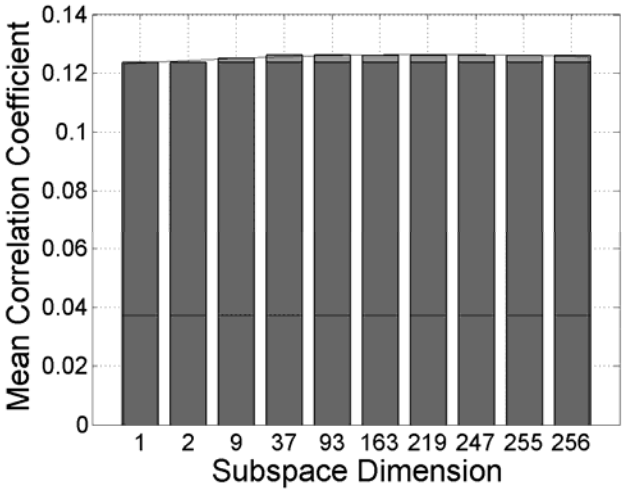
The SBox design affects the power consumption considerably. In our experiments we considered three different design principles:

- AES-TBL: Look-up table-based SBox Design (LUT-based)
- AES-PPRM3: Circuit-based SBox Design
- AES-COMP: Composite-field-based SBox Design

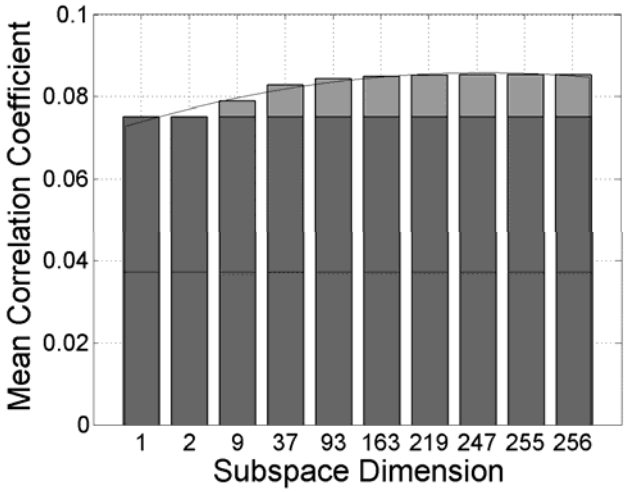


Correlation coefficients (averaged over all key bytes)

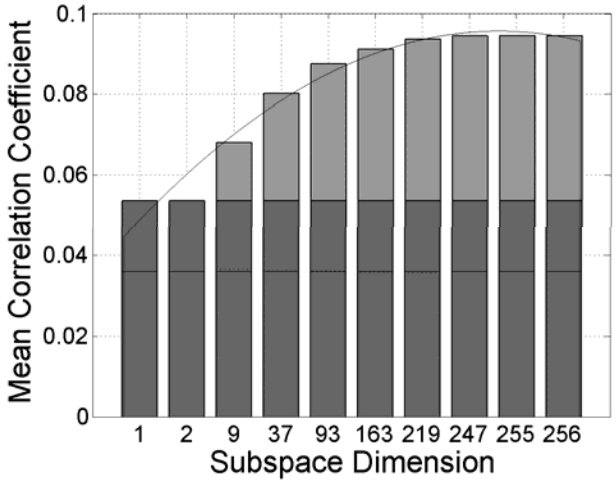
AES-TBL



AES-PPRM3



AES-COMP



1 : ‘Standard CPA’ (Hamming distance model)

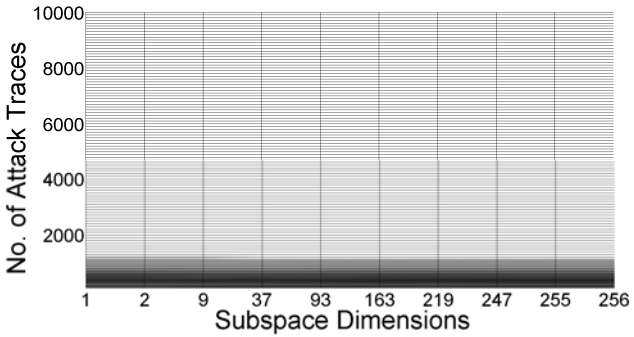
2 – 256: Optimal CPA (high-dimensional stochastic models; u = 2, …, 256)

Information gain

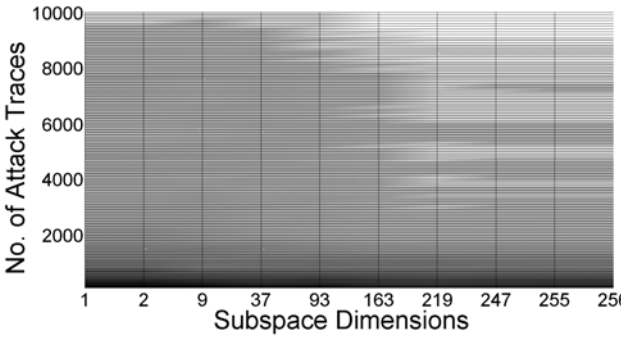
- The ratios of the averaged correlation coefficients (optimal CPA to ‘standard’ CPA) are 1.0190, 1.1350, and 1.7634 (for AES-TBL, AES-PPRM3, AES-COMP)
- *For AES-COMP the **information*** **gain** of the optimal CPA compared to ‘standard’ CPA (HD model) is maximal.
- For AES-COMP the high-dimensional subspaces $\mathcal{F}_{u,t;k}$ capture the leakage much better than the standard CPA.

Global success rate

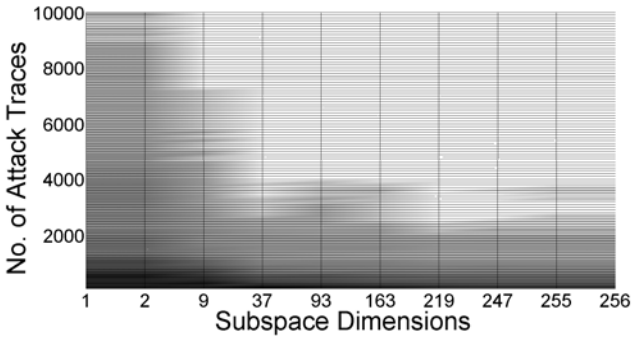
AES-TBL



AES-PPRM3



AES-COMP

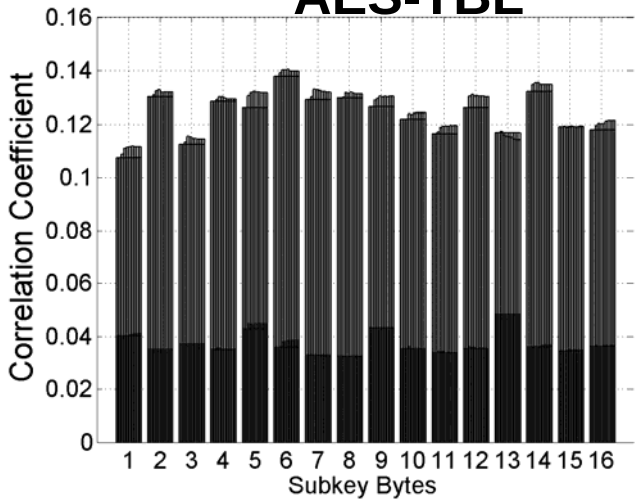


White colouring means ‘attack successful’.

- For AES-COMP the minimum number of power traces needed for stable GSR decreases from 10.000 (standard CPA) to 2.300 traces (optimal CPA).

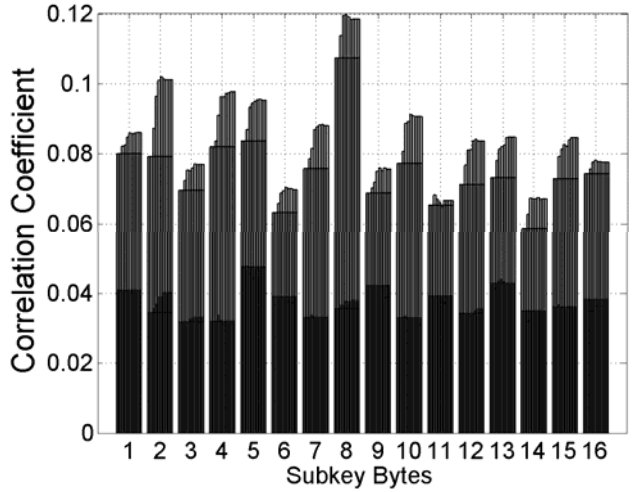
Correlation coefficients (fine grained)

AES-TBL

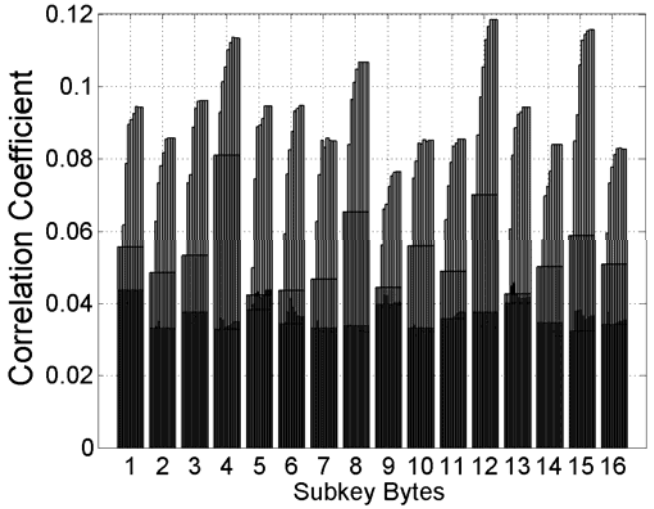


Each vertical bar consists of 10 vertical sub-bars, which correspond to the standard CPA and to the optimal CPA for u = 2,…,256, resp.

AES-PPRM3



AES-COMP



Design complexity and information gain

	no. of LUTs	Max. freq. (MHz)	max. timing (ns)
AES-TBL	1409	257.966	**4.558**
AES-PPRM3	2283	134.228	**8.138**
AES-COMP	2066	135.888	**11.274**

- *In our experiments the information gain of the optimal CPA over the standard CPA increased with the complexity of the Sbox design.*
- One might assume that high design complexity generally implies higher information gain. However, this requires further experiments with different designs.

Conclusion

- It has been well-known that the stochastic approach can be applied as an efficient attack tool and to obtain design information.
- In this talk we used the stochastic approach to derive the optimal CPA function to (arbitrary) given leakage distributions and for varying complexity of the leakage model.
- Experiments with three FPGA implementations of the AES cipher showed that the ‘information gain’ depends on the concrete implementation.

Federal Office
for Information Security

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