Bundesamt
für Sicherheit in der
Informationstechnik

hgi
Horst-Görtz Institut
für IT Sicherheit

A Stochastic Model for Differential Side Channel Cryptanalysis

Werner Schindler[1], Kerstin Lemke[2], Christof Paar[2]

[1] Bundesamt für Sicherheit in der Informationstechnik (BSI)
53175 Bonn, Germany

[2] Horst Görtz Institute
Ruhr University Bochum
44780 Bochum, Germany

Edinburgh, August 30, 2005

Outline

- Introduction
- A new stochastic approach
  - Fundamental ideas and benefits
  - Experimental results
  - Comparison with other attacks
  - Generalizations
- Conclusion

Comparison with other Techniques

Method	Profiling Step (Training Device)	Key Extraction Step (Target Device)
DPA/DEMA	no	yes
Template Attack	yes	yes
New Stochastic Approach	yes (..., but can be skipped)	yes

Our new approach combines

- engineer's insight (Which properties / features of the physical device have (significant) impact on the side-channel signal? (qualitative assessment))
- with efficient stochastic methods (exploiting this information in an optimal way)

Profiling: much more efficient than template attacks

Key Extraction: The efficiency is

- determined by the engineer's skills
- limited by the efficiency of template attacks

The Stochastic Model

target algorithm: block cipher (no masking)

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

$k \in \{0,1\}^s$ subkey

t time

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

Random variable
(depends on x and k)

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

deterministic part
(depends on x and k)

Random variable
$E(R_t) = 0$

Noise

Profiling, Step 1: Approximating the Deterministic Part

- Task: Estimate the function h_t for all

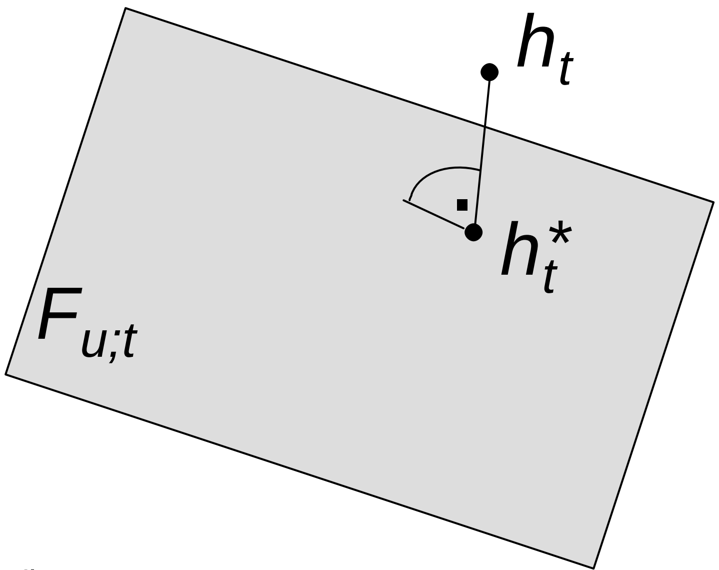
  $t \in \{ t_1, t_2, \ldots, t_m \}$ (measurement times)

- Naïve Approach: Estimate $h_t(x,k) = E\,(I_t(x,k))$ independently for each $(x,k) \in \{0,1\}^p \times \{0,1\}^s$

- Drawback: Giantic number of measurements

More favourable procedure (I)

- The unknown function h_t is interpreted as an element in a real vector space F.
- Approximate h_t by its orthogonal projection h^*_t onto a suitably chosen low-dimensional vector subspace $F_{u;t}$



geometric

visualization

More favourable procedure (II)

The subspace

$$\mathcal{F}_{u;t} := \{h' : \{0,1\}^p \times \{0,1\}^s \to \mathrm{IR} \mid \sum_{\mathrm{j}=0}^{\mathrm{u}-1} \beta'_{\mathrm{j}} \mathrm{g}_{\mathrm{jt}} \text{ with } \beta'_{\mathrm{j}} \in \mathrm{IR}\}$$

is spanned by known functions $g_{jt} : \{0,1\}^p \times \{0,1\}^s \to \mathrm{IR}$

Select functions $g_{0t},\ldots,g_{(u-1)t}$ under consideration of the attacked device.

The projection h^*_t is the best approximator of h_t in $F_{u;t}$ (= nearest element of $F_{u;t}$).

1st Minimum Principle (I)

Theorem: The image h^*_t of h_t under the orthogonal projection meets a minimum property:
For each subkey k and random plaintext X the expectation

$$E\left(\left(I_t(X,k) - h'(X,k) \right)^2 \right)$$

attains its minimum on $F_{u;t}$ for $h'=h^*_t$

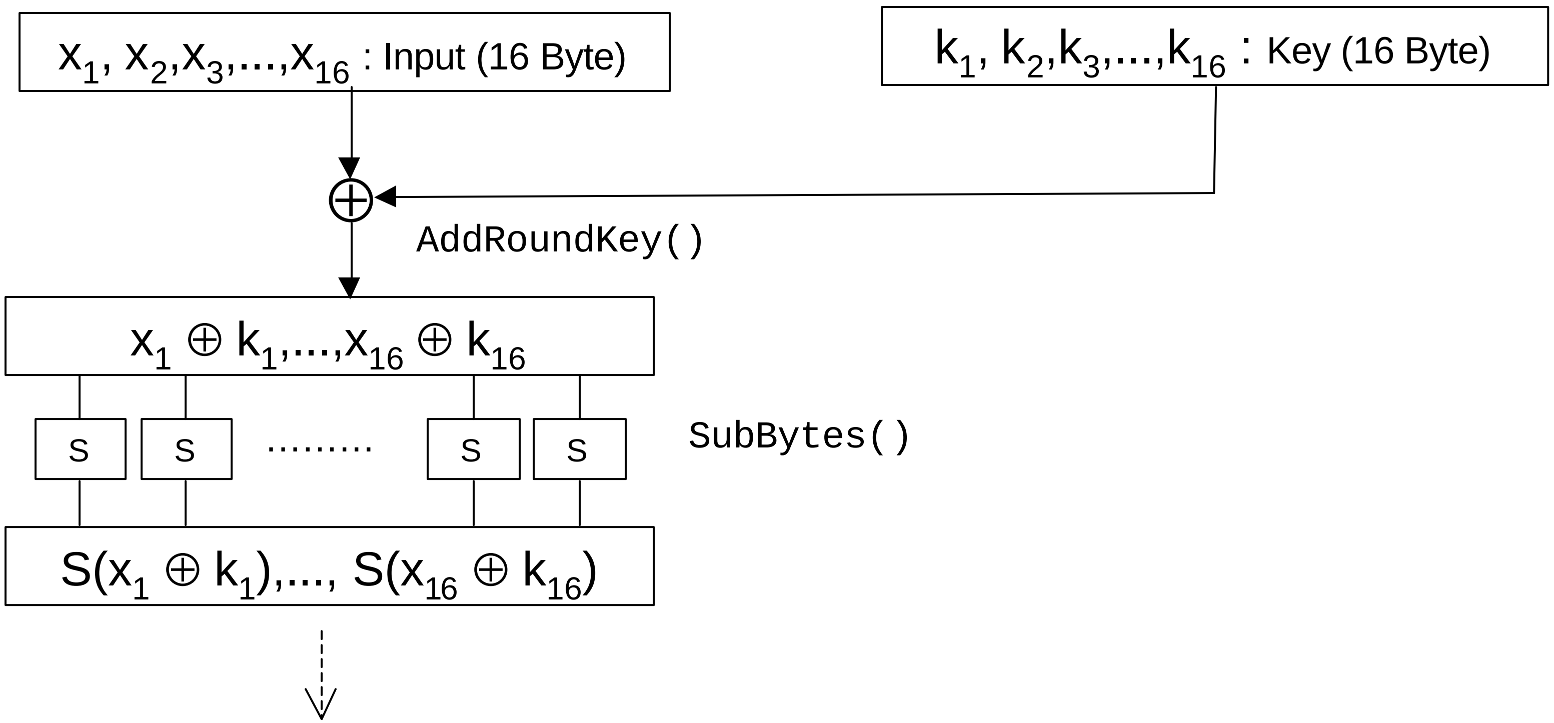
1st Minimum Principle (II)

Note: The image under the orthogonal projection, $h^*_t \in F_{u;t}$, can be determined **without the knowledge of h_t !**

In other words:

The estimation of h^*_t can completely be moved to the low-dimensional subspace $F_{u;t}$.

Example: AES (I)



- here: $x\,,\,k \in \{\,0,1\,\}^8$
- $h_t(x,k)$ depends only on the sum $x \oplus k$
- ® It is sufficient to determine $h_t(x,k)$ for any single subkey k.

Example: AES (II)

Reasonable candidates for the functions $g_{jt}(x,k)$:

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

$g_{jt}(x,k) = j^{th}$ bit of $S(x \oplus k)$ for $1 \le j \le 8$

....

interpreted as a real-valued function $\{0,1\}^8 \to IR$

$F_{9;t} = < g_{0t}, g_{1t}, \ldots, g_{8t} >$

vector subspace generated by $g_{0t}, g_{1t}, \ldots, g_{8t}$

Note: $\dim(F_{9;t}) = 9$ while $\dim(F) = 256$

no information on h_t

Profiling, Step 1: Approximating the Deterministic Part

- Task: Estimate the coefficients $\beta^*_{0t}, \ldots, \beta^*_{(u-1)t}$ of h^*_t with respect to the base $g_{0t}, \ldots, g_{(u-1)t}$ for each $t \in \{ t_1, \ldots, t_m \}$

  (measurement times)

- Procedure:
  1. perform N_1 measurements (i.e. observe N_1 encryptions) at the training device
  2. calculate the least-square-estimator (requires no more than elementary linear algebra)

Profiling, Step 2: Modelling the noise

- Assumption: The random vector $(R_{t1}, \ldots, R_{tm})$ is multi-variate normally distributed with covariance matrix C
- $h_{t1},\ldots,h_{tm}$ and C yield the conditional density $f(\cdot \mid x,k)$ for $(I_{t1}(x,k), \ldots, I_{tm}(x,k))$.

- Profiling, Step 2:
  - Perform N_2 further measurements (i.e., observe N_2 further encryptions at the times $t_1, \ldots, t_m$)
  - Determine estimators $\tilde{C}$ and $\tilde{f}(\cdot \mid x,k)$ for C and $f(\cdot \mid x,k)$

Key Extraction: Maximum Likelihood Method

- The adversary
  - performs N_3 measurements at the target device
  - substitutes the measured data into the estimated densities $\tilde{f}(\cdot \mid x,k)$ for each subkey k
  - decides for that subkey k° that maximizes this term (maximum-likelihood principle)

details: paper

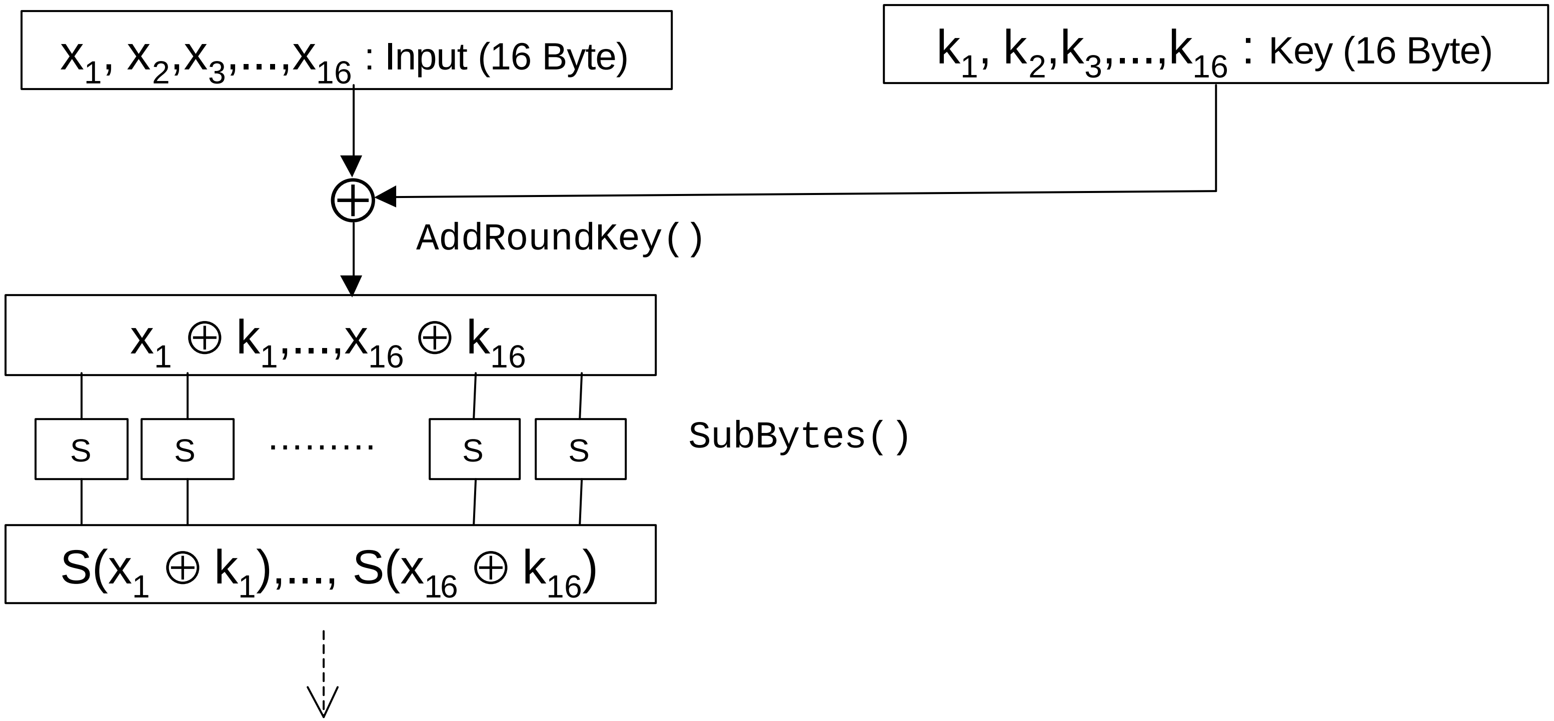
Key Extraction: Minimum Principle

- Alternative key extraction strategy: based on a **2nd minimum property**
- Properties:
  - Key extraction efficiency: smaller than for the maximum-likelihood method
  - Profiling: saves Step 2 (modelling the noise)

details: paper

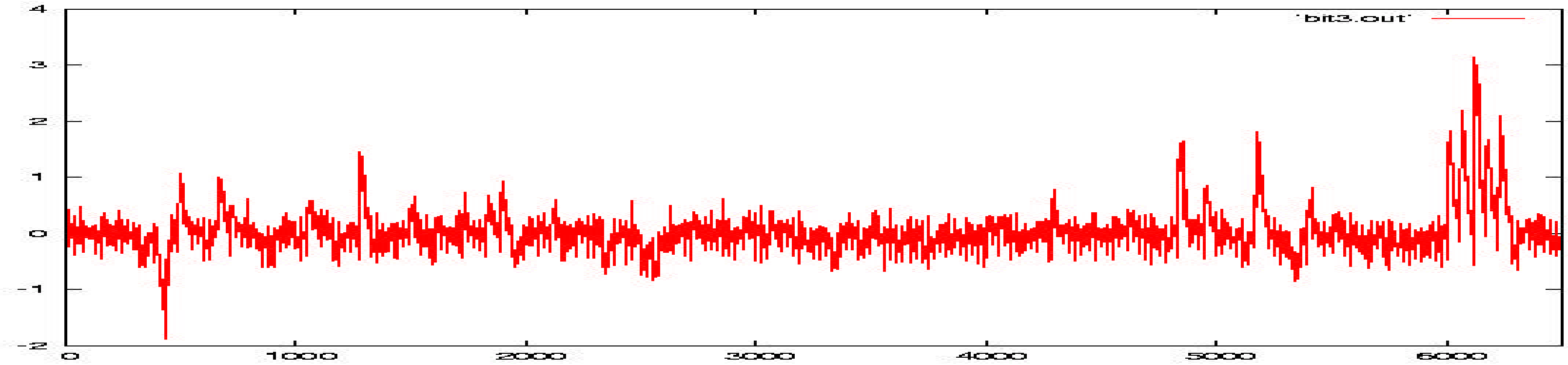
Experimental Results (I)

Power analysis at an unprotected AES implementation on an ATM163 microcontroller

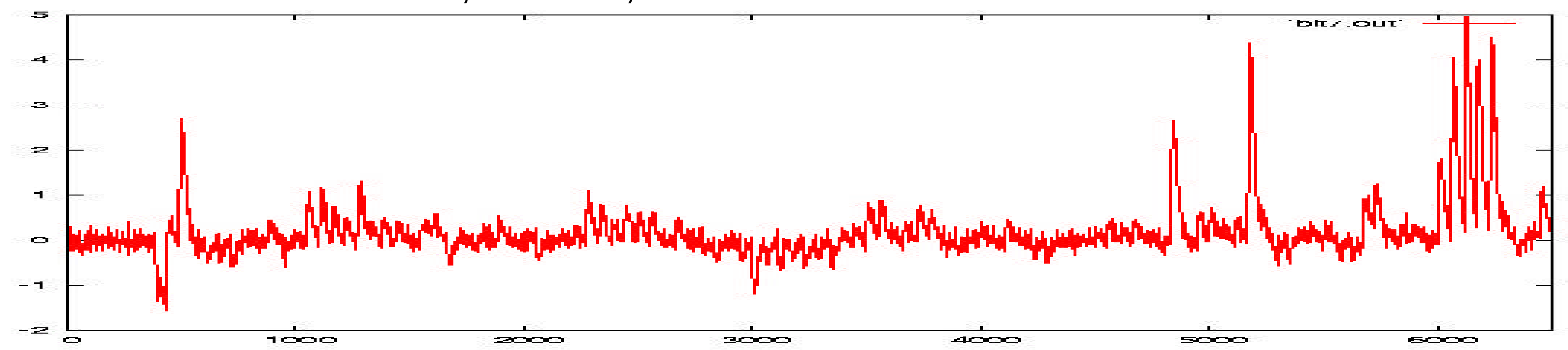


Experimental Results (II)

coefficient $ß_{3,t}$ in $F_{9;t}$



coefficient $ß_{7,t}$ in $F_{9;t}$



Time t

Empirical probabilities for the correctness of the rank 1-candidate

- For all instants t we used the vector subspace $F_{9;t} = F_9 := <\ 1, j^{th}$ bit of $S(x \oplus k)$ for $1 \leq j \leq 8\ >$

N_3	DPA (HW model)	Minimum Principle (N_1=2000) m=7	Minimum Principle (N_1=2000) m=21	Maximum-likelihood (N_1=1000) m=7(N_2=1000)	Maximum-likelihood (N_1=1000) m=21(N_2=5000)
5	0.82 %	28.47 %	33.40 %	36.30 %	41.43 %
7	1.31 %	48.20 %	53.88 %	61.12 %	68.34 %
10	2.74 %	73.45 %	78.69 %	84.12 %	90.17 %
15	6.04 %	92.92 %	95.15 %	97.97 %	99.25 %
20	9.70 %	98.31 %	98.82 %	99.85 %	99.96 %
30	19.67 %	99.89 %	99.95 %	99.99 %	> 99.99 %

Impact of Different Subspaces

$F_{2;t} = F_2 := \langle 1, \mathrm{HW}(S(x \oplus k)) \rangle$

$F_{10;t} = F_{10} := \langle F_9$, most significant 2nd order monomial $\rangle$

$F_{16;t} = F_{16} := \langle F_9$, all consecutive 2nd order monomials $\rangle$

Key Extraction: **Minimum Principle**

N_3	$N_1 = 2000$				$N_1 = 5000$	
	F_2	F_9	F_{10}	F_{16}	F_9	F_{10}
10	37.77 %	75.29 %	72.94 %	65.05 %	77.31 %	80.19 %

N_1 is too small

Example AES

No. of profiling series (exploiting symmetry):

- template attack: **256**
- new stochastic method: **1 - 2**

Generalizations and Remarks

- Our approach can be generalized in a natural way
  - to masking
  - to multi-channel attacks

  (details: paper).
- Profiling:
  - usually: known test key.
  - also works with unknown test keys (additional computations)
  - may completely be skipped (reduces the efficiency at key extraction)

Conclusion

We introduced a new methodology for differential side-channel attacks that

- combines engineer's insight with stochastic methods
- enables to determine those properties that have significant impact on the side-channel signal
- enables efficient assessment of the risk potential of a side-channel attack
- profiling: much more efficient than for template attacks
- key extraction efficiency: determined by the suitability of the chosen vector subspace $F_{u;t}$

Contact

Werner Schindler
Bundesamt für Sicherheit in der
Informationstechnik (BSI), Germany
Werner.Schindler@bsi.bund.de

Kerstin Lemke
Horst Görtz Institute
Ruhr University Bochum, Germany
lemke@crypto.rub.de

Christof Paar
Horst Görtz Institute
Ruhr University Bochum, Germany
cpaar@crypto.rub.de