Federal Office
for Information Security

The Stochastic Approach in Power Analysis – A Synthesis between Engineer's Expertise and Advanced Stochastics

Werner Schindler

Federal Office for Information Security (BSI), Bonn

Prague, June 26, 2009

Outline

- dpa
- Template attacks
- The stochastic approach
  - Description and mathematical background
  - Examples and experimental results
  - Comparison with other attacks
  - Advantages and useful properties
- Final Remarks

Preliminary remark

- In this talk we concentrate on power attacks on block ciphers

- Example: AES
  - The key is guessed byte by byte (= 8-bit subkeys)
  - The 16 subkeys are guessed independently.

DPA (Differential Power Analysis)

Pioneer work: Kocher, Jaffe, Jun (1999)

Basic idea: dpa exploits correlations between a function of a subkey (e.g. its Hamming weight) and the electrical current at time t (2nd order dpa: 2 time instants)

- + preparatory work: moderate
- - attacking efficiency: moderate

Disadvantages / Problems:

- Usually dpa only exploits a small fraction of the available information.
- It is not clear how to combine information from different time instants.

Template attacks (I)

Pionier work: Chari, Rao, Rohatgi (2002)

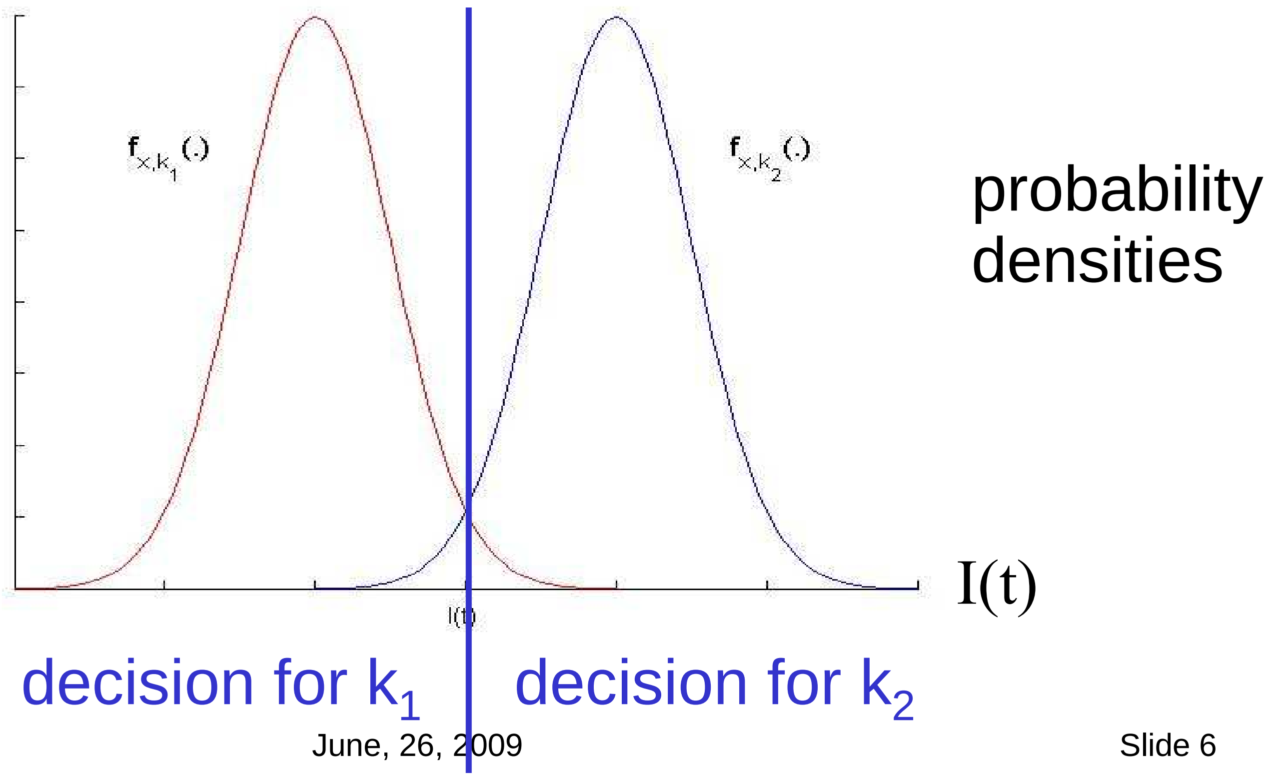
Basic idea:

- For each plaintext byte x, each subkey k, (possibly for each masking value z) the measurement values (~ electrical current) at time instants $t_1<\ldots<t_m$ are interpreted as values that are are assumed by random variables $I_{t\_1}(x,k), \ldots, I_{t\_m}(x,k)$, resp. of $I_{t\_1}(x,z,k), \ldots, I_{t\_m}(x,z,k)$.
- The distributions of these random variables are unknown.
- Profiling: Estimates for the probability densities (for all pairs (x,k), resp. for all triplets (x,z,k)) are gained from measurements performed at an identical training device.

Template attacks (II)

Attack: Measurement values from the target device are substituted into the estimated densities ($\rightarrow$ maximum likelihood estimator)

Example: m=1



Template attacks (III)

Advantages / disadvantages

- + attacking efficiency: maximal (for given $t_1<\ldots<t_m$)
- - profiling workload: gigantic, especially for strong implementations and in case of masking
- - A successful template attack shows that the implementation is vulnerable but does not identify the weakness

The stochastic approach (I)

- Theoretical foundations:
  Schindler, Lemke, Paar (2005), Schindler (2008)
- Experimental work:
  Schindler, Lemke, Paar (2005), Gierlichs, Lemke, Paar (2006), Lemke-Rust, Paar (2007), Standaert, Koeune, Schindler (2009, simulation studies)
- Target: block ciphers
- Similarities with template attacks
  - uses information from several time instants $t_1 < t_2 < \ldots < t_m$
  - interprets measurement values as values that are assumed by random variables

The stochastic approach (II)

The stochastic approach combines

- engineers‘ expertise
  - Question: Which properties of the implementation / hardware may have significant impact on the side channel leakage? (qualitative assessment)
- with advanced stochastic methods
  - Goal: exploit the available information in an optimal way

The stochastic model (basic variant)

Target: block cipher (e.g. AES), no masking

$x \in \{0,1\}^p$ (known) part of the plaintext or ciphertext (AES: p =8)

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

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

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

Random variable
(depends on x and k)

quantifies the
“randomness” of the
leakage at time t

Deterministic part
(depends on x and k)

Random variable
$E(R_t) = 0$

noise

The stochastic model (considers masking)

Target: block cipher (e.g. AES)

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

$z \in M$ masking value

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

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

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

Random variable
(depends on x,z,k)

Quantifies the
"randomness of the
leakage at time t

deterministic part
(depends on x,z,k)

Random variable
$E(R_t) = 0$

noise

Remark

- Note:
  - The 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“)

  are unknown.

- <u>Profiling:</u> The functions and the probability distributions are estimated on basis of measurements at an identical training device.

Profiling, Step 1: Naive Approach

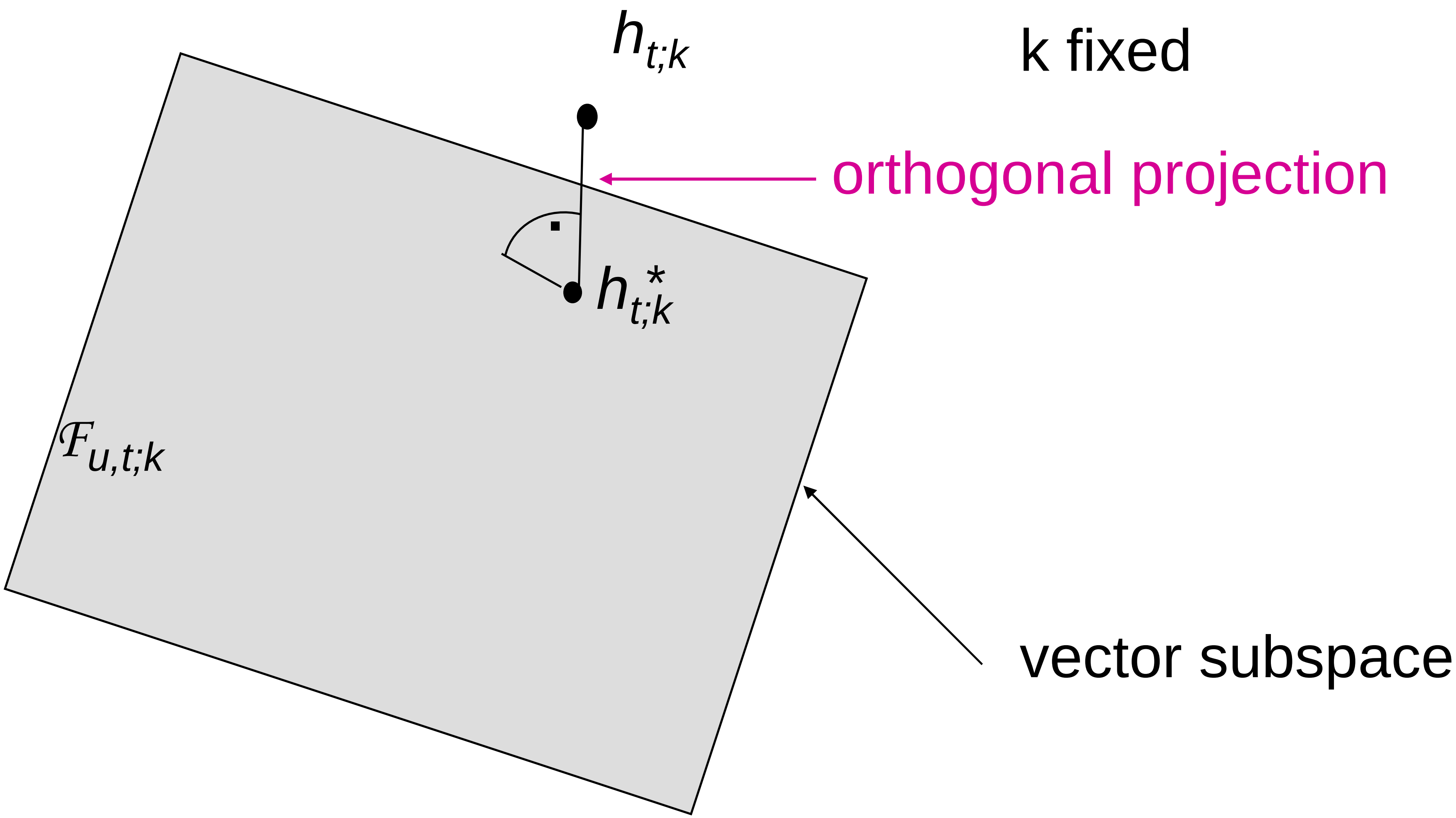
- Estimate $h_t(x,z;k) = E(I_t(x,z;k))$ *independently* for each triplet $(x,z;k) \in \{0,1\}^p \times M \times \{0,1\}^s$ ($\rightarrow$ template attack)
- Disadvantage: requires $2^{p+s}|M|$ measurement series $\rightarrow$ gigantic number of measurements (= power traces), especially for strong implementations (some reduction for chosen input attacks)

Much better … (central idea)

- Fix a subkey $k \in \{0,1\}^s$.
- Interpret the unknown function $h_{t;k}: \in \{0,1\}^p \times M \times \{k\} \to R$, $h_{t;k}(x,z;k) := h_t(x,z,k)$ as an element of a $2^p\,|M|$-dimensional real vector space $\mathcal{F}$.

- Idea: approximate $h_{t;k}$ by its image $h^*_{t;k}$ under an orthogonal projection on a suitably selected low-dimensional vector subspace $\mathcal{F}_{u,t;k}$

Geometric visualisation



The vector subspace $\mathcal{F}_{u,t;k}$

The u-dimensional vector subspace

$$F_{u,t;k} := \{h': \{0,1\}^p \times M \times \{k\} \to R \mid \sum_{j=0}^{u-1} \beta'_{j,t;k} g_{j,t;k} \quad mit \quad \beta'_{j,t;k} \in R\}$$

is spanned by u linear independent vectors (functions)

$g_{j,t;k} : \{0,1\}^p \times M \times \{k\} \to R \qquad 0 \leq j \leq u-1$

The image $h^*_{t,k}$ is the best approximator of h_t in $\mathcal{F}_{u,t;k}$ (= closest element in $\mathcal{F}_{u,t;k}$).

Main Theorem

Theorem: For any fixed subkey k the image $h^*_{t;k}$ of $h_{t;k}$ under the orthogonal projection meets the following minimum property:

For random plaintext X the mean value

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

attains ist minimum on $\mathcal{F}_{u,t;k}$ at $h'=h^*_{t;k}$.

Consequences

It is possible to determine the image $h^*_{t,k} \in \mathcal{F}_{u,t;k}$ without knowledge of the pre-image $\mathbf{h_{t;k}}$!!

The estimation of $h^*_{t,k}$ is completely moved to the low-dimensional subspace $\mathcal{F}_{u,t;k}$.

This property reduces the number of profiling measurements to a small fraction.

Of course, the basis $g_{0,t;k},\ldots,g_{(u-1),t;k}$ of the vector subspace $\mathcal{F}_{u,t;k}$ should be selected under consideration of the attacked device (→ engineer's expertise)

Example: AES (no masking / CHES 2005) (I)

- $t_1, t_2, \ldots, t_m$: time instants after the S- box evaluation

- Reasonable candidates for the functions $g_{j,t;k}(\cdot,k)$:

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

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

  ....

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

  $\mathcal{F}_{9,t;k} = < g_{0,t;k}, g_{1,t;k}, \ldots, g_{8,t;k} >$

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

Example: AES (no masking / CHES 2005) (II)

Note: $\dim(\mathcal{F}_{9,t;k}) = 9$ but $\dim(\mathcal{F}) = 256$

no pre-information on h_t

- The vector basis may be extended, e.g. to capture crossover effects:

$$g_{8+j,t;k}(x,k) = g_{j,t;k}(x,k)\, g_{j+1,t;k}(x,k) \quad \text{for } 1 \le j \le 7$$

$\rightarrow$ 16-dimensional vector subspace $\mathcal{F}_{16,t;k}$

Example AES (masked implementation)

- 8-bit bus implementation
- $t \in \{t_1,\ldots,t_m\}$: instants before the S-box evaluation
- Masking: $x \rightarrow (x \oplus z) \rightarrow (x \oplus z \oplus k) \rightarrow \ldots$

- Plausible candidates for the basis (depending on t)

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

$g_{j,t;k}(x,z,k) = j^{th}$ bit of $(x \oplus z \oplus k)$ für $1 \leq j \leq 8$

....

Interpreted as a real-valued function $\{0,1\}^{16} \rightarrow R$

Here: $\dim(\mathcal{F}_{9,t;k}) = 9$ but $\dim(\mathcal{F}) = 2^{16}$

Profiling, Step 1: Approximation of the Deterministic Part

- Task: Let $t \in \{t_1, ..., t_m\}$. For each admissible subkey k estimate the coefficients $\beta^*_{0,t;k}, \ldots, \beta^*_{(u-1),t;k}$ of the best approximator $h^*_{t;k}$ of $h_{t;k}$ with respect to the basis $g_{0,t;k}, \ldots, g_{(u-1),t;k}$
- Procedure:
  1. perform N_1 measurements (i.e. observe N_1 encryptions) at the training device
  2. calculate the least-square estimate
- Note: This procedure may be performed separately for all $t \in \{ t_1, ..., t_m \}$

Profiling, Step 2: Modelling the noise

- Assumption: The random vector $(R_{t1}, \ldots, R_{tm})$ is multivariate normally distributed with covariance matrix C
- Note: $h_{t1},\ldots,h_{tm}$ and C yield the parameter-dependent densities $f_{x,z;k}(\cdot)$ for $(I_{t1}(x,z,k), \ldots, I_{tm}(x,z,k))$.

- Profiling, Step 2:
  1. Perform N_2 new measurements (i.e., observe N_2 further encryptions at the instants $t_1, \ldots, t_m$)
  2. Determine estimates for $\tilde{C}$ and $\tilde{f}_{x,z;k}(\cdot)$ for C and $f_{x,z;k}(\cdot)$

Attacking phase

- Conduct N_3 measurements at the target device
- Decide for that subkey k that maximizes the term

$$\prod_{j=1}^{N_3} \sum_{z\_j \in M} \Pr(Z_j = z_j)\, \tilde{f}_{x\_j, z\_j; k}(i_j)$$

(maximum likelihood estimator)

Empirical probabilities for the correctness of the rank 1-candidate

- Reference: Schindler, Lemke, Paar (CHES 2005)
- For all instants t the 9-dimensional vector subspace

$\mathcal{F}_{9;t} = \mathcal{F}_9 := \langle 1, j^{th}$ bit of $S(x \oplus k)$ for $1 \leq j \leq 8 \rangle$ was used

N_3	**DPA** (HW model)	**Stochastic approach** (N_1=1000) m=7 (N_2=1000)	**Stochastic approach** (N_1=1000) m=21 (N_2=5000)
5	0.82 %	36.30 %	41.43 %
7	1.31 %	61.12 %	68.34 %
10	2.74 %	84.12 %	90.17 %
15	6.04 %	97.97 %	99.25 %
20	9.70 %	99.85 %	99.96 %
30	19.67 %	99.99 %	> 99.99 %

Comparison with template attacks (no masking, empirical results)

- Gierlichs, Lemke, Paar (2006):
  Exemplary implementation: Even a reduction of the profiling measurements to 2% (relative to a template attack) preserved acceptable attacking efficiency.
- Standaert, Koeune, Schindler (2009)
  Simulation studies: The stochastic approach required only 4% of the profiling measurements of template attacks with comparable attacking efficiency.

The degree of the advantage depends on the concrete implementation.

Comparison with template attacks (masking, empirical results)

- K. Lemke-Rust verified the applicability of the stochastic approach in presence of masking by many experiments.
- For masked implementations (relative to template attacks) the advantage in profiling efficiency is at least one order of magnitude larger than for implementations without masking.
- Example: AES (masked implementation)

  template attacks:

  profiling: 256*|M|*256 measurement series

  (reduction for chosen input attacks)

  stochastic approach:

  profiling: 256 +1 measurement series

Constructiveness of the stochastic approach

- If the absolute value of the coefficient $\beta^*_{j,t;k}$ is large, the „direction“ of the basis vector $g_{j,t;k}$ has significant impact on the subkey-dependent part of the leakage $h_{t;k}$ ($\rightarrow$ quantitative description of the weakness)

This property can be used to support aimed (re-)design.

Final Remarks

The stochastic approach has several interesting properties. It

- combines engineers' expertise ($\rightarrow$ selection of a suitable vector subspace $\mathcal{F}_{u,t;k}$) with stochastic methods
- can principally be applied to any masking scheme
- profiling: by order(s) of magnitude more efficient than for template attacks
- attacking efficiency: $\leq$ template attacks (depends on the choice of $\mathcal{F}_{u,t;k}$)
- identifies und quantifies properties / weaknesses that have significant influence on the side-channel leakage
- can be used to support aimed design / redesign

Contact

Federal Office for Information Security (BSI)

Werner Schindler
Godesberger Allee 185-189
53175 Bonn

Tel: +49 (0)22899 - 9582-5652
Fax: +49 (0)22899 - 10-9582-5652

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