Differential Cluster Analysis

Lejla Batina[1,3], Benedikt Gierlichs[1], and Kerstin Lemke-Rust[2]

[1] K.U. Leuven ESAT/SCD-COSIC and IBBT, Belgium
[2] University of Applied Sciences Bonn-Rhein-Sieg, Germany
[3] Radboud University Nijmegen, Netherlands

September 7, 2009

CHES Workshop 2009, Lausanne, Switzerland

Outline

Presentation Outline

- Introduction
- Differential Cluster Analysis
- Applications
- Experimental Results
- Conclusion

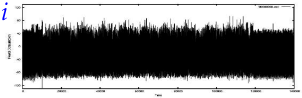
Differential Side Channel Adversary

Input x

Physical Leakage $\boldsymbol{i}$

while computing $f_{k^\circ}(x)$

i



Time t

Function f_{k° with

secret subkey k°

Crypto Device

Adversary Success

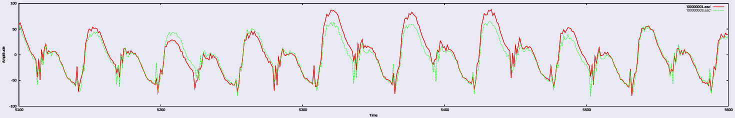
The adversary is successful if she recovers the secret subkey k°.

Introduction: Differential Power Analysis in a Nutshell

The DPA (Differential Power Analysis) Problem

Given measurements $\boldsymbol{i}_n$ ($n \in \{1, \ldots, N\}$) while the crypto device computes function f_{k^0} with random inputs x_n:

- Does a statistics prove significant differences of the measurements according to a partitioning function $g(f_k(x_n))$?



DPA Partitioning Functions g:

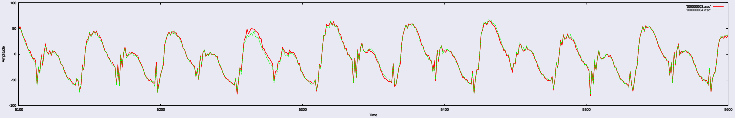
- Single-bit partitioning function (Kocher *et al.*, 1999)
- Multi-bit partitioning/comparison function:
  - "All-or-Nothing" with a leakage model (e.g., Hamming weight) (Messerges *et al.*, 2000)
  - CPA (Correlation Power Analysis) with a leakage model (e.g., Hamming distance) (Brier *et al.*, 2002)

Introduction: Collision Analysis in a Nutshell

The CA (Collision Analysis) Problem

Given measurements $\boldsymbol{i}_n$ ($n \in \{1, \ldots, N\}$) while the crypto device computes the many-to-one function f_{k^0} with random inputs x_n:

- Does a statistics prove high similarity of measurements with two inputs $x_i \neq x_j$?



Collision Detection:

- Euclidean distance of measurement vectors over some t. (Schramm *et al.*, 2003)

Our Approach

Objectives

- Combination of leakage detection functions for DPA (*Separation*) and CA (*Cohesion*).
- Sensitivity to general leakage features
- Multi-bit approach
  - using all measurements and
  - without the need for a good power model.
- Multivariate approach

Idea

- Our basic approach: Cluster Analysis
- Our basic question: Do clusters of measurements exist?

Differential Cluster Analysis

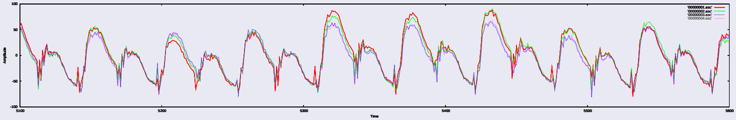
Requirement

- $f_k : \{0,1\}^u \mapsto \{0,1\}^w$ is a many-to-one collision function, i.e., at least two inputs $x_i, x_{i'} \in \{0,1\}^u$ with $x_i \neq x_{i'}$ collide in one state $\Delta \in \{0,1\}^w$.

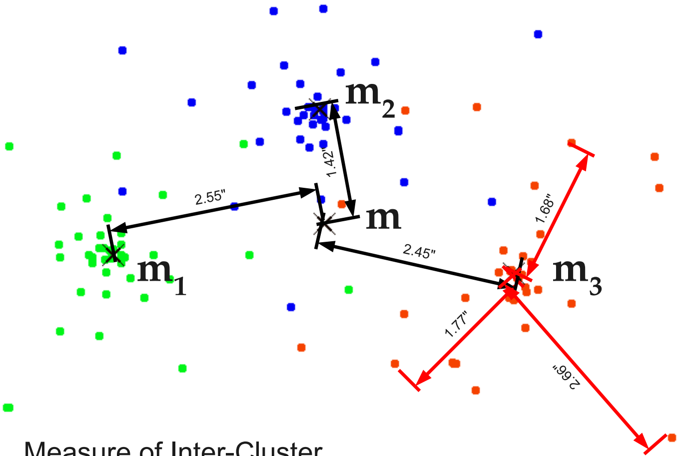
The DCA (Differential Cluster Analysis) Problem

Given measurements $\boldsymbol{i}_n$ ($n \in \{1, \ldots, N\}$) while the crypto device computes the many-to-one function f_{k^0} with random inputs x_n:

- Does a cluster criterion function prove the existence of clusters of measurements according to partitioning function $f_k(x_n)$?



Measuring Clustering Quality



Measure of Inter-Cluster Separation

Measure of Intra-Cluster Cohesion

Cluster Criterion Functions

Sum-of-Squared-Error:

$$J_{SSE} = \sum_{i=1}^{c} \sum_{\boldsymbol{x} \in \mathcal{D}_i} \| \boldsymbol{x} - \boldsymbol{m}_i \|^2$$

J_{SSE} evaluates intra-cluster cohesion. The optimal partition minimizes J_{SSE}.

Sum-of-Squares:

$$J_{SOS} = \sum_{i=1}^{c} n_i \| \boldsymbol{m}_i - \boldsymbol{m} \|^2$$

J_{SOS} evaluates inter-cluster separation. The optimal partition maximizes J_{SOS}.

The sum of J_{SSE} and J_{SOS} is a constant.

Special Cluster Criterion Functions

Variance criterion (Standaert *et al.*: ICISC 2008):

$$J_{VAR} = \frac{\| \boldsymbol{v} \|^2}{\frac{1}{N}\sum_{i=1}^{c} n_i \| \boldsymbol{v}_i \|^2}$$

J_{VAR} evaluates overall variance vs. intra cluster variances. The optimal partition maximizes J_{VAR}.

T-test criterion (Gierlichs *et al.*: CHES 2006):

$$J_{STT} = \sum_{i,j=1;i\neq j}^{c} \frac{\| \boldsymbol{m}_i - \boldsymbol{m}_j \|^2}{\sqrt{\frac{\|\boldsymbol{v}_i\|^2}{n_i} + \frac{\|\boldsymbol{v}_j\|^2}{n_j}}}$$

J_{STT} evaluates inter cluster separation, normalized by intra cluster variances and cluster sizes. The optimal partition maximizes J_{STT}.

Differential Cluster Analysis (General Approach)

Differential Cluster Analysis (General Approach)

1. For each subkey hypothesis k:
   - Sort measurements into 2^w clusters $\mathcal{D}_0, \ldots, \mathcal{D}_{2^w-1}$ according to $\Delta_i = f_k(x_n)$ $(1 \leq n \leq N)$.
   - Compute a cluster criterion function: J_k.
2. Rank the pairs (k, J_k) according to J_k.
3. Output subkey candidate that leads to the best clustering quality.

Detailed Comparison with CA and DPA

	DPA	CA	DCA
Many-to-one function	not required	required	required
Leakage model	• none for single-bit DPA • required for multi-bit DPA	none	not required, can be integrated
Statistics	based on differences	based on similarity	based on both separation and cohesion
Detected Leakage Features	• differences of two states for single-bit and "all-or-nothing" multi-bit DPA • linearity of differences for CPA	general features	general features
Multivariate Leakage	original approach can be extended	yes	yes

Comparison with DPA: An Example

Example

Assume $f_k : \{0,1\}^u \mapsto \{0,1\}^2$ is a many-to-one function.

- Single-bit DPA fails and
- Multi-bit DPA (with Hamming weight model) fails.

Does this assure that there is no leakage at all?

Comparison with DPA: An Example

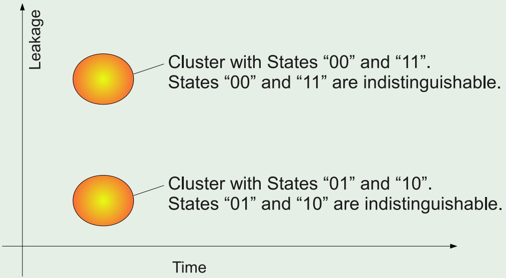
Example

Assume $f_k : \{0,1\}^u \mapsto \{0,1\}^2$ is a many-to-one function.

- Single-bit DPA fails and
- Multi-bit DPA (with Hamming weight model) fails.

Does this assure that there is no leakage at all?

No. Counter-Example:



Applications: Algorithmic Collisions

Algorithmic Collisions:

- f_k emerges from an abstract concept (e.g., cryptographic standard).
- f_k is implementation independent (despite of masking ...).

Example: DES

- DES S-box function is 4-to-1: $f_k : \{0,1\}^6 \mapsto \{0,1\}^4$ yields 2^4 clusters.

Example: AES

- AES S-box is not a collision function.
- Targeting only r-bit ($1 \leq r < 8$) of AES S-box outcome: $f_k : \{0,1\}^8 \mapsto \{0,1\}^r$ yields 2^r clusters.
- AES MixColumns transformation is $2^{24} - to - 1$: $f_k : \{0,1\}^{32} \mapsto \{0,1\}^8$ yields 2^8 clusters.

Applications: Implementation Specific Collisions

Implementation specific collisions:

- f_k emerges from implementation properties.
- f_k is not obvious in the algorithmic description.

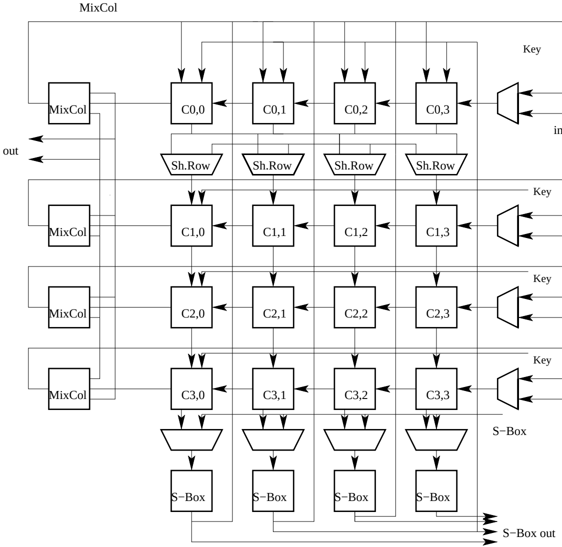
Example: AES Hardware Module

Differential of two adjacent data cells in the studied AES hardware architecture: $f_k : \{0,1\}^{16} \mapsto \{0,1\}^8$ yields 2^8 clusters.

$$f_k(x) = S(x_i \oplus k_i) \oplus S(x_{i'} \oplus k_{i'}) \tag{1}$$

General DCA Approach requires 2^{16} subkey hypotheses.

Application: AES Hardware Module



AES Hardware Module: A New Attack Strategy

If $f_k(x) = 0$ then $f_k(x) = S(x_i \oplus k_i) \oplus S(x_{i'} \oplus k_{i'})$ simplifies to

$$S(x_i \oplus k_i) = S(x_{i'} \oplus k_{i'}) \Rightarrow x_i \oplus k_i = x_{i'} \oplus k_{i'} \Rightarrow k_i \oplus k_{i'} = x_i \oplus x_{i'} .$$

The elements of one cluster are identical if $k_i \oplus k_{i'} = k_i^\circ \oplus k_{i'}^\circ$.

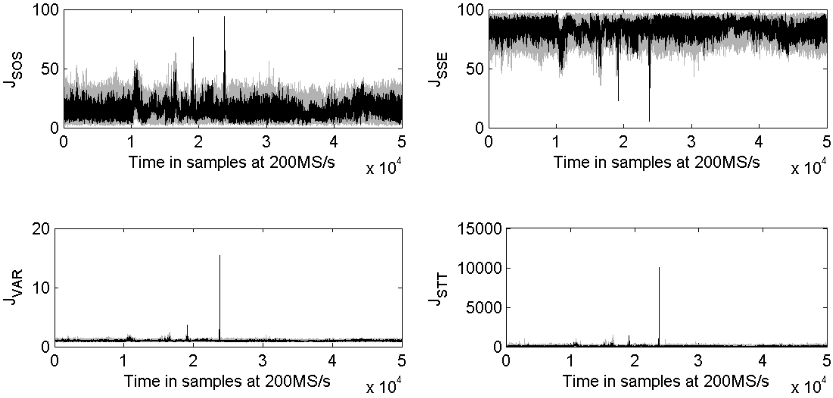
A new two-step key recovery attack:

1. Determine the correct xor-difference $k_i^\circ \oplus k_{i'}^\circ$ based on 2^8 hypotheses.
   - This is the difficult step that checks whether a special (small) cluster for $f_k(x) = 0$ exists.
2. Determine the correct pair $(k_i^\circ, k_{i'}^\circ)$ based on 2^8 hypotheses.
   - This is the easy step that checks whether up to 2^8 clusters exist.

- Attack strategy can also be applied with DPA.

DES Implementation in Software: Univariate DCA Results

Target Device: AVR Microcontroller



DES Implementation in Software: Comparison with CPA

Table: Success rates in % for various univariate and multivariate attack scenarios.

Test	Model	Time	N=15	N=20	N=25	N=30	N=40	N=50
CPA	LSB	overall	3	15	37	62	95	98
CPA	LSB	A	42	64	69	77	93	96
CPA	LSB	B	64	77	83	93	98	99
CPA	LSB	C	17	28	29	38	55	65
J_{SSE}	LSB	overall	3	15	37	62	95	98
J_{SSE}	LSB	A	42	64	70	77	93	96
J_{SSE}	LSB	B	64	78	82	93	98	99
J_{SSE}	LSB	C	18	28	31	38	56	65
CPA	LSB	AB	70	85	90	96	100	100
J_{SSE}	LSB	AB	70	83	91	97	100	100
CPA	LSB	ABC	76	90	96	99	100	100
J_{SSE}	LSB	ABC	78	94	96	99	100	100

DES Implementation in Software: Comparison with CPA

Table: Success rates in % for various univariate and multivariate attack scenarios.

Test	Model	Time	N=15	N=20	N=25	N=30	N=40	N=50
CPA	LSB	overall	3	15	37	62	95	98
CPA	LSB	A	42	64	69	77	93	96
CPA	LSB	B	64	77	83	93	98	99
CPA	LSB	C	17	28	29	38	55	65
J_{SSE}	LSB	overall	3	15	37	62	95	98
J_{SSE}	LSB	A	42	64	70	77	93	96
J_{SSE}	LSB	B	64	78	82	93	98	99
J_{SSE}	LSB	C	18	28	31	38	56	65
CPA	LSB	AB	70	85	90	96	100	100
J_{SSE}	LSB	AB	70	83	91	97	100	100
CPA	LSB	ABC	76	90	96	99	100	100
J_{SSE}	LSB	ABC	78	94	96	99	100	100

DES Implementation in Software: Comparison with CPA

Table: Success rates in % for various univariate and multivariate attack scenarios.

Test	Model	Time	N=15	N=20	N=25	N=30	N=40	N=50
CPA	LSB	overall	3	15	37	62	95	98
CPA	LSB	A	42	64	69	77	93	96
CPA	LSB	B	64	77	83	93	98	99
CPA	LSB	C	17	28	29	38	55	65
J_{SSE}	LSB	overall	3	15	37	62	95	98
J_{SSE}	LSB	A	42	64	70	77	93	96
J_{SSE}	LSB	B	64	78	82	93	98	99
J_{SSE}	LSB	C	18	28	31	38	56	65
CPA	LSB	AB	70	85	90	96	100	100
J_{SSE}	LSB	AB	70	83	91	97	100	100
CPA	LSB	ABC	76	90	96	99	100	100
J_{SSE}	LSB	ABC	78	94	96	99	100	100

AES Hardware Module: DCA Results

Target Device: Prototype chip which implements an AES-128 co-processor in 0.13 μm sCMOS technology without countermeasures.

General Approach with DCA:

- 2^{16} key hypotheses
- key recovery with approx. 5000 measurements

Two-Step Attack Strategy with DCA:

- 2^{9} key hypotheses
- key recovery with approx. 50 000 measurements
  - Step 1: 50 000 measurements
  - Step 2: 5000 measurements

Conclusion

- Introduction of Differential Cluster Analysis (DCA) – a new technique bridging the gap between Collision Analysis (CA) and Differential Power Analysis (DPA).
- Introduction of implementation specific collisions.
- Confirmation of DCA on both software (DES) and hardware (AES) implementation.
- New two-step attack strategy for an AES hardware module.

Collision attacks on AES are not constrained to 8-bit software implementations on simple controllers anymore...