Mutual Information Analysis

A Generic Side-Channel Distinguisher

Benedikt Gierlichs[1], Lejla Batina[1], Pim Tuyls[1,2], Bart Preneel[1]

[1] KU Leuven, Esat - Cosic, Belgium

[2] Philips Research Europe, The Netherlands

The key idea

- "All models are wrong, but some are useful." [George Box, 1979]
- Update to George Box's maxim: "All models are wrong, and increasingly you can succeed without them." [Peter Norvig, Google's research director, 2008]

- Google's founding philosophy is that we don't know why this page is better than that one: If the statistics of incoming links say it is, that's good enough. **No semantic or causal analysis is required.**
- […] We can analyze the data without hypotheses about what it might show. We can […] let statistical algorithms find patterns where science cannot.

[http://www.wired.com/science/discoveries/magazine/16-07/pb_theory]

Outline

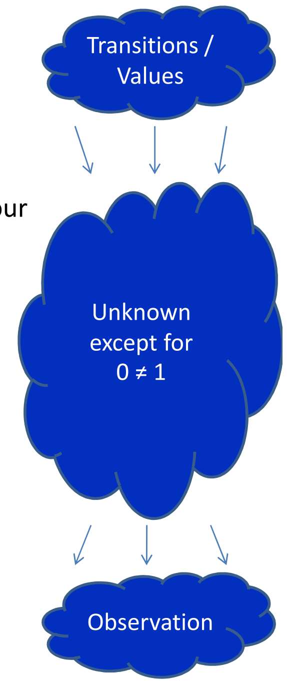
- Short history of Differential Power Analysis
- (Dis-) advantages of this direction

- Information-theoretic model for power analysis
- Our distinguisher: Mutual Information Analysis (MIA)
- Some examples

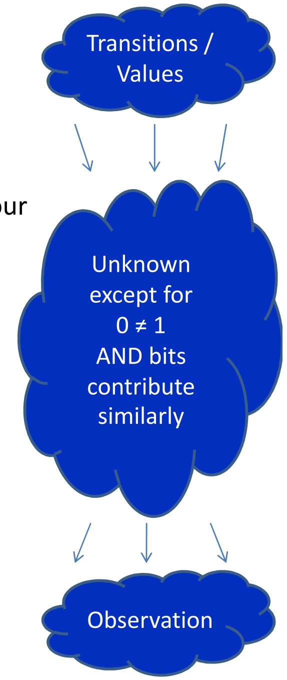
- Conclusion and future work

Background

- 1999: Single-bit DPA with DoM
  - PROs: requires few assumptions on leakage behaviour
  - CONs: algorithmic noise, PDF is reduced to its mean

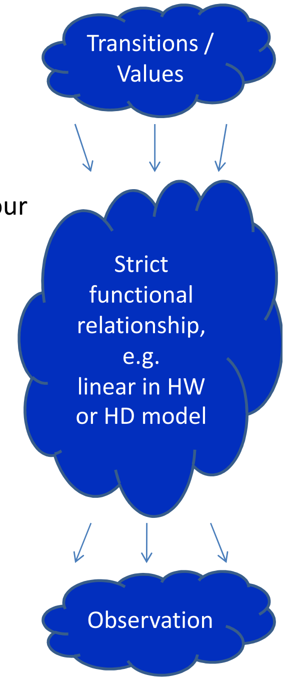


Background



- 1999: Single-bit DPA with DoM
  - PROs: requires few assumptions on leakage behaviour
  - CONs: algorithmic noise, PDF is reduced to its mean
- 2000: Multi-bit DPA with DoM
  - PROs: reduces algorithmic noise, higher SNR
  - CONs: requires more assumptions on leakage behaviour (e.g. all or nothing DPA), inefficient use of measurements

Background



- 1999: Single-bit DPA with DoM
  - PROs: requires few assumptions on leakage behaviour
  - CONs: algorithmic noise, PDF is reduced to its mean
- 2000: Multi-bit DPA with DoM
  - PROs: reduces algorithmic noise, higher SNR
  - CONs: requires more assumptions on leakage behaviour (e.g. all or nothing DPA), issue of unused measurements
- 2004: Power Model + Pearson Correlation
  - PROs: efficient
  - CONs: requires strong assumptions on leakage behaviour

PROs & CONs of this direction

+ Gradually, our models got closer to (CMOS) reality
+ A sound model allows efficient attacks and many conclusions
+ Power analysis with standard power models (HW,HD,...)

– Power model is part of the adversarial context
– Significance of negative results
  – Attack judges both: key hypotheses **and** the power model
  – Negative results are meaningless, if the power model is wrong
  – May we conclude *'secure'* if an attack doesn't work?
– What if it is hard (impossible) to set-up a reasonable model?

→ There exist no reasonable adversaries? Certainly not.

A challenging case

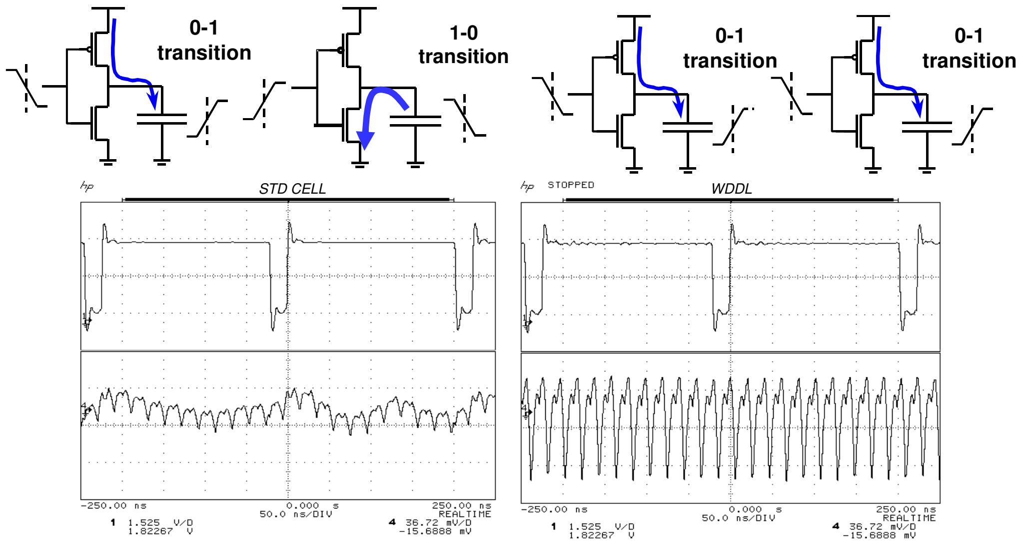
- Dynamic and differential logic (pre-charged dual rail)

1. Duplicate logic
   - Bits are encoded as tuples, e.g. 0 = (1,0) and 1 = (0,1)
2. Circuit is pre-charged, e.g. to all zero (0,0)

➢ Each DRP gate toggles exactly once per evaluation

A challenging case

CMOS vs. WDDL



Tiri, DATE 2004

A challenging case

- The number of bit flips is **constant** and **data independent**
  - ➢ Power models based on toggle count are meaningless
- Problem: imbalanced load capacitances per bit
  - – Which transition needs more power? (0,0) → (1,0) or (0,0) → (0,1) ?
  - – Random decision during Place&Route (also process variations)
  - ➢ For each single bit: 0 and 1 may be distinguishable via power consumption (but not identifiable)
  - ➢ The effect is **not** symmetric over several bits
  - ➢ Difficult to model the combined leakage of two or more bits

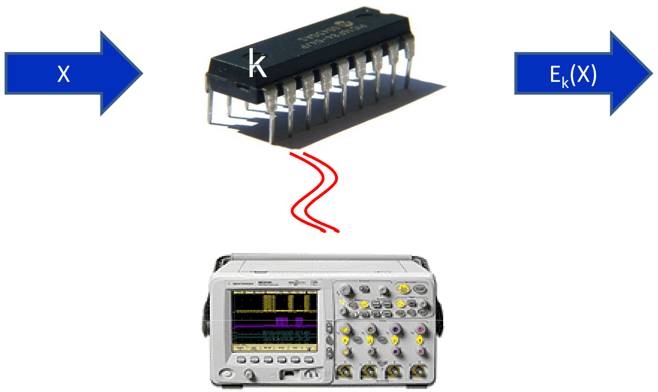
Differential Power Analysis without a restrictive power model?

- 2003, 2005, 1999: Template Attacks and the like
  - – Obtain power signature for each key dependency, attack with Bayesian inference
  - – PROs: no way to be wrong, highly efficient in attack phase
  - – CONs: requires training device and profiling step, profiling may be expensive and inefficient
- Can we do something similar without a profiling step?
  - + Attacking a single bit requires only the assumption 0 ≠ 1
  - – But ignoring other bits yields algorithmic noise
  - – Problem: how to model the combined leakage of several bits without a restrictive power model?

Information Theory Preliminaries

- Let **X** and **Y** be RV on discrete spaces $\mathcal{X}$ and $\mathcal{Y}$
- Entropy H(**X**): uncertainty about value of **X** (e.g. in bits)
- Conditional entropy H(**X**|**Y**): uncertainty about the value of **X** given the value of **Y**; cond. entropy ≤ entropy
- Mutual Information I(**X**;**Y**): reduction in uncertainty about **X** given the value of **Y**
  - – Lower bound: **X** and **Y** independent; Upper bound: **Y** fully determines **X**
  - – More Mutual Information → relation of **X** and **Y** is closer to 1:1

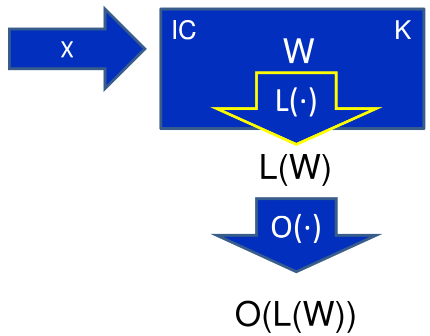
Information-Theoretic Model



Does the side-channel reduce an adversary's uncertainty about the secret key?

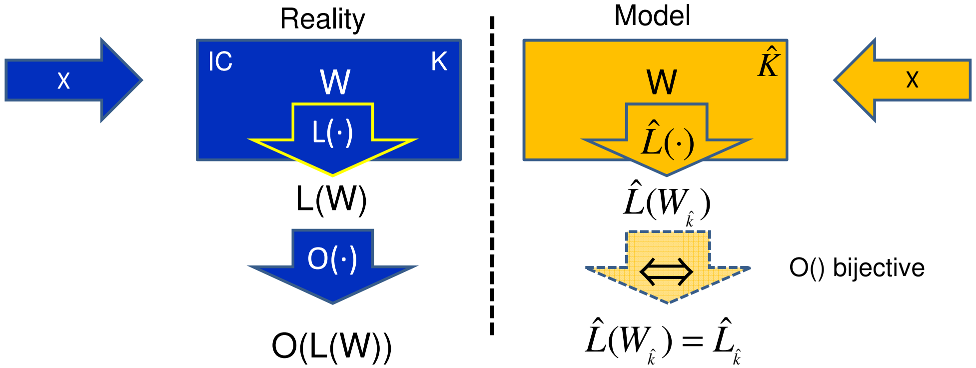
Information-Theoretic Model

- W: Transition
  given by two words (depending on X and k)
- L(·): Leakage function
  given by device properties
- L(W): Leaked values
  information that **leaks out** of the device
- O(·): Noisy observation channel
  given by measurement equipment etc.
- O(L(W)): Observations
  measurements of physical observables



- O(L(W)) depends on O, L and W (thus X and k)

Mutual Information Analysis

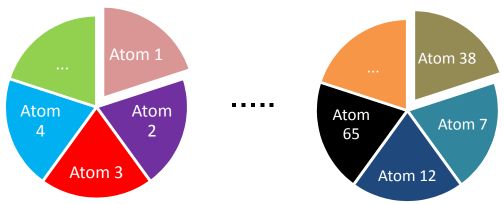


- A priori: $H(\hat{L}_{\hat{k}})$
- A posteriori: $H(\hat{L}_{\hat{k}} \mid O)$
- We learned: $I(\hat{L}_{\hat{k}}; O)$

Mutual Information Analysis

Why and how does it work?

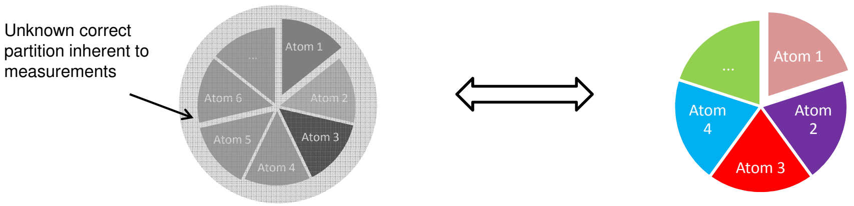
- Mutual Information compares two RV on nominal level
  - Not ratio: double L → double O
  - Not ordinal: increase L → increase O
  - Nominal: a distinct value of L → a distinct value of O
- To each key guess $\hat{k}$, we associate a partition of the space $\mathcal{L}$ of leaked values: All inputs X=x that leak the same $\hat{L}_{\hat{k}} = i$ belong to atom *i*
- Changing key guess means to re-shuffle



Mutual Information Analysis

Why and how does it work?

- A partition of $\mathcal{L}$ imposes a subdivision of $\mathcal{O}$ because each measurement is associated to an input
- Compute Mutual Information of partition and observations
  - Assess whether such partitioning leads to 1:1 relation *(order vs chaos)*



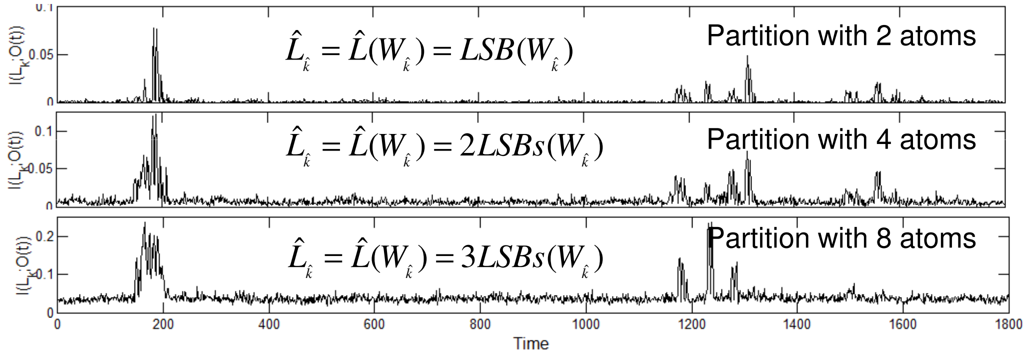
- Correct key guess leads to correct partition and maximises Mutual Information (L uniquely determines O)
- Wrong keys lead to (ideally) independent RVs

Mutual Information Analysis

Example

- AES-128, SW, 8bit µc sCMOS, Transition $W_{\hat{k}} := Z_{fix} \oplus Sbox(X \oplus \hat{k})$
- Mutual Information traces for correct key guess
- Generic leakage assumption for 1,2,3 LSBs

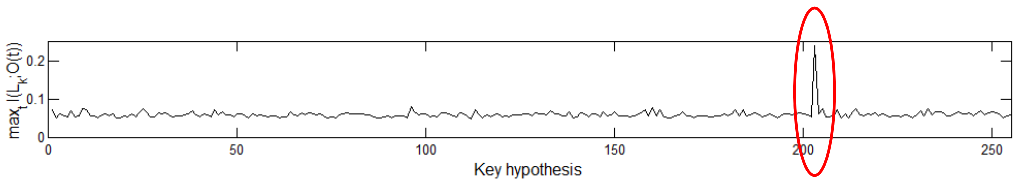
(Constant reference states are transparent in this particular case.)



Mutual Information Analysis

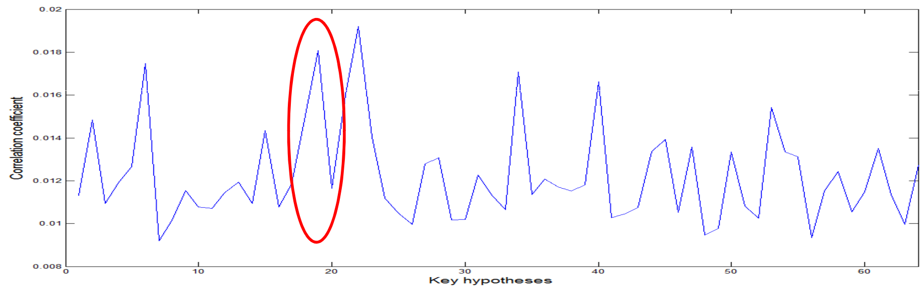
Example

- AES-128, SW, 8bit µc sCMOS, Transition $W_{\hat{k}} := Z_{fix} \oplus Sbox(X \oplus \hat{k})$
- Predictions := $\hat{L}_{\hat{k}} = \hat{L}(W_{\hat{k}}) = 3LSBs(W_{\hat{k}})$



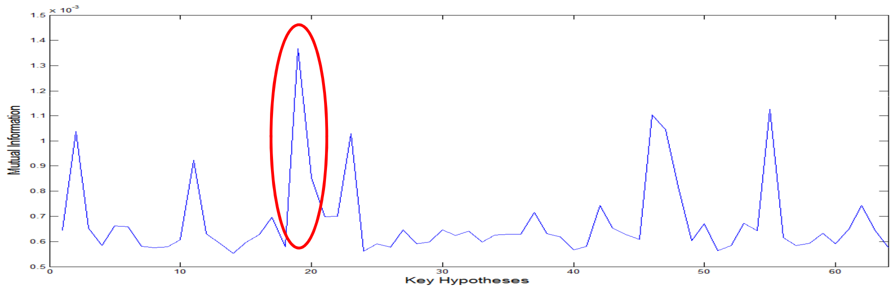
Back to the challenging case...

- 8bit µc in DRP-logic, DES Sbox S1 in software
- Targeted transition $W_{\hat{k}} = 0 \rightarrow S1(X \oplus \hat{k})$
- Correlation attack using the HW of $S1(X \oplus \hat{k})$
- 100.000 power traces



Back to the challenging case...

- 8bit µc in DRP-logic, DES Sbox S1 in software
- Targeted transition $W_{\hat{k}} = 0 \rightarrow S1(X \oplus \hat{k})$
- $\hat{L}(W_{\hat{k}}) = S1(X \oplus \hat{k})$ (every Sbox output value leaks a distinct value)
- 100.000 power traces



Conclusions

- **MIA** is a generic distinguisher for differential Side-Channel analysis
- It does not require
  - Restrictive assumptions about the device's leakage behaviour
  - The assumption that noise is Gaussian
- The price for this freedom
  - Analysis requires more data and computational power (limited increase)
- Clues about leakage behaviour and noise can be plugged in
  - ➢ Increases the efficiency
- Future work: better estimation of probability densities

Thanks for your attention!

Questions

?

benedikt.gierlichs@esat.kuleuven.be