Templates vs. Stochastic Methods

A Performance Analysis For Side Channel Cryptanalysis

B. Gierlichs[1,2,*] K. Lemke-Rust[2,**] C. Paar[2]

[1]ESAT / COSIC, Katholieke Universiteit Leuven, Belgium

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

*The research was done in cooperation with gemalto.

**Supported by the European Commission through the IST Contract IST-2002-507932 ECRYPT, the European Network of Excellence in Cryptology.

CHES 2006, Yokohama

- Given one or few power traces from an unknown implementation, what's the method of choice?

- Attacks with profiling step, previous work...
  - Inferential Power Analysis, Fahn, Pearson, CHES 1999
  - Template Attacks, Chari, Rao, Rohatgi, CHES 2002
  - Stochastic Model, Schindler, Lemke, Paar, CHES 2005

"The strongest form of side channel attack possible in an information theoretic sense" [1]

"More efficient than Templates in the profiling step but less precise in the classification step" [2]

- Given one or few power traces from an unknown implementation, what's the method of choice?

- Attacks with profiling step, previous work...
  - Inferential Power Analysis, Fahn, Pearson, CHES 1999
  - Template Attacks, Chari, Rao, Rohatgi, CHES 2002
  - Stochastic Model, Schindler, Lemke, Paar, CHES 2005

"The strongest form of side channel attack possible in an information theoretic sense" [1]

"More efficient than Templates in the profiling step but less precise in the classification step" [2]

- (sub-)key dependent operation O_i $(i = 1 \ldots K)$
- Template T_i characterization of noise in the side-channel assuming a multivariate Gaussian distribution:
- $\mathcal{P}_{O_i}(z) = \frac{1}{\sqrt{(2\pi)^p |C_i|}} \exp -\frac{1}{2}(z - m_i)^T C_i^{-1}(z - m_i)$
- Profiling (device characterization)
  - m_i by averaging
  - compute $\sum_{i,j=1}^{K} m_i - m_j$ $(j > i)$ to select p points of interest
  - C_i as empirical $(p \times p)$ covariance matrix
- Classification of a sample S
  - maximum likelihood hypothesis test
  - best candidate $O_i^* = \text{argmax}_{O_i} \mathcal{P}_{O_i}(S)$

- Choose a (small) vector subspace, e.g., $\mathcal{F}_9 \rightarrow$ linear, bitwise coefficient model [2]
- $\mathcal{P}_k(z) = \frac{1}{\sqrt{(2\pi)^p |C|}} \exp -\frac{1}{2}(z - \tilde{h}^*(x,k))^T C^{-1}(z - \tilde{h}^*(x,k))$
- Profiling (device characterization)
  - compile a system of linear equations: $b_0 \cdot \beta_0 + \cdots + b_7 \cdot \beta_7 +$ const = $\tilde{h}^*(x,k)$
  - solving the system yields a power consumption coefficient for each bit and the constant term at each instant
  - compute differential trace to select p points of interest
  - C as empirical $(p \times p)$ covariance matrix
- Classification of a sample S
  - maximum likelihood hypothesis test
  - best candidate $k^* = \text{argmax}_k \, \mathcal{P}_k(S)$

- Choose a (small) vector subspace, e.g., $\mathcal{F}_9 \rightarrow$ linear, bitwise coefficient model [2]
- $\mathcal{P}_k(z) = \frac{1}{\sqrt{(2\pi)^p |C|}} \exp -\frac{1}{2}(z - \tilde{h}^*(x,k))^T C^{-1}(z - \tilde{h}^*(x,k))$
- Profiling (device characterization)
  - compile a system of linear equations: $b_0 \cdot \beta_0 + \cdots + b_7 \cdot \beta_7 +$ const = $\tilde{h}^*(x,k)$

Example

Sample represents x = 113, k = 1, x $\oplus$ k = 112
Selection Function Sbox(x $\oplus$ k) = 81 = 01010001_2
$\tilde{h}^*(x,k) = b_6 \cdot \beta_6 + b_4 \cdot \beta_4 + b_0 \cdot \beta_0 +$ const

  - solving the system yields a power consumption coefficient for each bit and the constant term at each instant
  - compute differential trace to select p points of interest
  - C as empirical $(p \times p)$ covariance matrix
- Classification of a sample S
  - maximum likelihood hypothesis test
  - best candidate $k^* = \text{argmax}_k \mathcal{P}_k(S)$

- Choose a (small) vector subspace, e.g., $\mathcal{F}_9 \rightarrow$ linear, bitwise coefficient model [2]
- $\mathcal{P}_k(z) = \frac{1}{\sqrt{(2\pi)^p|C|}} \exp -\frac{1}{2}(z - \tilde{h}^*(x,k))^T C^{-1}(z - \tilde{h}^*(x,k))$
- Profiling (device characterization)
  - compile a system of linear equations: $b_0 \cdot \beta_0 + \cdots + b_7 \cdot \beta_7 +$ const = $\tilde{h}^*(x,k)$
  - solving the system yields a power consumption coefficient for each bit and the constant term at each instant
  - compute differential trace to select p points of interest
  - C as empirical $(p \times p)$ covariance matrix
- Classification of a sample S
  - maximum likelihood hypothesis test
  - best candidate $k^* = \text{argmax}_k \, \mathcal{P}_k(S)$

Template Attack

- signal: estimation of key-dependent signal
  $\longrightarrow$ 256 averaged signals
- noise: assumed to be key-dependent, characterized
  $\longrightarrow$ 256 covariance matrices

Stochastic Model

- signal: linear approximation of key-dependent signal in chosen subspace $\mathcal{F}_9$
  $\longrightarrow$ 9 sub-signals (8 bits + 1 non data-dependent)
- noise: assumed to be non key-dependent, characterized
  $\longrightarrow$ 1 covariance matrix

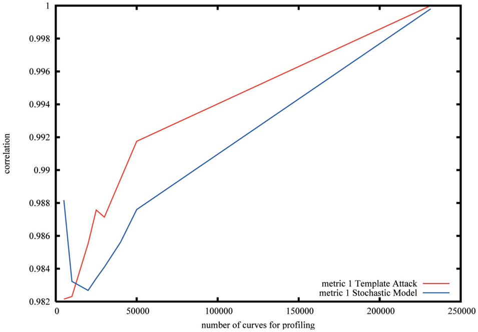
- Attack efficiency depends on (amongst others)
  - the quantity of the leakage (chip dependent)
  - the quality of the measurement setup (lab dependent)
  - the attack's ability to extract information (attack dependent)
- Selected parameters:
  - Methodical approach
  - Number of curves in the profiling step
  - Number of curves in the classification step
  - Number and composition of points of interest for multivariate noise probability density

- Metrics:
  1) Profiling, before point selection: Correlation coefficient
     $\rho_N = \frac{1}{K}\sum_{i=1}^{K} \mathrm{Corr}_e(m_{i,N}, m_{i,N_{max}})$
     ($m_{i,N}$ is approximated using $\tilde{h}_N^*(\cdot,\cdot)$ for Stochastic Methods)
  2) Profiling, at point selection: Compares the set of selected points obtained using N samples to the reference set obtained from N_{max} samples; returns the percentage of points located in the correct clock cycle
  3) Classification: success rate to obtain the correct key value

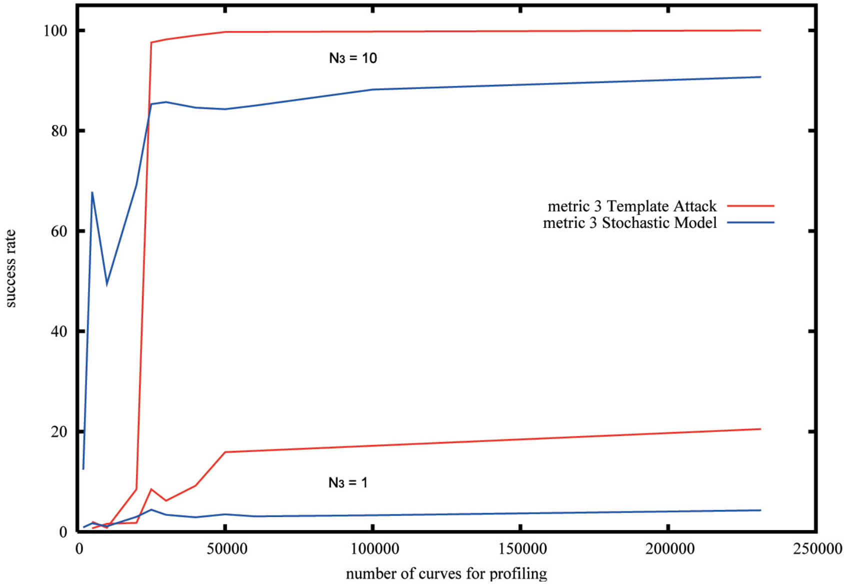
Setup	A	B (low-noise)
μc	ATMega163	Industrial Smartcard μc
Algorithm	AES-128 (software)	AES-128 (software)
Countermeasures	–	–
# of curves for		
Profiling	231k, 50k, 40k, 30k, 25k 20k, 10k, 5k, 2k[2], 1k[2], 200[2]	50k[1], 10k, 5k, 500[2], 100[2]
Classification	10, 5, 2, 1 randomly selected from 3000	5, 2, 1 randomly selected from 100
Points of interest	9, 6, 3, optimal	optimal

[1] Template attack only

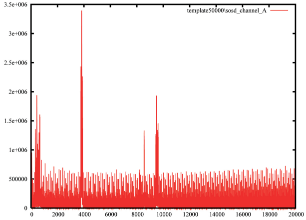
[2] Stochastic Model only, Template Attack caused numerical problems



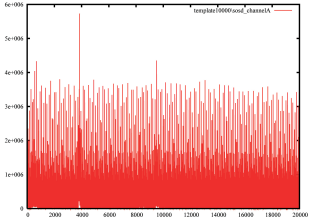
metric 2	231k	50k	40k	30k	25k	20k	10k	5k
Template Attack	1	0.89	0.89	0.78	0.67	0.56	0.23	0.23
Stochastic Model	1	1	1	1	1	1	0.67	0.78


100
80
60
40
20
0
success rate
N3 = 10
N3 = 1
metric 3 Template Attack
metric 3 Stochastic Model
0
50000
100000
150000
200000
250000
number of curves for profiling

$\sum_{i,j=1}^{K}(m_i - m_j)^{2}$ for $j \geq i$ | # samples



50.000

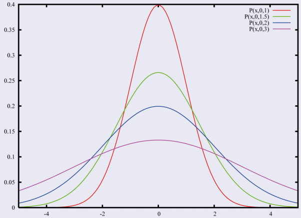


10.000

time $\longrightarrow$

$\sum_{i,j=1}^{K}(m_i - m_j)^{2}$ for $j \geq i$ # samples

T-Test

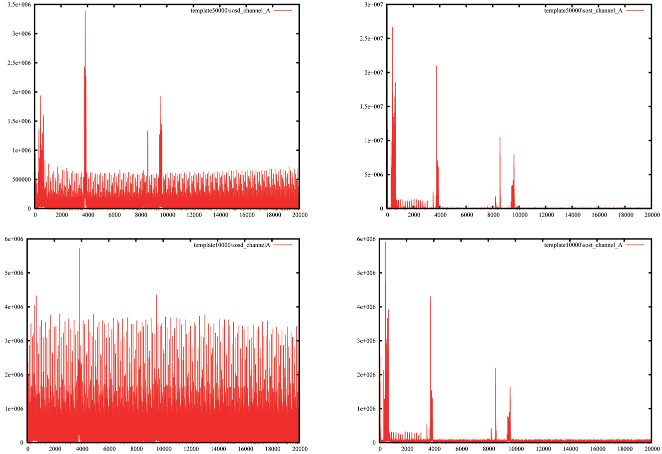


$$t = \frac{m_i - m_j}{\sqrt{\frac{\sigma_i^2}{n_i} + \frac{\sigma_j^2}{n_j}}} \approx \frac{\text{difference between group means}}{\text{variability of groups}} \approx \frac{\text{signal}}{\text{noise}}$$

$\sum_{i,j=1}^{K}(m_i - m_j)^2$ for $j \geq i$

$\sum_{i,j=1}^{K}\left(\frac{m_i - m_j}{\sqrt{\frac{\sigma_i^2}{n_i}+\frac{\sigma_j^2}{n_j}}}\right)^2$

samples



50.000

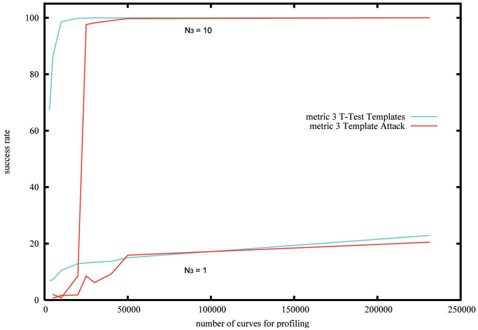
10.000

time $\longrightarrow$ time $\longrightarrow$

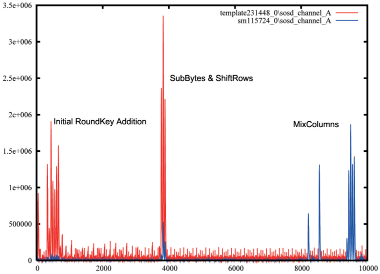
- Profiling

metric 2	231k	50k	40k	30k	20k	10k	5k
Template Attack	1	0.89	0.89	0.78	0.56	0.23	0.23
T-Test Templates	1	1	1	1	1	1	1

- Classification

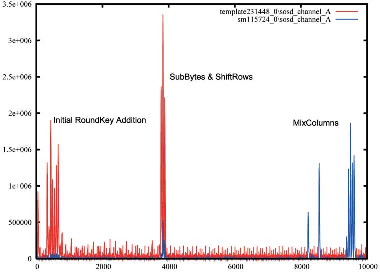


$\mathcal{F}_9$

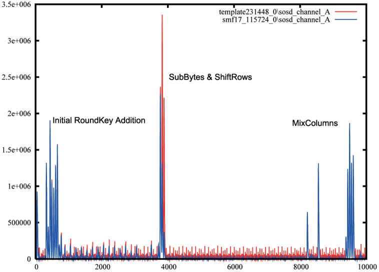


$$g_l(x \oplus k) = \left\{ \begin{array}{ll} 1 & \text{if } l = 0 \\ l\text{-th bit of S-box}(x \oplus k) & \text{if } 1 \leq l \leq 8 \end{array} \right\}$$

$\mathcal{F}_9$



$\mathcal{F}_{17}$



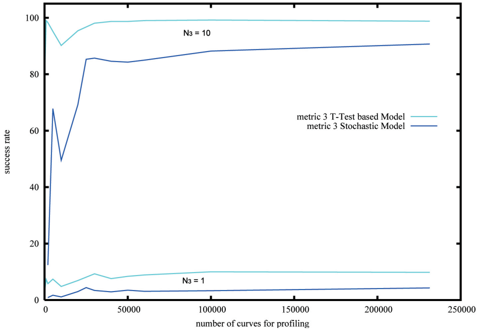
$$g_l(x \oplus k) = \begin{cases} 1 & \text{if } l = 0 \\ l\text{-th bit of S-box}(x \oplus k) & \text{if } 1 \leq l \leq 8 \\ (l-8)\text{-th bit of } x \oplus k & \text{if } 9 \leq l \leq 16 \end{cases}$$

- and T-Test based approach

- Profiling

metric 2	231k	50k	40k	30k	25k	20k	10k	5k
Stochastic Model	1	1	1	1	1	1	0.67	0.78
T-Test based Model	1	1	1	1	1	1	1	0.9

- Classification



Platform A vs. Platform B

The small print!

T-Test based Templates

metric 3		50k	10k	5k	500	100
Platform A	$N_3 = 1$	17.6	9.4	-	-	-
	$N_3 = 5$	96.7	83.0	-	-	-
Platform B	$N_3 = 1$	94.8	93.0	88.2	-	-
	$N_3 = 5$	100.0	100.0	100.0	-	-

T-Test based Stochastic Model

metric 3		50k	10k	5k	500	100
Platform A	$N_3 = 1$	-	7.2	7.7	7.3	2.8
	$N_3 = 5$	-	63.2	73.9	78.9	40.7
Platform B	$N_3 = 1$	-	57.5	60.1	46.8	27.1
	$N_3 = 5$	-	100.0	99.9	100.0	96.5

Conclusion

- Identified parameters with impact on attack efficiency
- Defined experimental framework for selected parameters
- Systematic experimental performance analysis of Template Attacks and Stochastic Model
- Experimentally verified optimizations
  - T-Test based Templates
    $\rightarrow$ increased performance towards low number of profiling samples
  - High-order T-Test based Stochastic Methods
    $\rightarrow$ increased overall performance

$\rightarrow$ T-Test based Templates are method of choice

- Work in progress:
  - what is the optimal vector subspace in an 8-bit context ?
  - efficient selection of points of interest

Thank you for your attention.

Questions?

{gierlichs, lemke, cpaar}@crypto.rub.de
benedikt.gierlichs@esat.kuleuven.be

Bibliography I

- S. Chari, J.R. Rao, P. Rohatgi: Template Attacks. In: B.S. Kaliski Jr., Ç.K. Koç, C. Paar (eds.): Cryptographic Hardware and Embedded Systems — CHES 2002, Springer, LNCS 2523, 2003, 13–28.
- W. Schindler, K. Lemke, C. Paar: A Stochastic Model for Differential Side Channel Cryptanalysis. In: J.R. Rao, B. Sunar (eds.): Cryptographic Hardware and Embedded Systems — CHES 2005, Springer, LNCS 3659, 2005, 30–46.