Exact Template Attacks with Spectral Computation

M. MAHAR M. OULADJ S. GUILLEY H. BELBACHIR F. MOKRANE

Université Paris 8, LAGA, UMR 7539, Paris, France.

CERIST, Algiers, Algeria.

Secure-IC S.A.S. , Rennes, France,
and École Normale Supérieure (ENS), Paris, France.

CATI, RECITS Laboratory, USTHB, Algiers, Algeria.

December 16-18, 2024

Plan

- Introduction;

Plan

- Introduction;
- Formal proof;

Plan

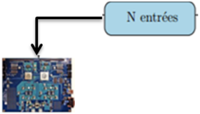
- Introduction;
- Formal proof;
- Spectral expression;

Plan

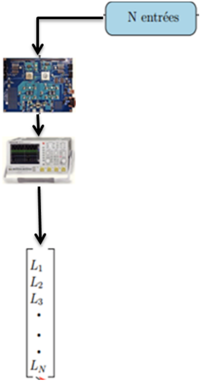
- Introduction;
- Formal proof;
- Spectral expression;
- Results and experimental validation;

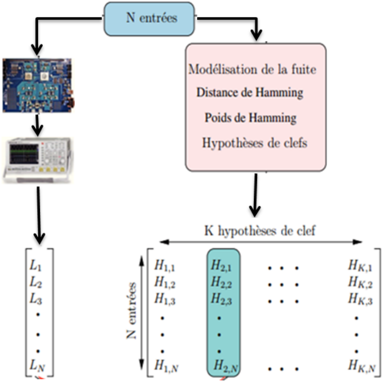
Plan

- Introduction;
- Formal proof;
- Spectral expression;
- Results and experimental validation;
- Conclusion and perspectives.

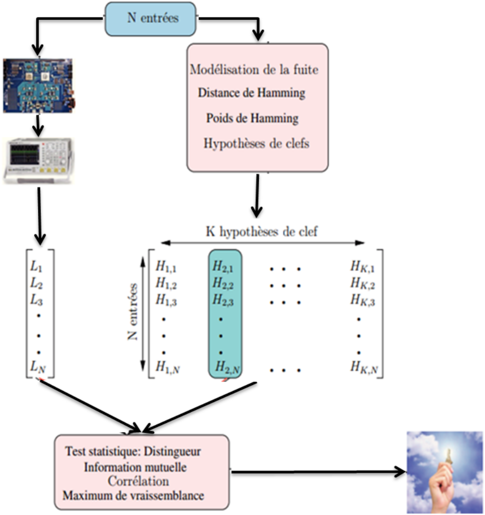


Side-Channel attacks


N entrées
L_1
L_2
L_3
L_N



Side-Channel attacks



N entrées
Modélisation de la fuite
Distance de Hamming
Poids de Hamming
Hypothèses de clefs
K hypothèses de clef
N entrées
L_1 L_2 L_3 ... L_N
$H_{1,1}$ $H_{1,2}$ $H_{1,3}$ ... $H_{1,N}$
$H_{2,1}$ $H_{2,2}$ $H_{2,3}$ ... $H_{2,N}$
$H_{K,1}$ $H_{K,2}$ $H_{K,3}$ ... $H_{K,N}$
Test statistique: Distingueur
Information mutuelle
Corrélation
Maximum de vraissemblance

SCA

Introduction: Correlation Power Attack

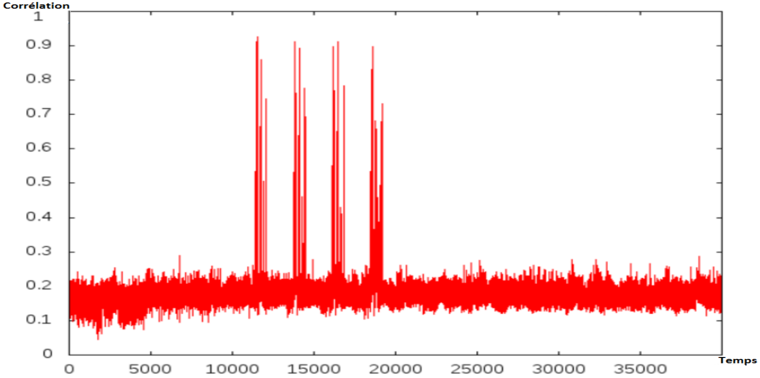
The correlation between the real leakage M and the leakage model V applied to the key k is:

$$\rho_{M,V_k} = \frac{cov(M,V_k)}{\sigma_M.\sigma_{V_k}} = \frac{\langle M,V_k\rangle}{\|M\|\|V_k\|}$$

Introduction: Correlation Power Attack

The correlation between the real leakage M and the leakage model V applied to the key k is:

$$\rho_{M,V_k} = \frac{cov(M,V_k)}{\sigma_M.\sigma_{V_k}} = \frac{\langle M,V_k \rangle}{\|M\|\|V_k\|}$$



Result of a Correlation Power Attack

Mathematical model and notations

We model the side-channel problem as follows:

- T is the digital part known by the attacker, typically some plaintext or ciphertext;

Mathematical model and notations

We model the side-channel problem as follows:

- T is the digital part known by the attacker, typically some plaintext or ciphertext;
- k is the digital part unknown by the attacker, typically some part of key, which is fixed;

Mathematical model and notations

We model the side-channel problem as follows:

- T is the digital part known by the attacker, typically some plaintext or ciphertext;
- k is the digital part unknown by the attacker, typically some part of key, which is fixed;
- Z is the vectorial Boolean function of T and k in which there is a leakage;

Mathematical model and notations

We model the side-channel problem as follows:

- T is the digital part known by the attacker, typically some plaintext or ciphertext;
- k is the digital part unknown by the attacker, typically some part of key, which is fixed;
- Z is the vectorial Boolean function of T and k in which there is a leakage;
- Y is the leakage model corresponding to Z;

Mathematical model and notations

We model the side-channel problem as follows:

- T is the digital part known by the attacker, typically some plaintext or ciphertext;
- k is the digital part unknown by the attacker, typically some part of key, which is fixed;
- Z is the vectorial Boolean function of T and k in which there is a leakage;
- Y is the leakage model corresponding to Z;
- X is the side-channel leakage measured by the attacker;

Mathematical model and notations

We model the side-channel problem as follows:

- T is the digital part known by the attacker, typically some plaintext or ciphertext;
- k is the digital part unknown by the attacker, typically some part of key, which is fixed;
- Z is the vectorial Boolean function of T and k in which there is a leakage;
- Y is the leakage model corresponding to Z;
- X is the side-channel leakage measured by the attacker;
- N is the noise.

Mathematical model and notations

We model the side-channel problem as follows:

- T is the digital part known by the attacker, typically some plaintext or ciphertext;
- k is the digital part unknown by the attacker, typically some part of key, which is fixed;
- Z is the vectorial Boolean function of T and k in which there is a leakage;
- Y is the leakage model corresponding to Z;
- X is the side-channel leakage measured by the attacker;
- N is the noise.

  Such that:

$$T, k \to Z$$

Mathematical model and notations

We model the side-channel problem as follows:

- T is the digital part known by the attacker, typically some plaintext or ciphertext;
- k is the digital part unknown by the attacker, typically some part of key, which is fixed;
- Z is the vectorial Boolean function of T and k in which there is a leakage;
- Y is the leakage model corresponding to Z;
- X is the side-channel leakage measured by the attacker;
- N is the noise.

  Such that:

$$T, k \to Z \to Y(Z) = Y$$

Mathematical model and notations

We model the side-channel problem as follows:

- T is the digital part known by the attacker, typically some plaintext or ciphertext;
- k is the digital part unknown by the attacker, typically some part of key, which is fixed;
- Z is the vectorial Boolean function of T and k in which there is a leakage;
- Y is the leakage model corresponding to Z;
- X is the side-channel leakage measured by the attacker;
- N is the noise.

  Such that:

$$T, k \to Z \to Y(Z) = Y \to X = Y + N.$$

Mathematical model and notations

- Those variables are measured many times (Q times);

Mathematical model and notations

- Those variables are measured many times (Q times);
- The noise N are all i.i.d, such that $N \sim \mathcal{N}(0, \Sigma)$;

Mathematical model and notations

- Those variables are measured many times (Q times);
- The noise N are all i.i.d, such that $N \sim \mathcal{N}(0, \Sigma)$;
- The measurements are multidimensional of dimensionality D;

Mathematical model and notations

- Those variables are measured many times (Q times);
- The noise N are all i.i.d, such that $N \sim \mathscr{N}(0, \Sigma)$;
- The measurements are multidimensional of dimensionality D;
- Oscilloscopes: D samples per trace and several samples per clock period;

Mathematical model and notations

- Those variables are measured many times (Q times);
- The noise N are all i.i.d, such that $N \sim \mathcal{N}(0, \Sigma)$;
- The measurements are multidimensional of dimensionality D;
- Oscilloscopes: D samples per trace and several samples per clock period;
- Σ: the $D \times D$ covariance matrix of the noise.

If the adversary has an identical copy of the target crypto system:

If the adversary has an identical copy of the target crypto system:
Template attack=Profiling stage + Matching stage.

Mathematical model and notations

If the adversary has an identical copy of the target crypto system:
Template attack=Profiling stage + Matching stage.

- Profiling stage: according to the the maximum likelihood principle, the model matrix Y_k and the covariance matrix Σ are estimated as:
  $Y_k = average(X_k)$; $\qquad \Sigma = \frac{1}{Q}(XX^\mathsf{T} - YY^\mathsf{T})$;

Mathematical model and notations

If the adversary has an identical copy of the target crypto system:
Template attack=Profiling stage + Matching stage.

- Profiling stage: according to the the maximum likelihood principle, the model matrix Y_k and the covariance matrix Σ are estimated as:
  $Y_k = average(X_k);\qquad \Sigma = \frac{1}{Q}(XX^\mathsf{T} - YY^\mathsf{T});$
  - $N_q = X_q - Y_{q,k}$
  - $p_{N_q}(X_q - Y_{q,k}) = \frac{1}{\sqrt{(2\pi)^Q|\Sigma|}} e^{-\frac{1}{2}(X_q - Y_{q,k})^T \Sigma^{-1}(X_q - Y_{q,k})}$

Mathematical model and notations

Template attack=Profiling stage + Matching stage.

Mathematical model and notations

Template attack=Profiling stage + Matching stage.

- Matching stage:
  - $X = Y_{k^*} + N$
  - $p_{N_q}(X_q - Y_{q,k}) = \frac{1}{\sqrt{(2\pi)^Q |\Sigma|}} e^{-\frac{1}{2}(X_q - Y_{q,k})^T \Sigma^{-1} (X_q - Y_{q,k})}$
  - Since N is independent of Y_k:

$$p(X|Y_k) = p_N(X - Y_k) = \prod_q p_{N_q}(X_q - Y_{q,k})$$

Mathematical model and notations

Template attack=Profiling stage + Matching stage.

- Matching stage:
  - $X = Y_{k^*} + N$
  - $p_{N_q}(X_q - Y_{q,k}) = \frac{1}{\sqrt{(2\pi)^Q |\Sigma|}} e^{-\frac{1}{2}(X_q - Y_{q,k})^T \Sigma^{-1} (X_q - Y_{q,k})}$
  - Since N is independent of Y_k:

$$p(X|Y_k) = p_N(X - Y_k) = \prod_q p_{N_q}(X_q - Y_{q,k})$$

Theorem (Theorem 1 of [1])

Template attacks guess the key as:

$$\hat{k} = \operatorname*{argmin}_k \operatorname{tr}((X - Y_k)^{\mathsf{T}} \Sigma^{-1} (X - Y_k)).$$

Mathematical model and notations

Template attack=Profiling stage + Matching stage.

- Matching stage:
  - $X = Y_{k^*} + N$
  - $p_{N_q}(X_q - Y_{q,k}) = \frac{1}{\sqrt{(2\pi)^Q|\Sigma|}} e^{-\frac{1}{2}(X_q - Y_{q,k})^T \Sigma^{-1} (X_q - Y_{q,k})}$
  - Since N is independent of Y_k:

$$p(X|Y_k) = p_N(X - Y_k) = \prod_q p_{N_q}(X_q - Y_{q,k})$$

Theorem (Theorem 1 of [1])

Template attacks guess the key as:

$$\hat{k} = \operatorname*{argmin}_k \operatorname{tr}((X - Y_k)^{\mathsf{T}} \Sigma^{-1} (X - Y_k)).$$

A scalability Problem!

Formal proof

the guessed key can be carried out by [5, 3]:

$$\hat{k} = \underset{k}{\operatorname{argmin}} \sum_{x=0}^{2^n-1} n_x(\tilde{L}_x - \tilde{M}_{x,k})^{\mathsf{T}} \Sigma^{-1} (\tilde{L}_x - \tilde{M}_{x,k}) \ . \tag{1}$$

Such that:

- n_x is the number of times the message x is involved,
- $\tilde{L}_x$ is the average trace over over all the traces corresponding to the same message x,
- $\tilde{M}_{x,k}$ is leakage model corresponding to the couple (x, k).
- $\Sigma = \frac{1}{N} L L^{\mathsf{T}} - \frac{1}{2^n} \tilde{M} \tilde{M}^{\mathsf{T}}$
- $\tilde{\mathbb{M}} = \Sigma^{-1} \tilde{M}$

Formal proof

the guessed key can be carried out by [5, 3]:

$$\hat{k} = \underset{k}{\operatorname{argmin}} \sum_{x=0}^{2^n-1} n_x (\tilde{L}_x - \tilde{M}_{x,k})^{\mathsf{T}} \Sigma^{-1} (\tilde{L}_x - \tilde{M}_{x,k}) \ . \tag{1}$$

Such that:

- n_x is the number of times the message x is involved,
- $\tilde{L}_x$ is the average trace over over all the traces corresponding to the same message x,
- $\tilde{M}_{x,k}$ is leakage model corresponding to the couple (x, k).
- $\Sigma = \frac{1}{N} L L^{\mathsf{T}} - \frac{1}{2^n} \tilde{M} \tilde{M}^{\mathsf{T}}$
- $\tilde{\mathbb{M}} = \Sigma^{-1} \tilde{M}$

The attack (1) is more efficient in terms of computation, and memory space, than the theorem 1, as soon as the number of traces N is greater than the number of plaintexts involved in the leakage model (e.g. for AES it is $2^n = 256$).

Formal proof

To compute (1) without using the approximation by the LLN, contrary to the state of the art, one can consider.

Proposition (Exact Template Attack – Expression of the Maximum Likelihood Distinguisher)

$$\hat{k} = \underset{k}{\operatorname{argmin}} \sum_{x=0}^{2^n-1} n_x \tilde{M}_{x\oplus k}^{\mathsf{T}} \Sigma^{-1} \tilde{M}_{x\oplus k} - 2 \sum_{x=0}^{2^n-1} (n_x \tilde{L}_x^{\mathsf{T}})(\Sigma^{-1} \tilde{M}_{x\oplus k}) .$$

Spectral expression

Recalling that, for any pair of pseudo-Boolean functions f and g, we have:

$$\sum_{x=0}^{2^n-1} f(x).g(x \oplus k) = (f \otimes g)(k) = WHT(WHT(f) \bullet WHT(g))(k),$$

where

1. "$\bullet$" denotes the direct product between two pseudo-Boolean functions (that is, the term-to-term product),
2. "$\otimes$" denotes the convolution product between two pseudo-Boolean functions,
3. WHT denotes the Walsh-Hadamard Transform.

$$WHT(f)(u) = \sum_x (-1)^{u \cdot x} f(x).$$

Spectral expression

$$
\begin{aligned}
\hat{k} &= \underset{k}{\operatorname{argmin}} \sum_{x=0}^{2^n-1} n_x \tilde{M}_{x\oplus k}^{\mathsf{T}} \Sigma^{-1} \tilde{M}_{x\oplus k} - 2 \sum_{x=0}^{2^n-1} (n_x \tilde{L}_x^{\mathsf{T}})(\Sigma^{-1} \tilde{M}_{x\oplus k}) \\
&= \underset{k}{\operatorname{argmin}}\ n(.) \otimes \mathscr{M}(.)(k) - 2 \sum_{u=1}^{D} L_{cumul}[u] \otimes \tilde{\mathbb{M}}[u](k) \\
&= \underset{k}{\operatorname{argmin}}\ WHT\Big[WHT(n) \bullet WHT(\mathscr{M}) - 2 \sum_{u=1}^{D} WHT(L_{cumul}[u]) \bullet WHT(\tilde{\mathbb{M}}[u
\end{aligned}
$$

Spectral expression

$$
\begin{aligned}
\hat{k} &= \underset{k}{\operatorname{argmin}} \sum_{x=0}^{2^n-1} n_x \tilde{M}_{x\oplus k}^{\mathsf{T}} \Sigma^{-1} \tilde{M}_{x\oplus k} - 2 \sum_{x=0}^{2^n-1} (n_x \tilde{L}_x^{\mathsf{T}})(\Sigma^{-1} \tilde{M}_{x\oplus k}) \\
&= \underset{k}{\operatorname{argmin}}\ n(.) \otimes \mathscr{M}(.)(k) - 2 \sum_{u=1}^{D} L_{cumul}[u] \otimes \tilde{\mathbb{M}}[u](k) \\
&= \underset{k}{\operatorname{argmin}}\ WHT \Big[WHT(n) \bullet WHT(\mathscr{M}) - 2 \sum_{u=1}^{D} WHT(L_{cumul}[u]) \bullet WHT(\tilde{\mathbb{M}}[
\end{aligned}
$$

So, we can carry out an exact template attack, by pre-processing $WHT(\mathscr{M})$, $WHT(\tilde{\mathbb{M}}[u])$ (for each u value), during the profiling phase, then guessing the key $\hat{k}$ accordingly.

- We employed raw traces from the SCA database (ASCAD) of the French National Agency for Information Systems Security (ANSSI) [2].

Results and experimental validation

- We employed raw traces from the SCA database (ASCAD) of the French National Agency for Information Systems Security (ANSSI) [2].
- The encryption algorithm target is a protected software implementation of AES running on an $\mathrm{ATMEGA\text{-}8515}$ μ-processor.

Results and experimental validation

- We employed raw traces from the SCA database (ASCAD) of the French National Agency for Information Systems Security (ANSSI) [2].
- The encryption algorithm target is a protected software implementation of AES running on an $\mathrm{ATMEGA\text{-}8515}$ μ-processor.
- The target variable is $Z = \mathrm{SBox}(x[2] \oplus k[2])$.

Results and experimental validation: Success rate /#traces

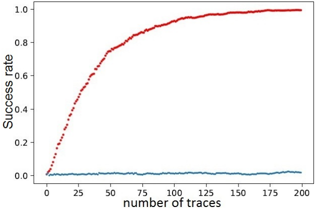


Figure: $D = 1$

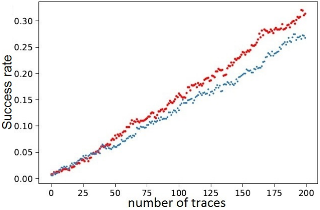


Figure: $D = 2$

Results and experimental validation: Success rate /#traces

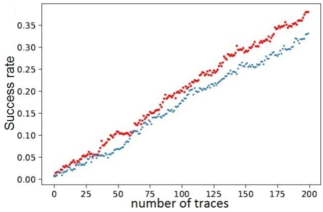


Figure: $D = 3$

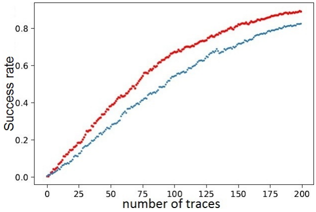


Figure: $D = 4$

Results and experimental validation: Success rate /#traces

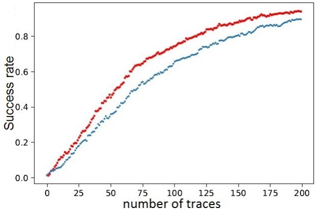


Figure: $D = 5$

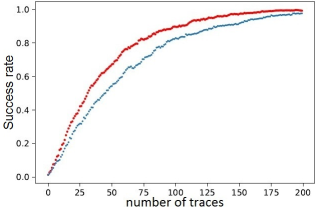


Figure: $D = 10$

Results and experimental validation: Success rate /#traces

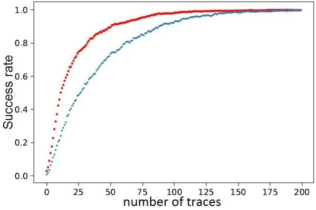


Figure: $D = 20$

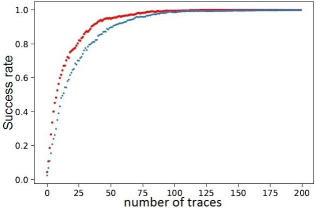
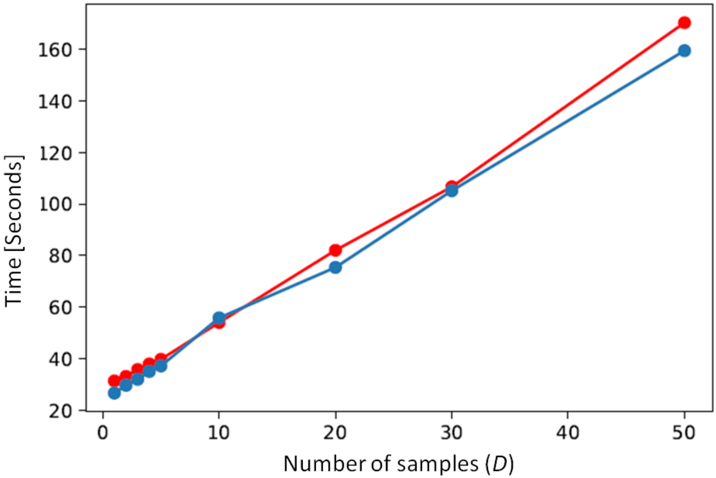


Figure: $D = 50$

Results and experimental validation: Computation time



Conclusions and perspectives

Conclusions:

- new improvement in template attacks' success rate, thanks to a spectral computation;

Conclusions and perspectives

Conclusions:

- new improvement in template attacks' success rate, thanks to a spectral computation;
- Can be applied to any algorithms that involve SBox whose input is the *XOR*;

Conclusions and perspectives

Conclusions:

- new improvement in template attacks' success rate, thanks to a spectral computation;
- Can be applied to any algorithms that involve SBox whose input is the *XOR*;
- A quasilinear instead of a quadratic time complexity (32*x* faster);
- Considerable gain in success rate comes at the expense of a marginal loss in computation time, which is explained in terms of complexity.

Conclusions and perspectives

Perspectives:

- How these improvements behave with countermeasures,

Conclusions and perspectives

Perspectives:

- How these improvements behave with countermeasures,
- This approach should be extended to the Linear Regression Analysis (LRA) in [4].

Thank you for your attention

Nicolas Bruneau, Sylvain Guilley, Annelie Heuser, Damien Marion, and Olivier Rioul.
Optimal side-channel attacks for multivariate leakages and multiple models.
J. Cryptographic Engineering, 7(4):331–341, 2017.

Prouff Emmanuel, Strullu Remi, Benadjila Ryad, Cagli Eleonora, and Dumas Cecile.
Study of deep learning techniques for side-channel analysis and introduction to ascad database.
CoRR, pages 1–45, 2018.

Maamar Ouladj and Sylvain Guilley.
Side-Channel Analysis of Embedded Systems.
Springer, 2021.
ISBN: 978-3-030-77221-5.

Maamar Ouladj, Sylvain Guilley, and Emmanuel Prouff.
On the implementation efficiency of linear regression-based side-channel attacks.

In *Constructive Side-Channel Analysis and Secure Design - 11th International Workshop, COSADE 2020, Lugano, Switzerland, October 5-7, 2020, Proceedings (LNCS 12244)*, pages 147–172, 2020.

Maamar Ouladj, Nadia El Mrabet, Sylvain Guilley, Philippe Guillot, and Gilles Millérioux.
On the power of template attacks in highly multivariate context.
Journal of Cryptographic Engineering - JCEN, 2020.