Side-Channel Expectation-Maximization Attacks

CHES 2022 Leuven

Julien Béguinot[1,2] Wei Cheng[1,2] Sylvain Guilley[1,2] Olivier Rioul[1]

[1] LTCI, Télécom Paris, Institut Polytechnique de Paris
[2] Secure-IC SAS

A Talk About Side-Channel Analysis

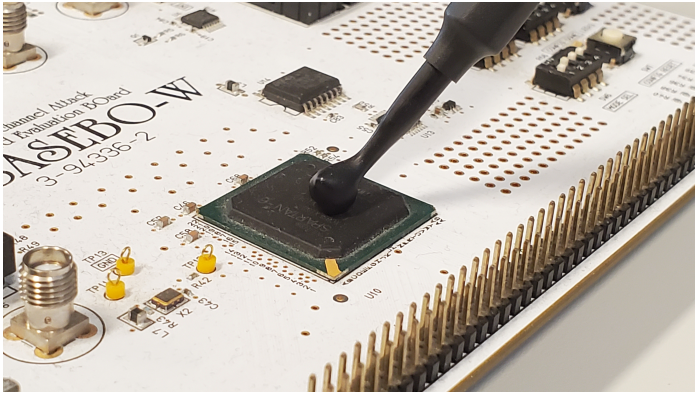


Figure: An Electromagnetic Capture for SCA

State of the Art

- Supervised
- Profiling Phase
- Parametric Model

Properties	Supervised	Profiling Phase	Parametric Model	Flexible Model
Template [Chari et al.]	✓	✓	✓	✓
P-EM [Lemke et al.]	–	✓	✓	✓
Proposed U-EM [This talk]	–	–	✓	✓
2O-CPA [Prouff et al.]	–	–	✓	–
MIA/KSA [Whitnall et al.]	–	–	–	–

We bridge the following gap. How to leverage parametric attack in presence of masking without prior profiling ? In particular the proposed U-EM outperforms 2O-CPA .

Notations and Setup

- An n-bit secret key k encrypts Q plaintext bytes $\mathrm{t} = (t_1, t_2, \ldots, t_Q)$
- The random masks $\mathrm{M} = (M_1, M_2, \ldots, M_Q)$ are i.i.d. unifom in $\mathbb{F}_2^n$
- The sensitive variable $\mathrm{X} = (X_1, X_2, \ldots, X_Q)$ is such that

$$X_q = S(k \oplus t_q) \oplus M_q \qquad (q = 1, 2, \ldots, Q) \tag{1}$$

  where S is the substitution box
- The noise is assumed Addditive White Gaussian (AWGN)
- A bivariate leakage model $\mathrm{Y} = (Y_1, \ldots, Y_Q)$ where for each $q \in \{1, 2, \ldots, Q\}$ one has $Y_q = (Y_q^{(1)}, Y_q^{(2)})$ with noise variance $\sigma^2 = (\sigma_1^2, \sigma_2^2)$

Leakage Models

- **Hamming Weight Leakage Model.**

$$\begin{cases} Y_q^{(1)} &= a^{*,(1)} w_H(X_q) + b^{*,(1)} + N_q^{(1)} \\ Y_q^{(2)} &= a^{*,(2)} w_H(M_q) + b^{*,(2)} + N_q^{(2)} \end{cases} \qquad (q = 1, 2, \ldots, Q) \tag{2}$$

  where $a^{*,(i)} \in \mathbb{R}$ and $b^{*,(i)}$ are unknown parameters.
- **Linear Leakage Model.**

$$\begin{cases} Y_q^{(1)} &= \langle a^{*(1)}, X_q \rangle + b^{*(1)} + N_q^{(1)} \\ Y_q^{(2)} &= \langle a^{*(2)}, M_q \rangle + b^{*(2)} + N_q^{(2)} \end{cases} \qquad (q = 1, 2, \ldots, Q) \tag{3}$$

  where $\langle \cdot, \cdot \rangle$ denotes a bitwise scalar product over the reals and where vector $a^{*(i)} \in \mathbb{R}^n$ and $b^{*(i)} \in \mathbb{R}$ are unknown parameters.
- **Quadratic Leakage Model.** More general leakage model, with interaction between bits:

$$\begin{cases} Y_q^{(1)} &= X_q^\top a^{*(1)} X_q + b^{*(1)} + N_q^{(1)} \\ Y_q^{(2)} &= M_q^\top a^{*(2)} M_q + b^{*(2)} + N_q^{(2)} \end{cases} \qquad (q = 1, 2, \ldots, Q), \tag{4}$$

.

Generic Notation for the Sensitive Variable

To simplify the derivations we introduce the notations.

- **Hamming Weight Model**

$$\begin{cases} x^{(1)}(a,b,k,t,m) = a \cdot w_H(m) + b \\ x^{(2)}(a,b,k,t,m) = a \cdot w_H(S(k \oplus t) \oplus m) + b \end{cases} \tag{5}$$

- **Linear Model**

$$\begin{cases} x^{(1)}(a,b,k,t,m) = \langle a, m \rangle + b \\ x^{(2)}(a,b,k,t,m) = \langle a, S(k \oplus t) \oplus m \rangle + b \end{cases} \tag{6}$$

- **Quadratic Model**

$$\begin{cases} x^{(1)}(a,b,k,t,m) = m^\top a m + b \\ x^{(2)}(a,b,k,t,m) = (S(k \oplus t) \oplus m)^\top a (S(k \oplus t) \oplus m) + b \end{cases} \tag{7}$$

Theoretical Optimal Distinguisher [Heuser et al.]

The Maximum Likelihood (ML)-based distinguisher is

$$\hat{k}(\mathrm{y}) = \arg\max_k \sum_{q=1}^{Q} \log\left[\sum_{m_q \in \mathbb{F}_2^n} \exp\left(-\frac{1}{2}\left\|y_q - x(a^*, b^*, k, t_q, m_q)\right\|^2\right)\right]. \tag{8}$$

The template based distinguisher [Chari et al.] uses the ML-based distinguisher with estimation $\hat{a}, \hat{b}$ of the parameters a^*, b^*.

2O-CPA [Prouff et al.]

For a given key hypothesis k, we write $\mathrm{x}(k) = (X(k)_1, \ldots, X(k)_Q)$ where

$$x(k)_q = \frac{1}{2^n} \sum_{m \in \mathbb{F}_2^n} w_H(m) w_H(S(k \oplus t_q) \oplus m) \qquad (q = 1, 2, \ldots, Q). \tag{9}$$

The distinguisher is then

$$\hat{k}_{\text{2O-CPA}}(\mathrm{y}) = \arg\max_k |\rho(\mathrm{x}(k), \widetilde{\mathrm{y}^{(1)}\mathrm{y}^{(2)}})| = \arg\max_k \left| \frac{\mathrm{Cov}(\mathrm{x}(k), \widetilde{\mathrm{y}^{(1)}\mathrm{y}^{(2)}})}{\sigma_{\mathrm{x}(k)} \sigma_{\mathrm{y}^{(1)}\mathrm{y}^{(2)}}} \right| \tag{10}$$

where ρ is the empirical Pearson correlation coefficient and $\widetilde{y}$ denotes the centered version of the vector y.

Expectation Maximization for SCA

$$a_{p+1}, b_{p+1} \longleftarrow \underbrace{\arg\max_{a,b}}_{\text{M-Step}} \underbrace{\mathbb{E}_{\mathsf{M}\sim\mathcal{U}(\mathbb{F}_2^n)^Q}\left[\log(\mathbb{P}(\mathsf{Y}=\mathsf{y},\mathsf{M}|k,a,b))\right]}_{\text{E-Step}} \tag{11}$$

Bayes Posterior

For fixed k and q, let

$$\beta_q^{(p)}(m) = \mathbb{P}(M_q = m) \exp\Big(-\frac{1}{2}\|y_q - x(a_p, b_p, k, t_q, m)\|^2\Big).$$

Then the Bayes posterior of the mask m being used for the q-th traces given a_p, b_p is

$$\alpha_q^{(p)}(m)(= \mathbb{P}(M_q = m|y, a_p, b_p)) = \frac{\beta_q^{(p)}(m)}{\sum_{m'} \beta_q(m')^{(p)}}.$$

.

Explicit E-Step

The E-Step at the p-th iteration is derived at follows.

$$\begin{aligned}\mathbb{E}\big[\log(\mathbb{P}(\mathrm{Y}=\mathrm{y},\mathrm{M}|a,b))\big] &= \mathbb{E}\Big[\sum_q \log(\mathbb{P}(Y_q=y_q,M_q|a,b))\Big] \\ &= \sum_q \sum_{m_q} \mathbb{P}(M_q=m_q|Y_q=y_q,a_p,b_p)\log(\mathbb{P}(Y_q=y_q,M_q=m_q|a,b)) \\ &= \sum_{q,m} \alpha_q^{(p)}(m)\log(\mathbb{P}(Y_q=y_q|M_q=m,a,b)) + cst\end{aligned} \tag{12}$$

where the constant cst is independent of a and b. Thus the E-Step of the EM reduces to

$$(a_{p+1},b_{p+1}) \leftarrow \arg\min_{(a,b)} \sum_q \sum_{m_q} \alpha_q^{(p)}(m_q)\|y_q - x(a,b,k,t_q,m_q)\|^2. \tag{13}$$

M-Step for Hamming Weight Model

Then the empirical covariance and variance are:

$$\widehat{\mathrm{Cov}_{\mathrm{xy}}^{(i)}} = \frac{1}{Q}\sum_{q,m}\alpha_q(m)y_q^{(i)}x_{k,t_q,m}^{(i)}$$

$$\widehat{\mathrm{Var}_{\mathrm{x}}^{(i)}} = \frac{1}{Q}\sum_{q,m}\alpha_q(m)y_q^{(i)2}$$

.

The M-Step is given by the following update rule ($i = 1, 2$):

$$a^{(i)} = \frac{\widehat{\mathrm{Cov}_{\mathrm{xy}}^{(i)}}}{\widehat{\mathrm{Var}_{\mathrm{x}}^{(i)}}} \qquad \textit{and} \qquad b^{(i)} = -a^{(i)}\bar{\mathrm{x}}^{(i)} \tag{14}$$

M-Step for Linear Model

The empirical autocorrelation matrix and intercorrelation vector are

$$\widehat{R_{\mathrm{xx}}^{(i)}} = \sum_{q,m} \alpha_q(m)(x_{k,t_q,m}^{(i)} - \bar{\mathrm{x}}^{(i)})(x_{k,t_q,m}^{(i)} - \bar{\mathrm{x}}^{(i)})^{\perp} \in \mathbb{R}^{n \times n}.$$

$$\widehat{R_{\mathrm{xy}}^{(i)}} = \sum_{q,m} \alpha_q(m)(x_{k,t_q,m}^{(i)} - \bar{\mathrm{x}}^{(i)}) y_q^{(i)^{\perp}} \in \mathbb{R}^{n}$$

The M-Step is given by the following update rule ($i = 1, 2$):

$$a^{(i)} = \left(\widehat{R_{\mathrm{xx}}^{(i)}}\right)^{-1} \widehat{R_{\mathrm{xy}}^{(i)}} \qquad \text{and} \qquad b^{(i)} = -\langle a^{(i)}, \bar{\mathrm{x}}^{(i)} \rangle \tag{15}$$

For the quadratic model the M-Step is performed with a gradient based optimizer.

Pseudo-Code of U-EM

Algorithm 2: Pseudo-code: U-EM-LIN.

Data: The traces $\mathbf{y} = (y_1, \ldots, y_Q)$ and the noise standard deviation $\sigma = (\sigma^{(1)}, \sigma^{(2)})$

Input: Convergence threshold ϵ

Output: Estimated key $\hat{k}$

$\bar{\mathbf{y}} \leftarrow \frac{1}{Q} \sum_{q=1}^{Q} y_q$; // Precompute the mean of the traces

$y^{(i)} \leftarrow \frac{y^{(i)} - \bar{\mathbf{y}^{(i)}}}{\sigma^{(i)}}$ for $i = 1, 2$; // Center and normalise the traces by σ

forall *key hypothesis* $k \in \mathbb{F}_2^n$ **do**

/* Initializations of the parameters a and b */

$a, b \leftarrow ((1, \ldots, 1), (1, \ldots, 1)), (0, 0)$; // Arbitrary, but could be chosen

while True **do**

/* E-Step */

forall q,m *compute* **do**

  $x_{k,t_q,m} \leftarrow (S(k \oplus t_q) \oplus m, m)$

forall q **do**

  $c_q \leftarrow \max_m -\frac{1}{2}\|y_q - \langle a, x_{k,t_q,m}\rangle - b\|^2$

forall q,m **do**

  $\beta_q(m) \leftarrow p(m)\exp(c_q - \frac{1}{2}\|y_q - \langle a, x_{k,t_q,m}\rangle - b\|^2)$

forall q *compute* **do**

  forall m *compute* **do**

    $\alpha_q(m) \leftarrow \frac{\beta_q(m)}{\sum \beta_q(m')}$

  $\widetilde{x}_q \leftarrow \sum_m \alpha_q(m) x_{k,t_q,m}$

$\bar{\mathbf{x}} \leftarrow \frac{1}{Q}\sum_{q=1}^{Q} \widetilde{x}_q$

/* M-Step */

for $i = 1,2$ **do**

$\widehat{R^{(i)}_{\mathbf{xx}}} \leftarrow \sum_{q,m} \alpha_q(m)(x^{(i)}_{k,t_q,m} - \bar{\mathbf{x}}^{(i)})(x^{(i)}_{k,t_q,m} - \bar{\mathbf{x}}^{(i)})^{\perp} \in \mathbb{R}^{n \times n}$

$\widehat{R^{(i)}_{\mathbf{xy}}} \leftarrow \sum_{q} \alpha_q(m)(\widetilde{x}^{(i)}_q - \bar{\mathbf{x}}^{(i)}) {y^{(i)}_q}^{\perp} \in \mathbb{R}^{n}$

$a'^{(i)} \leftarrow \widehat{R^{(i)}_{\mathbf{xx}}}^{-1} \widehat{R^{(i)}_{\mathbf{xy}}}, \quad b'^{(i)} \leftarrow -\langle a'^{(i)}, \bar{\mathbf{x}}^{(i)} \rangle$

if $(\|a - a'\|^2 + \|b - b'\|^2) < \epsilon$ **then**

Break ; // Exit condition

$a, b \leftarrow a', b'$

$LogLikelihood(k) \leftarrow \sum_{q=1}^{Q} \log(\sum \beta_q) + c_q$

Result: $\hat{k} = \arg\max_k \ LogLikelihood(k)$

Profiled EM [Lemke-Rust et al.]

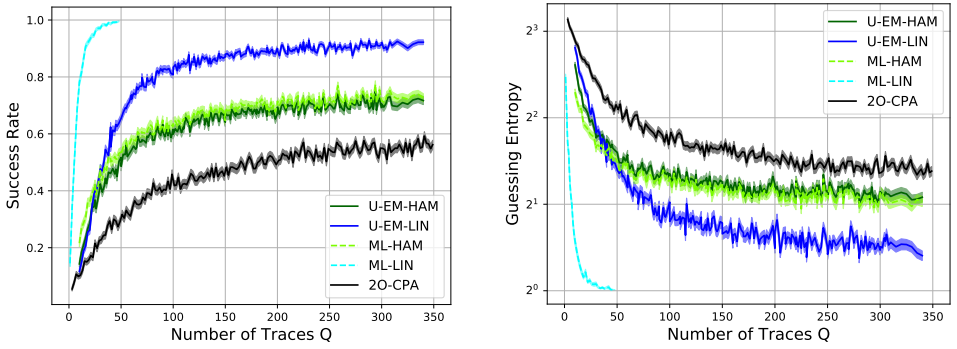
- Use EM to profile a template for each key Hypothesis.
- Require to have the necessary traces for profiling for each key hypothesis.
- Perform Maximum Likelihood with the derived templates.

Numerical Evaluation Framework and Epistemic Noise

- 4 bits PRESENT substitution box.
- Noise with standard deviation σ
- *Epistemic Noise* with standard deviation σ_a
- 10^3 independant repetitions. All attacks are performed on the same traces.
- Real Data: DPA Contest V4.2 with 8 bits Sub-Byte look up table

The epistemic noise is AWGN on the leakage coefficients and represents the physical pecularities of a leaking device.

Numerical Evaluations



(a) Success rate.

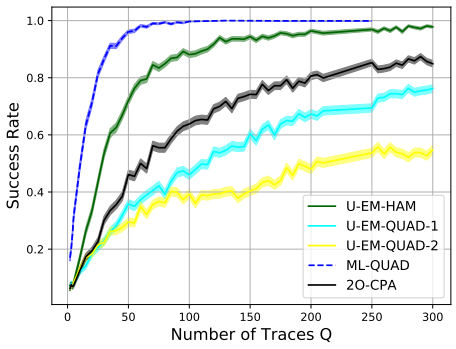
(b) Guessing entropy.

Figure: Attack metrics evaluated with $\sigma = 0.3$ and $\sigma_a = 0.8$.

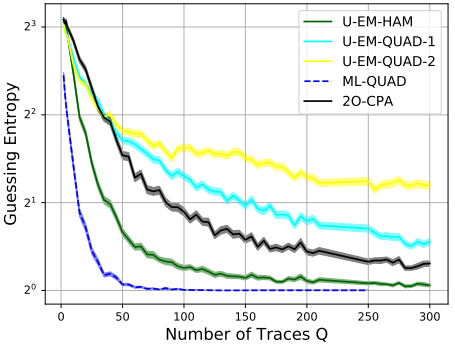
Tikhonov regularization in M-Step

The model complexity can be a burden for the U-EM. Hence it can be interesting to consider the well-known Tikhonov regularization (a.k.a ridge regression) as already suggested by [Wang et al.]. In this case we add this regularization term in the M-step.

Numerical Evaluations with Quadratic Model



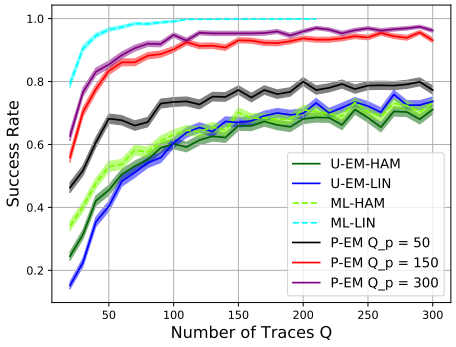
(a) Success rate.



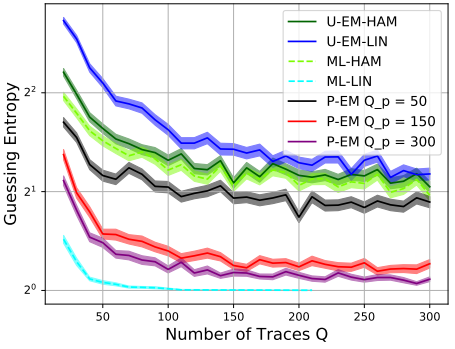
(b) Guessing entropy.

Figure: Attack metrics evaluated with $\sigma = 1$ and $\sigma_a = 0.8$.

Comparison with P-EM



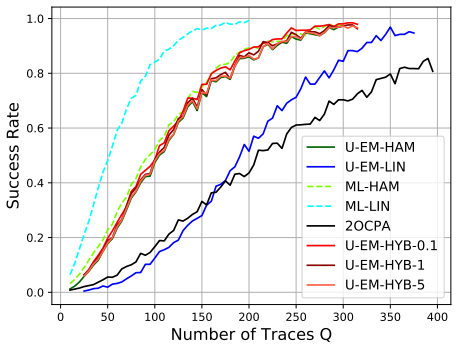
(a) Success rate on the DPA Contest.



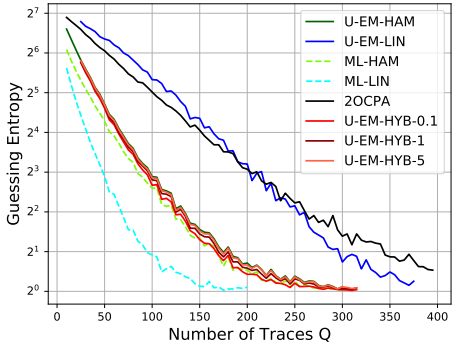
(b) Guessing entropy on the DPA Contest.

Figure: Attack metrics evaluated with $\sigma = 0.4$ and $\sigma_a = 0.4$.

Results on the Real Data from DPA Contest



(a) Success rate on the DPA Contest.



(b) Guessing entropy on the DPA Contest.

Figure: Attack metrics on the traces of the DPA Contest.

Conclusion and Perspectives

- U-EM is interesting when profiling is not feasible
- U-EM is computationally efficient
- U-EM can be easily adapted to higher order attacks and is flexible
- When profiling is possible template or P-EM are better than our U-EM as it suffers from "overfitting"
- U-EM fails to generalize to the quadratic model. Can unprofiled attack handle more complex model ? Choose best leakage basis ?

References

- Weijia Wang, Yu Yu, François-Xavier Standaert, Junrong Liu, Zheng Guo, and Dawu Gu, *Ridge-Based DPA: Improvement of Differential Power Analysis For Nanoscale Chips*, IEEE Trans. 2018
- Suresh Chari, Josyula R. Rao, and Pankaj Rohatgi, *Template attacks*, CHES 2002
- Benedikt Gierlichs, Lejla Batina, Pim Tuyls, and Bart Preneel, *Mutual information analysis*, CHES
- Kerstin Lemke-Rust and Christof Paar, *Gaussian mixture models for higher-order side channel analysis*, CHES 2007
- Houssem Maghrebi, Claude Carlet, Sylvain Guilley, and Jean-Luc Danger, *Optimal first-order masking with linear and non-linear bijections*, AFRICACRYPT 2012
- Emmanuel Prouff, Matthieu Rivain, and Régis Bevan, *Statistical Analysis of Second Order Differential Power Analysis*, IEEE Trans. Computers 58 (2009)
- Carolyn Whitnall, Elisabeth Oswald, and Luke Mather, *An Exploration of the Kolmogorov-Smirnov Test as a Competitor to Mutual Information Analysis*, CARDIS

Thank You For Your Attention

CHES 2022 Leuven

Julien Béguinot[1,2] Wei Cheng[1,2] Sylvain Guilley[1,2] Olivier Rioul[1]

[1] LTCI, Télécom Paris, Institut Polytechnique de Paris
[2] Secure-IC SAS