Maximizing the success of a side-channel attack

Authors

Annelie Heuser
Sylvain Guilley
Olivier Rioul

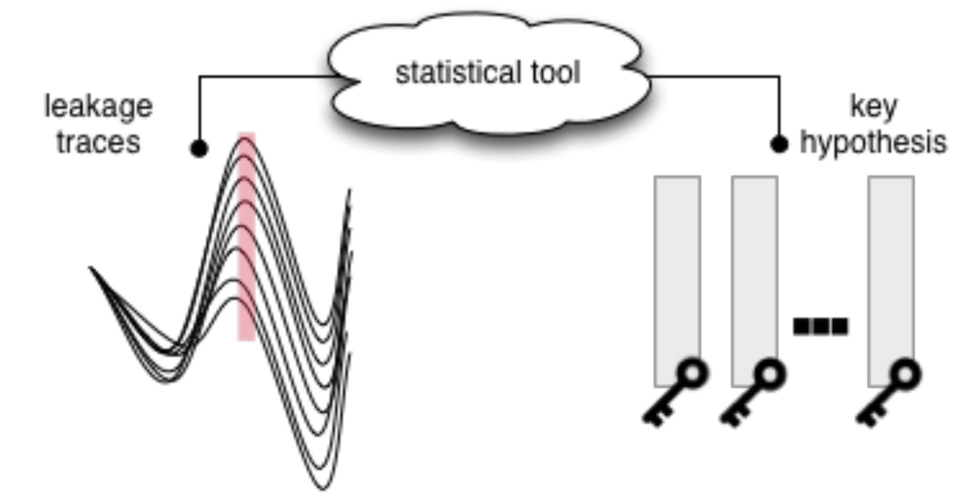
Télécom ParisTech / Institut Mines-Télécom / CNRS LTCI

Annelie Heuser is a recipient of the Google Europe Fellowship in Privacy.

State of the Art

✧ What distinguishes known distinguishers, in terms of distinctive features?
✧ Given a side-channel context, what is the best distinguisher amongst all known ones?

- Distinguishers were chosen as (arbitrary) **statistical tools** (correlation, difference of means, linear regression, etc.)
- [1] highlights that proposed distinguishers behave **equivalent** when using the same leakage model, only “statistical artifacts” can explain different behavior [2]
- The **estimation** of the statistical tools (esp. mutual information) is very crucial and effective on the success [3]

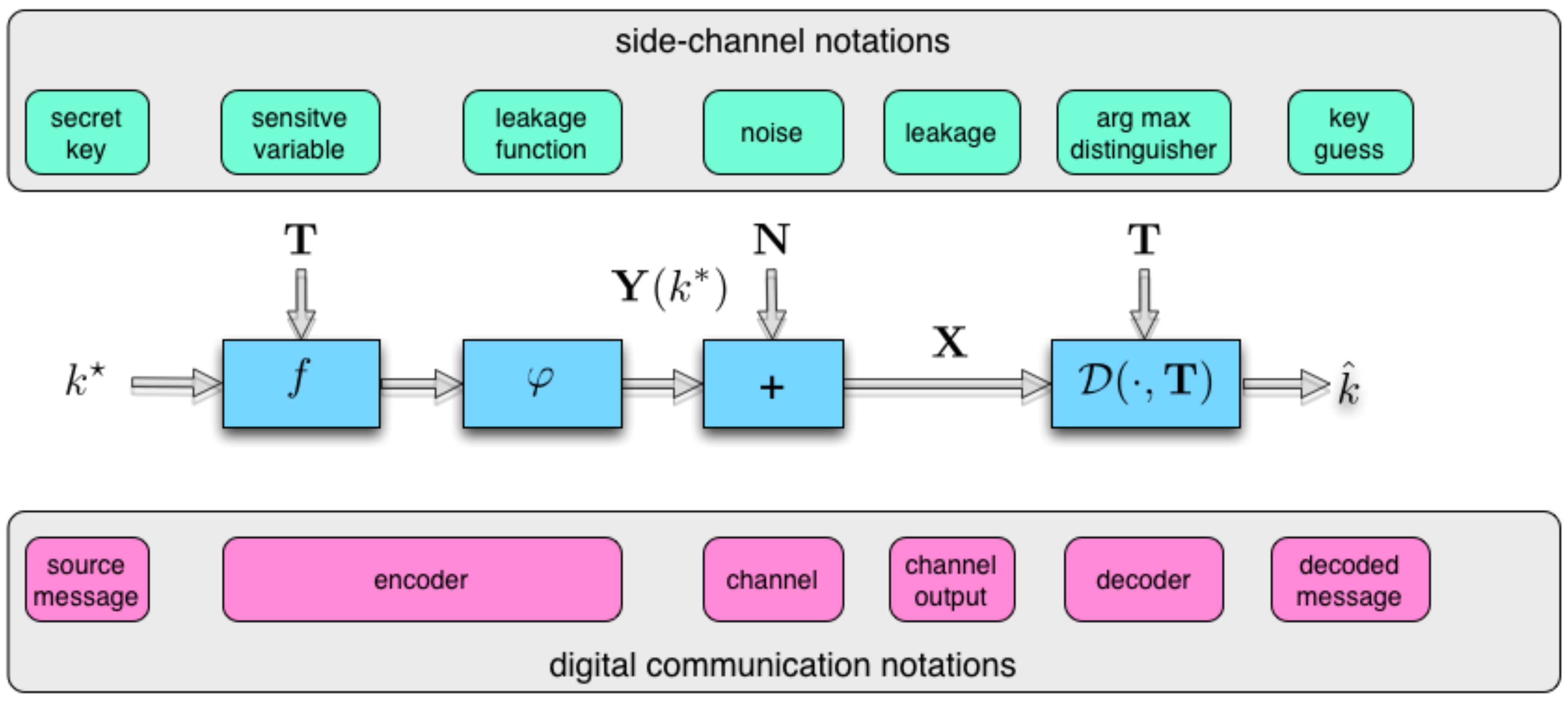


[1] Doget, Prouff, Rivain, and Standaert, JCEN, 2011
[2] Mangard, Oswald, and Standaert. IET, 2011
[3] Prouff and Rivain, IJACT, 2010.
[4] Heuser, Rioul, and Guilley, under submission

Side-channel analysis as a communication problem [4]

✧ Given a side-channel scenario, what is the best distinguisher, amongst all possible ones?

- Idea: Translate the problem of side-channel analysis into a problem of communication theory ➔ derive **optimal distinguisher**: maximize the success rate
- Leakage model is known to the attacker (**Theorem 1**)
  - Only statistical noise
  - Optimal decoding rule $\arg\max_k \Big(\mathbb{P}\{k\} \cdot p(\mathbf{x}|\mathbf{y}(k))\Big)$ (template attack, profiling is possible)
  - The optimal distinguisher only depends on the noise distribution (e.g., Laplacian, uniform, Gaussian)
- Leakage model is partially unknown to the attacker (**Theorem 2**)
  - Statistical and epistemic noise
  - Leakage arises due to a weighted sum of bits, where the weights follow a normal distribution



Theorem 1: optimal distinguisher when the leakage model is known

If the leakage arises from $X = Y(k^\star) + N$ with known leakage model $Y(k) = \varphi(f(k,T))$ then the optimal distinguishing rule are

- *Gaussian noise distribution:* $\mathcal{D}^{M,G}_{opt}(\mathbf{x},\mathbf{t}) = \arg\max_k \, \langle \mathbf{x}|\mathbf{y}(k)\rangle - \frac{1}{2}\|\mathbf{y}(k)\|_2^2$,
- *Uniform noise distribution:* $\mathcal{D}^{M,U}_{opt}(\mathbf{x},\mathbf{t}) = \arg\max_k \, -\|\mathbf{x}-\mathbf{y}(k)\|_\infty$,
- *Laplace noise distribution:* $\mathcal{D}^{M,L}_{opt}(\mathbf{x},\mathbf{t}) = \arg\max_k \, -\|\mathbf{x}-\mathbf{y}(k)\|_1$.

Theorem 2: optimal distinguisher when the leakage model is partially unknown

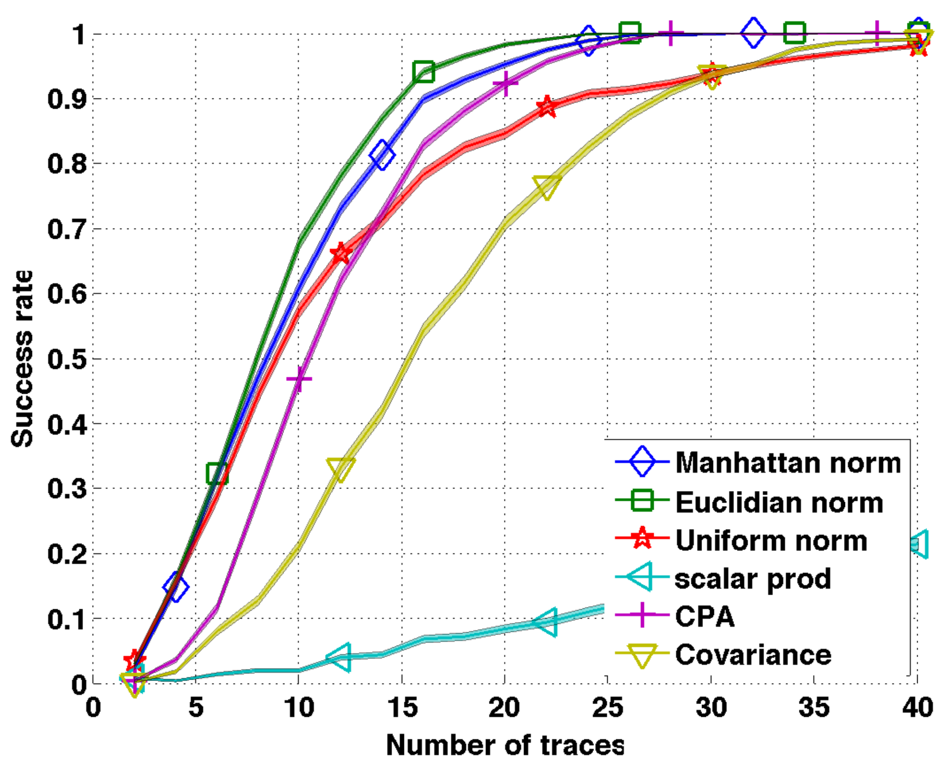
Let $\mathbf{Y}_{\boldsymbol{\alpha}}(k) = \sum_{j=1}^{n} \alpha_j [f(\mathbf{T},k)]_j$, $\mathbf{Y}_j(k) = [f(\mathbf{T},k)]_j$ *and* $\mathbf{X} = \sum_{j=1}^{n} \alpha_j [f(T,k^\star)]_j + N$ *with* $N \sim \mathcal{N}(0,\sigma^2)$. *Assuming weights are independently deviating normally from the Hamming weight model, then the optimal distinguishing rule is*

$$\mathcal{D}^{\alpha,G}(\mathbf{x},\mathbf{t}) = \arg\max_k \, (\gamma\langle \mathbf{x}|\mathbf{y}(k)\rangle + \mathbf{1})^t \cdot (\gamma Z(k) + I)^{-1} \cdot (\gamma\langle \mathbf{x}|\mathbf{y}(k)\rangle + \mathbf{1}) - \sigma_\alpha^2 \ln\det(\gamma Z(k) + I) \ ,$$

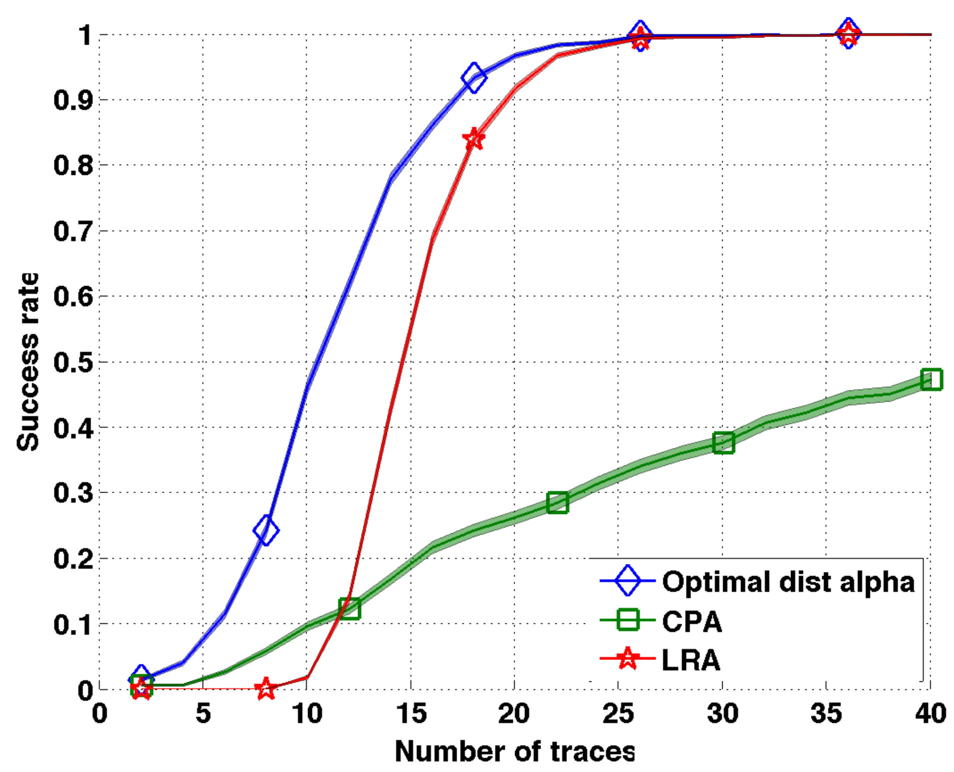
where $\gamma = \frac{\sigma_\alpha^2}{\sigma^2}$ *is the* **epistemic-to-stochastic-noise-ratio** *(ESNR).*

Correlation
Covariance
Linear regression

Our novel **optimal** distinguishers **outperform** all state-of-the-art distinguishers depending on statistical tools in terms of the **success rate**!



Known model



Partially unknown model

Contact annelie.heuser@telecom-paristech.fr Site web perso.telecom-paristech.fr/heuser