All You Ever Wanted to Know About Side-Channel Attacks and Protections

FSiC 2023, Paris, July 11, 2023

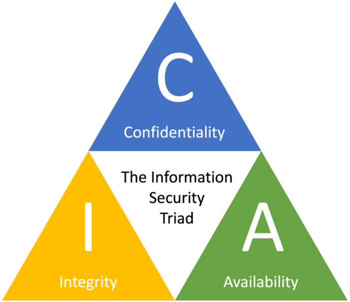
Wei Cheng[1,2], Sylvain Guilley[1,2], Olivier Rioul[1]

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

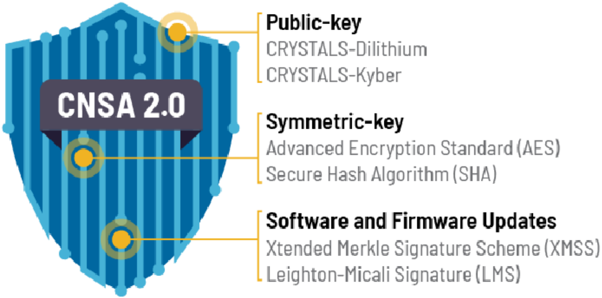
<firstname.secondname@telecom-paris.fr> or <firstname.secondname@secure-ic.fr>

Information Security

Needs

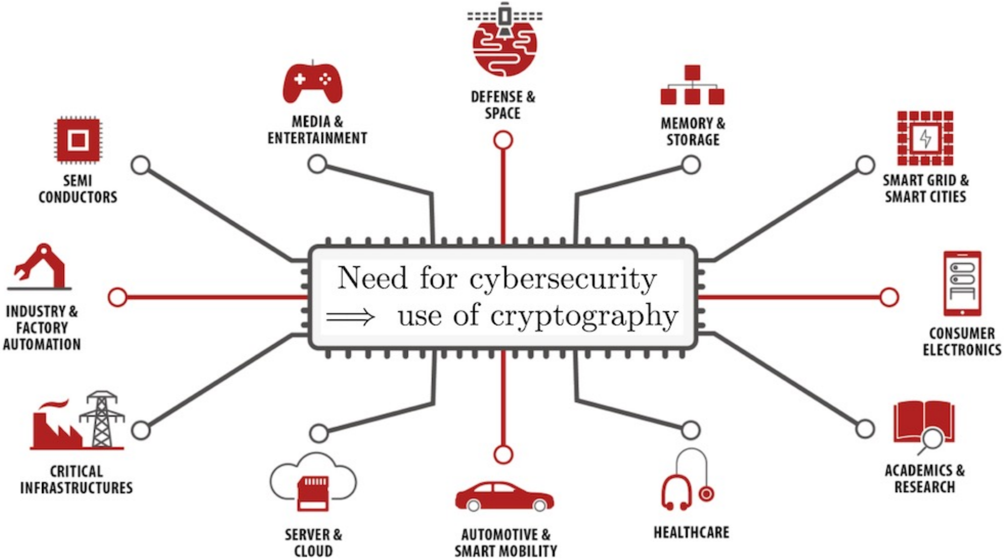


Trends



$\Longrightarrow$ Cryptographic algorithms evolve, and must be **implemented** securely.

Cryptography Is Pervasive



Information Leakage: ....Extracting RSA Keys

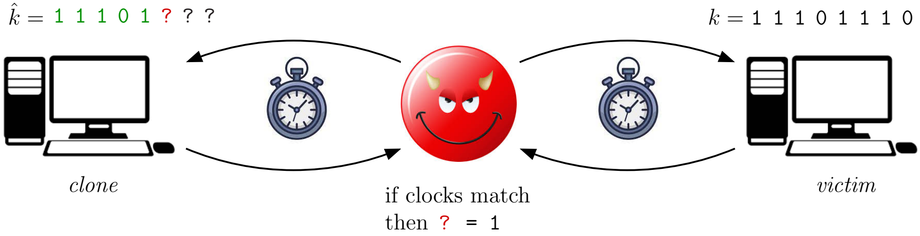
Seminal CRYPTO'96 paper: 6612 citations, till June 2023

Timing Attacks on Implementations of Diffie-Hellman, RSA, DSS, and Other Systems

Paul C. Kocher

Cryptography Consultant
P.O. Box 8243, Stanford, CA 94309, USA.
E-mail: pck@cryptography.com.

Abstract. By carefully measuring the amount of time required to perform private key operations, attackers may be able to find fixed Diffie-Hellman exponents, factor RSA keys, and break other cryptosystems.



$\Longrightarrow$ Modern cryptographic implementation is now **constant-time**.

Information Leakage: ....Extracting DES Keys

Seminal CRYPTO'99 paper: 10351 citations, till June 2023

Differential Power Analysis

Paul Kocher, Joshua Jaffe, and Benjamin Jun

Cryptography Research, Inc.
870 Market Street, Suite 1088
San Francisco, CA 94102, USA.
`http://www.cryptography.com`
E-mail: {`paul,josh,ben`}@`cryptography.com`.

Abstract. Cryptosystem designers frequently assume that secrets will be manipulated in closed, reliable computing environments. Unfortunately, actual computers and microchips leak information about the operations they process. This paper examines specific methods for analyzing power consumption measurements to find secret keys from tamper resistant devices. We also discuss approaches for building cryptosystems that can operate securely in existing hardware that leaks information.

$\Longrightarrow$ Modern cryptographic implementation should be **protected against SCAs**.

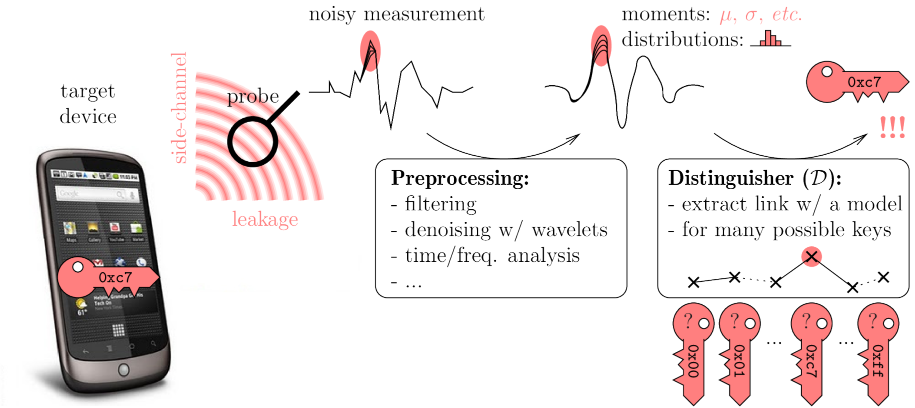
Information Leakage: ....Device Analysis

Side-channel attacks against ECDH implementation [1].

$\Longrightarrow$ Modern cryptographic implementation should be **protected against SCAs**.

[1] Genkin et al. *ECDH key-extraction via low-bandwidth electromagnetic attacks on PCs*. CT-RSA 2016.

Information Leakage: ....Device Analysis



$\Longrightarrow$ Information leakage through **power consumption**, **radiated electromagnetic field**, **clock frequency**, etc.

TELECOM ParisTech

Recent side-channel analyses ....From the remote!

Logo	Vuln. ID	Description
	CVE-2020-8694 CVE-2020-8694	With PLATYPUS, we present novel software-based power side-channel attacks on Intel server, desktop and laptop CPUs. We exploit the unprivileged access to the Intel RAPL interface exposing the processor's power consumption to infer data and extract cryptographic keys.
	CVE-2022-23823	Hertzbleed is a new family of side-channel attacks: frequency side channels. In the worst case, these attacks can allow an attacker to extract cryptographic keys from remote servers that were previously believed to be secure.
	CVE-2019-11090	They are practical. A local adversary can recover the ECDSA key from Intel fTPM in 4-20 minutes depending on the access level. We even show that these attacks can be performed remotely on fast networks, by recovering the authentication key of a virtual private network (VPN) server in 5 hours.
	CVE-2019-15809 CVE-2019-13627 CVE-2019-13627 CVE-2019-13629 CVE-2019-14318	This page describes our discovery of a group of side-channel vulnerabilities in implementations of ECDSA in programmable smart cards and cryptographic software libraries. Our attack allows for practical recovery of the long-term private key.
	CVE-2020-0549	We present CacheOut, a new speculative execution attack that is capable of leaking data from Intel CPUs across many security boundaries. SGAxe is an evolution of CacheOut, specifically targeting SGX enclaves.

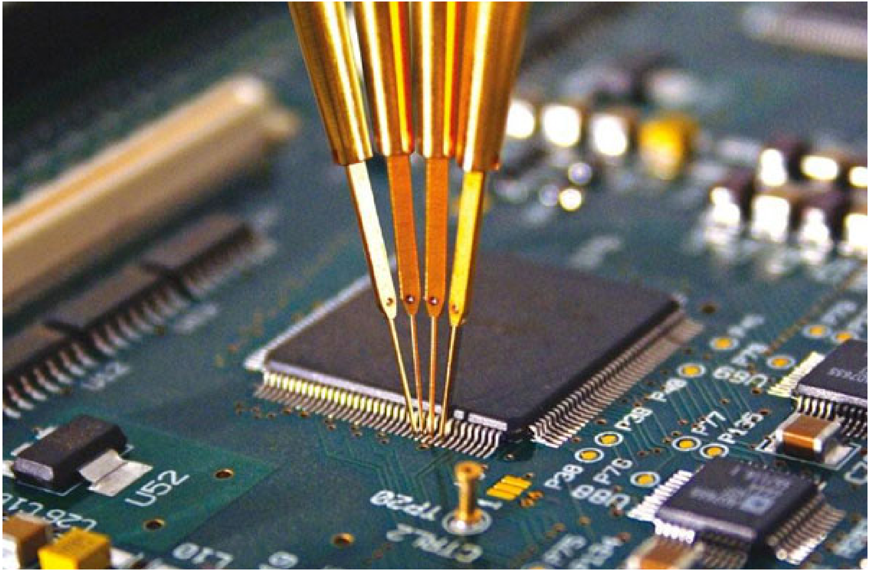
Principle

We are always "insecure" : it's a matter of time

The question is not whether you are secure or not,

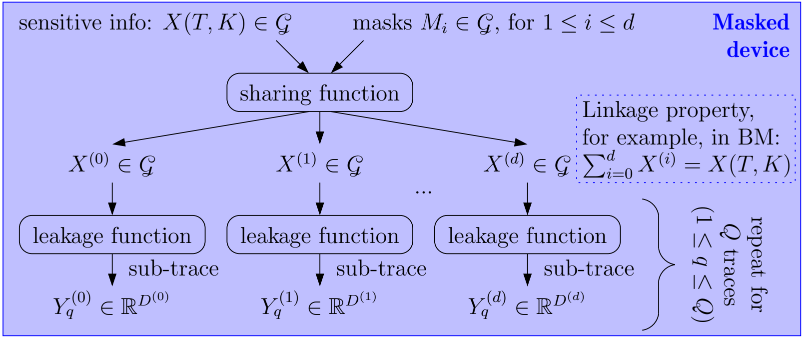
but: how much are you (in)secure?

In Practice



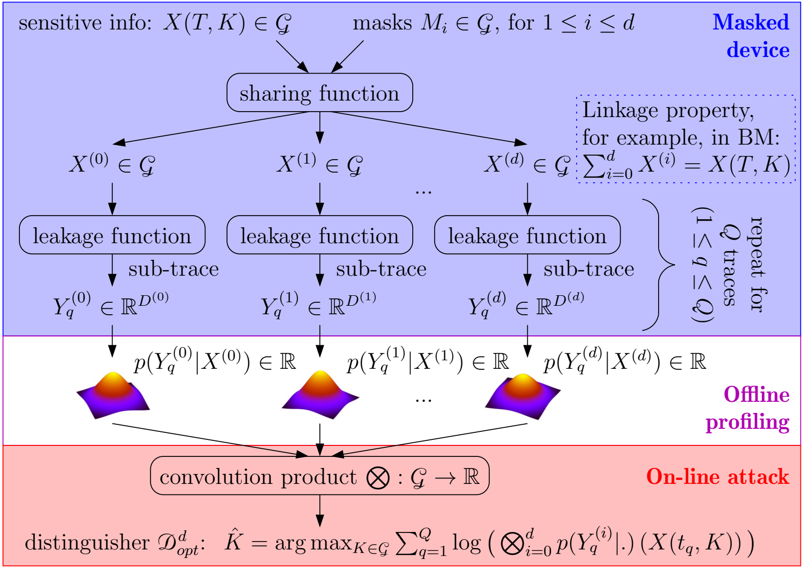
Masking as a Countermeasure

Example of Boolean Masking (BM) in $\mathcal{G} = \mathbb{Z}_{2^n}$



Masking as a Countermeasure

Example of Boolean Masking (BM) in $\mathcal{G} = \mathbb{Z}_{2^n}$



Construction of Secure Components

Security models

- probing model
- robust probing model (extended for physical effects)

Designs and proof

- follow a bottom-up design strategy
- composition strategy:
  - (S)NI: (Strong) Non-Interference
  - PINI: Probe Isolating Non-Interference

Construction of Secure Components

Security verification

- manual proof
- automated verification: maskVerif, SILVER, IronMask, etc
- automated generation of components: GHPC (Generic Hardware Private Circuits)

Security evaluation

- leakage assessment/detection
- attack-based evaluation

Security Certifications and Standards

International standards

- CC: Common Criteria: for information technology security evaluation
- ISO/IEC 19790: security requirements for a cryptographic module
- ISO/IEC 17825: specifies the non-invasive attack mitigation test metrics
- FIPS-140-2 & FIPS-140-3: security requirements for cryptographic module (US)
- etc.

Security Evaluation

Attacker's perspective

Devising the best attack:

- Optimizing success rate
- In various contexts:
  - Supervised
  - Unsupervised
- Depending on the scale of measurement
- Depending on the apriori knowledge on the Target Of Evaluation (TOE)

Defender's perspective

Normative “Vulnerability Assessment”.

Quotations, in terms of various factors:

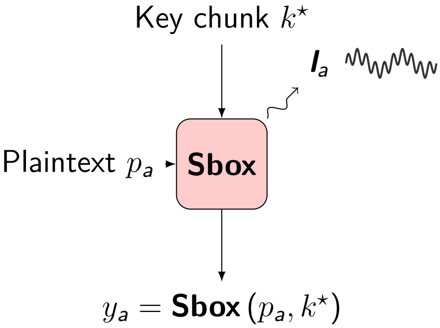
- Elapsed time
- Expertise
- Knowledge of TOE
- Window of Opportunity
- Equipment

Common Criteria ®

(ISO/IEC 15408)

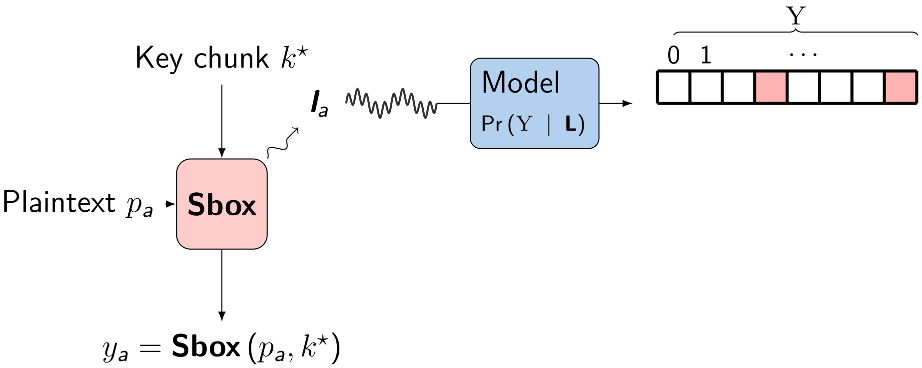
TELECOM ParisTech

In Theory



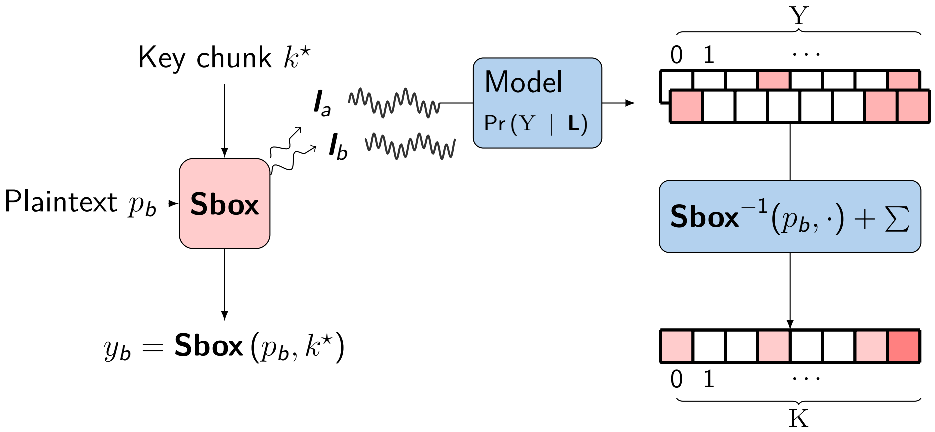
Key chunk $k^\star$
$\boldsymbol{l}_a$
Plaintext p_a
Sbox
$y_a = \textbf{Sbox}(p_a, k^\star)$

In Theory



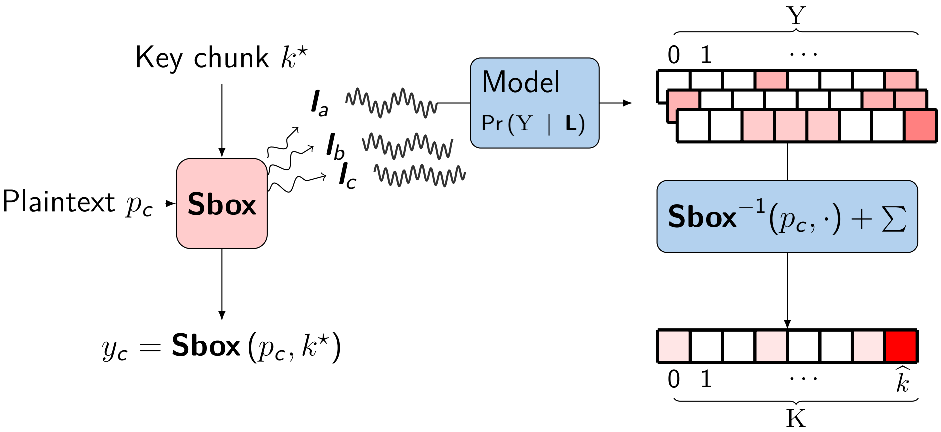
Key chunk $k^\star$
l_a
Model
$\Pr(Y \mid \mathbf{L})$
Y
0 1 ...
Plaintext p_a
Sbox
$y_a = \mathbf{Sbox}(p_a, k^\star)$

In Theory



Key chunk $k^\star$
l_a
l_b
Plaintext p_b
Sbox
$y_b = \textbf{Sbox}(p_b, k^\star)$
Model
$\Pr(Y \mid \mathbf{L})$
Y
0 1 ...
$\textbf{Sbox}^{-1}(p_b, \cdot) + \sum$
0 1 ...
K

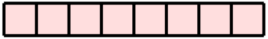
In Theory



Successful attack iff $\hat{k} = k^\star$

From Scores to Metrics

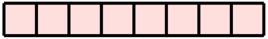
If, the adversary gets:



Sensitive computation unpredictable
SCA not more powerful than cryptanalysis
Device fully secure

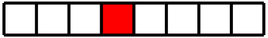
From Scores to Metrics

If, the adversary gets:



Sensitive computation unpredictable
SCA not more powerful than cryptanalysis
Device fully secure

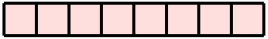
If, the adversary gets:



Exact prediction of the sensitive computation
Success rate of 100% with *one* trace
Device not secure at all

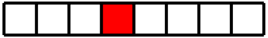
From Scores to Metrics

If, the adversary gets:



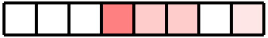
Sensitive computation unpredictable
SCA not more powerful than cryptanalysis
Device fully secure

If, the adversary gets:



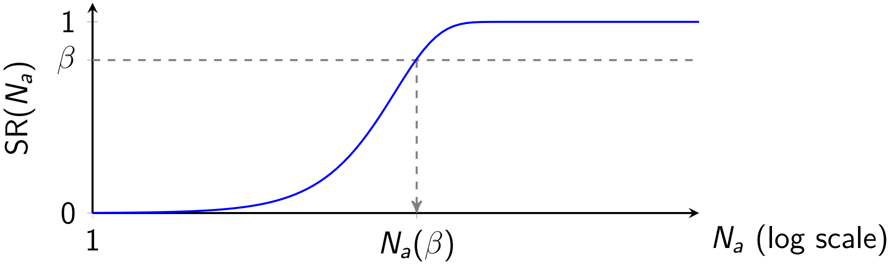
Exact prediction of the sensitive computation
Success rate of 100% with *one* trace
Device not secure at all

In general, the adversary gets:



How does this translate into SCA security metrics ?

Concrete SCA Metric: Success Rate (SR)



SR: probability to succeed the attack within N_a queries to the target

Secured device with prob. $\geq 1 - \beta$, $\implies$ refresh secret every $N_a(\beta)$ use ✓

Naive est. of $N_a(\beta)$ is expensive: complexity depends on $N_a(\beta)$ itself ✗

Concrete SCA Metric: Success Rate (SR)

Can we find surrogate metrics characterizing $N_a(\beta)$?

[1] Mangard, Oswald, and Popp, *Power analysis attacks - revealing the secrets of smart cards*
[2] Chérisey et al., "Best Information is Most Successful: Mutual Information and Success Rate in Side-Channel Analysis"

Concrete SCA Metric: Success Rate (SR)

Can we find surrogate metrics characterizing $N_a(\beta)$?

CPA [1]

Using correlation coeff.

$$N_a(\beta) \approx \frac{f(\beta)}{\rho^2}$$

- Easy to estimate ρ ✓
- Only for univariate, linear ✗

[1] Mangard, Oswald, and Popp, *Power analysis attacks - revealing the secrets of smart cards*

[2] Chérisey et al., "Best Information is Most Successful: Mutual Information and Success Rate in Side-Channel Analysis"

Concrete SCA Metric: Success Rate (SR)

Can we find surrogate metrics characterizing $N_a(\beta)$?

CPA [1]

Using correlation coeff.

$$N_a(\beta) \approx \frac{f(\beta)}{\rho^2}$$

Easy to estimate ρ ✓

Only for univariate, linear ✗

General case [2]

Using the Mutual Information (MI),

$$N_a(\beta) \geq \frac{f(\beta)}{\mathrm{MI}\,(\mathrm{Y}; \mathbf{L})}$$

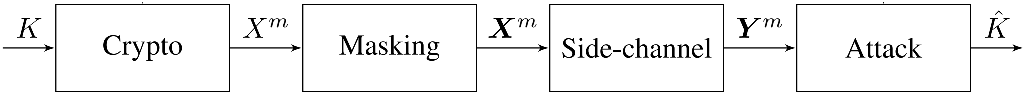
MI generalizes ρ ✓

MI hard to estimate ✗

[1] Mangard, Oswald, and Popp, *Power analysis attacks - revealing the secrets of smart cards*

[2] Chérisey et al., "Best Information is Most Successful: Mutual Information and Success Rate in Side-Channel Analysis"

Theoretical Problem



- compute sensitive values $X \sim \mathcal{U}(M)$ in an Abelian group $\mathcal{G}$ of order $M = |\mathcal{G}|$, which depends on some secret K;

Theoretical Problem

$$K \longrightarrow \boxed{\text{Crypto}} \xrightarrow{X^m} \boxed{\text{Masking}} \xrightarrow{\boldsymbol{X}^m} \boxed{\text{Side-channel}} \xrightarrow{\boldsymbol{Y}^m} \boxed{\text{Attack}} \longrightarrow \hat{K}$$

- compute sensitive values $X \sim \mathcal{U}(M)$ in an Abelian group $\mathcal{G}$ of order $M = |\mathcal{G}|$, which depends on some secret K;
- secret sharing computation: X is split into $d+1$ random shares $X_i \sim \mathcal{U}(M)$: $X = X_0 \oplus X_1 \oplus \cdots \oplus X_d$ in $\mathcal{G}$ with group operation $\oplus$;

Theoretical Problem

$$\xrightarrow{K} \boxed{\text{Crypto}} \xrightarrow{X^m} \boxed{\text{Masking}} \xrightarrow{\boldsymbol{X}^m} \boxed{\text{Side-channel}} \xrightarrow{\boldsymbol{Y}^m} \boxed{\text{Attack}} \xrightarrow{\hat{K}}$$

- compute sensitive values $X \sim \mathcal{U}(M)$ in an Abelian group $\mathcal{G}$ of order $M = |\mathcal{G}|$, which depends on some secret K;
- secret sharing computation: X is split into $d+1$ random shares $X_i \sim \mathcal{U}(M)$: $X = X_0 \oplus X_1 \oplus \cdots \oplus X_d$ in $\mathcal{G}$ with group operation $\oplus$;
- this is a dth-order masking countermeasure against noisy leakages $Y_0, \ldots, Y_d$, where the side channel $\mathbf{X} = (X_0, X_1, \ldots, X_d) \longmapsto \mathbf{Y} = (Y_0, Y_1, \ldots, Y_d)$ is memoryless;

Theoretical Problem

$$K \longrightarrow \boxed{\text{Crypto}} \xrightarrow{X^m} \boxed{\text{Masking}} \xrightarrow{\boldsymbol{X}^m} \boxed{\text{Side-channel}} \xrightarrow{\boldsymbol{Y}^m} \boxed{\text{Attack}} \xrightarrow{\hat{K}}$$

- compute sensitive values $X \sim \mathcal{U}(M)$ in an Abelian group $\mathcal{G}$ of order $M = |\mathcal{G}|$, which depends on some secret K;
- secret sharing computation: X is split into $d+1$ random shares $X_i \sim \mathcal{U}(M)$: $X = X_0 \oplus X_1 \oplus \cdots \oplus X_d$ in $\mathcal{G}$ with group operation $\oplus$;
- this is a dth-order masking countermeasure against noisy leakages $Y_0, \ldots, Y_d$, where the side channel $\mathbf{X} = (X_0, X_1, \ldots, X_d) \longmapsto \mathbf{Y} = (Y_0, Y_1, \ldots, Y_d)$ is memoryless;
- the adversary performs N_a measurements to achieve a given success rate (SR) β;

Theoretical Problem

$$\xrightarrow{K} \boxed{\text{Crypto}} \xrightarrow{X^m} \boxed{\text{Masking}} \xrightarrow{\boldsymbol{X}^m} \boxed{\text{Side-channel}} \xrightarrow{\boldsymbol{Y}^m} \boxed{\text{Attack}} \xrightarrow{\hat{K}}$$

- compute sensitive values $X \sim \mathcal{U}(M)$ in an Abelian group $\mathcal{G}$ of order $M = |\mathcal{G}|$, which depends on some secret K;
- secret sharing computation: X is split into $d+1$ random shares $X_i \sim \mathcal{U}(M)$: $X = X_0 \oplus X_1 \oplus \cdots \oplus X_d$ in $\mathcal{G}$ with group operation $\oplus$;
- this is a dth-order masking countermeasure against noisy leakages $Y_0, \ldots, Y_d$, where the side channel $\mathbf{X} = (X_0, X_1, \ldots, X_d) \longmapsto \mathbf{Y} = (Y_0, Y_1, \ldots, Y_d)$ is memoryless;
- the adversary performs N_a measurements to achieve a given success rate (SR) β;
- defender's (worst case) problem: Evaluate the *minimum number of measurements* $N_a(\beta)$ that can achieve the *best possible performance* (SR), i.e., probability of success $\beta = \mathbb{P}_s(K|\mathbf{Y}^m)$ given by the MAP rule.

Various Metrics $\Delta(X, Y)$

- how noisy is the leakage Y w.r.t. $X \sim \mathcal{U}(M)$?
- i.e., how close on average is $p_{X|Y}$ from $p_X = u$ (uniform $= \frac{1}{M}$) ?

Information Theory:

- KL divergence $D(p\|u) = \log M - H(p)$

  mutual information:

  $I(X; Y) = \mathbb{E}_Y D(p_{X|Y}\|p_X) = D(p_{XY}\|p_X \otimes p_Y)$

- Rényi divergence $D_\alpha(p\|u) = \log M - H_\alpha(p)$

  Sibson's α-information:

  $I_\alpha(X; Y) = \min_{q_Y} D_\alpha(p_{XY}\|p_X \otimes q_Y)$

  "Rényi" α-information:

  $I'_\alpha(X; Y) = D_\alpha(p_{XY}\|p_X \otimes p_Y) \geqslant I_\alpha(X; Y)$

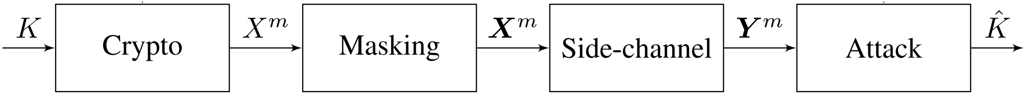
Various Metrics $\Delta(X, Y)$

- how noisy is the leakage Y w.r.t. $X \sim \mathcal{U}(M)$?
- i.e., how close on average is $p_{X|Y}$ from $p_X = u$ (uniform $= \frac{1}{M}$) ?

Statistics:

- total variation distance $\Delta_1(p, u) = \frac{1}{2}\|p - u\|_1 = \max_T |P(T) - U(T)|$
  – indistinguishability: no adversary can distinguish between p and u with advantage better than Δ_1.
  – statistical distance: $\Delta_1(X; Y) = \mathbb{E}_Y \Delta_1(p_{X|Y}, p_X) = \Delta_1(p_{XY}, p_X \otimes p_Y)$
- Euclidean bias $\Delta_2(p, u) = \|p - u\|_2^2$
  – mean-squared distance $\Delta_2(X; Y) = \mathbb{E}_Y \Delta_2(p_{X|Y}, p_X)$

Evaluation Context



- worst case security (Kerckhoffs's principle): all the implementation details are assumed known to the attacker who can even *profile* (estimate the statistical distribution of the leakage);
- with $d+1$ shares, this requires the characterization of high-order and multivariate distributions $\mathbf{Y}$, which is too expensive for high noise;
- to mitigate this difficulty, concrete evaluation practice is on

$$\Delta(X_i; Y_i) \text{ for } \textit{each share } i = 0, \ldots, d$$

instead of $\Delta(X; \mathbf{Y}) = \Delta(X_0 \oplus \cdots \oplus X_d; \mathbf{Y})$. In this way, security bounds can be derived without having to mount the complete attack.

Duc+*al* Evaluation Bound

"Making Masking Security Proofs Concrete," Duc, Faust, Standaert, Eurocrypt 2015.

Theorem (Duc+*al*, revisited)

Let $\epsilon(X_i;Y_i) = \epsilon_i$ *for each share* $i = 0, \ldots, d$*. Then*

$$N_a(\beta) \geqslant \frac{\log \frac{1-1/M}{1-\beta}}{-\log(1 - (\frac{M}{\sqrt{2\log e}})^{d+1} \prod_{i=0}^{d} \sqrt{I(X_i;Y_i)})}$$

For high noise, the denominator is $\approx (\frac{M}{\sqrt{2\log e}})^{d+1} \prod_{i=0}^{d} I(X_i;Y_i)^{1/2}$ which is too large even for moderate SNR.

Masure+*al* Evaluation Bound

"A Nearly Tight Proof of Duc et al.'s Conjectured Security Bound for Masked Implementations," Masure, Rioul, & Standaert CARDIS 2022.

Theorem (Masure+*al*)

$$N_a(\beta) \geqslant \frac{\log M - (1-\beta)\log(M-1) - h(\beta)}{\log(1 + \frac{M}{2}\prod_{i=0}^{d}\frac{2}{\log e} I(X_i;Y_i))}$$

- for high noise, the denominator is $\approx {\color{red}M}(\frac{2}{\log e})^d \prod_{i=0}^{d} I(X_i;Y_i)$ which is much improved compared to the previous one $(\frac{{\color{red}M}}{\sqrt{2\log e}})^{{\color{red}d+1}} \prod_{i=0}^{d} I(X_i;Y_i)^{{\color{red}1/2}}$
- independently, Ito et al.[1] derived the same expression with $M-1$ instead of $M/2$. Their proof uses Pinsker inequality and the Fourier transform on $\mathcal{G} = \mathbb{Z}_2^n$ (Parseval).
- still gives loose security guarantees compared to actual attacks (factor 256)

[1] Ito et al. *On the success rate of side-channel attacks on masked implementations*. CCS 2022.

Liu+*al* Evaluation Bound

“Improved Alpha-Information Bounds for Higher-Order Masked Cryptographic Implementations,” Liu, Béguinot, Cheng, Guilley, Masure, Rioul, Standaert, **ITW 2023 (St Malo, France)**

Theorem (Liu+*al*)

$$N_a(\beta) \geqslant \frac{\log M + \log(\beta^2 + (1-\beta)^2 (M-1)^{-1})}{\log\left(1 + \prod_{i=0}^{d} (\exp I'_2(X_i; Y_i) - 1)\right)}$$

- for high noise, the denominator is $\approx (\frac{1}{\log e})^d \prod_{i=0}^{d} I'_2(X_i;Y_i)$ where the alphabet size M no longer appears: improved by a large factor compared to the previous one $M(\frac{2}{\log e})^d \prod_{i=0}^{d} I(X_i;Y_i)$ although $I'_2(X_i;Y_i) \geqslant I(X_i;Y_i)$.

Béguinot+*al* Evaluation Bound

"Removing the Field Size Loss from Duc et al.'s Conjectured Bound for Masked Encodings," Béguinot, Cheng, Guilley, Liu, Masure, Rioul, Standaert, **COSADE 2023 (Munich, Germany)**

Based on **Mrs. Gerber's Lemma**, with the condition that there exists at least one $I(X_i;Y_i) < \log(2)$:

Theorem (Béguinot+*al*)

For alphabet size $M = 2^n$,

$$N_a(\beta) \geqslant \frac{\log M - (1-\beta)\log(M-1) - h(\beta)}{\varphi\left(\prod_i \varphi^{-1}(I(X_i;Y_i))\right)}$$

- for high noise (all $I(X_i;Y_i) < \log(2)$), since $\varphi(x) \approx (\frac{\log e}{2})x^2$ as $x \to 0$, the denominator is $\approx (\frac{1}{\log e})^d \prod_{i=0}^d I(X_i;Y_i)$, which is again improved compared to the previous one $(\frac{1}{\log e})^d \prod_{i=0}^d I'_2(X_i;Y_i)$.
- however, the numerator $d(\beta \| 1/M)$ is less than the previous one $d_2(\beta \| 1/M)$.

Maximal Leakage Evaluation Bound

"Maximal Leakage of Masked Implementations Using MGL for Min-Entropy," Béguinot, Liu, Rioul, Cheng, Guilley, **ISIT 2023 (Taibei, China)**

Based of a new "**Mrs. Gerber's Lemma**" for I_∞,

Theorem

For any Abelian group $\mathcal{G}$,

$$N_a(\beta) \geqslant \frac{\log(M\beta)}{\log\left(1 + c\prod_{i=0}^{d} \exp(I_\infty(X_i;Y_i)) - 1\right)}$$

- for high noise and even d, the denominator is $\approx (\frac{1}{\log e})^d \prod_{i=0}^{d} I_\infty(X_i;Y_i)$;
- the numerator $d_\infty(\beta \| 1/M)$ improves upon the preceding ones $d_2(\beta\|1/M)$ and $d(\beta\|1/M)$.

Why Do We Care?

Practical Example:

Bitslice masking: $|\mathcal{Y}| = 2$, Leakage model: $\mathbf{L}_i = \mathrm{hw}(Y_i) +$ Noise$(0, \sigma^2)$

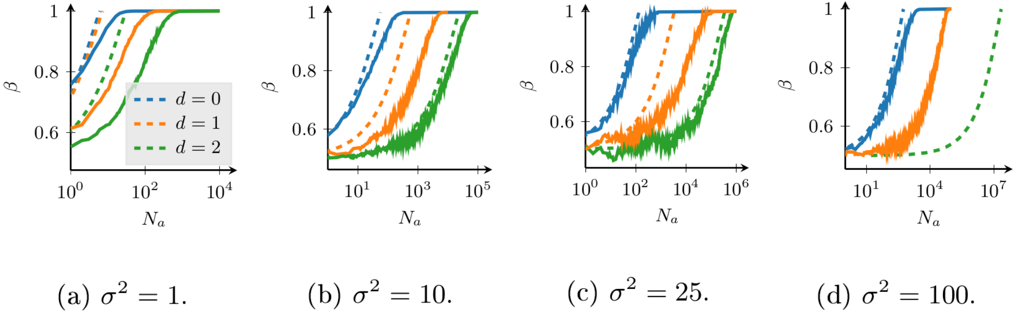


Figure: Success rate of concrete bit recoveries and MI-based upper bounds.

Conclusions

- Crypto-Analysis is mathematical
- Side-Channel Analysis is physical
- Thanks to information theory, we manage to provide formal guarantees
- The key to certification is security-by-design
- A book to be published in Q1 2024 at Springer/Nature:
  - mathematical foundation of security guarantees
  - derivation of optimal attacks
  - easy evaluations for side-channel resilience.

All You Ever Wanted to Know About Side-Channel Attacks and Protections

Thank you!

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

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

<firstname.secondname@telecom-paris.fr> or <firstname.secondname@secure-ic.fr>