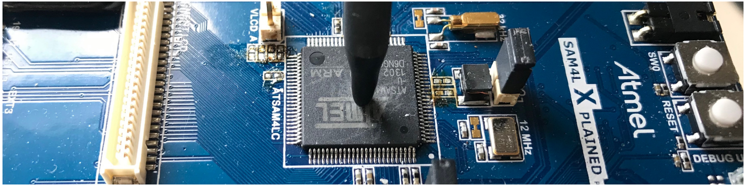


A Nearly Tight Proof of Duc *et al.*'s Conjectured Security Bound for Masked Implementations

Loïc Masure Olivier Rioul François-Xavier Standaert

CARDIS 2022, Birmingham, November 7th
https://ia.cr/2022/600

European Research Council
Established by the European Commission
GROUPE CRYPTO
UCLouvain

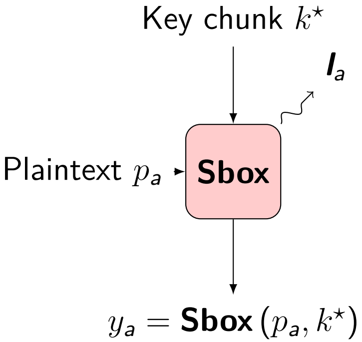
If You Are Interested...

Same result obtained independently by Akira Ito, Rei Ueno, Naofumi Homma.
To be presented at CCS 2022 (`https://ia.cr/2022/576`)

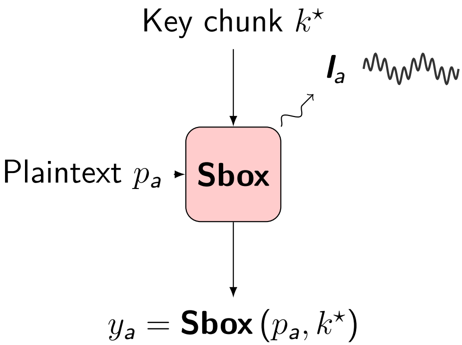
Table of Contents

Content

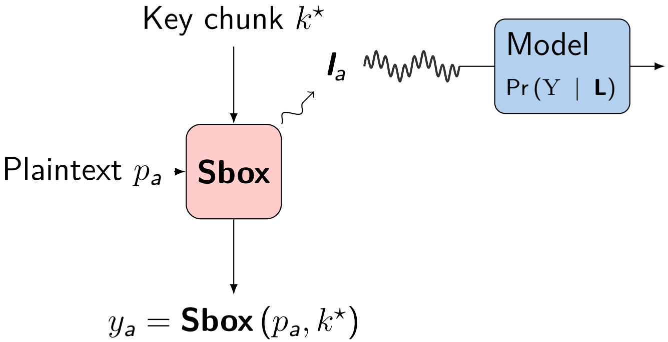
How does an Side-Chanel Analysis (SCA) work



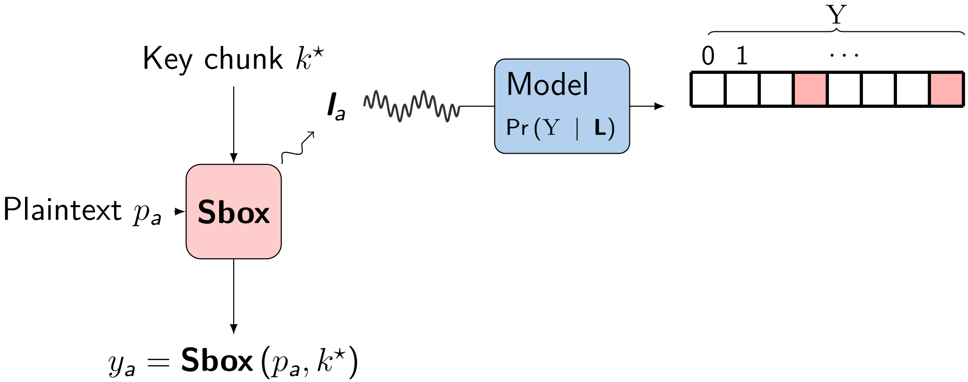
How does an SCA work



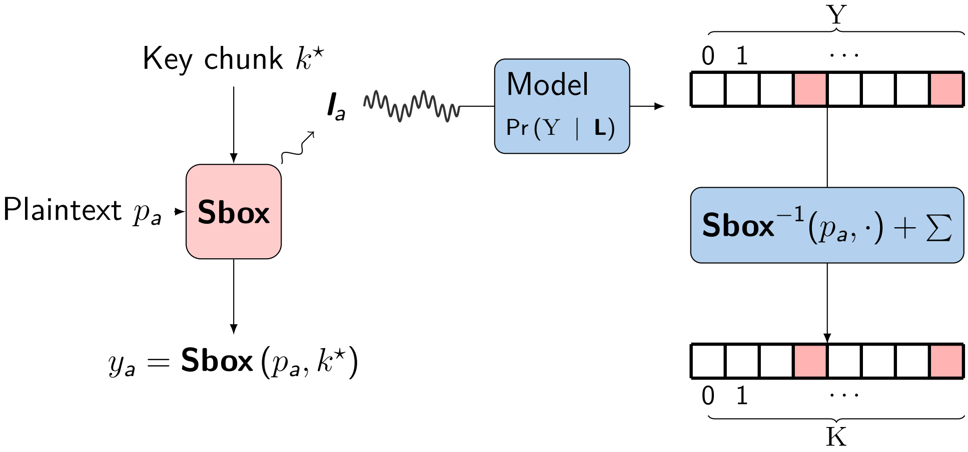
How does an SCA work



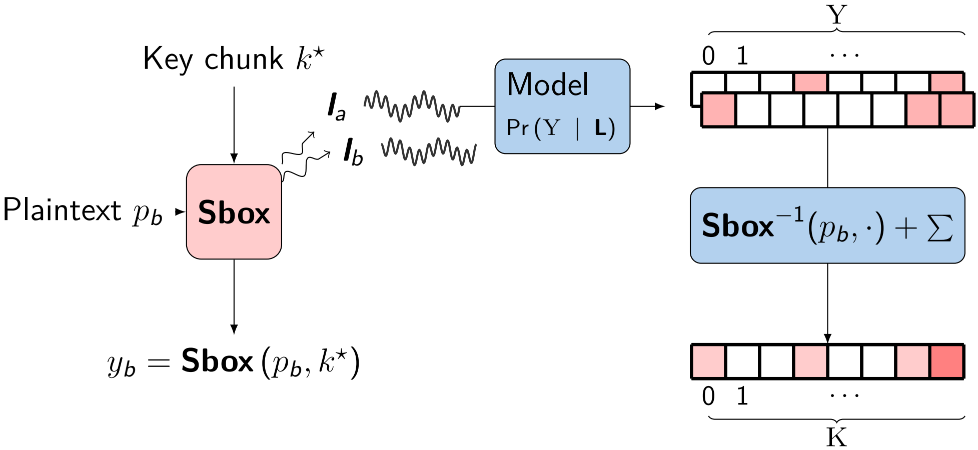
How does an SCA work



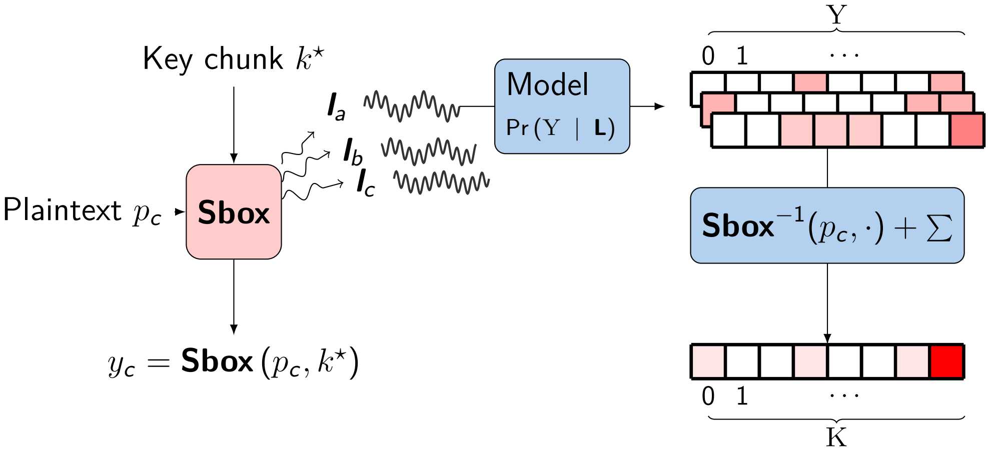
How does an SCA work



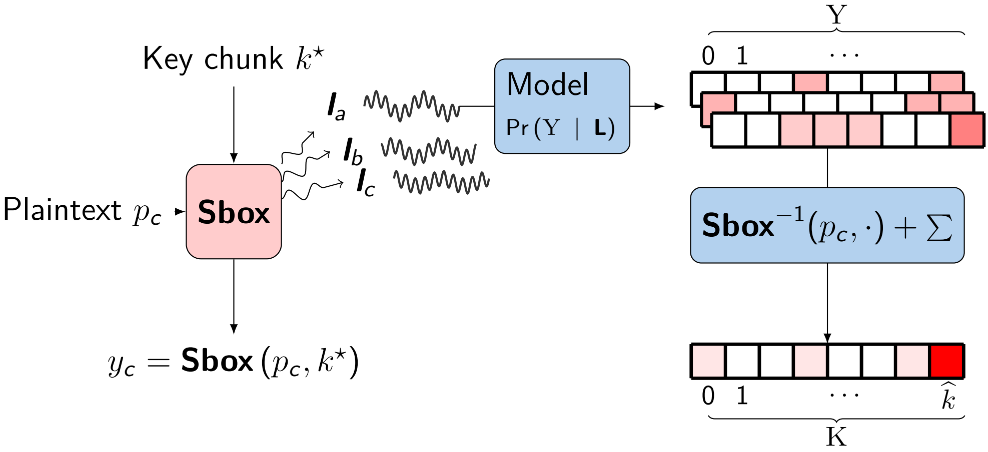
How does an SCA work



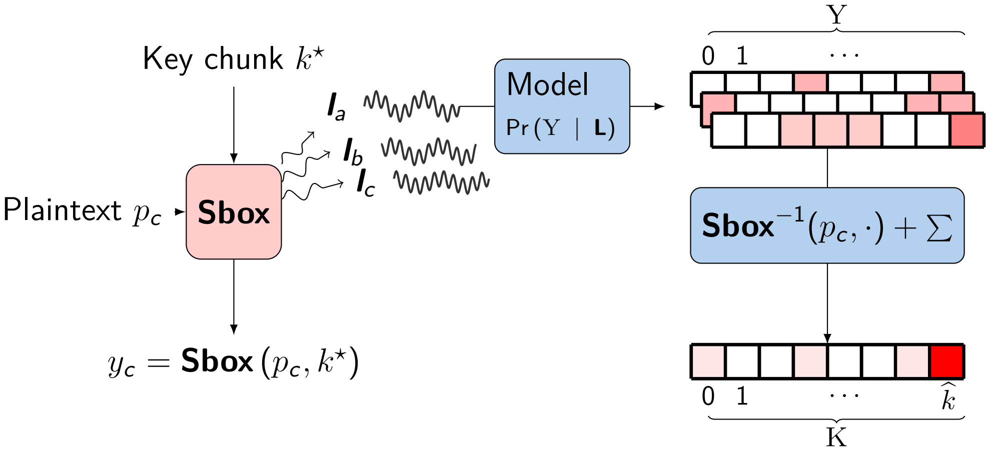
How does an SCA work



How does an SCA work



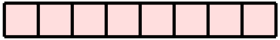
How does an SCA work



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

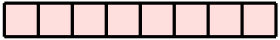
From scores to Metrics

If, the adversary gets:



From scores to Metrics

If, the adversary gets:



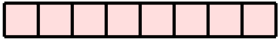
Sensitive computation unpredictable

SCA not more powerful than cryptanalysis

Device fully secure

From scores to Metrics

If, the adversary gets:



Sensitive computation unpredictable

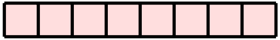
SCA not more powerful than cryptanalysis

Device fully secure

If, the adversary gets:

From scores to Metrics

If, the adversary gets:



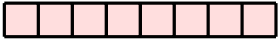
Sensitive computation unpredictable
SCA not more powerful than cryptanalysis
Device fully secure

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

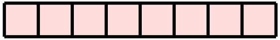
If, the adversary gets:

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

In general, the adversary gets:

From scores to Metrics

If, the adversary gets:



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

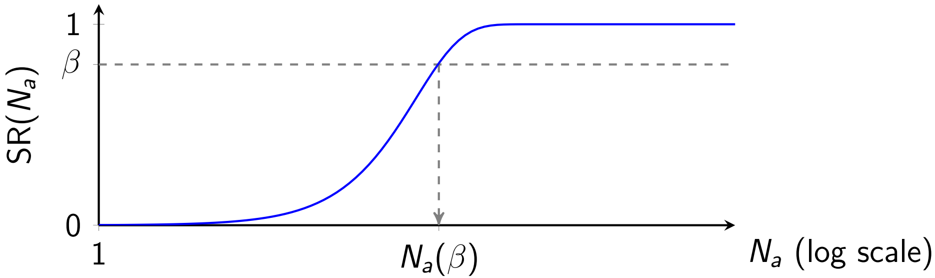
If, the adversary gets:

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

In general, the adversary gets:

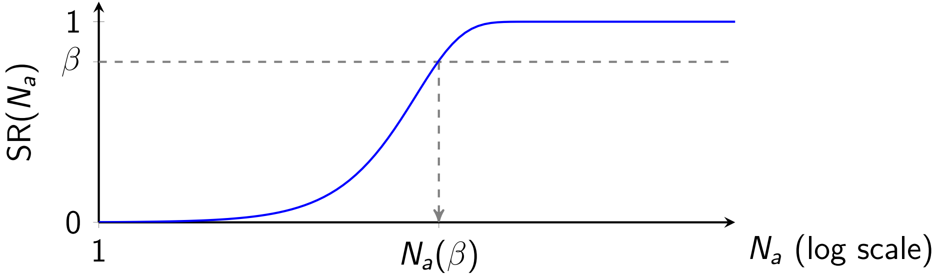
How does this translate into SCA security metrics ?

Concrete SCA Metrics: the Success Rate (SR)



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

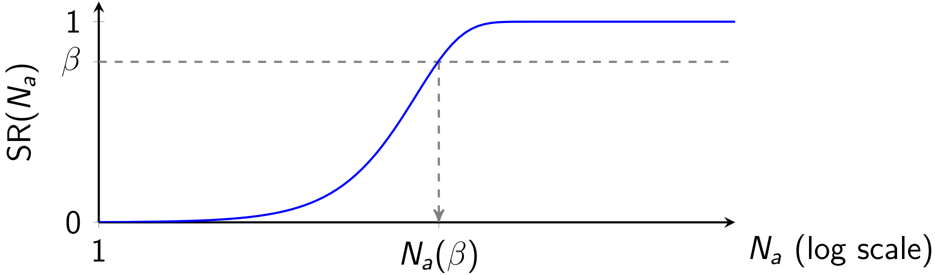
Concrete SCA Metrics: the Success Rate (SR)



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 ✓

Concrete SCA Metrics: the 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 ✗

Circumventing the Drawbacks of the 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"

Circumventing the Drawbacks of the 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"

Circumventing the Drawbacks of the 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),

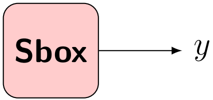
$$N_a(\beta) \geq \frac{f(\beta)}{\mathrm{MI}\left(\mathrm{Y}; \mathbf{L}\right)}$$

MI generalizes ρ ✓
MI hard to estimate ✗

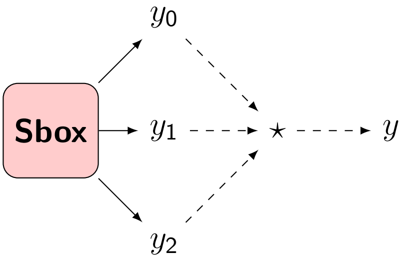
[1] Mangard, Oswald, and Popp, *Power analysis attacks - revealing the secrets of smart cards*
[2] Chérisey et al., "Best Information is Most Successful: Mutual Information and Success Rate in Side-Channel Analysis"

Content

How to protect against SCA: Masking



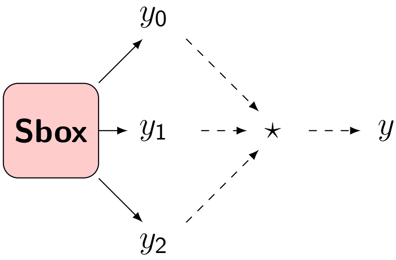
(a) Unprotected



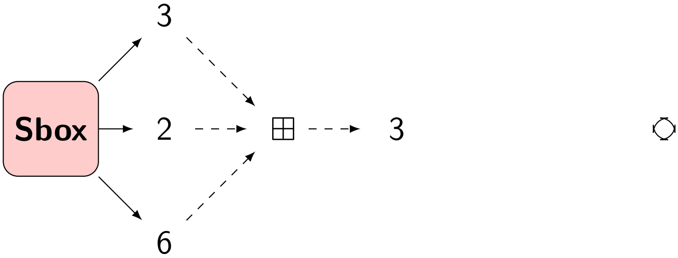
(b) Masking with $d + 1 = 3$ shares

Each share y_i drawn uniformly, such that $y = y_0 \star \ldots \star y_d$

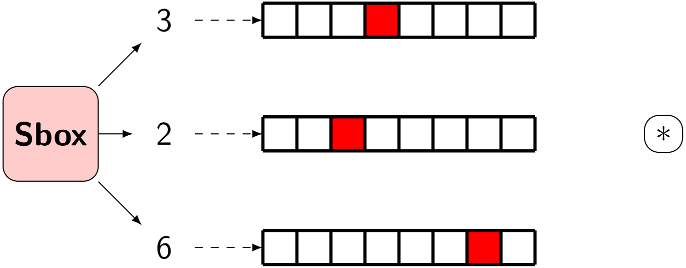
Masking = Convolutions



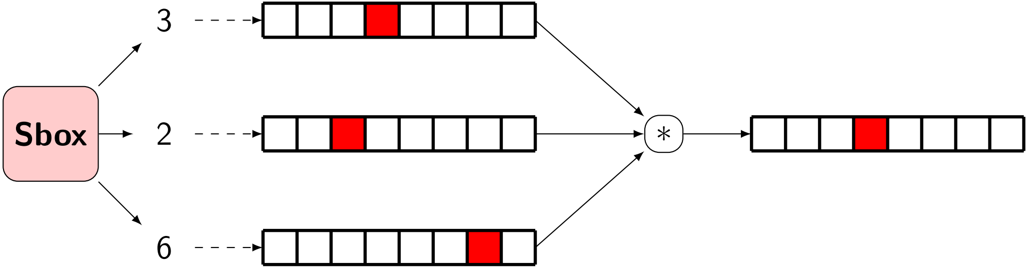
Masking = Convolutions



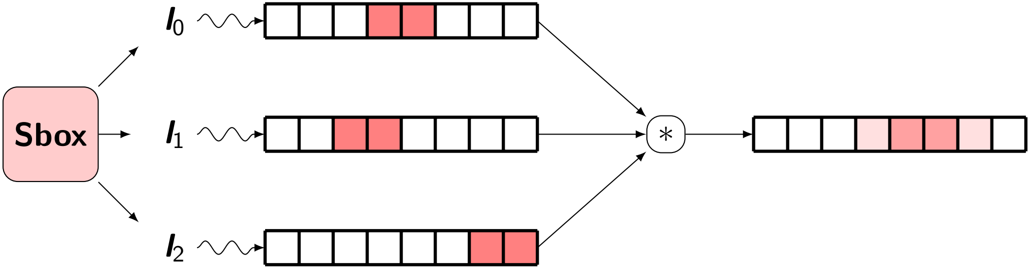
Masking = Convolutions



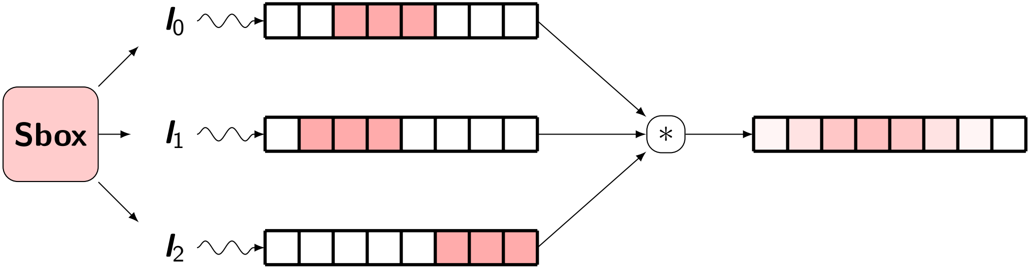
Masking = Convolutions



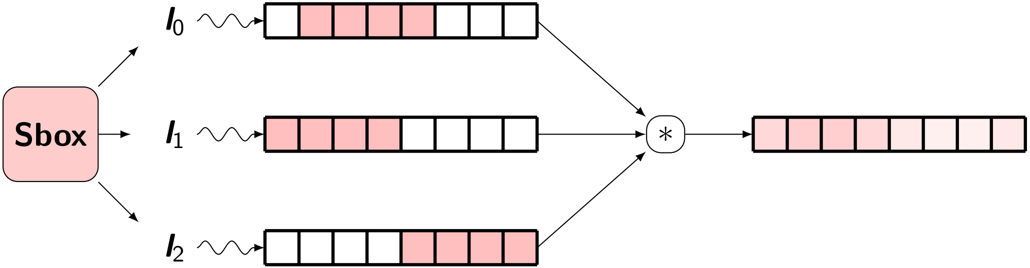
Masking = Convolutions



Masking = Convolutions



Masking = Convolutions



Masking amplifies the noise . . . exponentially with #shares

MI *very* hard to compute naively with masking

Curse of dimensionality increases with #shares

Higher #shares $\implies$ lower MI $\implies$ harder est.

Content

Duc et al's Conjecture [3]

"Can we infer the whole MI using the MI on each share (much easier) ?"

[3] Duc, Faust, and Standaert, "Making Masking Security Proofs Concrete (Or How to Evaluate the Security of Any Leaking Device), Extended Version"

Duc et al's Conjecture [3]

"Can we infer the whole MI using the MI on each share (much easier) ?"

$$N_a(\beta) \approx \frac{f(\beta)}{\prod_{i=0}^{d} \left(\mathsf{MI}\left(\mathrm{Y}_i; \mathrm{L}_i\right)/\tau\right)^r}$$

τ: noise amplification threshold

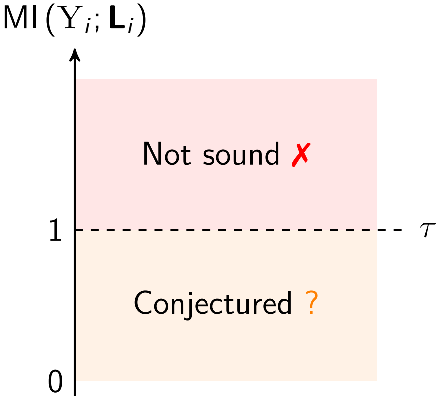
$d \cdot r$: "*effective*" security parameter

Duc *et al.* conjectured:

$\tau = 1$,

$r = 1$,

$f(\beta) \perp |\mathcal{Y}|$,



[3] Duc, Faust, and Standaert, "Making Masking Security Proofs Concrete (Or How to Evaluate the Security of Any Leaking Device), Extended Version"

Duc et al's Conjecture [3]

"Can we infer the whole MI using the MI on each share (much easier) ?"

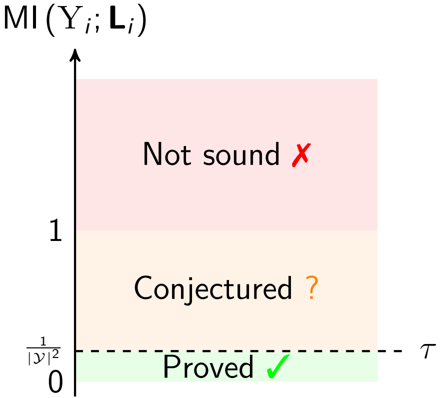
$$N_a(\beta) \geq \frac{f(\beta)}{\prod_{i=0}^{d} \left(\mathrm{MI}\left(\mathrm{Y}_i; \mathrm{L}_i\right)/\tau\right)^r}$$

τ: noise amplification threshold

$d \cdot r$: "*effective*" security parameter

Duc *et al.* only proved:

- $\tau = 0.7/|\mathcal{Y}|^2$,
- $r = 1/2$,
- $f(\beta) \perp |\mathcal{Y}|$,



[3] Duc, Faust, and Standaert, "Making Masking Security Proofs Concrete (Or How to Evaluate the Security of Any Leaking Device), Extended Version"

Duc et al's Conjecture [3]

"Can we infer the whole MI using the MI on each share (much easier) ?"

$$N_a(\beta) \geq \frac{f(\beta)}{\prod_{i=0}^{d} \left(\mathrm{MI}\left(\mathrm{Y}_i; \mathrm{L}_i\right)/\tau\right)^r}$$

τ: noise amplification threshold

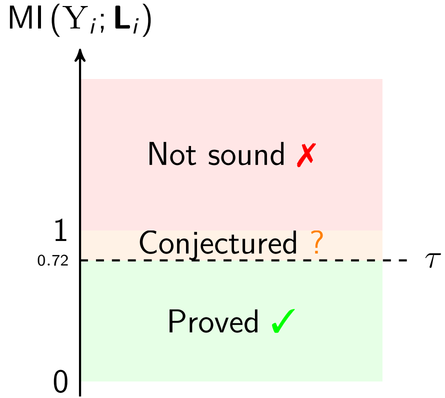
$d \cdot r$: "*effective*" security parameter

In our paper we prove:

$\tau = 0.72,$

$r = 1,$

$f(\beta) \propto \log(|\mathcal{Y}|)/|\mathcal{Y}|,$



[3] Duc, Faust, and Standaert, "Making Masking Security Proofs Concrete (Or How to Evaluate the Security of Any Leaking Device), Extended Version"

Illustration on Simulations

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

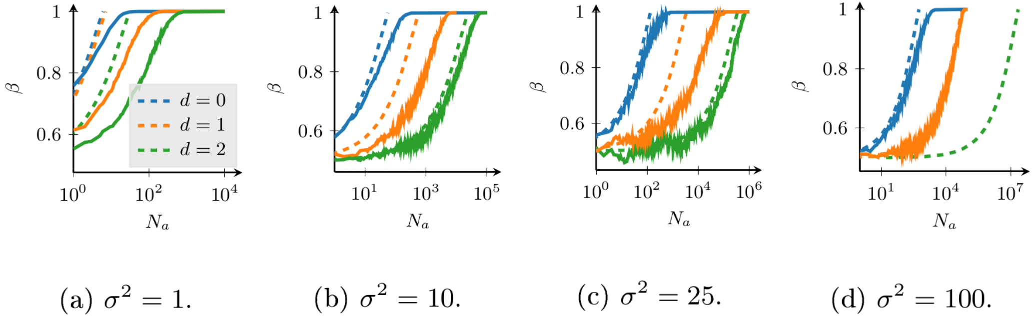
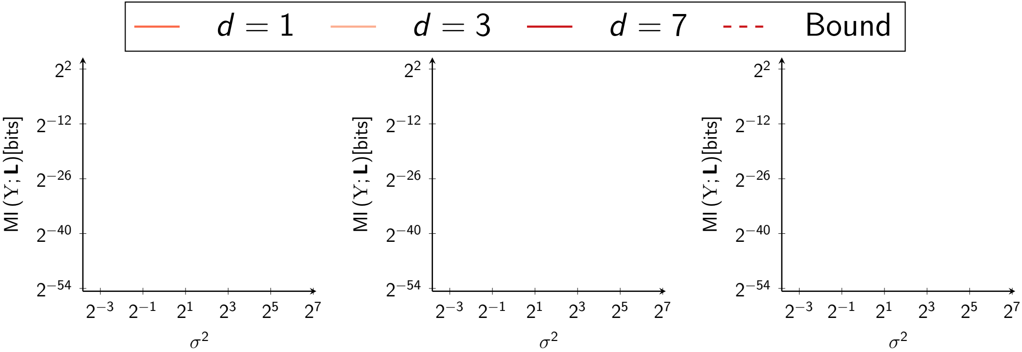


Figure: Success rate of concrete bit recoveries and MI-based upper bounds.

Content

Is Our Nearly-Tight Proof Actually Tight ?

Leakage model: $\mathbf{L}_i = \mathrm{hw}(\mathrm{Y}_i) + \text{ Noise}(0, \sigma^2)$. MI estimated with Monte-Carlo (MC) methods



(a) $|\mathcal{Y}| = 2$ (bit-slice). (b) $|\mathcal{Y}| = 16$. (c) $|\mathcal{Y}| = 256$.

Is Our Nearly-Tight Proof Actually Tight ?

Leakage model: $\mathbf{L}_i = \mathrm{hw}(\mathrm{Y}_i) +$ Noise$(0, \sigma^2)$. MI estimated with MC methods

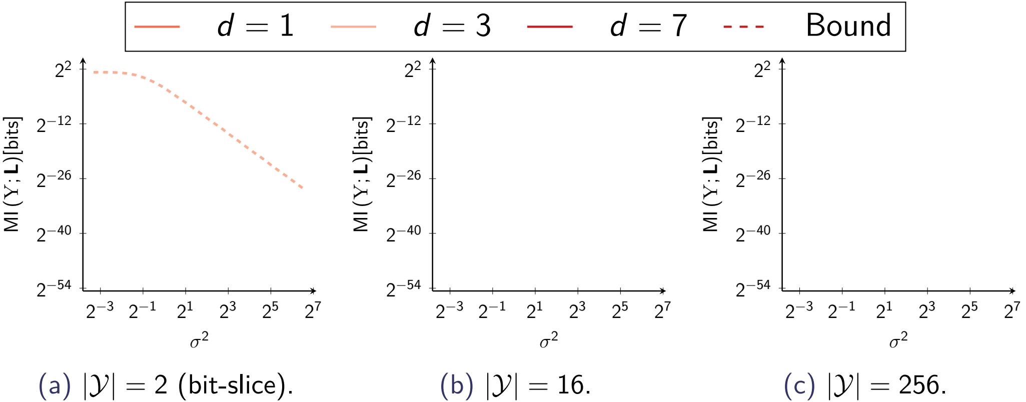
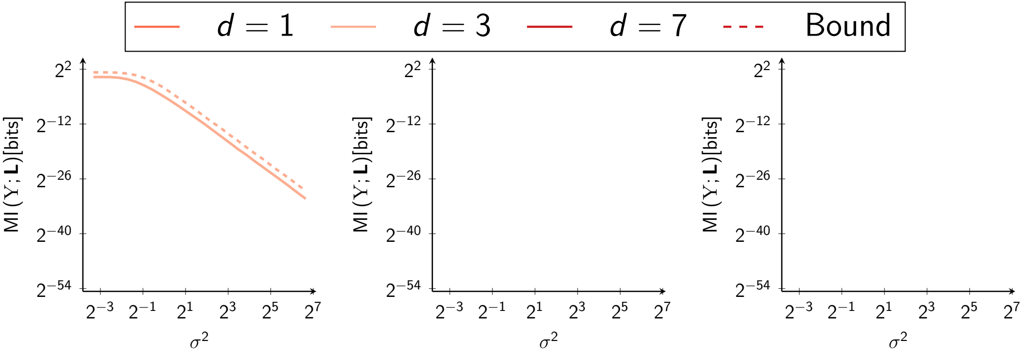


Figure: MI (plain) and new MI upper bound (dashed) for different field sizes.

Is Our Nearly-Tight Proof Actually Tight ?

Leakage model: $\mathbf{L}_i = \mathrm{hw}(Y_i) + \text{ Noise}(0, \sigma^2)$. MI estimated with MC methods

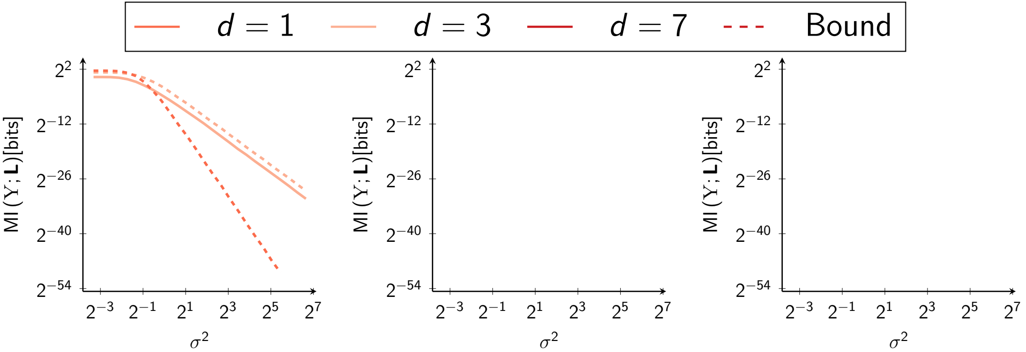


(a) $|\mathcal{Y}| = 2$ (bit-slice). (b) $|\mathcal{Y}| = 16$. (c) $|\mathcal{Y}| = 256$.

Figure: MI (plain) and new MI upper bound (dashed) for different field sizes.

Is Our Nearly-Tight Proof Actually Tight ?

Leakage model: $\mathbf{L}_i = \mathsf{hw}(\mathrm{Y}_i) +$ Noise$(0, \sigma^2)$. MI estimated with MC methods

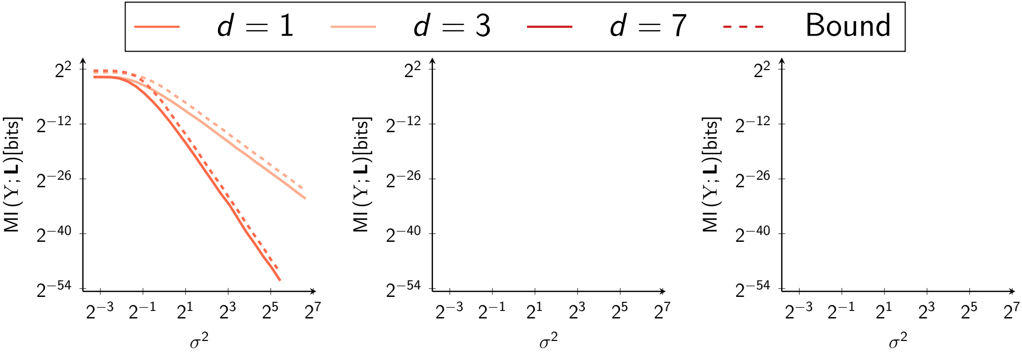


(a) $|\mathcal{Y}| = 2$ (bit-slice). (b) $|\mathcal{Y}| = 16$. (c) $|\mathcal{Y}| = 256$.

Figure: MI (plain) and new MI upper bound (dashed) for different field sizes.

Is Our Nearly-Tight Proof Actually Tight ?

Leakage model: $\mathbf{L}_i = \mathrm{hw}(Y_i) + \mathrm{Noise}(0, \sigma^2)$. MI estimated with MC methods

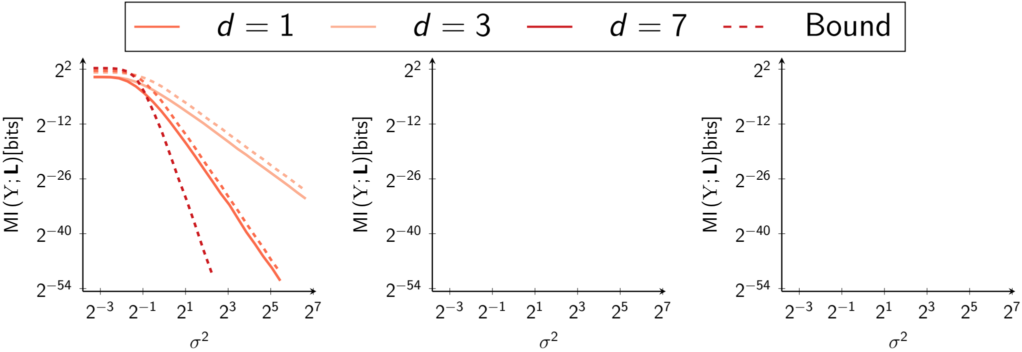


(a) $|\mathcal{Y}| = 2$ (bit-slice). (b) $|\mathcal{Y}| = 16$. (c) $|\mathcal{Y}| = 256$.

Figure: MI (plain) and new MI upper bound (dashed) for different field sizes.

Is Our Nearly-Tight Proof Actually Tight ?

Leakage model: $\mathbf{L}_i = \mathrm{hw}(\mathrm{Y}_i) + \mathrm{Noise}(0, \sigma^2)$. MI estimated with MC methods

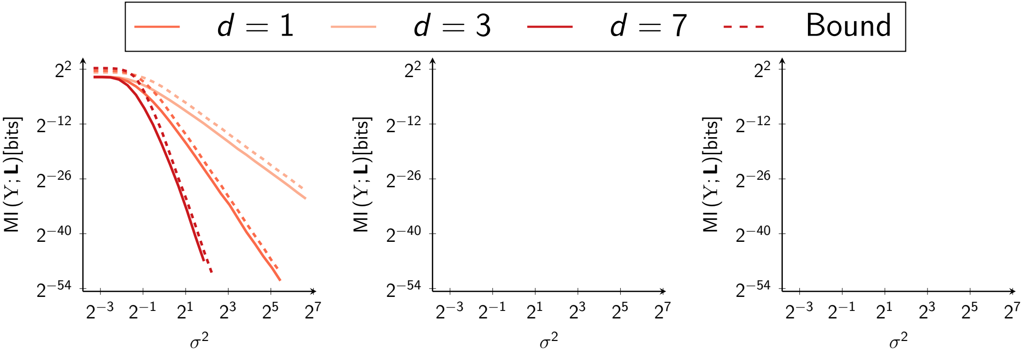


(a) $|\mathcal{Y}| = 2$ (bit-slice). (b) $|\mathcal{Y}| = 16$. (c) $|\mathcal{Y}| = 256$.

Figure: MI (plain) and new MI upper bound (dashed) for different field sizes.

Is Our Nearly-Tight Proof Actually Tight ?

Leakage model: $\mathbf{L}_i = \mathrm{hw}(Y_i) + \mathrm{Noise}(0, \sigma^2)$. MI estimated with MC methods

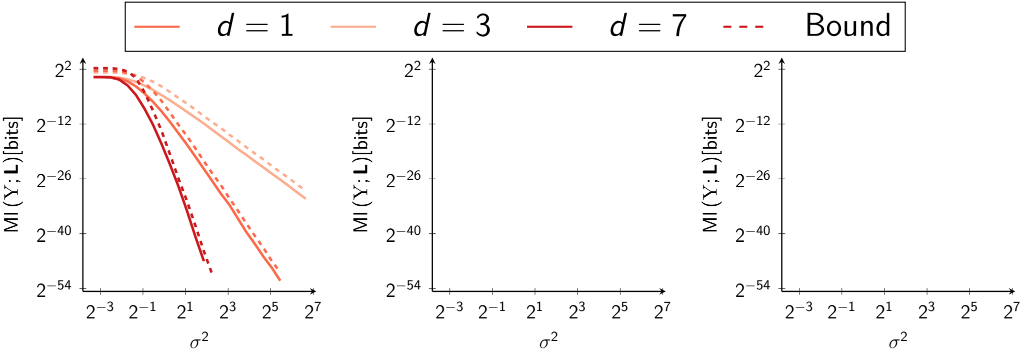


(a) $|\mathcal{Y}| = 2$ (bit-slice). (b) $|\mathcal{Y}| = 16$. (c) $|\mathcal{Y}| = 256$.

Figure: MI (plain) and new MI upper bound (dashed) for different field sizes.

Is Our Nearly-Tight Proof Actually Tight ?

Leakage model: $\mathbf{L}_i = \mathrm{hw}(\mathrm{Y}_i) + \text{ Noise}(0, \sigma^2)$. MI estimated with MC methods

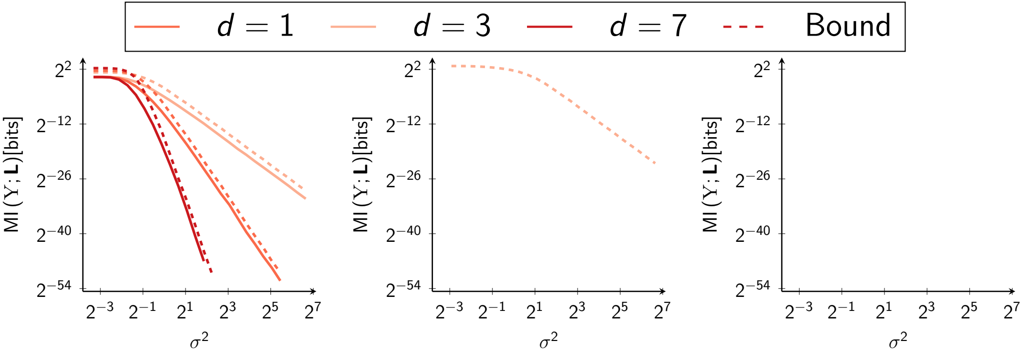


(a) $|\mathcal{Y}| = 2$ (bit-slice). (b) $|\mathcal{Y}| = 16$. (c) $|\mathcal{Y}| = 256$.

Figure: MI (plain) and new MI upper bound (dashed) for different field sizes.

Is Our Nearly-Tight Proof Actually Tight ?

Leakage model: $\mathbf{L}_i = \mathrm{hw}(Y_i) +$ Noise$(0, \sigma^2)$. MI estimated with MC methods



(a) $|\mathcal{Y}| = 2$ (bit-slice). (b) $|\mathcal{Y}| = 16$. (c) $|\mathcal{Y}| = 256$.

Figure: MI (plain) and new MI upper bound (dashed) for different field sizes.

Is Our Nearly-Tight Proof Actually Tight ?

Leakage model: $\mathbf{L}_i = \mathrm{hw}(Y_i) + \mathrm{Noise}(0, \sigma^2)$. MI estimated with MC methods

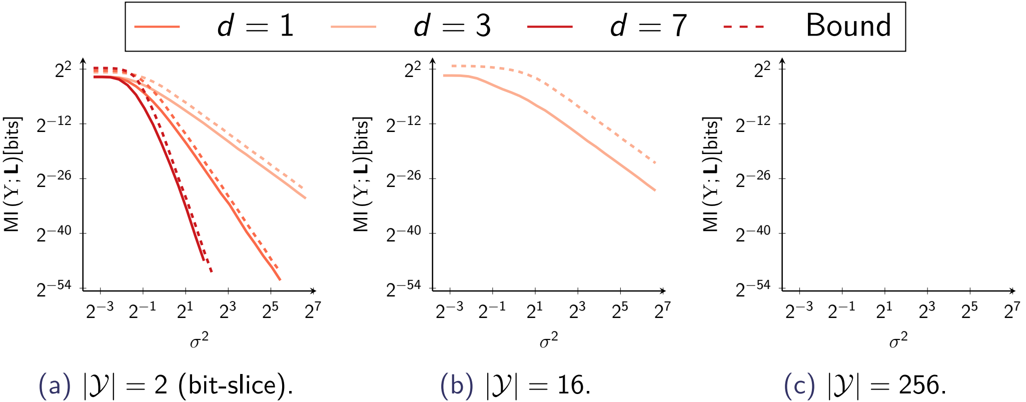
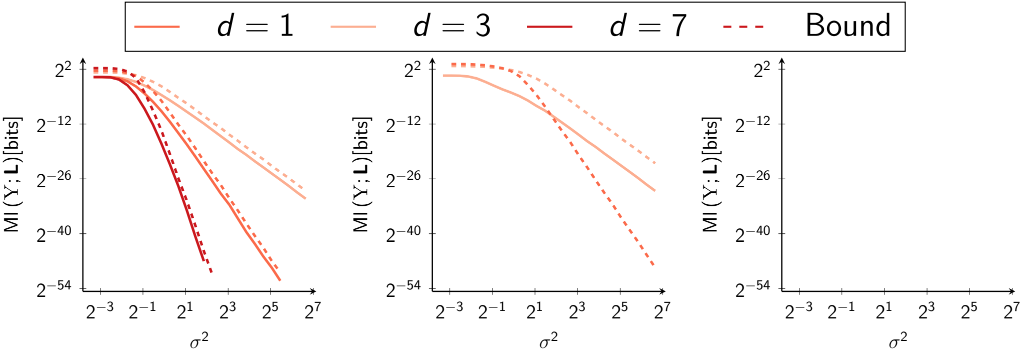


Figure: MI (plain) and new MI upper bound (dashed) for different field sizes.

Is Our Nearly-Tight Proof Actually Tight ?

Leakage model: $\mathbf{L}_i = \mathrm{hw}(\mathrm{Y}_i) + \text{Noise}(0, \sigma^2)$. MI estimated with MC methods

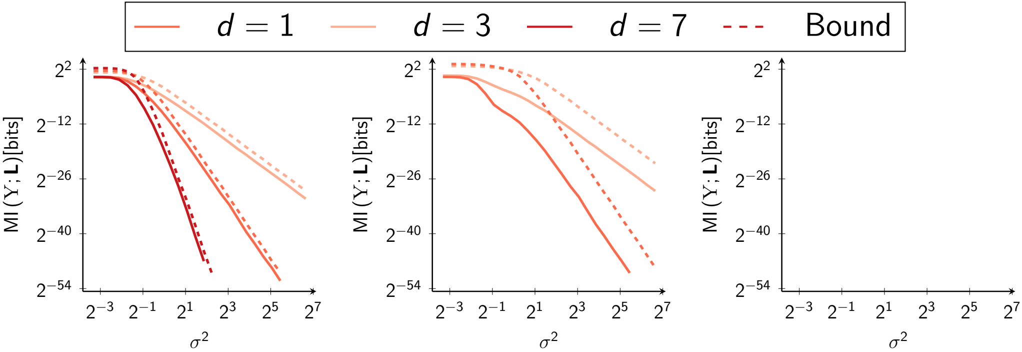


(a) $|\mathcal{Y}| = 2$ (bit-slice). (b) $|\mathcal{Y}| = 16$. (c) $|\mathcal{Y}| = 256$.

Figure: MI (plain) and new MI upper bound (dashed) for different field sizes.

Is Our Nearly-Tight Proof Actually Tight ?

Leakage model: $\mathbf{L}_i = \mathrm{hw}(\mathrm{Y}_i) + \mathrm{Noise}(0, \sigma^2)$. MI estimated with MC methods

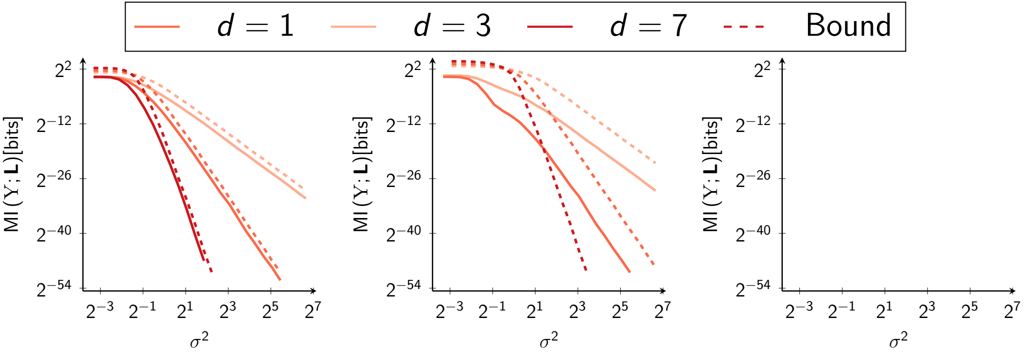


(a) $|\mathcal{Y}| = 2$ (bit-slice). (b) $|\mathcal{Y}| = 16$. (c) $|\mathcal{Y}| = 256$.

Figure: MI (plain) and new MI upper bound (dashed) for different field sizes.

Is Our Nearly-Tight Proof Actually Tight ?

Leakage model: $\mathbf{L}_i = \mathrm{hw}(Y_i) + \text{Noise}(0, \sigma^2)$. MI estimated with MC methods

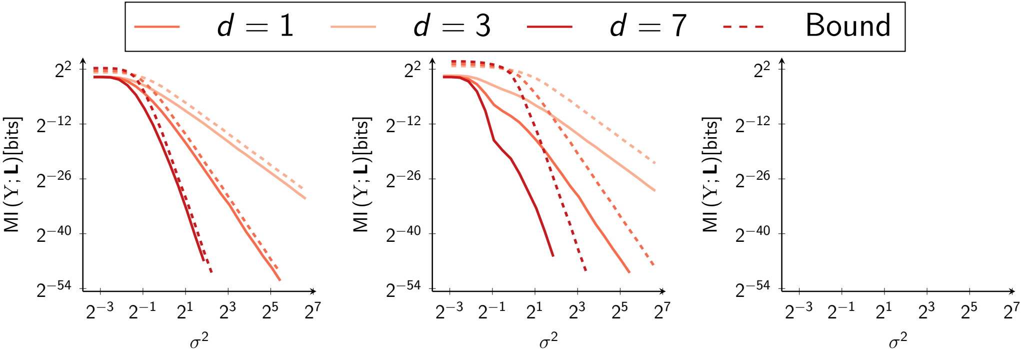


(a) $|\mathcal{Y}| = 2$ (bit-slice).   (b) $|\mathcal{Y}| = 16$.   (c) $|\mathcal{Y}| = 256$.

Figure: MI (plain) and new MI upper bound (dashed) for different field sizes.

Is Our Nearly-Tight Proof Actually Tight ?

Leakage model: $\mathbf{L}_i = \mathrm{hw}(Y_i) + \mathrm{Noise}(0, \sigma^2)$. MI estimated with MC methods

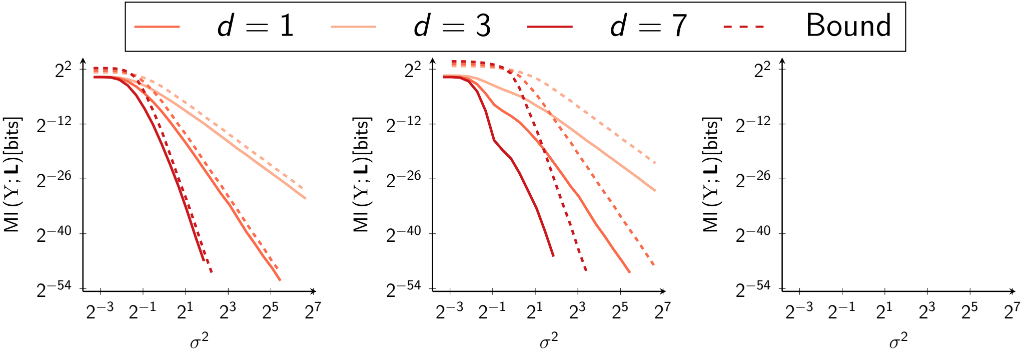


(a) $|\mathcal{Y}| = 2$ (bit-slice). (b) $|\mathcal{Y}| = 16$. (c) $|\mathcal{Y}| = 256$.

Figure: MI (plain) and new MI upper bound (dashed) for different field sizes.

Is Our Nearly-Tight Proof Actually Tight ?

Leakage model: $\mathbf{L}_i = \mathrm{hw}(Y_i) + \text{Noise}(0, \sigma^2)$. MI estimated with MC methods

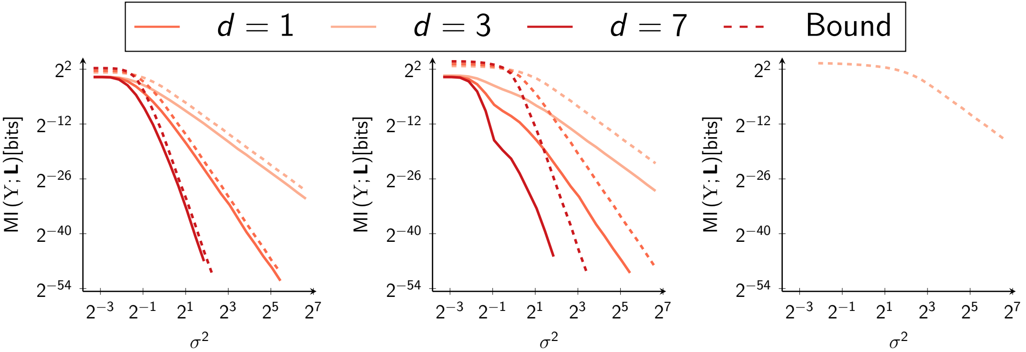


(a) $|\mathcal{Y}| = 2$ (bit-slice). (b) $|\mathcal{Y}| = 16$. (c) $|\mathcal{Y}| = 256$.

Figure: MI (plain) and new MI upper bound (dashed) for different field sizes.

Is Our Nearly-Tight Proof Actually Tight ?

Leakage model: $\mathbf{L}_i = \mathrm{hw}(Y_i) +$ Noise$(0, \sigma^2)$. MI estimated with MC methods

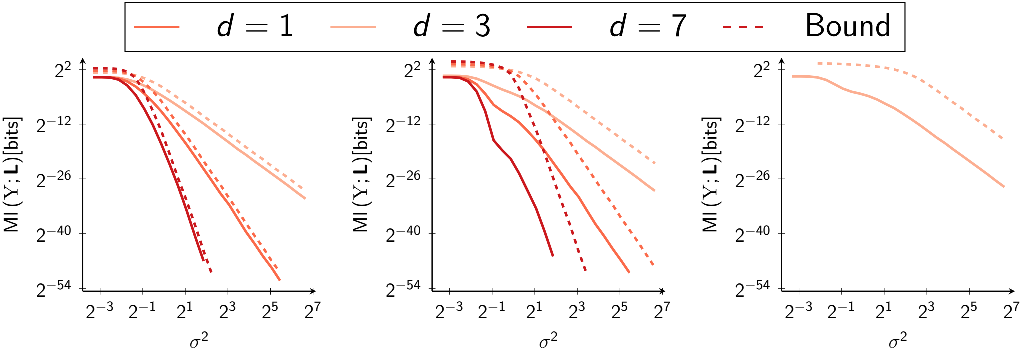


(a) $|\mathcal{Y}| = 2$ (bit-slice). (b) $|\mathcal{Y}| = 16$. (c) $|\mathcal{Y}| = 256$.

Figure: MI (plain) and new MI upper bound (dashed) for different field sizes.

Is Our Nearly-Tight Proof Actually Tight ?

Leakage model: $\mathbf{L}_i = \mathsf{hw}(\mathrm{Y}_i) + \mathsf{Noise}(0, \sigma^2)$. MI estimated with MC methods

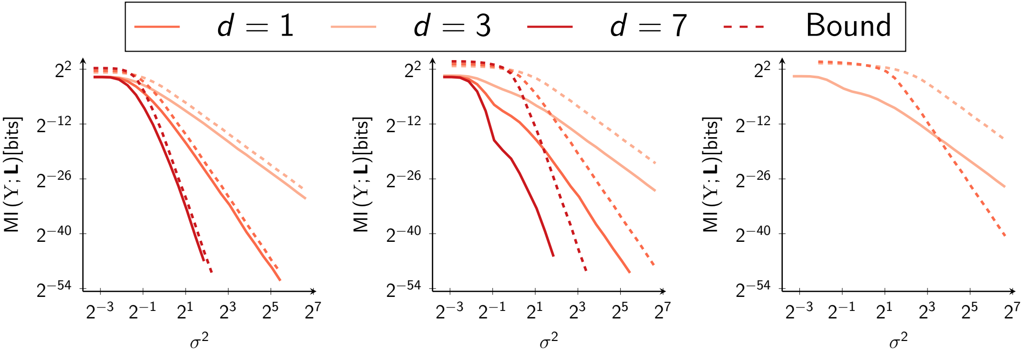


(a) $|\mathcal{Y}| = 2$ (bit-slice). (b) $|\mathcal{Y}| = 16$. (c) $|\mathcal{Y}| = 256$.

Figure: MI (plain) and new MI upper bound (dashed) for different field sizes.

Is Our Nearly-Tight Proof Actually Tight ?

Leakage model: $\mathbf{L}_i = \mathsf{hw}(\mathrm{Y}_i) + \mathsf{Noise}(0, \sigma^2)$. MI estimated with MC methods

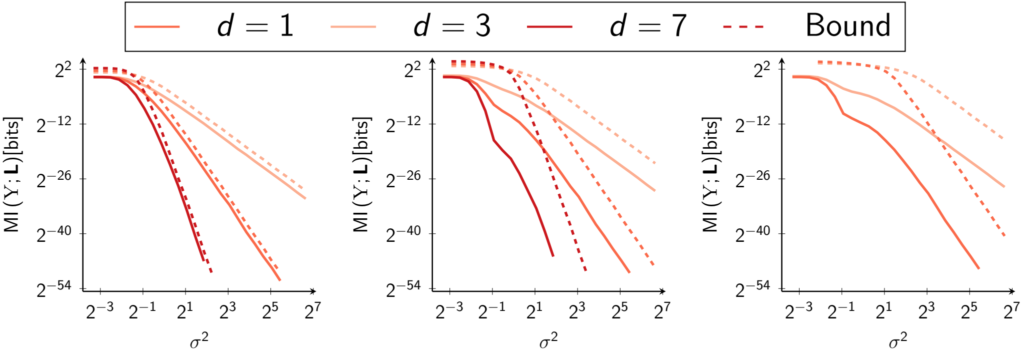


(a) $|\mathcal{Y}| = 2$ (bit-slice). (b) $|\mathcal{Y}| = 16$. (c) $|\mathcal{Y}| = 256$.

Figure: MI (plain) and new MI upper bound (dashed) for different field sizes.

Is Our Nearly-Tight Proof Actually Tight ?

Leakage model: $\mathbf{L}_i = \mathrm{hw}(\mathrm{Y}_i) + \mathrm{Noise}(0, \sigma^2)$. MI estimated with MC methods

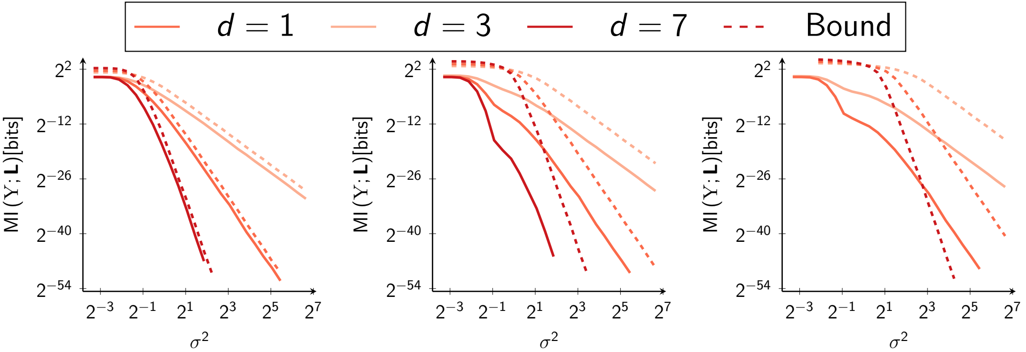


(a) $|\mathcal{Y}| = 2$ (bit-slice). (b) $|\mathcal{Y}| = 16$. (c) $|\mathcal{Y}| = 256$.

Figure: MI (plain) and new MI upper bound (dashed) for different field sizes.

Is Our Nearly-Tight Proof Actually Tight ?

Leakage model: $\mathbf{L}_i = \mathrm{hw}(Y_i) + \text{Noise}(0, \sigma^2)$. MI estimated with MC methods

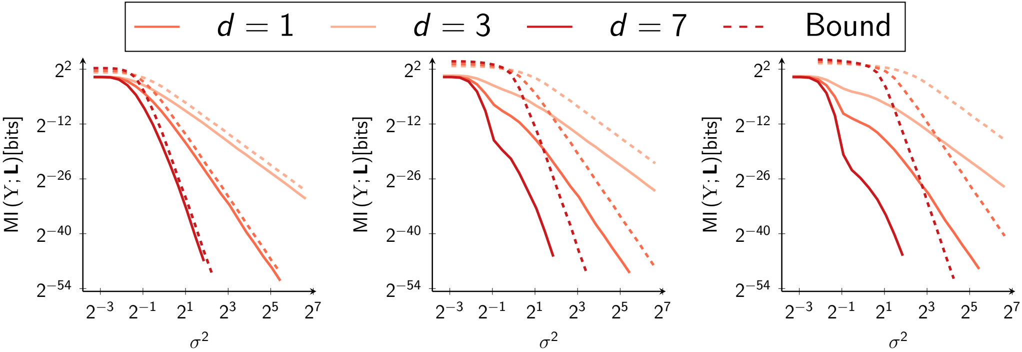


(a) $|\mathcal{Y}| = 2$ (bit-slice). (b) $|\mathcal{Y}| = 16$. (c) $|\mathcal{Y}| = 256$.

Figure: MI (plain) and new MI upper bound (dashed) for different field sizes.

Is Our Nearly-Tight Proof Actually Tight ?

Leakage model: $\mathbf{L}_i = \mathrm{hw}(Y_i) + \mathrm{Noise}(0, \sigma^2)$. MI estimated with MC methods

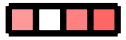
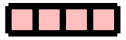


(a) $|\mathcal{Y}| = 2$ (bit-slice). (b) $|\mathcal{Y}| = 16$. (c) $|\mathcal{Y}| = 256$.

Figure: MI (plain) and new MI upper bound (dashed) for different field sizes.

Content

Outline of the Demonstration

How to measure the discrepancy between $\mathrm{p}_{\mathrm{Y}_i \mid \boldsymbol{l}_i} \approx$  and $\mathrm{p}_{\mathrm{Y}_i} =$  ?

Outline of the Demonstration

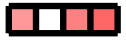
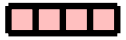
How to measure the discrepancy between $\mathsf{p}_{\mathrm{Y}_i \mid \boldsymbol{l}_i} \approx$ and $\mathsf{p}_{\mathrm{Y}_i} =$?

Using IT metrics: KL divergence, MI

$$\mathsf{D}(\mathsf{p} \parallel \mathsf{m}) = \sum_{y \in \mathcal{Y}} \mathsf{p}(y) \log \left(\frac{\mathsf{p}(y)}{\mathsf{m}(y)} \right)$$

$$\mathsf{MI}(\mathrm{Y}; \mathbf{L}) = \mathop{\mathbb{E}}_{\mathbf{L}} \left[\mathsf{D}(\mathsf{p}_{\mathrm{Y} \mid \mathbf{L}} \parallel \mathsf{p}_{\mathrm{Y}}) \right]$$

Outline of the Demonstration

How to measure the discrepancy between $\mathsf{p}_{\mathrm{Y}_i \mid \boldsymbol{l}_i} \approx$  and $\mathsf{p}_{\mathrm{Y}_i} =$  ?

Using IT metrics: KL divergence, MI

$$\mathsf{D}(\mathsf{p} \parallel \mathsf{m}) = \sum_{y \in \mathcal{Y}} \mathsf{p}(y) \log \left(\frac{\mathsf{p}(y)}{\mathsf{m}(y)} \right)$$

$$\mathsf{MI}(\mathrm{Y}; \mathbf{L}) = \mathop{\mathbb{E}}_{\mathbf{L}} \left[\mathsf{D}(\mathsf{p}_{\mathrm{Y} \mid \mathbf{L}} \parallel \mathsf{p}_{\mathrm{Y}}) \right]$$

Using Total Variation (TV) & SD

$$\mathsf{TV}(\mathsf{p}; \mathsf{m}) = \frac{1}{2} \sum_{y \in \mathcal{Y}} |\mathsf{p}(y) - \mathsf{m}(y)|$$

$$\mathsf{SD}(\mathrm{Y}; \mathbf{L}) = \mathop{\mathbb{E}}_{\mathbf{L}} \left[\mathsf{TV}\left(\mathsf{p}_{\mathrm{Y} \mid \mathbf{L}}; \mathsf{p}_{\mathrm{Y}} \right) \right]$$

Outline of the Demonstration

How to measure the discrepancy between $\mathsf{p}_{\mathrm{Y}_i \mid \boldsymbol{l}_i} \approx$ and $\mathsf{p}_{\mathrm{Y}_i} =$?

Using IT metrics: KL divergence, MI

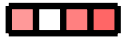
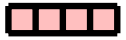
$$\begin{aligned} \mathsf{D}(\mathsf{p} \parallel \mathsf{m}) &= \sum_{y \in \mathcal{Y}} \mathsf{p}(y) \log\left(\frac{\mathsf{p}(y)}{\mathsf{m}(y)}\right) \\ \mathsf{MI}(\mathrm{Y}; \mathbf{L}) &= \mathop{\mathbb{E}}_{\mathbf{L}}\left[\mathsf{D}(\mathsf{p}_{\mathrm{Y} \mid \mathbf{L}} \parallel \mathsf{p}_{\mathrm{Y}})\right] \end{aligned}$$

MI relates well to SR ✓
Not convenient with convolutions ✗

Using Total Variation (TV) & SD

$$\begin{aligned} \mathsf{TV}(\mathsf{p}; \mathsf{m}) &= \frac{1}{2} \sum_{y \in \mathcal{Y}} \left|\mathsf{p}(y) - \mathsf{m}(y)\right| \\ \mathsf{SD}(\mathrm{Y}; \mathbf{L}) &= \mathop{\mathbb{E}}_{\mathbf{L}}\left[\mathsf{TV}\left(\mathsf{p}_{\mathrm{Y} \mid \mathbf{L}}; \mathsf{p}_{\mathrm{Y}}\right)\right] \end{aligned}$$

Outline of the Demonstration

How to measure the discrepancy between $\mathrm{p}_{\mathrm{Y}_i \mid \boldsymbol{l}_i} \approx$  and $\mathrm{p}_{\mathrm{Y}_i} =$  ?

Using IT metrics: KL divergence, MI

$$\begin{aligned} \mathrm{D}(\mathrm{p} \parallel \mathrm{m}) &= \sum_{y \in \mathcal{Y}} \mathrm{p}(y) \log \left(\frac{\mathrm{p}(y)}{\mathrm{m}(y)} \right) \\ \mathrm{MI}(\mathrm{Y}; \mathbf{L}) &= \mathbb{E}_{\mathbf{L}} \left[\mathrm{D}(\mathrm{p}_{\mathrm{Y} \mid \mathbf{L}} \parallel \mathrm{p}_{\mathrm{Y}}) \right] \end{aligned}$$

MI relates well to SR ✓

Not convenient with convolutions ✗

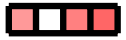
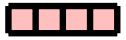
Using Total Variation (TV) & SD

$$\begin{aligned} \mathrm{TV}(\mathrm{p}; \mathrm{m}) &= \frac{1}{2} \sum_{y \in \mathcal{Y}} |\mathrm{p}(y) - \mathrm{m}(y)| \\ \mathrm{SD}(\mathrm{Y}; \mathbf{L}) &= \mathbb{E}_{\mathbf{L}} \left[\mathrm{TV}\left(\mathrm{p}_{\mathrm{Y} \mid \mathbf{L}}; \mathrm{p}_{\mathrm{Y}} \right) \right] \end{aligned}$$

SD relates poorly to SR ✗

Very convenient with convolutions ✓

Outline of the Demonstration

How to measure the discrepancy between $\mathsf{p}_{\mathrm{Y}_i \mid \boldsymbol{l}_i} \approx$  and $\mathsf{p}_{\mathrm{Y}_i} =$  ?

Using IT metrics: KL divergence, MI

$$\mathsf{D}(\mathsf{p} \parallel \mathsf{m}) = \sum_{y \in \mathcal{Y}} \mathsf{p}(y) \log \left(\frac{\mathsf{p}(y)}{\mathsf{m}(y)} \right)$$

$$\mathsf{MI}(\mathrm{Y}; \mathbf{L}) = \mathop{\mathbb{E}}_{\mathbf{L}} \left[\mathsf{D}(\mathsf{p}_{\mathrm{Y} \mid \mathbf{L}} \parallel \mathsf{p}_{\mathrm{Y}}) \right]$$

MI relates well to SR ✓

Not convenient with convolutions ✗

Using Total Variation (TV) & SD

$$\mathsf{TV}(\mathsf{p}; \mathsf{m}) = \frac{1}{2} \sum_{y \in \mathcal{Y}} |\mathsf{p}(y) - \mathsf{m}(y)|$$

$$\mathsf{SD}(\mathrm{Y}; \mathbf{L}) = \mathop{\mathbb{E}}_{\mathbf{L}} \left[\mathsf{TV}\left(\mathsf{p}_{\mathrm{Y} \mid \mathbf{L}}; \mathsf{p}_{\mathrm{Y}} \right) \right]$$

SD relates poorly to SR ✗

Very convenient with convolutions ✓

Can we leverage both advantages ?

Back and Forth Between Metrics

Theorem (Pinsker's Inequalities)

Allows to convert TV to KL divergence, and inversely:

$$2\log(2)\,\mathrm{TV}\,(\mathrm{p};\mathrm{m})^2 \underset{\textit{Pinsker}}{\leq} \mathrm{D}(\mathrm{p} \parallel \mathrm{m})$$

Back and Forth Between Metrics

Theorem (Pinsker's Inequalities)

Allows to convert TV to KL divergence, and inversely:

$$2\log(2)\,\mathsf{TV}\left(\mathsf{p};\mathsf{m}\right)^2 \underset{\text{Pinsker}}{\leq} \mathsf{D}(\mathsf{p}\parallel\mathsf{m}) \underset{\text{Reversed Pinsker}}{\leq} \log_2\left(1+2\left|\mathcal{Y}\right|\mathsf{TV}\left(\mathsf{p};\mathsf{m}\right)^2\right) \leq 2\log(2)\left|\mathcal{Y}\right|\mathsf{TV}\left(\mathsf{p};\mathsf{m}\right)^2$$

The Core Ingredient: the *Xor* Lemma, I

Theorem (Xor Lemma[4])

If $\mathrm{Y} = \mathrm{Y}_0 \star \ldots \star \mathrm{Y}_d$, *then*

$$\mathsf{TV}\left(\mathsf{p}_{\mathrm{Y} \mid \boldsymbol{l}}; \mathsf{p}_{\mathrm{Y}}\right) \leq 2^d \prod_{i=0}^{d} \mathsf{TV}\left(\mathsf{p}_{\mathrm{Y}_i \mid \boldsymbol{l}_i}; \mathsf{p}_{\mathrm{Y}_i}\right) \qquad \textit{(Local)}$$

$$\mathsf{SD}\left(\mathrm{Y}; \mathbf{L}\right) \leq 2^d \prod_{i=0}^{d} \mathsf{SD}\left(\mathrm{Y}_i; \mathbf{L}_i\right) \qquad \textit{(Average)}$$

This is where the noise **amplification** comes from

[4] Dziembowski, Faust, and Skórski, "Optimal Amplification of Noisy Leakages".

The Core Ingredient: the *Xor* Lemma, II

Theorem (Xor Lemma, KL-version)

$$\mathsf{D}(\mathsf{p}_{\mathrm{Y} \mid \boldsymbol{I}} \parallel \mathsf{p}_{\mathrm{Y}}) \leq \log_2 \left(1 + |\mathcal{Y}| \prod_{i=0}^{d} \left(2 \log(2) \mathsf{D}(\mathsf{p}_{\mathrm{Y}_i \mid \boldsymbol{I}_i} \parallel \mathsf{p}_{\mathrm{Y}_i}) \right) \right) \qquad \textit{(Local)}$$

Last Ingredient: Local vs. Average Metrics

By assumption, the leakages $\mathbf{L}_i$ on each share Y_i are independent
"The expectation of the product = the product of expectations"

$$\mathsf{MI}\left(\mathrm{Y}; \mathbf{L}\right) = \underset{\mathbf{L}}{\mathbb{E}}\left[\mathsf{D}(\mathsf{p}_{\mathrm{Y} \mid \mathbf{L}} \parallel \mathsf{p}_{\mathrm{Y}})\right]$$

Last Ingredient: Local vs. Average Metrics

By assumption, the leakages $\mathbf{L}_i$ on each share Y_i are independent
"The expectation of the product = the product of expectations"

$$\mathsf{MI}(\mathrm{Y};\mathbf{L}) = \qquad \mathop{\mathbb{E}}_{\mathbf{L}}\left[\mathsf{D}(\mathsf{p}_{\mathrm{Y}\,|\,\mathbf{L}} \parallel \mathsf{p}_{\mathrm{Y}})\right]$$

Last Ingredient: Local vs. Average Metrics

By assumption, the leakages $\mathbf{L}_i$ on each share Y_i are independent
"The expectation of the product = the product of expectations"

$$\mathsf{MI}\left(\mathrm{Y};\mathbf{L}\right) \underset{\text{Xor Lemma}}{\leq} \underset{\mathbf{L}}{\mathbb{E}}\left[\log_2\left(1+|\mathcal{Y}|\cdot\prod_{i=0}^{d}\left(C\cdot \mathsf{D}(\mathsf{p}_{\mathrm{Y}_i\mid\mathbf{L}_i}\parallel \mathsf{p}_{\mathrm{Y}_i})\right)\right)\right]$$

Last Ingredient: Local vs. Average Metrics

By assumption, the leakages $\mathbf{L}_i$ on each share Y_i are independent
"The expectation of the product = the product of expectations"

$$\mathsf{MI}\left(\mathrm{Y};\mathbf{L}\right) \leq \mathop{\mathbb{E}}_{\mathbf{L}}\left[\log_2\left(1+|\mathcal{Y}|\cdot\prod_{i=0}^{d}\left(C\cdot \mathsf{D}(\mathsf{p}_{\mathrm{Y}_i \mid \mathbf{L}_i} \parallel \mathsf{p}_{\mathrm{Y}_i})\right)\right)\right]$$

Last Ingredient: Local vs. Average Metrics

By assumption, the leakages $\mathbf{L}_i$ on each share Y_i are independent
"The expectation of the product = the product of expectations"

$$\mathsf{MI}\left(\mathrm{Y}; \mathbf{L}\right) \leq \mathop{\mathbb{E}}_{\mathbf{L}}\left[\log_2\left(1 + |\mathcal{Y}| \cdot \prod_{i=0}^{d}\left(C \cdot \mathsf{D}(\mathsf{p}_{\mathrm{Y}_i \mid \mathbf{L}_i} \parallel \mathsf{p}_{\mathrm{Y}_i})\right)\right)\right]$$

Last Ingredient: Local vs. Average Metrics

By assumption, the leakages $\mathbf{L}_i$ on each share Y_i are independent
"The expectation of the product = the product of expectations"

$$\mathsf{MI}\left(\mathrm{Y};\mathbf{L}\right) \underset{\text{Jensen}}{\leq} \log_2\left(1 + |\mathcal{Y}| \cdot \mathop{\mathbb{E}}_{\mathbf{L}}\left[\prod_{i=0}^{d}\left(C \cdot \mathsf{D}(\mathsf{p}_{\mathrm{Y}_i \mid \mathbf{L}_i} \parallel \mathsf{p}_{\mathrm{Y}_i})\right)\right]\right)$$

Last Ingredient: Local vs. Average Metrics

By assumption, the leakages $\mathbf{L}_i$ on each share Y_i are independent
"The expectation of the product = the product of expectations"

$$\mathsf{MI}\left(\mathrm{Y};\mathbf{L}\right) \leq \log_2\left(1 + |\mathcal{Y}| \cdot \mathbb{E}_{\mathbf{L}}\left[\prod_{i=0}^{d}\left(C \cdot \mathsf{D}(\mathsf{p}_{\mathrm{Y}_i \mid \mathbf{L}_i} \parallel \mathsf{p}_{\mathrm{Y}_i})\right)\right]\right)$$

Last Ingredient: Local vs. Average Metrics

By assumption, the leakages $\mathbf{L}_i$ on each share Y_i are independent
"The expectation of the product = the product of expectations"

$$\mathsf{MI}(\mathrm{Y};\mathbf{L}) \leq \log_2\left(1 + |\mathcal{Y}| \cdot \mathop{\mathbb{E}}_{\mathbf{L}}\left[\prod_{i=0}^{d}\left(C \cdot \mathsf{D}(\mathsf{p}_{\mathrm{Y}_i \mid \mathbf{L}_i} \parallel \mathsf{p}_{\mathrm{Y}_i})\right)\right]\right)$$

Last Ingredient: Local vs. Average Metrics

By assumption, the leakages $\mathbf{L}_i$ on each share Y_i are independent
"The expectation of the product $=$ the product of expectations"

$$\mathrm{MI}\left(\mathrm{Y};\mathbf{L}\right) \underset{\text{Indep.}}{\leq} \log_2\left(1 + |\mathcal{Y}| \cdot \prod_{i=0}^{d}\left(C \cdot \mathop{\mathbb{E}}_{\mathbf{L}}\left[\mathrm{D}(\mathrm{p}_{\mathrm{Y}_i \mid \mathbf{L}_i} \parallel \mathrm{p}_{\mathrm{Y}_i})\right]\right)\right)$$

Last Ingredient: Local vs. Average Metrics

By assumption, the leakages $\mathbf{L}_i$ on each share Y_i are independent
"The expectation of the product = the product of expectations"

$$\mathsf{MI}\left(\mathrm{Y};\mathbf{L}\right) \leq \log_2\left(1 + |\mathcal{Y}| \cdot \prod_{i=0}^{d}\left(C \cdot \mathop{\mathbb{E}}_{\mathbf{L}}\left[\mathsf{D}(\mathsf{p}_{\mathrm{Y}_i \mid \mathbf{L}_i} \parallel \mathsf{p}_{\mathrm{Y}_i})\right]\right)\right)$$

Last Ingredient: Local vs. Average Metrics

By assumption, the leakages $\mathbf{L}_i$ on each share Y_i are independent
"The expectation of the product $=$ the product of expectations"

$$\mathsf{MI}\left(\mathrm{Y};\mathbf{L}\right) \leq \log_2\left(1 + |\mathcal{Y}| \cdot \prod_{i=0}^{d}\left(C \cdot \mathop{\mathbb{E}}_{\mathbf{L}}\left[\mathsf{D}(\mathsf{p}_{\mathrm{Y}_i \mid \mathbf{L}_i} \parallel \mathsf{p}_{\mathrm{Y}_i})\right]\right)\right)$$

Theorem (Xor-Lemma, MI-version)

$$\mathsf{MI}\left(\mathrm{Y};\mathbf{L}\right) \leq 2\log(2)\,|\mathcal{Y}| \prod_{i=0}^{d}\left(\mathsf{MI}\left(\mathrm{Y}_i;\mathbf{L}_i\right)/\tau\right) , \qquad (1)$$

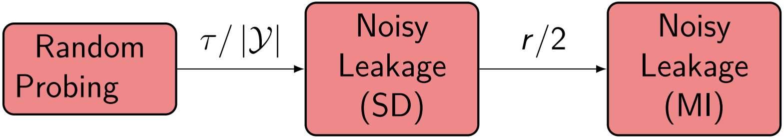
$\tau = \frac{1}{2\log(2)} \approx 0.72$

Why Former Papers are not Tight?

Warning ! Pinsker allows also to convert SD to MI, but not inversely:

$$2\log(2)\,\mathrm{SD}\left(\mathrm{Y};\mathbf{L}\right)^2 \underset{\text{Pinsker + Jensen}}{\leq} \mathrm{MI}\left(\mathrm{Y};\mathbf{L}\right) \not\lesssim 2\log(2)\,\mathrm{SD}\left(\mathrm{Y};\mathbf{L}\right)^2$$

Duc *et al.*'s result relies on the following reduction



Figure

References I

Chérisey, E. de et al. "Best Information is Most Successful: Mutual Information and Success Rate in Side-Channel Analysis". In: *IACR Transactions on Cryptographic Hardware and Embedded Systems* 2019.2 (2019), pp. 49–79. DOI: 10.13154/tches.v2019.i2.49-79. URL: https://tches.iacr.org/index.php/TCHES/article/view/7385.

Duc, A., S. Faust, and F. Standaert. "Making Masking Security Proofs Concrete (Or How to Evaluate the Security of Any Leaking Device), Extended Version". In: *J. Cryptology* 32.4 (2019), pp. 1263–1297. DOI: 10.1007/s00145-018-9277-0. URL: https://doi.org/10.1007/s00145-018-9277-0.

References II

- Dziembowski, S., S. Faust, and M. Skórski. "Optimal Amplification of Noisy Leakages". In: *Theory of Cryptography - 13th International Conference, TCC 2016-A, Tel Aviv, Israel, January 10-13, 2016, Proceedings, Part II*. Ed. by E. Kushilevitz and T. Malkin. Vol. 9563. Lecture Notes in Computer Science. Springer, 2016, pp. 291–318. DOI: 10.1007/978-3-662-49099-0_11. URL: https://doi.org/10.1007/978-3-662-49099-0_11.
- Mangard, S., E. Oswald, and T. Popp. *Power analysis attacks - revealing the secrets of smart cards*. Springer, 2007. ISBN: 978-0-387-30857-9.