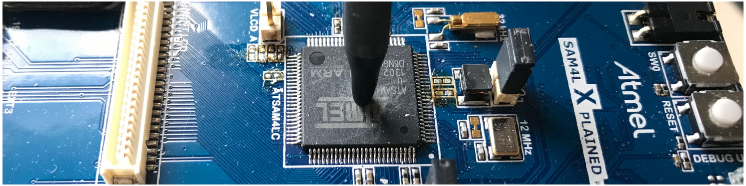


Information Bounds and Convergence Rates for Side-Channel Security Evaluators

Loïc Masure Olivier Bronchain Gaëtan Cassiers François Durvaux
Julien Hendrickx François-Xavier Standaert

Gardanne, May 18^{th}

European Research Council
Established by the European Commission

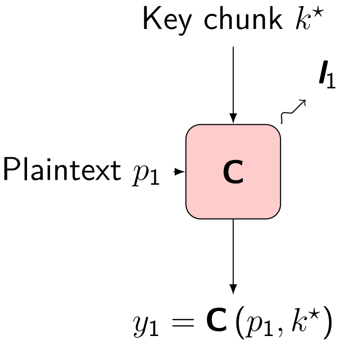
GROUPE CRYPTO

UCLouvain

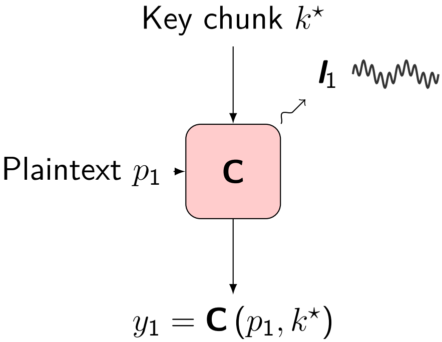
Table of Contents

Content

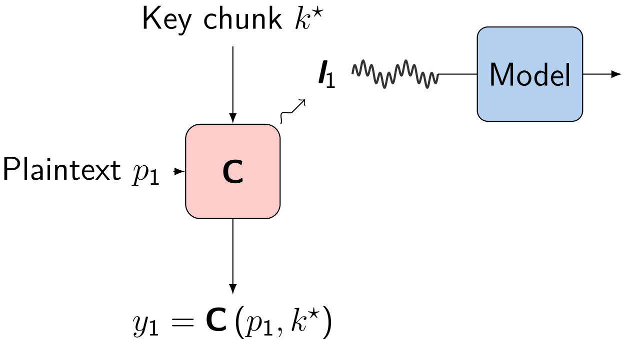
How an SCA works



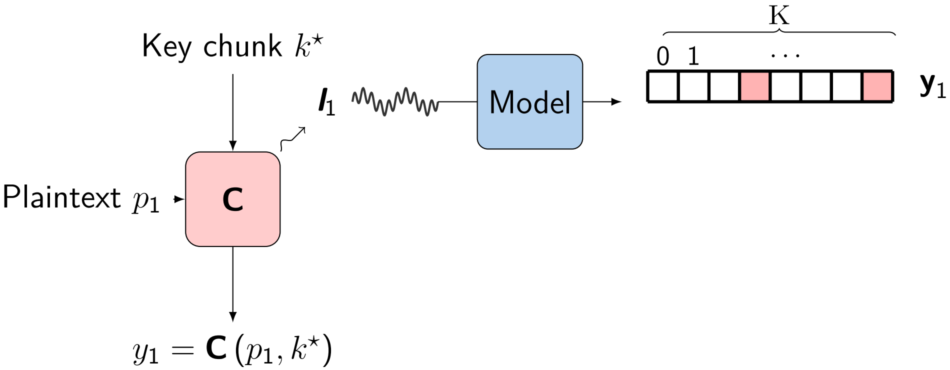
How an SCA works



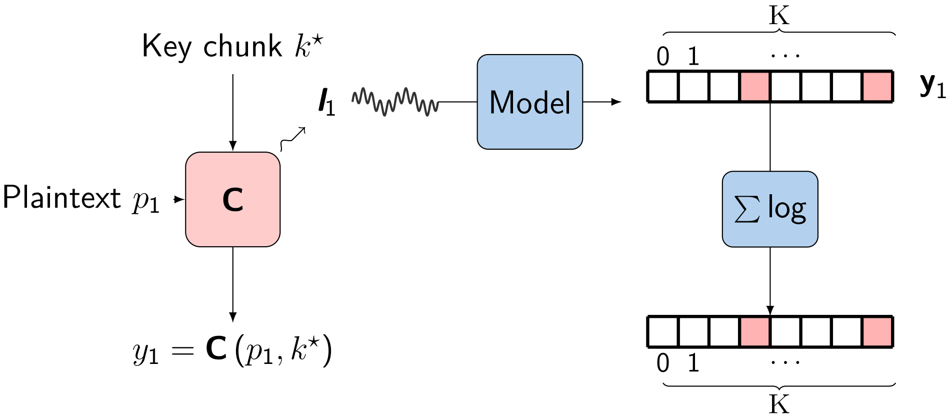
How an SCA works



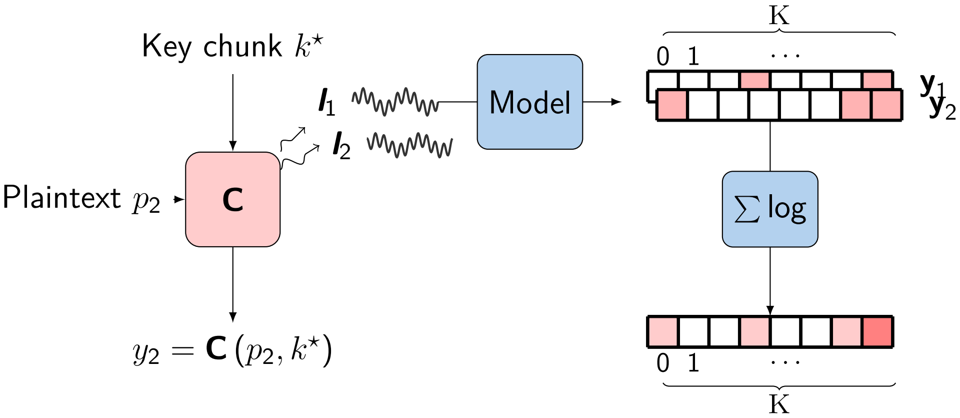
How an SCA works



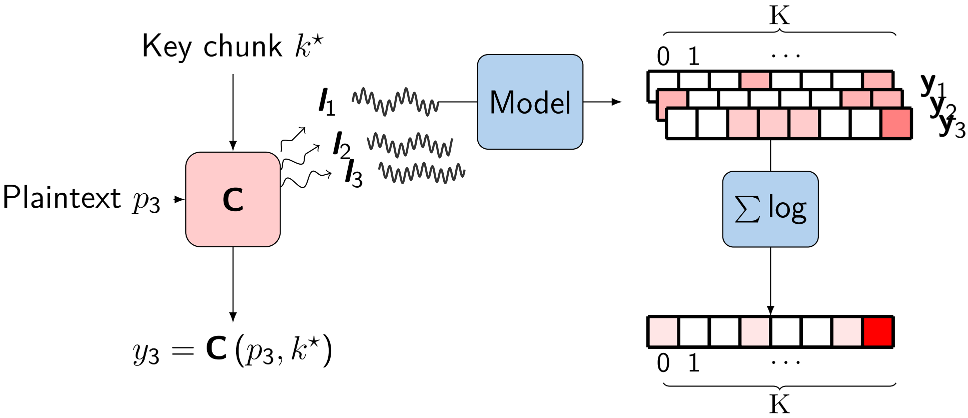
How an SCA works



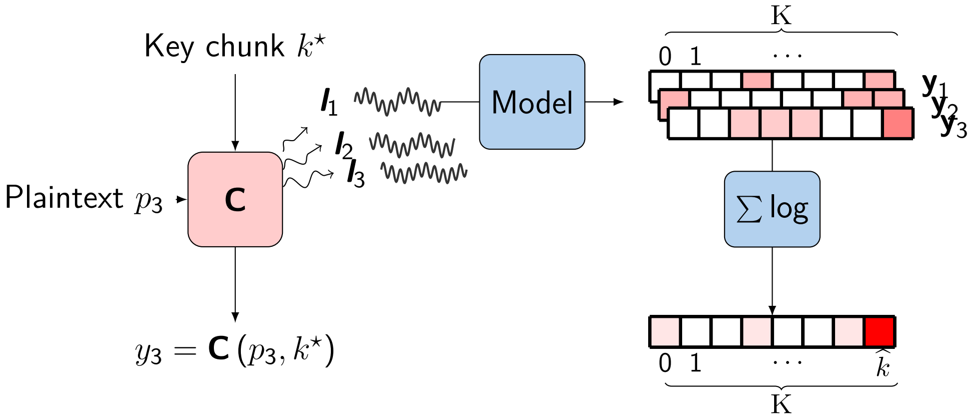
How an SCA works



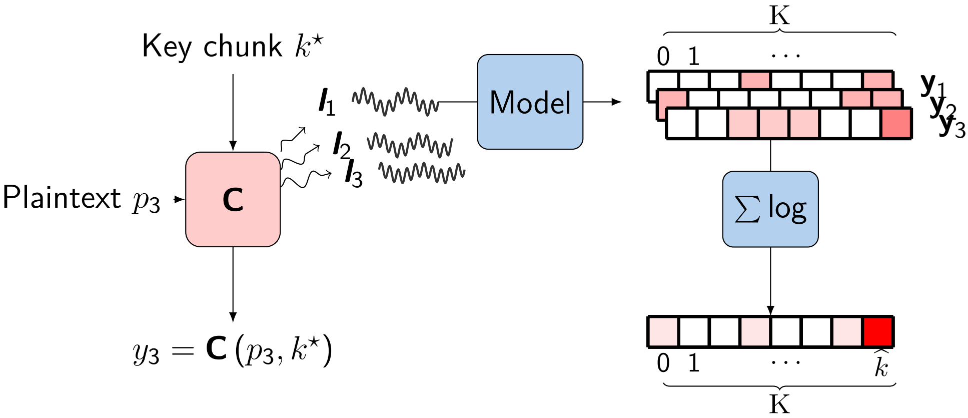
How an SCA works



How an SCA works



How an SCA works



Successful attack i.f.f. $\widehat{k} = k^\star$

What is behind Model ?

What is behind Model ?

Optimal Adversary (security proofs):

Unbounded profiling power

What is behind Model ?

Optimal Adversary (security proofs):
Unbounded profiling power
$\implies \Pr(Y \mid \mathbf{L})$

What is behind Model ?

Optimal Adversary (security proofs):

Unbounded profiling power

$\implies \Pr(Y \mid \mathbf{L})$

Actual Adversary:

Bounded profiling power

What is behind Model ?

Optimal Adversary (security proofs):

Unbounded profiling power

$\implies \Pr(Y \mid \mathbf{L})$

Actual Adversary:

Bounded profiling power

$\implies$ estimation with a model F

What is behind Model ?

Optimal Adversary (security proofs):

Unbounded profiling power

$\implies \Pr(\mathrm{Y} \mid \mathbf{L})$

Actual Adversary:

Bounded profiling power

$\implies$ estimation with a model F

Evaluator/Developer

What is the minimal amount of queries $N_a^\star$ needed for the best adversary to succeed with proba. $\geq \beta$?

What is behind Model ?

Optimal Adversary (security proofs):

Unbounded profiling power

$\implies \Pr(\mathrm{Y} \mid \mathbf{L})$

Actual Adversary:

Bounded profiling power

$\implies$ estimation with a model F

Evaluator/Developer

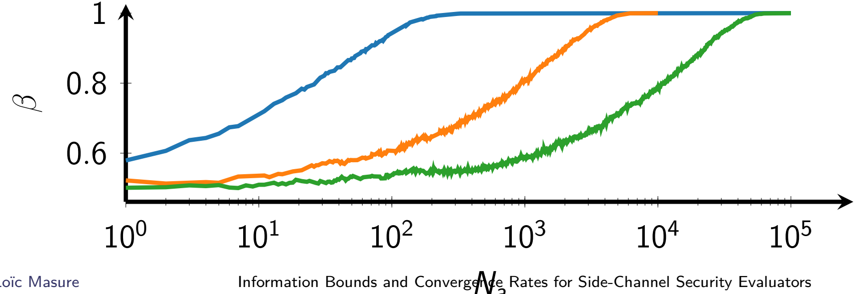
What is the minimal amount of queries $N_a^\star$ needed for the best adversary to succeed with proba. $\geq \beta$?

Adversary

What is the minimal amount of queries $N_a(F)$ needed for F to succeed with proba. $\geq \beta$?

Guessing Security Bounds with IT Metrics

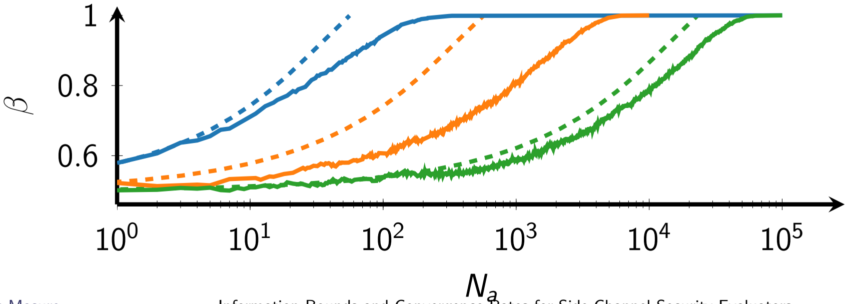
Estimating $N_a^\star$ requires many traces, especially for high values



Guessing Security Bounds with IT Metrics

Estimating $N_a^\star$ requires many traces, especially for high values
Shortcut to evaluate the security against SCA [CHES 2019]:

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



MI: shortcut, but often hard to estimate

Fact: any estimator for $\mathsf{MI}(\mathrm{Y}; \mathbf{L})$ is *biased*

MI: shortcut, but often hard to estimate

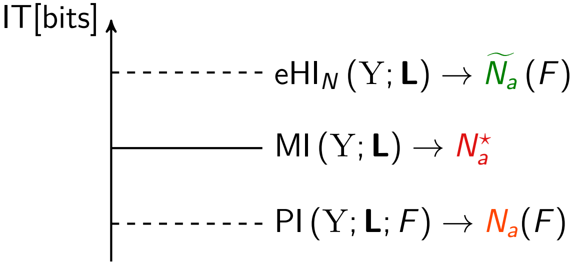
Fact: any estimator for $\mathsf{MI}(\mathrm{Y}; \mathbf{L})$ is *biased*
Bad news for statisticians ...

MI: shortcut, but often hard to estimate

Fact: any estimator for $\mathsf{MI}(Y; \mathbf{L})$ is *biased*
Bad news for statisticians ... Not necessarily for evaluators !

MI: shortcut, but often hard to estimate

Fact: any estimator for $\mathsf{MI}(\mathrm{Y};\mathbf{L})$ is *biased*
Bad news for statisticians ... Not necessarily for evaluators !



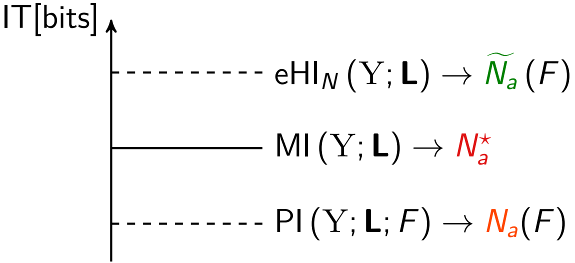
eHI: MI computed with empirical distribution
PI $\sim$ cross-entropy between model and true distribution

IT metrics measure the attack complexity

MI: shortcut, but often hard to estimate

Fact: any estimator for $\mathsf{MI}(\mathrm{Y}; \mathbf{L})$ is *biased*

Bad news for statisticians … Not necessarily for evaluators !



eHI: MI computed with empirical distribution

PI $\sim$ cross-entropy between model and true distribution

IT metrics measure the attack complexity

What about the *profiling* complexity?

Content

Why Profiling Complexity Matters?

Trace acquisition campaign: often the critical task ($\approx$ several days) ...

Why Profiling Complexity Matters?

Trace acquisition campaign: often the critical task ($\approx$ several days) ...
... But convergence of some metrics can be exponentially slow

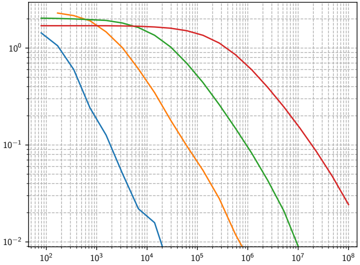


Figure: eHI $-$ MI (y-axis) vs. N (x-axis) for $D = 1$ (blue), 2 (orange), 3 (green), and 4 (red).

Variants of HI Don't Work

Counter-example on a 2-bit masked variable:

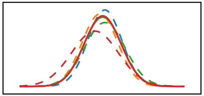


Figure: Plain: true Gaussian mixtures. Dashed: Gaussian templates.

Variants of HI Don't Work

Counter-example on a 2-bit masked variable:

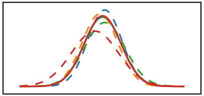
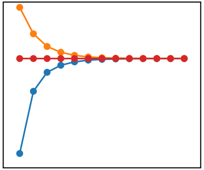


Figure: Plain: true Gaussian mixtures. Dashed: Gaussian templates.



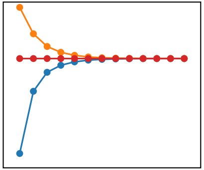
(a) $\mathrm{SNR} = 0.02$

Figure: PI (blue), HI (orange) and MI (red) vs. # profiling traces .

Variants of HI Don't Work

Counter-example on a 2-bit masked variable:

Figure: Plain: true Gaussian mixtures. Dashed: Gaussian templates.



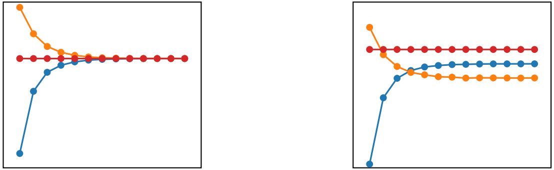
(a) $\mathrm{SNR} = 0.02$

Figure: PI (blue), HI (orange) and MI (red) vs. # profiling traces .

Variants of HI Don't Work

Counter-example on a 2-bit masked variable:

Figure: Plain: true Gaussian mixtures. Dashed: Gaussian templates.



(a) $\mathrm{SNR} = 0.02$ (b) $\mathrm{SNR} = 2$

Figure: PI (blue), HI (orange) and MI (red) vs. # profiling traces .

Variants of HI Don't Work

Counter-example on a 2-bit masked variable:

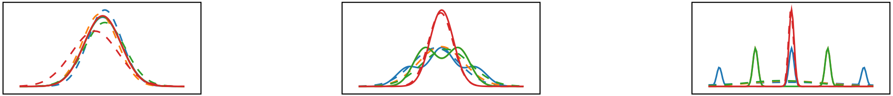
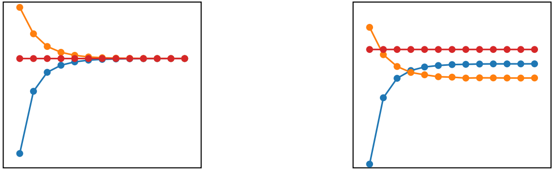


Figure: Plain: true Gaussian mixtures. Dashed: Gaussian templates.



(a) $\mathrm{SNR} = 0.02$ (b) $\mathrm{SNR} = 2$

Figure: PI (blue), HI (orange) and MI (red) vs. # profiling traces .

Variants of HI Don't Work

Counter-example on a 2-bit masked variable:

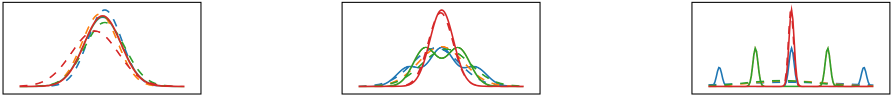
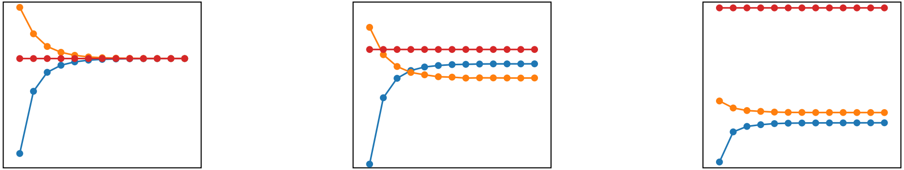


Figure: Plain: true Gaussian mixtures. Dashed: Gaussian templates.



(a) $\mathrm{SNR} = 0.02$ (b) $\mathrm{SNR} = 2$ (c) $\mathrm{SNR} = 200$

Figure: PI (blue), HI (orange) and MI (red) vs. # profiling traces .

Content

Towards New Metrics

Trade-off between evaluator and adversary point of view: restricts the adversary to a collection $\mathcal{H}$ of models

Towards New Metrics

Trade-off between evaluator and adversary point of view: restricts the adversary to a collection $\mathcal{H}$ of models (Gaussian templates, neural networks, ...)

Towards New Metrics

Trade-off between evaluator and adversary point of view: restricts the adversary to a collection $\mathcal{H}$ of models (Gaussian templates, neural networks, ...)

$\mathcal{H}$-Adversary

What is the highest PI reached by the best model from $\mathcal{H}$ to succeed with proba. $\geq \beta$?

$$\mathsf{LI}\left(\mathrm{Y}; \mathbf{L}; \mathcal{H}\right) = \sup_{\mathsf{m} \in \mathcal{H}} \mathsf{PI}\left(\mathrm{Y}; \mathbf{L}; \mathsf{m}\right) \leq \mathsf{MI}\left(\mathrm{Y}; \mathbf{L}\right)$$

Towards New Metrics

Trade-off between evaluator and adversary point of view: restricts the adversary to a collection $\mathcal{H}$ of models (Gaussian templates, neural networks, ...)

$\mathcal{H}$-Adversary

What is the highest PI reached by the best model from $\mathcal{H}$ to succeed with proba. $\geq \beta$?

$$\mathsf{LI}\left(\mathrm{Y};\mathbf{L};\mathcal{H}\right) = \sup_{\mathsf{m}\in\mathcal{H}} \mathsf{PI}\left(\mathrm{Y};\mathbf{L};\mathsf{m}\right) \leq \mathsf{MI}\left(\mathrm{Y};\mathbf{L}\right)$$

LI: surrogate to MI

Towards New Metrics

Trade-off between evaluator and adversary point of view: restricts the adversary to a collection $\mathcal{H}$ of models (Gaussian templates, neural networks, ...)

$\mathcal{H}$-Adversary

What is the highest PI reached by the best model from $\mathcal{H}$ to succeed with proba. $\geq \beta$?

$$\mathsf{LI}\left(\mathrm{Y};\mathbf{L};\mathcal{H}\right) = \sup_{\mathrm{m}\in\mathcal{H}} \mathsf{PI}\left(\mathrm{Y};\mathbf{L};\mathrm{m}\right) \leq \mathsf{MI}\left(\mathrm{Y};\mathbf{L}\right)$$

LI: surrogate to MI
PI: natural lower bound of LI

Towards New Metrics

Trade-off between evaluator and adversary point of view: restricts the adversary to a collection $\mathcal{H}$ of models (Gaussian templates, neural networks, ...)

$\mathcal{H}$-Adversary

What is the highest PI reached by the best model from $\mathcal{H}$ to succeed with proba. $\geq \beta$?

$$\mathsf{LI}\left(\mathrm{Y}; \mathbf{L}; \mathcal{H}\right) = \sup_{\mathsf{m} \in \mathcal{H}} \mathsf{PI}\left(\mathrm{Y}; \mathbf{L}; \mathsf{m}\right) \leq \mathsf{MI}\left(\mathrm{Y}; \mathbf{L}\right)$$

LI: surrogate to MI

PI: natural lower bound of LI What about an upper bound for LI?

Upper Bounds to LI

Training Information (TI)

Any $\mathcal{H}$-adversary applying Empirical Risk Minimization (ERM) using a profiling set $\mathcal{S}_p$:

$$\mathsf{TI}_N\left(\mathrm{Y};\mathbf{L};\mathcal{A}\right) = \max_{\mathsf{m}\in\mathcal{H}} \Delta^{\mathsf{m}}_{\tilde{\mathrm{e}}_{\mathcal{S}_p}}$$

Upper Bounds to LI

Training Information (TI)

Any $\mathcal{H}$-adversary applying Empirical Risk Minimization (ERM) using a profiling set $\mathcal{S}_p$:

$$\mathsf{TI}_N\left(\mathrm{Y};\mathbf{L};\mathcal{A}\right) = \max_{\mathrm{m}\in\mathcal{H}} \Delta^{\mathrm{m}}_{\tilde{\mathrm{e}}_{\mathcal{S}_p}}$$

$\iff$ *training loss when model pushed to fit the training set — w/o regularization, early-stopping, ...*

TI behaves like the empirical HI

TI behaves like the empirical HI

$$\mathsf{MI}\left(\mathrm{Y};\mathbf{L}\right) \leq \quad \mathbb{E}\left[\mathsf{eHI}_{N}\left(\mathrm{Y};\mathbf{L}\right)\right] \quad \leq \mathbb{E}\left[\mathsf{eHI}_{N-1}\left(\mathrm{Y};\mathbf{L}\right)\right]$$

TI behaves like the empirical HI

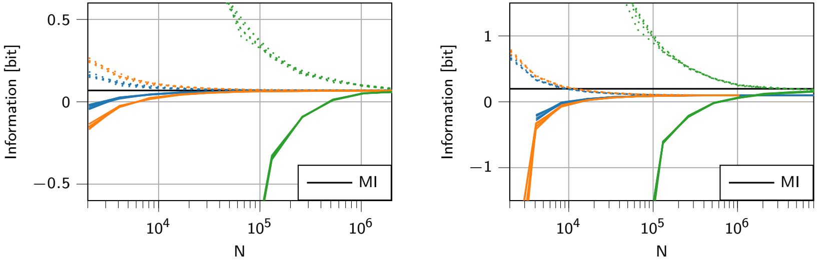
$$\mathsf{MI}\left(\mathrm{Y};\mathbf{L}\right) \leq \mathbb{E}\left[\mathsf{eHI}_N\left(\mathrm{Y};\mathbf{L}\right)\right] \leq \mathbb{E}\left[\mathsf{eHI}_{N-1}\left(\mathrm{Y};\mathbf{L}\right)\right]$$

$$\mathsf{LI}\left(\mathrm{Y};\mathbf{L};\mathcal{H}\right) \leq \mathbb{E}\left[\mathsf{TI}_N\left(\mathrm{Y};\mathbf{L};\mathcal{A}\right)\right] \leq \mathbb{E}\left[\mathsf{TI}_{N-1}\left(\mathrm{Y};\mathbf{L};\mathcal{A}_{\mathcal{H}}\right)\right]$$

TI behaves like the empirical HI

$$\mathsf{MI}\left(Y; \mathbf{L}\right) \leq \mathbb{E}\left[\mathsf{eHI}_N\left(Y; \mathbf{L}\right)\right] \leq \mathbb{E}\left[\mathsf{eHI}_{N-1}\left(Y; \mathbf{L}\right)\right]$$

$$\mathsf{LI}\left(Y; \mathbf{L}; \mathcal{H}\right) \leq \mathbb{E}\left[\mathsf{TI}_N\left(Y; \mathbf{L}; \mathcal{A}\right)\right] \leq \mathbb{E}\left[\mathsf{TI}_{N-1}\left(Y; \mathbf{L}; \mathcal{A}_{\mathcal{H}}\right)\right]$$

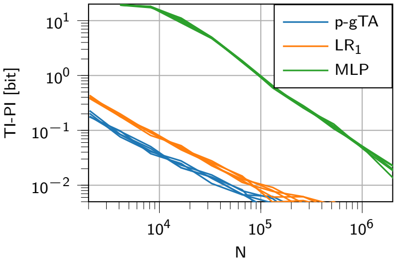


(a) HW: not masked, SNR=0.1 (b) SW: masked, SNR=10.

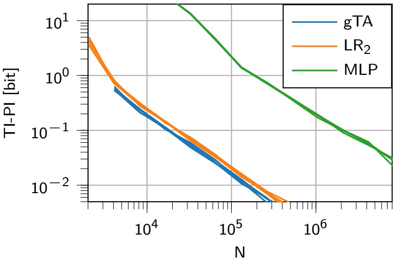
Figure: Convergence of information metrics. Dotted lines: TI. Solid lines: PI.

Content

Simulated Experiments I

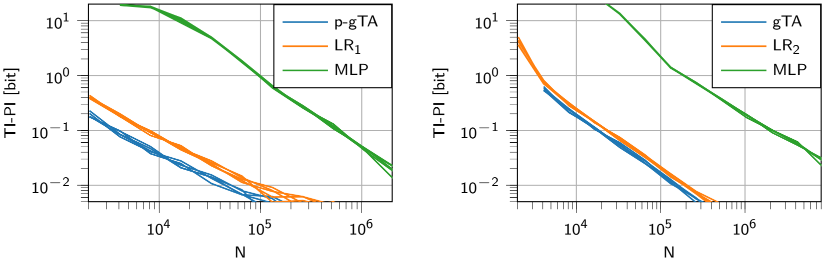


(a) HW: not masked, SNR=0.1



(b) SW: masked, SNR=10.

Simulated Experiments I



(a) HW: not masked, SNR=0.1 (b) SW: masked, SNR=10.

No curse of dimensionality, trend $\propto \frac{1}{N}$

Can we infer the profiling complexity N, without acquiring N traces ?

Simulated Experiments II

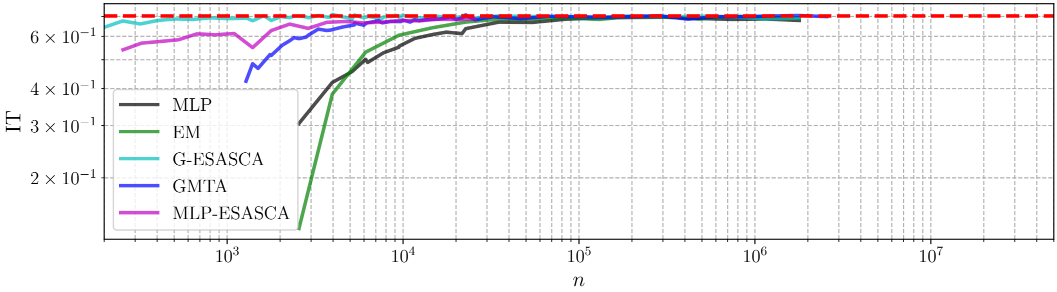


Figure: Learning curves, 1-o masking, 4-bit target, $SNR = 10$.

Simulated Experiments II

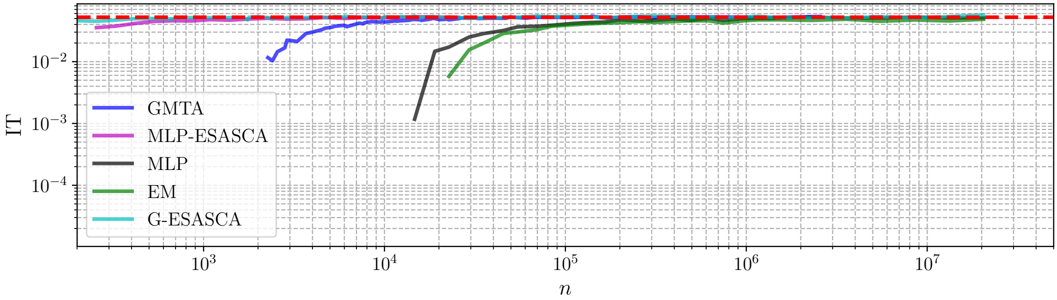


Figure: Learning curves, 1-o masking, 4-bit target, $SNR = 1$.

Simulated Experiments II

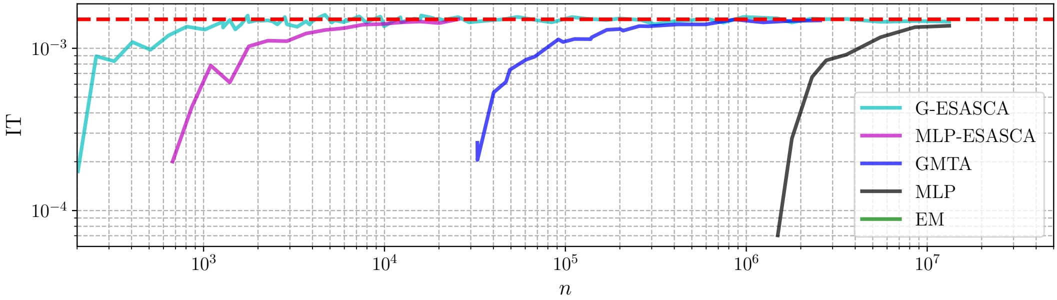


Figure: Learning curves, 1-o masking, 4-bit target, $SNR = 0.1$.

Bounds on the Convergence Rate

First result: PI converges to LI with a convergence rate $\widetilde{\mathcal{O}}\left(\frac{cst(\mathcal{H})}{N}\right)$, where cst is polynomial in the dimensions of $\mathcal{H} \implies$ much tighter lower bound

Bounds on the Convergence Rate

Second result: Training Information (TI) converges at most twice as slow as PI $\implies$ tight upper bound

Bounds on the Convergence Rate

Third result: convergence bounds for Template Attacks:

- Classical TA: $\mathcal{O}\left(\frac{QD^2}{N}\right)$
- *Pooled* TA: $\mathcal{O}\left(\frac{QD}{N}\right)$ (at least for $Q = 2$)

Content

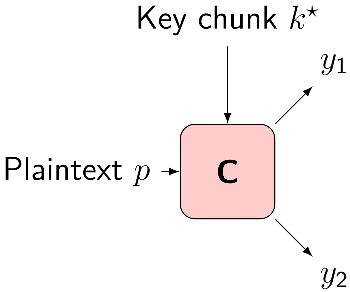
Profiling with Masking in White-Box

Masking: $\mathbf{C}(p, k^\star) = y_1 \star y_2$

Profiling with Masking in White-Box

Masking: $\mathbf{C}(p, k^\star) = y_1 \star y_2$

White-Box: the adversary knows the random shares during profiling

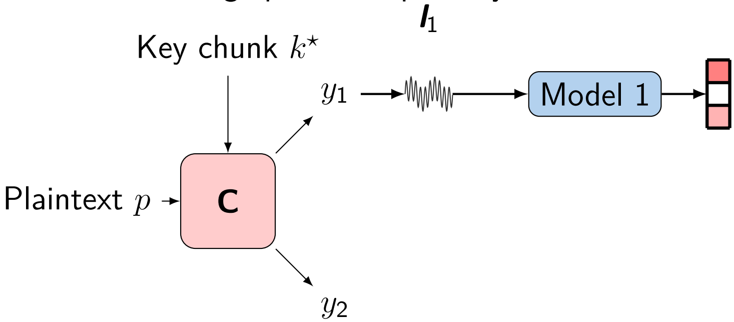


Profiling with Masking in White-Box

Masking: $\mathbf{C}(p, k^\star) = y_1 \star y_2$

White-Box: the adversary knows the random shares during profiling

$\implies$ each leakage profiled separately ...

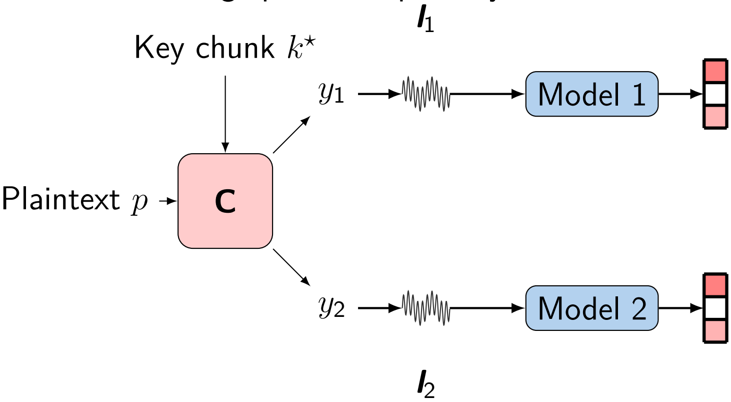


Profiling with Masking in White-Box

Masking: $\mathbf{C}(p, k^\star) = y_1 \star y_2$

White-Box: the adversary knows the random shares during profiling

$\implies$ each leakage profiled separately ...

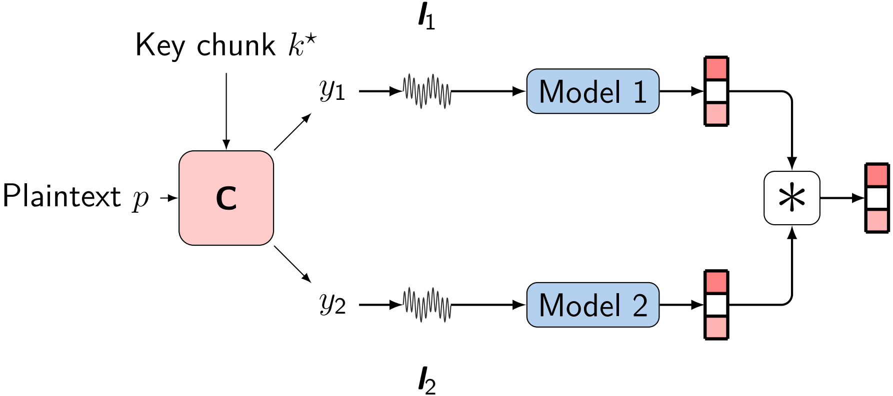


Profiling with Masking in White-Box

Masking: $\mathbf{C}\left(p, k^{\star}\right) = y_1 \star y_2$

White-Box: the adversary knows the random shares during profiling

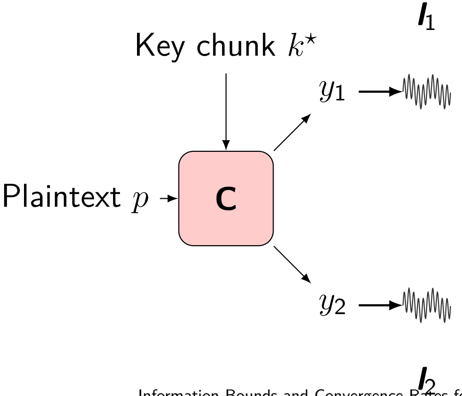
$\Longrightarrow$ each leakage profiled separately ... then scores are recombined



Profiling with Masking in Black-Box

Masking: $\mathbf{C}(p, k^\star) = y_1 \star y_2$

Black-Box profiling: the adversary does not know the random shares

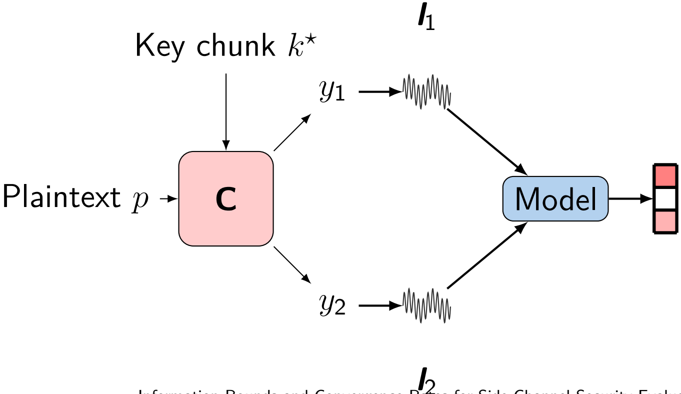


Profiling with Masking in Black-Box

Masking: $\mathbf{C}(p, k^\star) = y_1 \star y_2$

Black-Box profiling: the adversary does not know the random shares

$\implies$ directly profiles $\Pr(\mathrm{Y} \mid \mathbf{L})$ with a single model



Simulated Experiments III

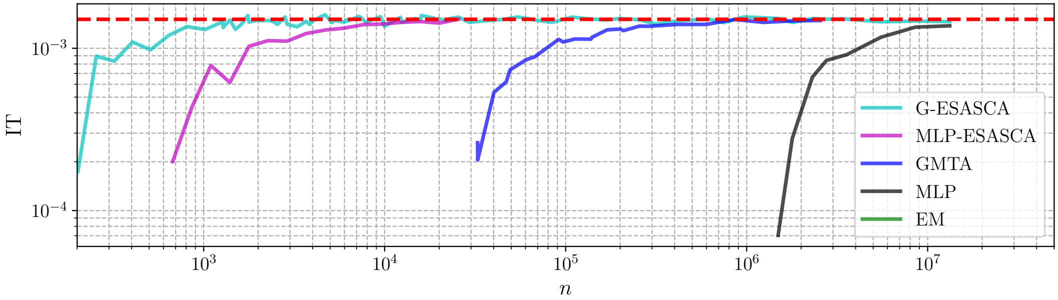


Figure: First-order masking, SNR = 0.1

White-box models converge faster than black-box counter-parts

Simulated Experiments III

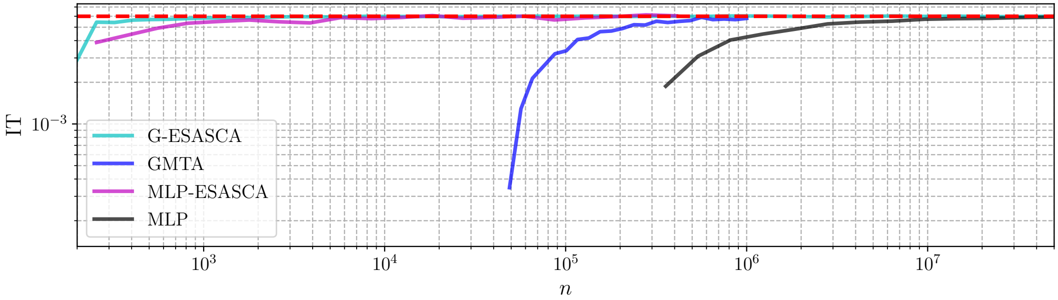


Figure: Second-order masking, SNR = 1

White-box models converge faster than black-box counter-parts

Simulated Experiments III

White-box models converge faster than black-box counter-parts

Can we theoretically explain all these observations?

Profiling Complexity of Black vs. White Box

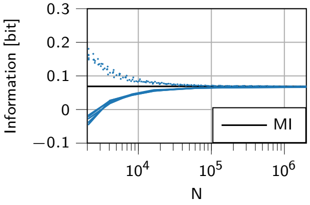
Sound model ≈ model with PI > 0[a]

Profiling Complexity of Black vs. White Box

Sound model $\approx$ model with PI > 0[a]

Ok if the profiling complexity $N \propto \frac{1}{LI-PI} \geq \frac{1}{MI}$

[a]Not always, see [**cryptoeprint:2021:1216**].

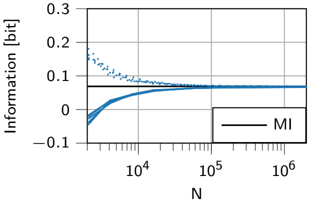


Profiling Complexity of Black vs. White Box

Sound model ≈ model with PI > 0[a]

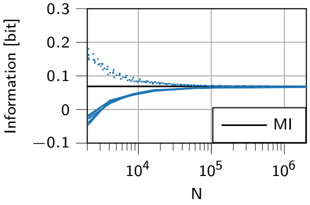
Ok if the profiling complexity $N \propto \frac{1}{LI-PI} \geq \frac{1}{MI}$

[a] Not always, see [**cryptoeprint:2021:1216**].



Example with masking:

Profiling Complexity of Black vs. White Box



Sound model ≈ model with PI > 0[a]

Ok if the profiling complexity $N \propto \frac{1}{LI - PI} \geq \frac{1}{MI}$

[a] Not always, see [**cryptoeprint:2021:1216**].

Example with masking:

Black-Box

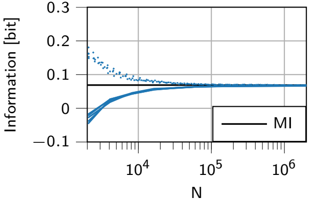
Profiles Y directly

Profiling Complexity of Black vs. White Box

Sound model ≈ model with PI > 0[a]

Ok if the profiling complexity $N \propto \frac{1}{LI-PI} \geq \frac{1}{MI}$

[a] Not always, see [**cryptoeprint:2021:1216**].



Example with masking:

Black-Box

Profiles Y directly

White-Box

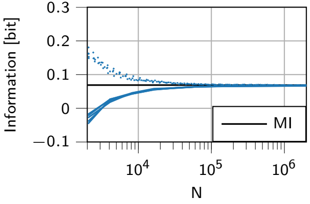
Profiles each share *separately*

Profiling Complexity of Black vs. White Box

Sound model ≈ model with PI > 0[a]

Ok if the profiling complexity $N \propto \frac{1}{LI-PI} \geq \frac{1}{MI}$

[a] Not always, see [**cryptoeprint:2021:1216**].



Example with masking:

Black-Box

Profiles Y directly

$\mathsf{MI}(\mathrm{Y}; \mathbf{L}) \propto \frac{1}{\sigma^{2d}}$

Prof. complexity ≈ Att. complexity

White-Box

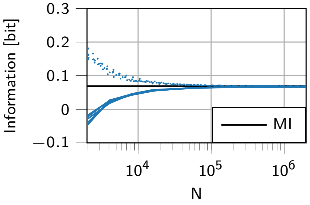
Profiles each share *separately*

Profiling Complexity of Black vs. White Box

Sound model ≈ model with PI > 0[a]

Ok if the profiling complexity $N \propto \frac{1}{LI - PI} \geq \frac{1}{MI}$

[a] Not always, see [**cryptoeprint:2021:1216**].



Example with masking:

Black-Box

Profiles Y directly

$\mathrm{MI}\left(\mathrm{Y}; \mathbf{L}\right) \propto \frac{1}{\sigma^{2d}}$

Prof. complexity ≈ Att. complexity

White-Box

Profiles each share *separately*

$\mathrm{MI}\left(\mathrm{Y}_i; \mathbf{L}_i\right) \propto \frac{1}{\sigma^{2}}$

Prof. complexity ≈ Att. complexity *without* masking

Content

Conclusion

Conclusion

We provide to the SCA evaluator some theoretical insights to assess the *profiling* complexity

Conclusion

We provide to the SCA evaluator some theoretical insights to assess the *profiling* complexity

HI can be replaced by a tighter metric: TI

Conclusion

- We provide to the SCA evaluator some theoretical insights to assess the *profiling* complexity
- HI can be replaced by a tighter metric: TI
- We explain why profiling in black-box may be much more difficult than in white-box, especially in presence of noise

Conclusion

We provide to the SCA evaluator some theoretical insights to assess the *profiling* complexity

HI can be replaced by a tighter metric: TI

We explain why profiling in black-box may be much more difficult than in white-box, especially in presence of noise

Open question: what about black-box profiling in low-noise settings (*e.g.* ASCAD datasets)?

Conclusion

- We provide to the SCA evaluator some theoretical insights to assess the *profiling* complexity
- HI can be replaced by a tighter metric: TI
- We explain why profiling in black-box may be much more difficult than in white-box, especially in presence of noise
- Open question: what about black-box profiling in low-noise settings (*e.g.* ASCAD datasets)?
  - Some evidences discussed in [**cryptoeprint:2022:493**]

References