Attacking masked cryptographic implementations: Information-theoretic bounds

Wei Cheng, Yi Liu, Sylvain Guilley, Olivier Rioul

Attacking Masked Cryptographic Implementations: Information-Theoretic Bounds

Wei Cheng*, Yi Liu*, Sylvain Guilley†*, and Olivier Rioul*

*LTCI, Télécom Paris, Institut Polytechnique de Paris, 91 120, Palaiseau, France, firstname.lastname@telecom-paris.fr
†Secure-IC S.A.S., 75 014, Paris, France, sylvain.guilley@secure-ic.com

***Abstract*—Measuring the information leakage is critical for evaluating the practical security of cryptographic devices against side-channel analysis. Information-theoretic measures can be used (along with Fano's inequality) to derive upper bounds on the success rate of any possible attack in terms of the number of side-channel measurements. Equivalently, this gives lower bounds on the number of queries for a given success probability of attack. In this paper, we consider cryptographic implementations protected by (first-order) masking schemes, and derive several information-theoretic bounds on the efficiency of any (second-order) attack. The obtained bounds are generic in that they do not depend on a specific attack but only on the leakage and masking models, through the mutual information between side-channel measurements and the secret key. Numerical evaluations confirm that our bounds reflect the practical performance of optimal maximum likelihood attacks.**

***Index Terms*—Side-Channel Analysis, Information-Theoretic Metric, Masking Scheme, Success Rate, Monte-Carlo Simulation.**

I. Introduction

Since the seminal work by Kocher et al. [1], side-channel analyses (SCAs) have been ones of the most powerful practical attacks against cryptographic devices. They exploit physically observable information leakage like instantaneous power consumption [1] or electromagnetic radiation [2] to extract secret keys as illustrated in Fig. 1.

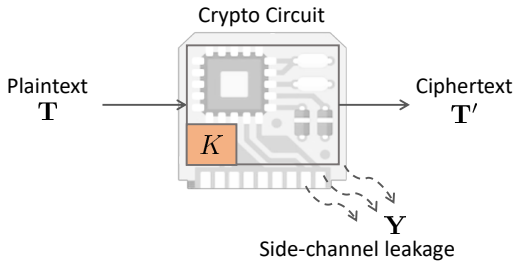


Fig. 1. Side-channel in a nutshell. An adversary attempts to recover the secret key K embedded in a cryptographic circuit by exploiting noisy side-channel leakage $\mathbf{Y}$ and public plaintext $\mathbf{T}$ (or ciphertext $\mathbf{T}'$).

In last two decades, many different types of attacks have been proposed to exploit various types of leakages. In particular, Heuser et al. [3] presented a channel representation of side-channel analysis to derive optimal (maximum likelihood) attacks that maximize success rate for a given leakage model. Other performance metrics such as guessing entropy also provide a fair comparison between different attacks [4].

To counteract SCAs, many countermeasures were proposed; *masking* is a well-established protection which provides provable security [5]–[7]. The idea is to split a sensitive (secret-dependent) variable into several shares and perform computations separately on each (secret-independent) share. Since the masks themselves are leaking, sound attacks against masked implementations must be multidimensional and require an exponentially high number of measurements in the number of shares to succeed [8].

A precise evaluation of the efficiency of *any* possible side-channel attack in the presence of countermeasures is an open problem. Given a set of side-channel measurements, can one establish a generic upper bound on the success rate of any attack? Several bounds have been proposed in [7], [8] by approximations and inequalities. The resulting lower bounds (on the number of traces needed for a given success rate) are quite loose. Chérisey et al. [9], [10] derived several upper bounds on the success rate using mutual information, which are tight in assessing *unprotected* cryptographic implementations. However, as we show in this paper, such bounds can also be very loose when targeting a *protected* cryptographic implementation.

In this paper, we aim at providing tight bounds on the success rate of any SCA by leveraging information-theoretic tools. To do so, we consider a channel framework similar to the ones proposed in [3], [9]–[11] but enhance it for masking schemes. The overview of the framework is shown in Fig. 2 with notations introduced in the following Subsection.

A. Notations

In the sequel, uppercase letters (e.g., X) denote random variables; lowercase letters (e.g., x) are for realizations (typically bytes); bold letters are for vectors, e.g., $\mathbf{X} = (X_1, X_2, \ldots, X_q)$. The cryptographic implementation typically works on bytes (e.g., of 8 bits) where the attacker, in a divide and conquer strategy, tries to recover each key byte K one by one. Let $T \oplus K$ be the bitwise exclusive or (XOR) operation between a text byte and a key. For a sequence of q text bytes $\mathbf{T}$ we write $\mathbf{T} \oplus K = (T_1 \oplus K, T_2 \oplus K, \ldots, T_q \oplus K)$. Also let $w_H(X)$ denote the *Hamming weight* of X and $w_H(\mathbf{X}) = (w_H(X_1), w_H(X_2), \ldots, w_H(X_q))$.

Throughout this paper we make the following notations as illustrated in Fig. 2:

- $K \in \mathbb{F}_{2^\ell}$ is the targeted key byte (typically $\ell = 8$, e.g., for AES);

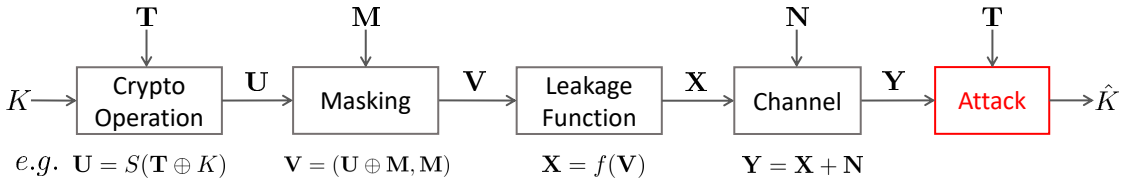


Fig. 2. Channel representation of side-channel analysis of a masked cryptographic operation.

- $\mathbf{T} \in \mathbb{F}_{2^\ell}^q$ denotes plaintext or ciphertext sequences, as vectors of length q;
- $\mathbf{U}$ is the *sensitive variable*, say $\mathbf{U} = S(\mathbf{T} \oplus K)$ where S denotes a cryptographic operation like the Sbox in AES;
- $\mathbf{V} = (\mathbf{U} \oplus \mathbf{M}, \mathbf{M})$ in a first-order Boolean masking with random mask $\mathbf{M} \in \mathbb{F}_{2^\ell}^q$; here $\mathbf{V} = (\mathbf{V}_1, \mathbf{V}_2) \in \mathbb{F}_{2^\ell}^{q \times 2}$ is a concatenation of $\mathbf{V}_1 = \mathbf{U} \oplus \mathbf{M}$ and $\mathbf{V}_2 = \mathbf{M}$; In the unprotected case (no masking) we would simply have $\mathbf{V} = \mathbf{U}$ as in [3], [10];
- $\mathbf{X} = f(\mathbf{V}) = f(\mathbf{V}_1) + f(\mathbf{V}_2)$ is the so-called deterministic leakage, where e.g., $f = w_H$ in well-known Hamming weight model as in [3]; more general models are possible;
- $\mathbf{Y} = \mathbf{X} + \mathbf{N}$ is the (noisy) leakage which models q measurements (a.k.a. traces) in practice, where $\mathbf{N}$ is an independent i.i.d. noise (memoryless additive channel); in particular $\mathbf{N} \sim \mathcal{N}(0, \sigma^2 \mathbf{I})$ for the AWGN channel.
- The attack is performed with a so-called distinguisher $\mathcal{D}$ which results in a guessed key $\hat{K} = \mathcal{D}(\mathbf{Y}, \mathbf{T})$.

From an information-theoretic perspective, it follows from Fig. 2 that conditionally on $\mathbf{T}$, we have a Markov chain:

$$K - \mathbf{U} - \mathbf{V} - \mathbf{X} - \mathbf{Y} - \hat{K}.$$

Remark 1: *It is important to note that Fig. 2 is* not *a genuine communication channel. The designer wants the secret key K to remain unknown and static (the same secret key is used for every side-channel use) as shown in Fig. 1. Therefore, there is no message to be intentionally encoded and transmitted: K leaks unintentionally. Besides, the actual (plain or encrypted) message $\mathbf{T}$ is public in our context and is supposedly known to the adversary. For all these reasons, our situation is totally different from problems such as those arising in a wiretap channel [12] for which a message is to be encoded, transmitted and decoded reliably in the presence of an eavesdropper.*

As recalled in [10] for a memoryless channel, we have the following relation to single-letter quantities: $I(\mathbf{X}; \mathbf{Y}|\mathbf{T}) \leq qI(X; Y|T)$. In particular, this explains why mutual information evaluation provides lower bounds on the number q of queries as in [10, §3.1]. Also, in [13, Theorem 4], the leakage metric is: $I(K; Y|T) = I(U; Y|T)$, which is implicitly connected to q [14].

B. Our Contributions

In this work, we derive security bounds for side-channel attacks in the presence of first-order masking countermeasures. Instead of using theoretical upper bounds on mutual information (MI) $I(\mathbf{X}; \mathbf{Y}|\mathbf{T})$ as in [9], [10], we numerically evaluate mutual information itself to derive bounds on the success rate thanks to Fano's inequality [15]. We also use $I(\mathbf{U}; \mathbf{Y}|\mathbf{T})$ in place of $I(\mathbf{X}; \mathbf{Y}|\mathbf{T})$ in the presence of masking because the resulting bounds are much tighter. Numerical results in a commonly used side-channel setting will confirm that our new bound provides more accurate security guarantees for the chip designer in the context of masked cryptographic implementations.

The remainder of this paper is organized as follows. Section II provides connections between mutual informations (MIs) for different pairs of variables in a side-channel setting. Section III presents several bounds on success rate. The numerical results for additive Gaussian noise are in Section IV. Finally, Section V concludes the paper.

II. Theoretical Preliminaries

A. Links between MIs of Different Variables

With the notations shown in Fig. 2 in the context of side-channel analysis, we have the following chain of equalities and inequalities for MIs on different pairs of variables.

Lemma 1: *With the above definitions and notations, one has*

$$I(K; \mathbf{Y}|\mathbf{T}) = I(\mathbf{U}; \mathbf{Y}|\mathbf{T}) \leq I(\mathbf{V}; \mathbf{Y}|\mathbf{T}) = I(\mathbf{X}; \mathbf{Y}|\mathbf{T}). \quad (1)$$

As a result, we shall restrict ourselves only on the two MIs $I(\mathbf{U}; \mathbf{Y}|\mathbf{T})$ and $I(\mathbf{X}; \mathbf{Y}|\mathbf{T})$, where the former will necessarily give a better bound than the latter.

Proof: Conditionally on $\mathbf{T}$, $K - \mathbf{U} - \mathbf{Y}$ is a Markov subchain; by the data processing inequality one has $I(K; \mathbf{Y}|\mathbf{T}) \leq I(\mathbf{U}; \mathbf{Y}|\mathbf{T})$. Now since $\mathbf{U} = S(\mathbf{T} \oplus K)$ is a deterministic function of K for fixed $\mathbf{T}$, $\mathbf{U} - K - \mathbf{Y}$ also forms a Markov chain conditionally on $\mathbf{T}$ and the converse inequality holds. This shows equality $I(K; \mathbf{Y}|\mathbf{T}) = I(\mathbf{U}; \mathbf{Y}|\mathbf{T})$. Similarly, conditionally on $\mathbf{T}$, $\mathbf{V} - \mathbf{X} - \mathbf{Y}$ is a Markov subchain, but since $\mathbf{X} = f(\mathbf{V})$, $\mathbf{X} - \mathbf{V} - \mathbf{Y}$ also forms a Markov chain. Then the data processing inequality in both directions implies equality $I(\mathbf{V}; \mathbf{Y}|\mathbf{T}) = I(\mathbf{X}; \mathbf{Y}|\mathbf{T})$. The data processing inequality applied to the Markov subchain $\mathbf{U} - \mathbf{V} - \mathbf{Y}$ gives $I(\mathbf{U}; \mathbf{Y}|\mathbf{T}) \leq I(\mathbf{V}; \mathbf{Y}|\mathbf{T})$ yet the converse is not true because of the presence of the unknown random mask $\mathbf{M}$. ■

Lemma 2: *With the above definitions and notations, for* any *attack,*

$$I(K; \hat{K}) \leq I(K; \hat{K}|\mathbf{T}) \leq I(K; \mathbf{Y}|\mathbf{T}). \quad (2)$$

Proof: Since conditioning reduces entropy, $H(K|\hat{K}) \geq H(K|\hat{K}, \mathbf{T})$. Then, since K is independent of $\mathbf{T}$, we have $I(K; \hat{K}|\mathbf{T}) = H(K|\mathbf{T}) - H(K|\hat{K}, \mathbf{T}) = H(K) - H(K|\hat{K}, \mathbf{T}) \geq H(K) - H(K|\hat{K}) = I(K; \hat{K})$. This proves the first inequality.

Secondly, given $\mathbf{T}$, we have a Markov chain: $K - \mathbf{Y} - \hat{K}$, since for fixed $\mathbf{T}$, $\hat{K} = \mathcal{D}(\mathbf{Y}, \mathbf{T})$ is a deterministic function of $\mathbf{Y}$. The data processing inequality ends the proof. ■

Remark 2: *The ML (maximum likelihood) rule* $\hat{k} = \mathcal{D}(\mathbf{y}, \mathbf{t}) = \arg\max_k \mathbb{P}(\mathbf{Y} = \mathbf{y}|k, \mathbf{T} = \mathbf{t})$ *gives the optimal distinguisher [3] when it coincides with MAP (Maximum A Posterior) rule for uniformly distributed* K *— a common assumption in SCA.*

A trivial upper bound on $I(K;\mathbf{Y}|\mathbf{T})$ is as follows.

Lemma 3: *With the above definitions and notations,*

$$I(K;\mathbf{Y}|\mathbf{T}) \leq H(K) \leq \ell. \tag{3}$$

where typically $\ell = 8$ *bits.*

Proof: $I(K;\mathbf{Y}|\mathbf{T}) = H(K|\mathbf{T}) - H(K|\mathbf{Y},\mathbf{T}) = H(K) - H(K|\mathbf{Y},\mathbf{T}) \leq H(K)$. ■

Lemma 3 simply reflects the fact that the total amount of information any adversary could extract cannot exceed the information carried by the secret key, as measured by the entropy $H(K)$. Notice that a common assumption in SCAs is that K is uniformly distributed, in which case $H(K) = \ell$.

B. Relation to Channel Capacity

Lemma 4: *With the above definitions and notations of Fig. 2,*

$$I(\mathbf{X};\mathbf{Y}) - I(\mathbf{T};\mathbf{Y}) = I(\mathbf{X};\mathbf{Y}|\mathbf{T}) \geq 0. \tag{4}$$

Proof: Since $\mathbf{T} - \mathbf{X} - \mathbf{Y}$ forms a Markov chain, one has $H(\mathbf{Y}|\mathbf{X},\mathbf{T}) = H(\mathbf{Y}|\mathbf{X})$ [1]. Hence $I(\mathbf{X};\mathbf{Y}|\mathbf{T}) = H(\mathbf{Y}|\mathbf{T}) - H(\mathbf{Y}|\mathbf{X},\mathbf{T}) = H(\mathbf{Y}|\mathbf{T}) - H(\mathbf{Y}|\mathbf{X}) = H(\mathbf{Y}) - H(\mathbf{Y}|\mathbf{X}) - \big(H(\mathbf{Y}) - H(\mathbf{Y}|\mathbf{T})\big) = I(\mathbf{X};\mathbf{Y}) - I(\mathbf{T};\mathbf{Y})$. ■

Note that the inequality $I(\mathbf{X};\mathbf{Y}) - I(\mathbf{T};\mathbf{Y}) \geq 0$ is also a direct consequence of the data processing inequality on the Markov chain $\mathbf{T} - \mathbf{X} - \mathbf{Y}$.

One is led to define the *capacity* of the side-channel (in bits per q channel uses) as

$$q\,C = \max_{\mathbf{T}-\mathbf{X}-\mathbf{Y}} I(\mathbf{X};\mathbf{Y}|\mathbf{T}) = \max_{\mathbf{T}-\mathbf{X}-\mathbf{Y}} I(\mathbf{X};\mathbf{Y}) - I(\mathbf{T};\mathbf{Y}), \tag{5}$$

where the maximum is taken over all distributions of $\mathbf{X}$ given $\mathbf{T}$ such that $\mathbf{T} - \mathbf{X} - \mathbf{Y}$ is a Markov chain. Because the "side information" $\mathbf{T}$ is known both at the "encoder" (leaking crypto) and "decoder" (attack), the capacity can be determined in the usual way:

Lemma 5: *With the above definitions and notations of Fig. 2 where the side-channel is independent of* $\mathbf{T}$*, one has*

$$q\,C = \max_{\mathbf{X}} I(\mathbf{X};\mathbf{Y}) \tag{6}$$

where the maximum is taken over all channel input distributions $\mathbf{X}$*.*

Proof: Since $I(\mathbf{X};\mathbf{Y}|\mathbf{T}) = \mathbb{E}_{\mathbf{T}} I(\mathbf{X};\mathbf{Y}|\mathbf{T} = \mathbf{t})$, we can choose $p(\mathbf{x}|\mathbf{t}) = p(\mathbf{x})$ to maximize each $I(\mathbf{X};\mathbf{Y}|\mathbf{T} = \mathbf{t})$ to achieve channel capacity in (5). As the optimal distribution does not depend on $\mathbf{t}$, it also maximizes the expectation $\mathbb{E}_{\mathbf{T}} I(\mathbf{X};\mathbf{Y}|\mathbf{T} = \mathbf{t}) = I(\mathbf{X};\mathbf{Y}|\mathbf{T})$ and thus $\max_{\mathbf{T}-\mathbf{X}-\mathbf{Y}} I(\mathbf{X};\mathbf{Y}|\mathbf{T}) = \max_{\mathbf{X}} I(\mathbf{X};\mathbf{Y})$. ■

Remark 3: *This result is also obtained by taking* $\mathbf{X}$ *(and thus* $\mathbf{Y}$*) independent of* $\mathbf{T}$ *such that* $I(\mathbf{T};\mathbf{Y}) = 0$ *in the preceding Lemma. We could also consider the more general situation where the channel also depends on* $\mathbf{T}$*. In this case we would have* $C = \mathbb{E}\{C_T\}$ *where* $q\,C_{\mathbf{t}} = \max_{\mathbf{X}} I(\mathbf{X};\mathbf{Y}|\mathbf{T} = \mathbf{t})$.

Remark 4: *As it turns out, capacity yields an upper bound on* $I(K;\mathbf{Y}|\mathbf{T})$ *which can improve the trivial upper bound of Lemma 3. This does not mean, however, that one is faced with a channel coding problem since the "encoder" hence* X*'s distribution cannot be chosen by the attacker.*

III. Bounds on the Success Probability of Attack

By combining Lemmas 1, 2 and 3, we have $I(K;\hat{K}) \leq I(\mathbf{U};\mathbf{Y}|\mathbf{T}) \leq H(K)$. Now the probability of success (estimated as the *success rate* in SCA) is defined as: $P_s = \mathbb{P}(\hat{K} = K)$. The corresponding "probability of error" (of attack failure) is $P_e = 1 - P_s$. Using Fano's inequality [15] we end up with the following theorem.

Theorem 1: *Given the side-channel setting as in Fig. 2, we have*

$$d_P(P_s) \leq I(\mathbf{U};\mathbf{Y}|\mathbf{T}), \tag{7}$$

where $d_P(p) = H(K) - H_2(p) - (1-p)\log(2^\ell - 1)$ *and* $H_2(p) = -p\log p - (1-p)\log(1-p)$*, for* $p \in [2^{-\ell}, 1]$*. (Recall that* ℓ *denotes the number of bits in* $K = k$*.)*

Proof: By Fano's inequality [15] and Lemma 2, we have $H(K) - H_2(P_s) - (1-P_s)\log(2^\ell - 1) \leq H(K) - H(K|\hat{K}) = I(K;\hat{K}) \leq I(\mathbf{U};\mathbf{Y}|\mathbf{T})$. ■

Since $d_P(p)$ is strictly increasing for $p \in [2^{-\ell}, 1]$ [16, §A] and $I(\mathbf{U};\mathbf{Y}|\mathbf{T})$ increases as q increases, Theorem 1 not only provides an upper bound on P_s, but also gives a *lower bound* on the number of queries q to obtain a specific value of P_s.

Remark 5: *A much looser bound on* P_s *can obtained from Lemmas 1 and 5. Using Theorem 1, one readily obtains*

$$d_P(P_s) \leq I(\mathbf{X};\mathbf{Y}|\mathbf{T}) \leq q\,C \tag{8}$$

where C *is the side-channel capacity, which is* $C = \frac{1}{2}\log(1 + \mathit{SNR})$ *for an AWGN channel* [2].

However, as we will show below, this bound is useless in evaluating masked implementations, particularly because $I(\mathbf{X};\mathbf{Y}|\mathbf{T})$ *is unbounded (compare with Lemma 3). In fact, we will show in next section that* $I(\mathbf{X};\mathbf{Y}|\mathbf{T})$ *is very close to the capacity* $q\,C$ *in the presence of a Boolean masking on an AWGN channel, hence it increases linearly in* q *without bound.*

IV. Application to Hamming Weight Leakages with Additive White Gaussian Noise

By the equalities of Lemma 1, the only two MIs that need to be evaluated are $I(\mathbf{X};\mathbf{Y}|\mathbf{T})$ and $I(\mathbf{U};\mathbf{Y}|\mathbf{T})$. Taking notations from Fig. 2, we calculate both MIs numerically. We have

$$\begin{aligned} I(\mathbf{X};\mathbf{Y}|\mathbf{T}) &= H(\mathbf{Y}|\mathbf{T}) - H(\mathbf{Y}|\mathbf{X},\mathbf{T}), \\ I(\mathbf{U};\mathbf{Y}|\mathbf{T}) &= H(\mathbf{Y}|\mathbf{T}) - H(\mathbf{Y}|\mathbf{U},\mathbf{T}), \end{aligned} \tag{9}$$

where for the AWGN channel

$$H(\mathbf{Y}|\mathbf{X},\mathbf{T}) = H(\mathbf{Y}|\mathbf{X}) = H(\mathbf{N}) = \frac{q}{2}\log\left(2\pi e\sigma^2\right), \tag{10}$$

and where $H(\mathbf{Y}|\mathbf{T})$ and $H(\mathbf{Y}|\mathbf{U},\mathbf{T}) = H(\mathbf{Y}|\mathbf{U})$ are estimated by Monte-Carlo simulations as shown next.

[1] We use H both discrete and continuous variables, even though h is used more frequently for differential entropy of a continuous variable.

[2] We use $\log_2$ to have mutual information and entropy expressed in bits.

A. Monte-Carlo Simulation

As the number of traces q gets very large, direct integration to evaluate mutual information becomes infeasible. Monte-Carlo simulation is a well-known method to estimate expectations of a function under certain distribution by repeated random sampling. We can then estimate the first term $H(\mathbf{Y}|\mathbf{T})$ in (9) by randomly drawing N_C samples:

$$\begin{aligned} H(\mathbf{Y}|\mathbf{T}) &= \int_{\mathbf{y}} \sum_{\mathbf{t}} p(\mathbf{y},\mathbf{t}) \log \frac{1}{p(\mathbf{y}|\mathbf{t})} \, \mathrm{d}\mathbf{y} \\ &= \lim_{N_C \to \infty} -\frac{1}{N_C} \sum_{j=1}^{N_C} \log p(\mathbf{y}^j|\mathbf{t}^j), \end{aligned} \tag{11}$$

where each $(\mathbf{t}^j, \mathbf{y}^j)$, for $1 \le j \le N_C$, is drawn randomly. The estimation in (11) is sound based on the law of large numbers [15, Chap. 3] and it has been numerically verified in [10]. Similarly, $H(\mathbf{Y}|\mathbf{U})$ can be estimated using Monte-Carlo simulation by $H(\mathbf{Y}|\mathbf{U}) = -\frac{1}{N_C} \sum_{j=1}^{N_C} \log p(\mathbf{y}^j|\mathbf{u}^j)$.

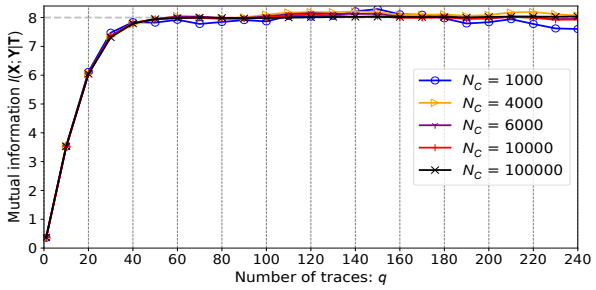


Fig. 3. Monte-Carlo simulation with various N_C draws where $\sigma^2 = 10.00$.

The accuracy of Monte-Carlo simulation highly depends on the number of samples. As an illustration, consider the unprotected case where there is no masking and for which $I(\mathbf{X};\mathbf{Y}|\mathbf{T})$ is bounded by $H(K) = 8$ bits. As shown in Fig. 3, the estimation of $I(\mathbf{X};\mathbf{Y}|\mathbf{T})$ gets more accurate by using larger N_C. In particular, this estimation on $I(\mathbf{X};\mathbf{Y}|\mathbf{T})$ is accurate enough by using only $N_C = 100,000$ draws. For all results in this paper we use $N_C = 1,000,000$ to obtain a very stable estimation.

B. Numerical Results for First-order Boolean Masking

Here $(\mathbf{t}^j, \mathbf{y}^j)$, for $1 \le j \le N_C$, is drawn i.i.d. according to this process:

- $\mathbf{t}^j \sim \mathcal{U}(\mathbb{F}_{2^\ell}^q)$,
- $\mathbf{m}^j \sim \mathcal{U}(\mathbb{F}_{2^\ell}^q)$,
- $k^j \sim \mathcal{U}(\mathbb{F}_{2^\ell})$, and
- $\mathbf{y}^j \sim \mathcal{N}(w_H(S(\mathbf{t}^j \oplus k^j) \oplus \mathbf{m}^j) + w_H(\mathbf{m}^j), \sigma^2 \mathbf{I}_q) \in \mathbb{R}^q$.

Note that we consider the zero-offset leakage [14] where the leakages of each share are summed together (see the sum of two Hamming weights above). For each draw $(\mathbf{t}, \mathbf{y})$, we have

$$\begin{aligned} p(\mathbf{y}|\mathbf{t}) &= \sum_k p(k) p(\mathbf{y}|\mathbf{t},k) = \sum_k p(k) \prod_{i=1}^{q} p(\mathbf{y}_i|\mathbf{t}_i,k) \\ &= \sum_k p(k) \prod_{i=1}^{q} \sum_{m_i} p(m_i) p(\mathbf{y}_i|\mathbf{t}_i,k,m_i) \\ &= \sum_k p(k) \prod_{i=1}^{q} \sum_{m_i} p(m_i) \frac{e^{\frac{-(\mathbf{y}_i - f(\mathbf{t}_i,k,m_i))^2}{2\sigma^2}}}{(2\pi\sigma^2)^{1/2}}, \end{aligned} \tag{12}$$

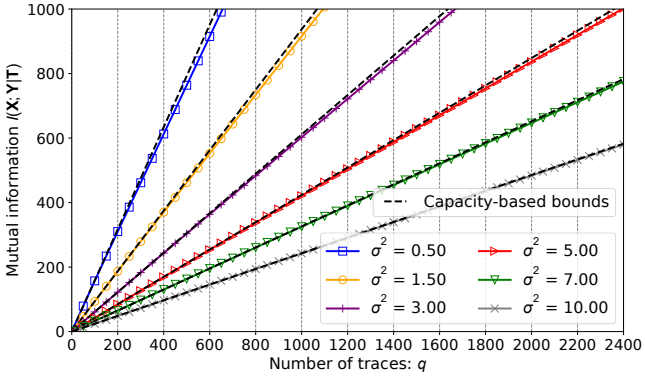
where $f(\mathbf{t}_i, k, m_i) = w_H(S(\mathbf{t}_i \oplus k) \oplus m_i) + w_H(m_i)$ is the zero-offset leakage under Hamming weight model. Again, taking $K \in \mathbb{F}_{2^\ell}$ uniformly, and considering that all masks are i.i.d. $\sim \mathcal{U}(\mathbb{F}_{2^\ell})$, we have

$$\begin{aligned} \log p(\mathbf{y}|\mathbf{t}) = -\ell(q+1) - \frac{q}{2} \log\left(2\pi\sigma^2\right) \\ + \log \sum_k \prod_{i=1}^{q} \sum_m e^{\frac{-(\mathbf{y}_i - f(\mathbf{t}_i,k,m))^2}{2\sigma^2}}. \end{aligned} \tag{13}$$

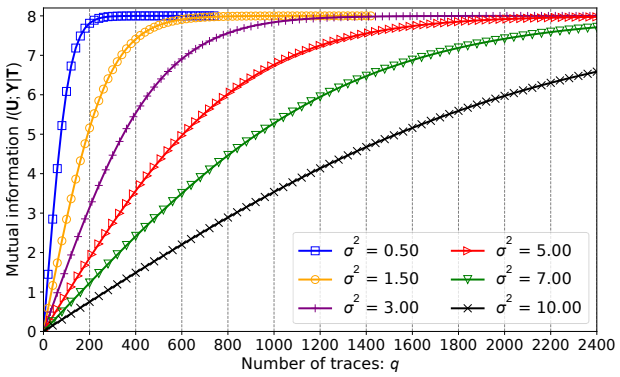
$$\begin{aligned} \log p(\mathbf{y}|\mathbf{u}) = -q\ell - \frac{q}{2} \log\left(2\pi\sigma^2\right) \\ + \log \prod_{i=1}^{q} \sum_m e^{\frac{-(\mathbf{y}_i - f'(\mathbf{u}_i,m))^2}{2\sigma^2}}. \end{aligned} \tag{14}$$

where $f'(\mathbf{u}_i, m) = w_H(\mathbf{u}_i \oplus m) + w_H(m)$.

The numerical results of $I(\mathbf{X};\mathbf{Y}|\mathbf{T})$ are depicted in Fig. 4(a). It clearly appears that the effect of masking is to increase the values of $I(\mathbf{X};\mathbf{Y}|\mathbf{T})$ without bound. This motivates our focus on $I(\mathbf{U};\mathbf{Y}|\mathbf{T})$. The dotted black lines in Fig. 4(a) show that upper bounds given by (8) are very tight.



(a) $I(\mathbf{X};\mathbf{Y}|\mathbf{T})$



(b) $I(\mathbf{U};\mathbf{Y}|\mathbf{T})$

Fig. 4. Evolution of mutual information $I(\mathbf{X};\mathbf{Y}|\mathbf{T})$ and $I(\mathbf{U};\mathbf{Y}|\mathbf{T})$ with the number of traces under different levels of noise in masked cases, with $N_C = 1,000,000$. Note that $I(\mathbf{X};\mathbf{Y}|\mathbf{T})$ is upper bounded by Shannon's channel capacity, while $I(\mathbf{U};\mathbf{Y}|\mathbf{T})$ is upper bounded by $H(K) = 8$ bits.

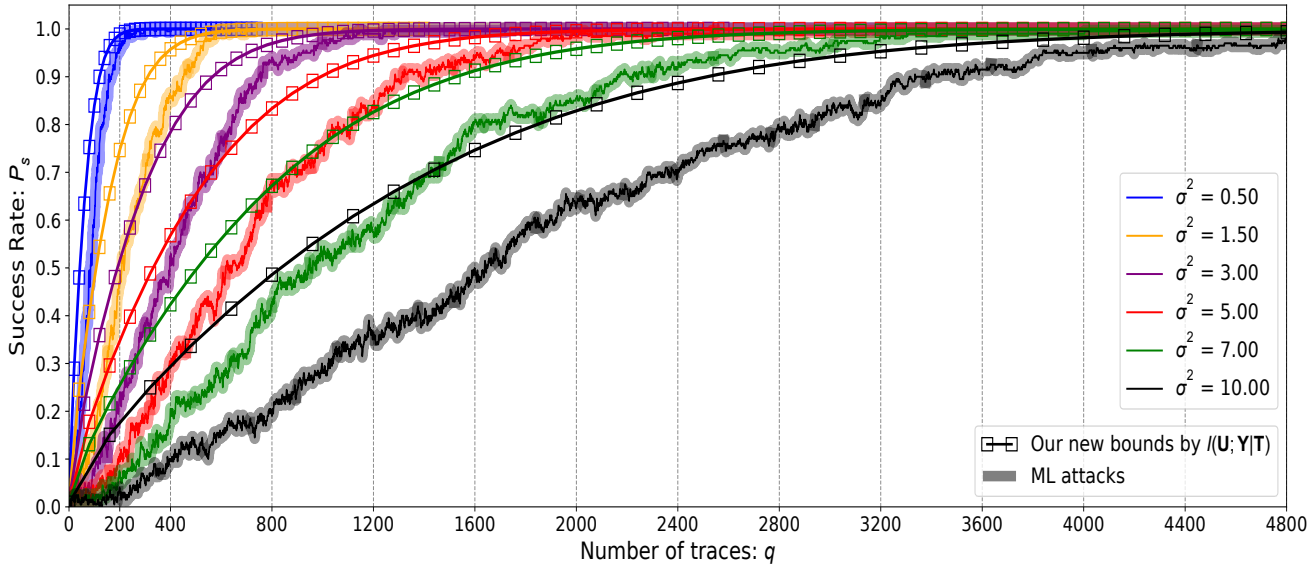


Fig. 5. Application and comparison of bounds on success rate. We present six instances with different noise levels by using $q_{\max} = 4800$ traces. Note that we omit the bounds given by $I(\mathbf{X};\mathbf{Y}|\mathbf{T})$ as they are invisible when plotted together with bounds given by $I(\mathbf{U};\mathbf{Y}|\mathbf{T})$.

As shown in Fig. 4(b), $I(\mathbf{U};\mathbf{Y}|\mathbf{T})$ is bounded as expected by $H(K)$ in Lemma 3. Particularly, given the same noise level, the number of traces needed to obtain $I(K;\mathbf{Y}|\mathbf{T}) = I(\mathbf{U};\mathbf{Y}|\mathbf{T}) \approx 8$ bits is much larger than in the unprotected case. The curves $I(\mathbf{U};\mathbf{Y}|\mathbf{T})$ vs σ^2 also look homothetic with a scale of σ^2. This is justified by a simple scaling argument: if the number of traces for a given set of $(\mathbf{T},\mathbf{U})$ is doubled, then the mutual information is the same as with the nominal number of queries, but with SNR doubled as well.

C. Bounds on Success Rate in Masked Implementations

By Theorem 1, we have an upper bound on probability of success P_s. This equivalently gives a lower bound on the minimum of q to get a specific P_s.

Numerical results are shown in Fig. 5 where we present several instances with different levels of Gaussian noises. In particular, the ML attacks utilize the higher-order distinguishers which have been demonstrated to be optimal in the presence of masking [17]. In order to evaluate P_s of ML attacks, each attack is repeated 200 times to have a more accurate success rate.

Figure 5 already shows the usefulness of the bound given by $I(\mathbf{U};\mathbf{Y}|\mathbf{T})$. Indeed, a commonly used metric on attacks is the minimum number of traces to reach $P_s \geq 95\%$. Considering $\sigma^2 = 3.00$ in Fig. 5, we set $P_s = 95\%$ and the ML attack needs around $q = 800$ traces, where our new bound gives $q = 720$, while the bound proposed in [10] by using $I(\mathbf{X};\mathbf{Y}|\mathbf{T})$ only gives $q = 12$. The comparison would be even worse for higher levels of noise.

Figure 6 provides a more detailed comparison by plotting the predicted minimum numbers of traces $q_{\min}$ reaching $P_s \geq 95\%$ given by both $I(\mathbf{U};\mathbf{Y}|\mathbf{T})$ and $I(\mathbf{X};\mathbf{Y}|\mathbf{T})$. These curves show that our new bound is much tighter than the previous one from the state-of-the-art [9], [10], as it captures the masking scheme — recall from Fig. 2 that the masking countermeasure step is between $\mathbf{U}$ and $\mathbf{Y}$ but not between $\mathbf{X}$ and $\mathbf{Y}$.

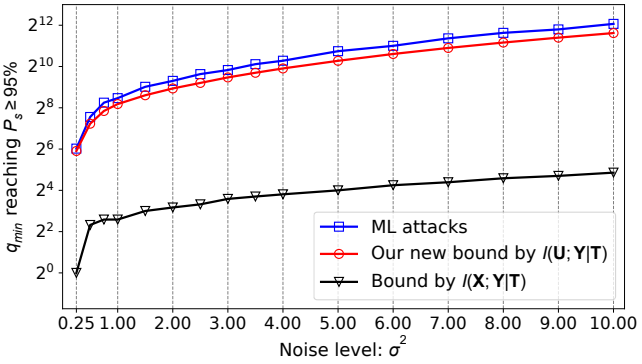


Fig. 6. Comparison of the minimum number of traces $q_{\min}$ to reach $P_s \geq 95\%$ predicted by our new bound, by $I(\mathbf{X};\mathbf{Y}|\mathbf{T})$ as in [10] and also the baseline given by an ML attack.

V. Conclusions

We derived security bounds for side-channel attacks in the presence of countermeasures (first-order masking). To do this, we leveraged the seminal framework from Chérisey et al. [9], [10] and extended it to the masking case of a protection aiming at randomizing the leakage.

The generalization not only enhances bounds compared to Chérisey et al., but also improves on the computation method for the security metric, by resorting to a powerful information estimation based on the Monte Carlo method. Our results provide quantitative bounds allowing for the theoretical ("pre-silicon") evaluation of protections applied on top of a given cryptographic algorithm in designing secure circuits. As a perspective, we will push forward the practical applications of our findings in evaluating concrete security level of cryptographic circuits.

REFERENCES

[1] P. C. Kocher, J. Jaffe, and B. Jun, "Differential Power Analysis," in *CRYPTO*, ser. Lecture Notes in Computer Science, M. J. Wiener, Ed., vol. 1666. Springer, 1999, pp. 388–397.

[2] K. Gandolfi, C. Mourtel, and F. Olivier, "Electromagnetic Analysis: Concrete Results," in *Proceedings of the Third International Workshop on Cryptographic Hardware and Embedded Systems*, ser. CHES '01. London, UK, UK: Springer-Verlag, 2001, pp. 251–261. [Online]. Available: http://dl.acm.org/citation.cfm?id=648254.752700

[3] A. Heuser, O. Rioul, and S. Guilley, "Good Is Not Good Enough - Deriving Optimal Distinguishers from Communication Theory," in *Cryptographic Hardware and Embedded Systems - CHES 2014 - 16th International Workshop, Busan, South Korea, September 23-26, 2014. Proceedings*, ser. Lecture Notes in Computer Science, L. Batina and M. Robshaw, Eds., vol. 8731. Springer, 2014, pp. 55–74. [Online]. Available: http://dx.doi.org/10.1007/978-3-662-44709-3_4

[4] F.-X. Standaert, T. Malkin, and M. Yung, "A Unified Framework for the Analysis of Side-Channel Key Recovery Attacks," in *EUROCRYPT*, ser. LNCS, vol. 5479. Springer, April 26-30 2009, pp. 443–461, Cologne, Germany.

[5] Y. Ishai, A. Sahai, and D. Wagner, "Private Circuits: Securing Hardware against Probing Attacks," in *CRYPTO*, ser. Lecture Notes in Computer Science, vol. 2729. Springer, August 17–21 2003, pp. 463–481, Santa Barbara, California, USA.

[6] M. Rivain and E. Prouff, "Provably Secure Higher-Order Masking of AES," in *CHES*, ser. LNCS, S. Mangard and F.-X. Standaert, Eds., vol. 6225. Springer, 2010, pp. 413–427.

[7] E. Prouff and M. Rivain, "Masking against Side-Channel Attacks: A Formal Security Proof," in *Advances in Cryptology - EUROCRYPT 2013, 32nd Annual International Conference on the Theory and Applications of Cryptographic Techniques, Athens, Greece, May 26-30, 2013. Proceedings*, ser. Lecture Notes in Computer Science, T. Johansson and P. Q. Nguyen, Eds., vol. 7881. Springer, 2013, pp. 142–159. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-38348-9_9

[8] A. Duc, S. Faust, and F. Standaert, "Making Masking Security Proofs Concrete - Or How to Evaluate the Security of Any Leaking Device," in *Advances in Cryptology - EUROCRYPT 2015 - 34th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Sofia, Bulgaria, April 26-30, 2015, Proceedings, Part I*, ser. Lecture Notes in Computer Science, E. Oswald and M. Fischlin, Eds., vol. 9056. Springer, 2015, pp. 401–429. [Online]. Available: http://dx.doi.org/10.1007/978-3-662-46800-5_16

[9] É. de Chérisey, S. Guilley, O. Rioul, and P. Piantanida, "An Information-Theoretic Model for Side-Channel Attacks in Embedded Hardware," in *IEEE International Symposium on Information Theory, ISIT 2019, Paris, France, July 7-12, 2019*. IEEE, 2019, pp. 310–315. [Online]. Available: https://doi.org/10.1109/ISIT.2019.8849763

[10] É. de Chérisey, S. Guilley, O. Rioul, and P. Piantanida, "Best Information is Most Successful — Mutual Information and Success Rate in Side-Channel Analysis," *IACR Trans. Cryptogr. Hardw. Embed. Syst.*, vol. 2019, no. 2, pp. 49–79, 2019. [Online]. Available: https://doi.org/10.13154/tches.v2019.i2.49-79

[11] S. Guilley, A. Heuser, and O. Rioul, "Codes for Side-Channel Attacks and Protections," in *Codes, Cryptology and Information Security - Second International Conference, C2SI 2017, Rabat, Morocco, April 10-12, 2017, Proceedings - In Honor of Claude Carlet*, ser. Lecture Notes in Computer Science, S. E. Hajji, A. Nitaj, and E. M. Souidi, Eds., vol. 10194. Springer, 2017, pp. 35–55. [Online]. Available: https://doi.org/10.1007/978-3-319-55589-8_3

[12] H. Tyagi and A. Vardy, "Semantically-Secure Coding Scheme Achieving the Capacity of a Gaussian Wiretap Channel," *CoRR*, vol. abs/1412.4958, 2014. [Online]. Available: http://arxiv.org/abs/1412.4958

[13] W. Cheng, S. Guilley, C. Carlet, S. Mesnager, and J.-L. Danger, "Optimizing Inner Product Masking Scheme by a Coding Theory Approach," *IEEE Trans. Inf. Forensics Secur.*, vol. 16, pp. 220–235, 2021. [Online]. Available: https://doi.org/10.1109/TIFS.2020.3009609

[14] C. Carlet and S. Guilley, "Statistical Properties of Side-Channel and Fault Injection Attacks Using Coding Theory," *Cryptography and Communications*, vol. 10, no. 5, pp. 909–933, 2018. [Online]. Available: https://doi.org/10.1007/s12095-017-0271-4

[16] É. de Chérisey, S. Guilley, O. Rioul, and P. Piantanida, "Best Information is Most Successful," Cryptology ePrint Archive, Report 2019/491, extended version of [10], 2019, https://eprint.iacr.org/2019/491.

[17] N. Bruneau, S. Guilley, A. Heuser, and O. Rioul, "Masks Will Fall Off – Higher-Order Optimal Distinguishers," in *Advances in Cryptology – ASIACRYPT 2014 - 20th International Conference on the Theory and Application of Cryptology and Information Security, Kaoshiung, Taiwan, R.O.C., December 7-11, 2014, Proceedings, Part II*, ser. Lecture Notes in Computer Science, P. Sarkar and T. Iwata, Eds., vol. 8874. Springer, 2014, pp. 344–365. [Online]. Available: http://dx.doi.org/10.1007/978-3-662-45608-8_19

[15] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley-Interscience, July 18 2006, ISBN-10: ISBN-10: 0471241954, ISBN-13: 978-0471241959, 2nd edition.