Tight and Scalable Side-Channel Attack Evaluations through Asymptotically Optimal Massey-like Inequalities on Guessing Entropy

Andrei Tănăsescu, Marios Choudary, Olivier Rioul, Pantelimon George Popescu

Article

Tight and Scalable Side-Channel Attack Evaluations through Asymptotically Optimal Massey-Like Inequalities on Guessing Entropy

Andrei Tănăsescu [1], Marios O. Choudary [1], Olivier Rioul [2], Pantelimon George Popescu [1,*]

[1] Department of Computer Science and Engineering, University Politehnica of Bucharest, Splaiul Independentei 313 (6), 060042 Bucharest, Romania; andrei.tanasescu@mail.ru (A.T.); marios.choudary@upb.ro (M.O.C.)
[2] LTCI, Télécom Paris, Institut Polytechnique de Paris, 91120 Palaiseau, France; rioul@enst.fr (O.R.)
* Correspondence: pgpopescu@yahoo.com

Abstract: The bounds presented at CHES 2017 based on Massey's guessing entropy represent the most scalable side-channel security evaluation method to date. In this paper, we present an improvement of this method, by determining the asymptotically optimal Massey-like inequality and then further refining it for finite support distributions. The impact of these results is highlighted for side-channel attack evaluations, demonstrating the improvements over the CHES 2017 bounds.

Keywords: guessing entropy; side-channel attacks; shannon entropy; massey inequality

1. Introduction

Side-channel attacks on electronic devices have become a very important threat for our society, as shown by European reports [1,2] as well as several recent publications [3,4]. To deal with such threats it is important that devices used in security–critical applications are well protected. Such protection is typically verified through security certifications.

A typical scenario for certifications on cryptographic algorithms such as AES is the estimation of attack success probability as a function of time or data availability. At the beginning of the last decade, a certification would usually include estimating the time necessary for recovering one secret byte from an AES implementation after running a side-channel attack. As such, security metrics such as Success Rate or empirical Guessing Entropy [5] have become popular in this context. But soon afterwards, the question arose how to estimate the probability of success in recovering the entire key (e.g., 16 key bytes for AES-128), not just one byte. In this situation, directly applying the existing Success Rate or Guessing Entropy metrics would not work, due to the impossibility of dealing with all 2^{128} possible values of a 16-byte key. Hence, several algorithmic approaches [6–11], were developed to estimate e.g., the Guessing Entropy when dealing with full cryptographic keys such as the 16-byte keys used with AES-128.

However, these methods could not scale to deal with very large cryptographic keys, beyond 128 bytes, such as 8912-bit (1024-byte) RSA keys. To deal with this problem, Choudary and Popescu [12] presented a new approach based on mathematical bounds for Massey's guessing entropy [13]. Their approach could easily handle very large keys, with even beyond 1024 bytes.

Nevertheless, a suggestion was made by Grosso [14] that the bounds of Choudary and Popescu could not be improved, hence this would provide a limitation of that method.

In this paper, we show that this is not the case, by actually tightening the results of Choudary and Popescu, through the derivation of new relations between Massey's guessing entropy and Shannon's entropy. These important mathematical results are then validated through concrete side-channel attack experiments.

In brief, the main contributions of this paper are as follows:

1. We demonstrate that a recent improvement on Massey's inequality between Massey' Guessing entropy and Shannon's entropy (Rioul's improved inequality) is asymptotically optimal (which is highly relevant to scalability).
2. We provide a new improvement on Massey's inequality that is even tighter than the above for all finite-size data distributions.
3. We extend and prove the above results when dealing with multiple lists of probabilities (distributions), as is the case when dealing with the results of side-channel attacks on multiple key bytes (proving scalability).
4. We apply our results on concrete side-channel attack datasets to demonstrate the improvements of the methods from this paper over the state of the art.

2. Preliminaries

The guessing entropy associated with a (positive descending) probability distribution $\mathbf{p} = (p_1, p_2, \ldots, p_n)$ with $p_1 \geq \cdots \geq p_n > 0$ is the expected value of the random variable $G(\mathbf{p})$ given by $\mathbb{P}[G(\mathbf{p}) = i] = p_i$ $(i = 1, \ldots, n)$, i.e., $\mathbb{E}[G(\mathbf{p})] = \sum_{i=1}^{n} i p_i$. It corresponds to the minimal average number of binary questions required to guess the value of a random variable distributed according to $\mathbf{p}$ [13]. J. Massey has provided a well-known relation between guessing entropy and the Shannon entropy (In this formula, as well as the remaining of the paper $\log()$ denotes logarithm to base 2.) $H(\mathbf{p}) = -\sum_{i=1}^{n} p_i \log p_i$, which reads [13]:

$$\mathbb{E}[G(\mathbf{p})] \geq 2^{H(\mathbf{p})-2} + 1, \tag{1}$$

when $H(\mathbf{p}) \geq 2$ bits.

Massey's inequality has been recently improved in various ways, yet all known refinements share the same shape. For instance, in an ISIT paper, Popescu and Choudary [15] proved

$$\mathbb{E}[G(\mathbf{p})] \geq 2^{H(\mathbf{p})+2p_n-2} + 1 - p_n \geq 2^{H(\mathbf{p})+p_n-2} + 1 - \frac{1}{2}p_n \geq 2^{H(\mathbf{p})-2} + 1,$$

subject to the same condition $H(\mathbf{p}) \geq 2$ bits as in the Massey inequality. Meanwhile, Rioul's inequality [16], published in a CHES paper [17] states that for all values of $H(\mathbf{p}) \geq 0$,

$$\mathbb{E}[G(\mathbf{p})] > \frac{1}{e}2^{H(\mathbf{p})}, \tag{2}$$

which refines Massey's inequality when $H(\mathbf{p}) \geq \log \frac{e}{1-e/4} \approx 3$. Finally, in an Entropy paper, Tanasescu and Popescu [18] found that under the same condition as in Massey's inequality (here $h(\alpha)$ is the binary Shannon entropy),

$$\mathbb{E}[G(\mathbf{p})] \geq \sup_{\alpha \in [0,1/2]} 2^{H(\mathbf{p})+\frac{h(\alpha)}{1-\alpha}p_n-2} + 1 - \frac{\alpha}{1-\alpha}p_n \geq 2^{H(\mathbf{p})+2p_n-2} + 1 - p_n > 2^{H(\mathbf{p})-2} + 1.$$

The authors of [18] hinted that a similar refinement can be found for inequality (2).

In the following section, we present one such refinement, by first optimising exponential relations between the guessing and Shannon entropies, i.e., lower bounds of the form $\mathbb{E}[G(\mathbf{p})] \geq a \cdot b^{H(\mathbf{p})} + c$ valid when the Shannon entropy lies above a given threshold. We thus arrive at an improved Rioul's inequality [19] by an additive constant of $1/2$, which is asymptotically optimal among other global lower bounds depending only on the Shannon entropy as $H(\mathbf{p}) \to \infty$ (such condition reflects variables that can take a very large number of values, e.g., when dealing with very large cryptographic keys in our examples). Then, using the techniques of [15,18] we further refine this inequality for finite support distributions allowing us to increase the multiplicative constant depending on the smallest probability p_n. Finally, then, we apply our results in the context of side-channel attacks, where guessing entropy is a key metric [20–22], showing that our results provide an improvement (tighter bounds) over the method of Choudary and Popescu [12], which is known as the most scalable full-key security evaluation method to date.

3. The Asymptotically Optimal Massey-Like Inequality

Considering the increasingly large key space of cryptographic systems, in this section we seek the best Massey-like inequality $\mathbb{E}[G(\mathbf{p})] \geq a \cdot b^{H(\mathbf{p})} + c$, in the sense that it is optimal for arbitrarily large entropy, i.e., when $H(\mathbf{p}) \to \infty$, as can be obtained for infinite support (infinitely large probability lists). Then, we show that this asymptotically optimal bound also holds for all possible distributions.

Recently, [19] proposed an improved version of Rioul's inequality

$$\mathbb{E}[G(\mathbf{p})] \geq \frac{1}{e} 2^{H(\mathbf{p})} + \frac{1}{2}. \tag{3}$$

Now we show that as $H(\mathbf{p}) \to \infty$ this inequality is in fact the optimal Massey-like inequality.

Theorem 1. *The Massey-like inequality* (3) *is asymptotically optimal.*

Proof. Following Massey's approach [13], the best lower bound on guessing entropy based on Shannon entropy is sharp for geometric distributions, i.e., for any guessing entropy value $\mathbb{E}[G(\mathbf{p})] = \mu$, the maximal Shannon entropy is obtained for the geometric distribution with mean μ, $p_i = \frac{1}{\mu}\left(1 - \frac{1}{\mu}\right)^{i-1}$,

$$H(\mathbf{p}) \leq \log(\mu - 1) - \mu \log(1 - 1/\mu),$$

as found by Massey [13]. Moreover, in practical applications where $\mathbf{p}$ has finite length, this inequality is actually strict, but the upper bound can be approached as closely as desired if the list of probabilities is long enough.

We seek bounds of the form $\mathbb{E}[G(\mathbf{p})] \geq a \cdot b^{H(\mathbf{p})} + c$, i.e., $H(\mathbf{p}) \leq \log_b \frac{\mu - c}{a}$. In order for this to be valid for all μ, we should necessarily have

$$\log_b \frac{\mu - c}{a} \geq \log(\mu - 1) - \mu \log(1 - 1/\mu).$$

In particular, as $\mu \to \infty$, the expression on the left has asymptotic

$$\log_b \frac{\mu - c}{a} = \log_b \mu - \log_b a - \frac{c \log_b e}{\mu} + o(1/\mu),$$

while the expression on the right has asymptotic

$$\log(\mu - 1) - \mu \log(1 - 1/\mu) = \log \mu + \log e - \frac{\log e}{2\mu} + o(1/\mu).$$

As a consequence, we necessarily have $\log_b \mu \geq \log \mu$, i.e., $\log b \leq 1$ or $b \leq 2$, so that the optimal (maximum) value of b is $b = 2$. Next, we should have $-\log a \geq \log e$, i.e., $a \leq 1/e$, so that the optimal (maximum) value of a is $1/e$. Finally, we should have $-c \log e \geq -(\log e)/2$, i.e., $c \leq 1/2$, so that the optimal (maximum) value of c is $c = 1/2$.

The asymptotically optimal bound then writes

$$\log(\mu - 1/2) + \log e \geq \log(\mu - 1) - \mu \log(1 - 1/\mu) \tag{4}$$

which readily gives (3) when μ or $H(\mathbf{p})$ tend to infinity.

A simple proof of (3) for all values of $H(\mathbf{p}) > 0$ can be found in [19]. □

We conclude this section by remarking that the optimal Massey-like inequality in Theorem 1 is very general, as it also holds for small support corresponding to a few bytes, and even for very small entropy, $H(\mathbf{p}) \to 0$, improving on the original Massey inequality which holds just when $H(\mathbf{p}) \geq 2$.

4. Refinement for Finite Support Distributions

In this section, we show a new relation between the Shannon and guessing entropy, dependent on the minimal probability of a given distribution, further refining Rioul's improved inequality (3).

We begin with a direct improvement of Theorem 1 following the technique [15,18] used to improve the Massey inequality. To this end, we make use of the binary Shannon entropy, $h(\alpha) = -\alpha \log \alpha - (1-\alpha)\log(1-\alpha)$ for $0 \leq \alpha \leq 1$.

Lemma 1. *For any positive descending probability distribution $\mathbf{p} \in \mathbb{R}^n$ such that $H(\mathbf{p}) \geq 1$ bit, we have*

$$\mathbb{E}[G(\mathbf{p})] \geq \sup_{\alpha \in [0,1/2]} \frac{1}{e} 2^{H(\mathbf{p})+p_n h(\alpha)} - \alpha p_n + \frac{1}{2} \geq \frac{1}{e} 2^{H(\mathbf{p})+p_n} - \frac{1}{2} p_n + \frac{1}{2} \geq \frac{1}{e} 2^{H(\mathbf{p})} + \frac{1}{2}.$$

Proof. Consider a positive decreasing distribution $\mathbf{p} = (p_1, p_2, \ldots, p_n)$ with $H(\mathbf{p}) \geq 2$. Following the approach in [15] we construct the new probability distribution $\mathbf{q} = (p_1, p_2, \ldots, p_{n-1}, (1-\alpha)p_n, \alpha p_n)$, which is decreasing and strictly positive if and only if $\alpha \in (0, 1/2]$. From the grouping property of entropy, $H(\mathbf{q}) = H(\mathbf{p}) + p_n h(\alpha)$, and moreover $\mathbb{E}[G(\mathbf{q})] = \mathbb{E}[G(\mathbf{p})] + \alpha p_n$. Then

$$\begin{aligned} \mathbb{E}[G(\mathbf{p})] =& \mathbb{E}[G(\mathbf{q})] - \alpha p_n \geq \frac{1}{e} 2^{H(\mathbf{q})} - \alpha p_n + \frac{1}{2} \\ =& \frac{1}{e} 2^{H(\mathbf{p})+p_n h(\alpha)} - \alpha p_n + \frac{1}{2}. \end{aligned} \tag{5}$$

The first desired inequality follows taking the supremum over α in eq. (5), the second by substituting $\alpha = 1/2$. To justify the third, we use $2^x > 1 + x \ln 2$ for $x = p_n$ obtaining

$$\frac{1}{e} 2^{H(\mathbf{p})+p_n} - \frac{1}{2} p_n \geq \frac{1}{e} 2^{H(\mathbf{p})} (1 + p_n \ln 2) - \frac{1}{2} p_n = \frac{1}{e} 2^{H(\mathbf{p})} + \left(\frac{2^{H(\mathbf{p})} \ln 2}{e} - \frac{1}{2} \right) p_n,$$

where p_n's coefficient is positive whenever $H(\mathbf{p}) \geq 1 \geq \log \frac{e}{2 \ln 2}$. This ends the proof. □

We can further refine this lemma using the generalization techniques of [15,18] as follows.

Theorem 2. *For any positive descending probability distributions $\mathbf{p} \in \mathbb{R}^n$ such that $H(\mathbf{p}) \geq 1$, we have*

$$\mathbb{E}[G(\mathbf{p})] \geq \sup_{\alpha \in [0,1/2]} \frac{1}{e} 2^{H(\mathbf{p}) + \frac{h(\alpha)}{1-\alpha} p_n} + \frac{1}{2} - \frac{\alpha}{1-\alpha} p_n \geq \frac{1}{e} 2^{H(\mathbf{p}) + \frac{1}{2} p_n} + \frac{1}{2} - p_n \geq \frac{1}{e} 2^{H(\mathbf{p})} + \frac{1}{2}.$$

Proof. Given the initial decreasing $\mathbf{p}$, we construct a sequence of probability distributions $\{\mathbf{Q}_k\}$, recursively defined using the procedure in the previous proof.

We begin by fixing an arbitrary parameter $\alpha \in [0, 1/2]$ as above. Denoting by $Q_{k,i}$ the i^{th} component of the sequence $\mathbf{Q}_k$, we define the terms of the list $\{\mathbf{Q}_k\}$ as follows. We let the support of the first term coincide with $\mathbf{p}$, i.e., $\mathbf{Q}_0 = (p_0, p_1, \ldots, p_n, 0, 0, \ldots, 0, \ldots)$, and we define the other terms by recurrence:

$$\begin{aligned} \mathbf{Q}_{k+1} =& (Q_{k,0}, Q_{k,1}, \ldots, Q_{k,n+k-1}, \\ & (1-\alpha) Q_{k,n+k}, \alpha Q_{k,n+k}, 0, 0, \ldots, 0, \ldots). \end{aligned}$$

and at each step of the construction, we have the inequality

$$\mathbb{E}[G(\mathbf{Q}_k)] = \mathbb{E}[G(\mathbf{Q}_{k+1})] - \alpha Q_{k,n+k} \geq \frac{2^{H(\mathbf{Q}_{k+1})}}{e} - \alpha Q_{k,n+k} > \frac{2^{H(\mathbf{Q}_k)}}{e} + \frac{1}{2}.$$

After the first k steps of the construction we find

$$\begin{aligned}\mathbb{E}[G(\mathbf{p})] =& \mathbb{E}[G(\mathbf{Q}_k)] - p_n\alpha\frac{1-\alpha^k}{1-\alpha}\\ =& \mathbb{E}[G(\mathbf{Q}_k)] + \sum_{j=0}^{k-1}\left(\mathbb{E}\left[G\left(\mathbf{Q}_j\right)\right] - \mathbb{E}\left[G\left(\mathbf{Q}_{j+1}\right)\right]\right)\\ \geq& \frac{1}{2}2^{H(\mathbf{Q}_k)} + \frac{1}{2} + \sum_{j=0}^{k-1}\left(\mathbb{E}\left[G\left(\mathbf{Q}_j\right)\right] - \mathbb{E}\left[G\left(\mathbf{Q}_{j+1}\right)\right]\right)\\ >& \frac{1}{e}2^{H(\mathbf{Q}_{k-1})} + \frac{1}{2} + \sum_{j=0}^{k-2}\left(\mathbb{E}\left[G\left(\mathbf{Q}_j\right)\right] - \mathbb{E}\left[G\left(\mathbf{Q}_{j+1}\right)\right]\right)\\ >& \cdots > \frac{1}{e}2^{H(\mathbf{Q}_0)} + \frac{1}{2} = \frac{1}{e}2^{H(\mathbf{p})} + \frac{1}{2},\end{aligned}$$

where the tightest of the enumerated bounds is

$$\mathbb{E}[G(\mathbf{p})] \geq \frac{1}{e}2^{H(\mathbf{Q}_k)} + \frac{1}{2} + \sum_{j=0}^{k-1}\left(\mathbb{E}\left[G\left(\mathbf{Q}_j\right)\right] - \mathbb{E}\left[G\left(\mathbf{Q}_{j+1}\right)\right]\right) = \frac{1}{e}2^{H(\mathbf{p})+p_n h(\alpha)\frac{1-\alpha^k}{1-\alpha}} + \frac{1}{2} - p_n\alpha\frac{1-\alpha^k}{1-\alpha},$$

which as we have shown increases with k up to the limit

$$\mathbb{E}[G(\mathbf{p})] \geq \frac{1}{e}2^{H(\mathbf{p})+p_n\frac{h(\alpha)}{1-\alpha}} + \frac{1}{2} - p_n\frac{\alpha}{1-\alpha}$$

valid for any $\alpha \in [0, 1/2]$. The first desired inequality now follows taking *supremum* over the last equation, the second by substituting $\alpha = 1/2$ and the third by noting that all bounds in the sequence are greater than the last one $\frac{1}{e}2^{H(\mathbf{p})} + \frac{1}{2}$. □

We conclude this section by remarking that Theorem 2 provides a very scalable result as both the Shannon entropy of a joint probability distribution and its minimal entry are very easy to compute, as we will show in the following section.

5. Scalability of Bounds

For side-channel attack evaluations on full cryptographic keys (e.g., 16-byte AES keys or 1024-byte RSA keys), we need to combine the attack results on each key byte to derive a security metric that estimates as well as possible the difficulty of recovering the entire key. For example, given the lists of probabilities $\mathbf{p}^1 = \{p_1^1, p_2^1, \ldots, p_{256}^1\}$, $\mathbf{p}^2 = \{p_1^2, p_2^2, \ldots, p_{256}^2\}$, …, $\mathbf{p}^{16} = \{p_1^{16}, p_2^{16}, \ldots, p_{256}^{16}\}$ obtained after applying a side-channel attack such as the Template Attack [21,23] for the 16 key bytes of AES, we need security metrics that can use this information efficiently.

In this context, Choudary and Popescu [12] have provided the following bounds ($\mathrm{LB_{GM}}$ and $\mathrm{UB_{GM}}$ for full key):

$$\frac{1}{1+\ln n^{N_s}}\prod_{i=1}^{N_s}\left[\sum_{k=1}^{n}\sqrt{p_k^i}\right]^2 \leq \mathbb{E}[G(\mathbf{p})]^f \leq \frac{1}{2}\prod_{i=1}^{N_s}\left[\sum_{k=1}^{n}\sqrt{p_k^i}\right]^2 + \frac{1}{2}, \tag{6}$$

where N_s is the number of bytes (e.g., $N_s = 16$ for AES-128), n represents the number of values per byte (list) and $\mathbb{E}[G(\mathbf{p})]^f$ represents the guessing entropy for the full-key (which cannot be computed for large N_s, e.g., $N_s >= 10$).

Below we show how to extend the bounds from this paper to apply them in the full-key context.

Theorem 3. *For any full list of probabilities* $\mathbf{p}$ *we have*

$$\mathbb{E}[G(\mathbf{p})]^f \geq \frac{1}{e}2^{\sum_{k=1}^{N_s}H(\mathbf{p}^k)} + \frac{1}{2}.$$

Theorem 4. *For any positive descending probability vectors* $\left\{\mathbf{p}^k\right\}_{k=1}^{N_s} \subseteq \mathbb{R}^n$ *such that* $H(\mathbf{p})^f \geq 1$, *we have*

$$\begin{aligned}
&\mathbb{E}[G(\mathbf{p})]^f \\
&\geq \sup_{\alpha\in[0,1/2]} \frac{1}{e} 2^{\left(\sum_{k=1}^{N_s} H(\mathbf{p}^k)\right)+\frac{h(\alpha)}{1-\alpha}\prod_{k=1}^{N_s} p_n^k} + \frac{1}{2} - \frac{\alpha}{1-\alpha}\prod_{k=1}^{N_s} p_n^k \\
&\geq \frac{1}{e} 2^{\left(\sum_{k=1}^{N_s} H(\mathbf{p}^k)\right)+\frac{1}{2}\prod_{k=1}^{N_s} p_n^k} + \frac{1}{2} - \prod_{k=1}^{N_s} p_n^k \geq \frac{1}{e} 2^{\sum_{k=1}^{N_s} H(\mathbf{p}^k)} + \frac{1}{2}.
\end{aligned}$$

Proofs. Both results follow immediately from Theorem 1 and Theorem 2 considering the additivity of the Shannon entropy,

$$H(\mathbf{p})^f = H\left(\bigotimes_{k=1}^{N_s} \mathbf{p}^k\right) = \sum_{k=1}^{N_s} H\left(\mathbf{p}^k\right),$$

and the fact that the minimal entry in the full list of probabilities $\mathbf{p}$ is the product of the individual minima p_n^k. □

In conclusion, we presented the optimal Massey-like inequality for the full-key context in the form of Theorem 3 as well as an improvement of it, Theorem 4, showing that our results are indeed highly scalable.

For practical purposes, given that limited representation of numbers may lead to zero values in large lists of probabilities, in our experiments we consider as p_n the least non-zero probability, i.e., $p_n > 0$, which leads to accurate results for the bounds presented in this paper.

As a final remark, we note here that our new improvement which most completely manifests in the form of Theorem 2 is tight whenever the smallest non-zero probability is significant, such as uniform distributions or geometric distributions with truncated tail, but can also be beneficial for other classes of distributions, such as those encountered in side-channel attack evaluation, as will be discussed in the next section.

6. Evaluation on Side-Channel Attack Data

As mentioned in the introduction, in many security-critical applications, such as banking or physical access control, it is imperative to use hardware that is security certified. In order to obtain a security certification such as those offered by Common Criteria [24] it is also typically necessary to prove that a device is resilient to side-channel attacks and this is generally done by showing that the guessing entropy or some other security metric is within certain thresholds. In this context, the bounds from this paper represent a very useful tool for a security evaluator, as they allow improving the tightness of the CHES2017 bounds, which are considered to be the most scalable tool to date for evaluating the security of cryptographic algorithms, allowing security estimation when dealing with very large keys. Hence, in this section we demonstrate the relevance of the results from this paper, by comparing the scalable versions of our bounds (see previous sections) against the bounds from CHES 2017.

6.1. Evaluation Data

To easily compare the bounds from this paper against the CHES 2017 method of Choudary and Popescu [12], we used the same datasets as in the CHES 2017 paper: (*a*) a simulated dataset (MATLAB generated power consumption from the execution of the AES S-box) and (*b*) a real dataset (power traces from the execution of AES in the AES hardware engine of an AVR XMEGA microcontroller).

In both experiments, the AES encryption algorithm is used with 128-bit (16-byte) keys. The AES state is composed of 16 bytes, which are processed sequentially within certain operations such as the Sub Bytes (S-Box) operation, which is the typical target of side-channel attacks, including ours.

The steps for our experiments are as follows:

1. For each dataset (power traces), we run a Template Attack [23] using the set of power traces to determine the most likely value of each of the 16 bytes of the AES key. The result of this attack is a list of probabilities $\mathbf{p}^k = \{p_1, p_2, \ldots, p_{256}\}$ for each of the 16 bytes of the AES key ($K = [k_1 k_2 \ldots k_{16}]$).
2. Using the lists of probabilities $\mathbf{p}^1, \mathbf{p}^2, \ldots, \mathbf{p}^{16}$, we compute the bounds (those from this paper as well as those from CHES 2017) first for each byte individually and then for attacks on two or more key bytes. Please note that a direct computation of the guessing entropy through the computation of the cross-product of several lists of probabilities (e.g., for more than 8 key bytes) is not feasible as we would have to process lists of more than 2^{64} elements. Instead, the bounds from this paper (as well as those from CHES 2017) use directly and very efficiently the lists of probabilities for each key byte, without performing the cross-product, to derive security metrics for attacks on many target bytes.

In the following, we present the results of our evaluations for three interesting cases: (1) application of the bounds on single lists of probabilities—this is equivalent to attacking a single key byte in side-channel attack evaluations; (2) application of the bounds when targeting two bytes—this is interesting to test the scalability of the bounds; (3) application of the bounds when attacking 16 bytes—this represents a complete attack on the full AES key and hence is a representative scenario of a full-fledged security evaluation, where scalability and tightness are very important.

6.2. Evaluation on a Single Byte

We show the bounds for a single key byte on the simulated and real datasets in Figure 1. Here we can see that while the CHES lower bound is tighter when the guessing entropy is low (below 4 bits), in the other (most) cases Rioul's lower bound is better. Furthermore, we can see that Theorem 1 provides a better (tighter) lower bound than Rioul's lower bound and Theorem 2 in turn provides an even better lower bound than Theorem 1.

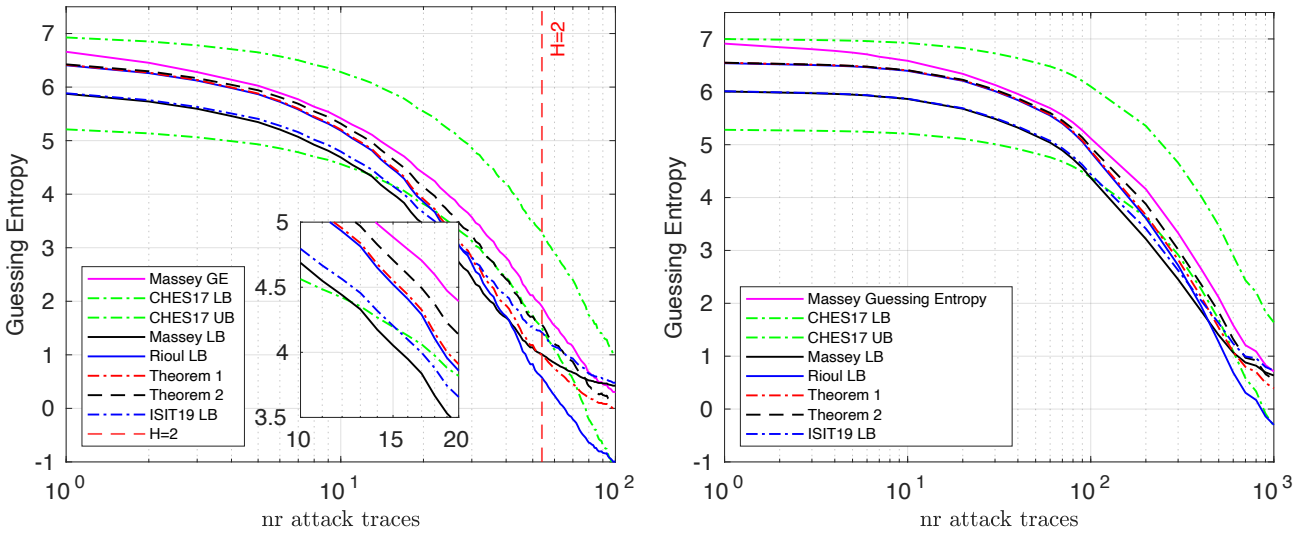

Figure 1. Bounds for the simulated (left) and real (right) datasets, when targeting a single subkey byte. These are averaged results over 100 experiments.

An interesting artifact appears when the guessing entropy decreases below two bits ($\log(G(\mathbf{p})) = 1$), where the Massey inequality (and the ones in ISIT 2019 [15]) does not necessarily hold (considering for example geometric distributions with $p_1 \geq 1/2$). In this case, most bounds do not hold anymore. Meanwhile, bounds based on Rioul's inequality (Rioul LB, Theorem 1, Theorem 2) all continue to hold in this regime, owing to the fact that it does not impose preconditions on the minimal value of $H(\mathbf{p})$.

For reference, in Figure 1 we included all present refinements of Massey's inequality. However, for clarity, in the following sections we will only compare our bounds with CHES 2017.

6.3. Evaluation on Two Bytes

We show the bounds when targeting two key bytes on the simulated and real datasets in Figure 2. Here we see again that Rioul's bound is tight when the guessing entropy is higher, but then the CHES lower bound becomes tighter, as the guessing entropy decreases. We can also confirm here that our theorems provide a better (tighter) lower bound than Rioul's lower bound.

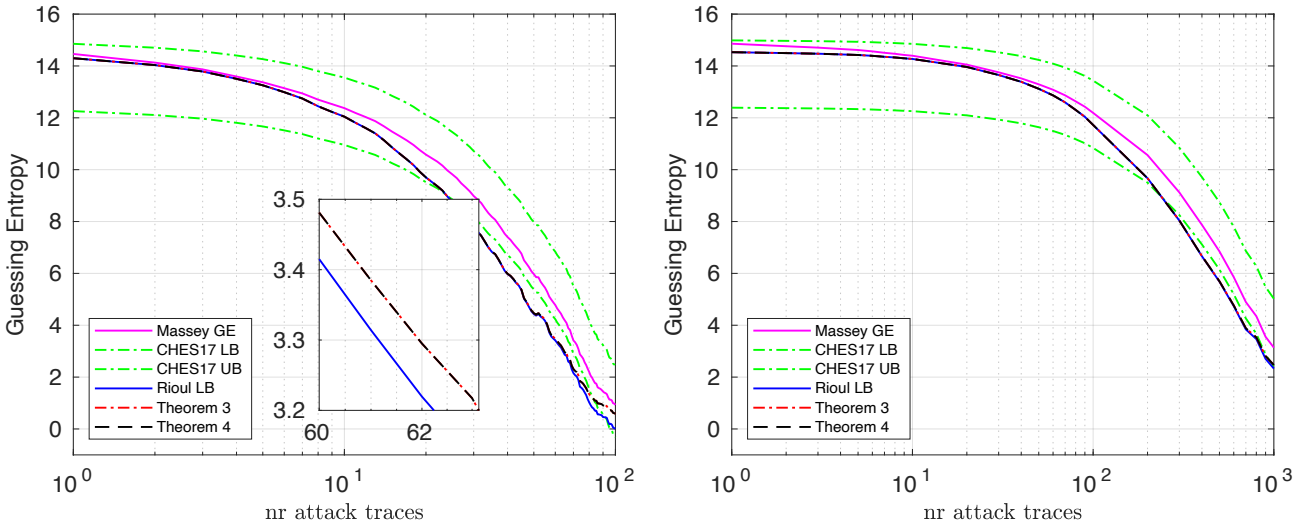

Figure 2. Bounds for the simulated (left) and real (right) datasets, when targeting two subkey bytes. These are averaged results over 100 experiments.

6.4. Evaluation on All 16 Bytes

Finally, we show the bounds when targeting all the 16 bytes of the full AES key on the simulated and real datasets in Figure 3. We did not plot the actual value of the guessing entropy in this case, because it is not possible to compute it: it would require the iteration over (and sorting of) a list of 2^{128} elements. Hence, in this case the computationally efficient bounds compared in this paper become very valuable. From the figure we see again that when the guessing entropy is very high (e.g., above 120 bits), all the lower bounds presented in this paper are tighter than the CHES 2017 lower bound (Theorems 3 and 4 provide numerically similar results to Rioul's lower bound), hence tightening the security evaluation results for larger values of the guessing entropy. This allows for an overall improved method than that of CHES 2017 (e.g., by taking the maximum between the CHES 2017 lower bound and Theorems 3/4).

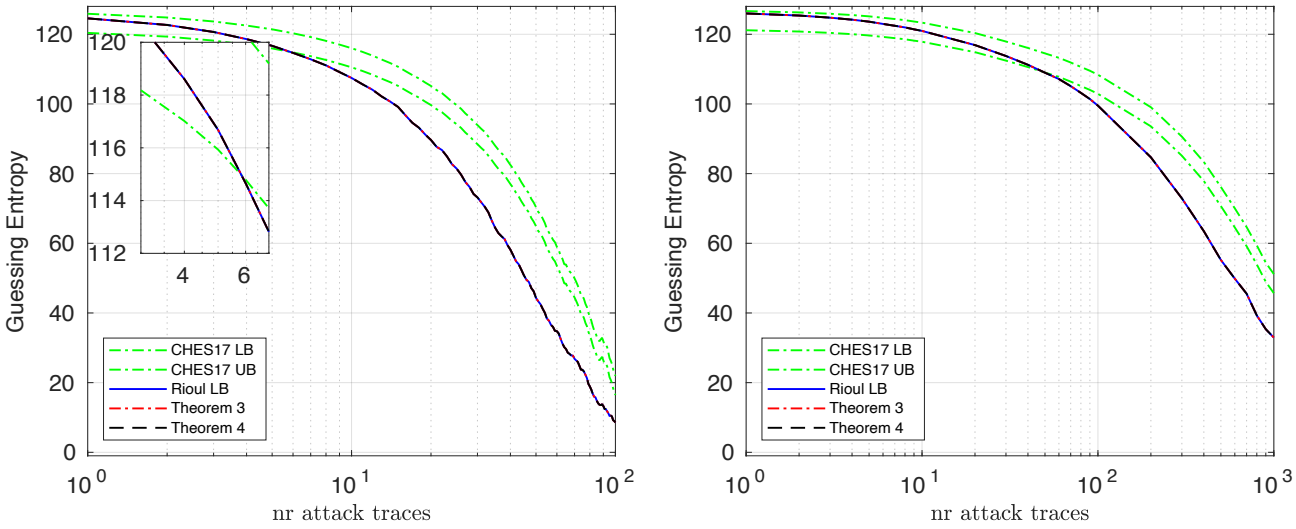

Figure 3. Bounds for the simulated (left) and real (right) datasets, when targeting all the 16 AES key bytes. These are averaged results over 100 experiments.

7. Conclusions

In this paper, we have improved the security evaluation metric of Choudary and Popescu from CHES 2017, which is considered the most scalable method to date, by tightening Massey's inequality even further. First, we have demonstrated that the improved Rioul's inequality is asymptotically optimal and showed how to scale this method for use in full-key evaluation methods. Then, using the techniques of [15,18], we further refined this inequality for finite support distributions allowing us to increase the multiplicative constant depending on the smallest probability p_n. We compared all our results to those of Choudary and Popescu from CHES 2017 using their datasets, demonstrating the usefulness of the improvements from this paper.

For future work we are very interested in further results based on other (additive) entropies, such as Rényi entropies where other guessing bounds are already investigated [19] past their original use in moment inequalities [25–27] and other derived problems such as guessing with limited (or no) memory [28].

Author Contributions: The authors contributed equally to this work. All authors have read and agreed to the published version of the manuscript.

Funding: This research was partially supported by the Romanian Ministry of Education and Research, CNCS – UEFISCDI, project number PN-III-P1-1.1-TE-2019-2245, within PNCDI III.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Network, E.U.A.F.; Security, I. *Cyber Security and Resilience of Smart Cars*; 2016.
2. Network, E.U.A.F.; Security, I. *Sectoral/Thematic Threat Analysis: ENISA Thread Lanscape*; 2020.
3. Garcia, F.D.; Oswald, D.; Kasper, T.; Pavlides, P. *Lock It and Still Lose It—on the (In)Security of Automotive Remote Keyless Entry Systems*; USENIX Security Symposium: Austin, TX, 2016.
4. Camurati, G.; Poeplau, S.; Muench, M.; Hayes, T.; Francillon, A. Screaming Channels : When Electromagnetic Side Channels Meet Radio Transceivers. *ACM SIGSAC Conf. Comput. Commun. Secur. (CCS)* **2018**, *2018*, 163–177.

5. Standaert, F.X.; Malkin, T.G.; Yung, M. A unified framework for the analysis of side-channel key recovery attacks. *Eurocrypt* **2009**, Cologne, Germany, April 26-30, 2009, 443–461, ISBN: 978-3-642-01001-9.
6. Veyrat-Charvillon, N.; Gerard, B.; Renauld, M.; Standaert, F.X. An Optimal Key Enumeration Algorithm and Its Application to Side-Channel Attacks; *Selected Areas of Cryptography:* 2012; Windsor, ON, Canada, August 15-16, 2012, pp. 390–406, ISBN: 978-3-642-35999-6.
7. Veyrat-Charvillon, N.; Gerard, B.; Standaert, F.X. Security evaluations beyond computing power. *Eurocrypt* **2013**, Athens, Greece, May 26-30, 2013, 126–141, ISBN: 978-3-642-38348-9.
8. Bernstein, D.J.; Lange, T.; van Vredendaal, C. Tighter, Faster, Simpler Side-Channel Security Evaluations beyond Computing Power. 2015. *ePrint Archive. Available online: https://eprint.iacr.org/2015/221 (accessed on June 2021).*
9. Glowacz, C.; Grosso, V.; Poussier, R.; Schüth, J.; Standaert, F.X. Simpler and more efficient rank estimation for side-channel security assessment. *Fast Softw. Encryption* **2015**, Istanbul, Turkey, March 8-11, 2015, 117–129, ISBN: 978-3-662-48116-5.
10. Poussier, R.; Standaert, F.X.; Grosso, V. Simple key enumeration (and rank estimation) using histograms: An integrated approach. *In Proceedings of the International Conference on Cryptographic Hardware and Embedded Systems (CHES)*, Santa Barbara, CA, USA, August 17-19, 2016; pp. 61–81, ISBN: 978-3-662-53140-2.
11. David, L.; Wool, A. A Bounded-Space Near-Optimal Key Enumeration Algorithm for Multi-subkey Side-Channel Attacks. *Top. Cryptolgy*, San Francisco, CA, USA, February 14–17, **2017**, 311–327, ISBN: 978-3-319-52153-4.
12. Choudary, M.O.; Popescu, P.G. Back to Massey: Impressively fast, scalable and tight security evaluation tools. *In Proceedings of the 2017 International Conference on Cryptographic Hardware and Embedded Systems (CHES)*, Taipei, Taiwan, September 25-28, 2017; pp. 367–386, ISBN: 978-3-319-66787-4.
13. Massey, J.L. Guessing and entropy. *In Proceedings of the 1994 IEEE International Symposium on Information Theory (ISIT)*, Trondheim, Norway, 27 June-1 July 1994; p. 204, ISBN:0-7803-2015-8.
14. Grosso, V. Scalable key rank estimation (and key enumeration) algorithm for large keys. *CARDIS*, Montpellier, France, November 12–14, **2018**, 80–94, ISBN: 978-3-030-15462-2.
15. Popescu, P.G.; Choudary, M.O. Refinement of Massey Inequality. *In Proceedings of the 2019 IEEE International Symposium on Information Theory (ISIT)*, Paris, France, 7-12 July 2019; pp. 495–496.
16. Rioul, O. On Guessing. *unpublished note*, 2013.
17. de Chérisey, E.; Guilley, S.; Rioul, O.; Piantanida, P. Best Information is Most Successful. *In Proceedings of the 2019 International Conference on Cryptographic Hardware and Embedded Systems (CHES)*, Atlanta, USA, August 25–28, 2019; pp. 49–79.
18. Tănăsescu, A.; Popescu, P.G. Exploiting the Massey Gap. *Entropy* **2020**, *22*, 1398.
19. Rioul, O. Variations on a Theme by Massey. *arXiv* **2021**, arXiv:2102.04200.
20. Mazumdar, B.; Mukhopadhyay, D.; Sengupta, I. Constrained Search for a Class of Good Bijective *S*-Boxes With Improved DPA Resistivity. *IEEE Trans. Inf. Forensics Secur.* **2013**, *8*, 2154–2163.
21. Choudary, M.O.; Kuhn, M.G. Efficient, portable template attacks. *IEEE Trans. Inf. Forensics Secur.* **2017**, *13*, 490–501.
22. Carré, S.; Guilley, S.; Rioul, O. Persistent fault analysis with few encryptions. *In Proceedings of the 2020 International Workshop on Constructive Side-Channel Analysis and Secure Design (COSADE)*, Lugano, Switzerland, April 1–3, 2020, ISBN: 978-3-030-68773-1.
23. Chari, S.; Rao, J.R.; Rohatgi, P. Template Attacks. *In Proceedings of the 2003 Cryptographic Hardware and Embedded Systems (CHES)*; Redwood Shores, CA, USA, August 13–15, 2002; pp. 13–28, ISBN: 978-3-540-36400-9.
24. The Common Criteria Web Site. 2021. Available online: https://www.commoncriteriaportal.org/ (accessed on June 2021).
25. Arikan, E. An inequality on guessing and its application to sequential decoding. *IEEE Trans. Inf. Theory* **1996**, *42*, 99–105.
26. Sason, I.; Verdú, S. Improved bounds on lossless source coding and guessing moments via Rényi measures. *IEEE Trans. Inf. Theory* **2018**, *64*, 4323–4346.
27. Kuzuoka, S. On the conditional smooth Rényi entropy and its applications in guessing and source coding. *IEEE Trans. Inf. Theory* **2019**, *66*, 1674–1690.
28. Huleihel, W.; Salamatian, S.; Médard, M. Guessing with limited memory. *In Proceedings of the 2017 IEEE International Symposium on Information Theory (ISIT)*, Aachen, Germany, 25-30 June 2017; pp. 2253–2257.