Asymptotically Optimal Massey-Like Inequality on Guessing Entropy With Application to Side-Channel Attack Evaluations

Andrei Tanasescu, Marios O. Choudary, Olivier Rioul, and Pantelimon George Popescu

Abstract—A Massey-like inequality is any useful lower bound on guessing entropy in terms of the computationally scalable Shannon entropy. The asymptotically optimal Massey-like inequality is determined and further refined for finite-support distributions. The impact of these results are highlighted for side-channel attack evaluation where guessing entropy is a key metric. In this context, the obtained bounds are compared to the state of the art.

Index Terms—Massey inequality, guessing entropy, Shannon entropy, side-channel attacks.

I. INTRODUCTION

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

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

subject to the same condition \( H(p) \geq 2 \) bits as in the Massey inequality. Meanwhile, Rioul’s inequality [6] published in a CHES paper [4] states that for all values of \( H(p) \geq 0 \),

\[
\mathbb{E}[G(p)] \geq \frac{1}{e} 2^{H(p)},
\]

which refines Massey’s inequality when \( H(p) \geq \log \frac{2}{1-1/e}. \)

Finally, in an Entropy paper, Tanasescu and Popescu [5] found that under the same condition as in Massey’s inequality,

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

\( \geq 2^{H(p)+2p_n-2} + 1 - p_n > 2^{H(p)-2} + 1. \)

The authors of [5] hinted that a similar refinement can be found for inequality [1].

In this paper, we optimize exponential relations between the guessing and Shannon entropies, i.e., lower bounds of the form \( \mathbb{E}[G(p)] \geq a \cdot b^{H(p)} + c \) valid when the Shannon entropy lies above a given threshold. We arrive at an improved Rioul’s inequality [6] by an additive constant of 1/2, which is asymptotically optimal among other global lower bounds depending only on the Shannon entropy as \( H(p) \to \infty \). Then, using the techniques of [2, 5] we further refine this inequality for finite support distributions allowing us to increase the multiplicative constant depending on the smallest probability \( p_n \). Finally, we apply our results to side-channel attack evaluation, where guessing entropy is a key metric [7]. [8, 9], comparing our results to the best on the market and showing that under certain conditions the Shannon entropy is indeed a precious quantifier of guessing entropy.

II. THE ASYMPTOTICALLY OPTIMAL MASSEY-LIKE INEQUALITY

In this section we consider bounds of the form \( \mathbb{E}[G(p)] \geq a \cdot b^{H(p)} + c \) with \( a > 0 \) and seek to determine the optimal coefficients \( a, b, c \) prioritizing the asymptotic shape as \( H(p) \to \infty \) holding whenever \( H(p) \) is larger then a given threshold.

Theorem 1. The optimal Massey-like inequality \( \mathbb{E}[G(p)] \geq a \cdot b^{H(p)} + c \) as \( H(p) \to \infty \) is Rioul’s improved inequality [6]

\[
\mathbb{E}[G(p)] \geq \frac{1}{e} 2^{H(p)} + \frac{1}{2},
\]

which holds for all values of \( H(p) \geq 0 \).

Proof. Following Massey’s approach [1], finding the best lower bound on guessing entropy is equivalent with the statement that among all probability distributions with guessing entropy \( \mu > 1 \), the maximal Shannon entropy is attained by the geometric distribution with mean \( \mu \), that is,

\[
H(p) \leq \log(\mu - 1) - \mu \log(1 - 1/\mu)
\]

where \( \log() \) denotes logarithm to base 2. The inequality is actually strict when \( p \) has finite length, but the upper bound can be approached as closely as desired.
We seek bounds of the form $E \left[ G (p) \right] \geq a \cdot e^{H(p)} + c$, i.e. $H (p) \leq \frac{\log b - c}{a}$. In order for this to be valid for all $\mu$, 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

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

while the expression on the right has asymptotic

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

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

The asymptotically optimal bound then writes

$$\log (\mu - 1/2) + \log e \geq \log (\mu - 1) - \mu \log (1 - 1/\mu) \quad (3)$$

which readily gives (2) when $\mu$ or $H(p)$ tend to infinity. A simple proof of (2) for all values of $H(p) > 0$ can be found in [6], but one can also prove directly that (2) holds for all values of $\mu > 1$ as follows. The first and second-order derivatives of the difference $f(\mu) = \log (\mu - 1/2) + 1 - \log (\mu - 1) + \mu\log (1 - 1/\mu)$ between the two sides of (1) (expressed in natural units) are

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

$$f''(\mu) = - \frac{1}{(\mu - 1/2)^2} + \frac{1}{\mu(\mu - 1)} = \frac{1}{4\mu(\mu - 1)(\mu - 1/2)^2}.$$ 

It follows that $f'' > 0$, so that $f'$ is increasing while also vanishing as $\mu \to +\infty$, hence $f' < 0$ for all $\mu > 1$. As a consequence, $f$ is decreasing for all $\mu > 1$. Therefore, since (3) holds when $\mu \to +\infty$, it also holds for all $\mu > 1.

We conclude this section by remarking that the obtained optimal inequality (2) only improves (1) by an additive constant 1/2. Riu’s strengthened inequality (6) now writes

$$E \left[ G (p) \right] \geq \frac{1}{e} e^{H(p)} + \frac{1}{2}. \quad (2)$$

It is further generalized to scalable Rényi entropies in [6].

III. REFINEMENT FOR FINITE SUPPORT DISTRIBUTIONS

In this section we find a new relation between the Shannon and guessing entropy, dependent on the minimal probability of a given distribution, further refining Riu’s improved inequality (2).

We begin with a direct improvement following the technique of [2], [5].

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

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

$$\geq \frac{1}{e} 2^{H(p) + p_n h(\alpha)} - \frac{1}{e} 2^{H(p)} + \frac{1}{2}.$$ 

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

$$E \left[ G (p) \right] = E \left[ G (q) \right] - \alpha p_n \geq \frac{1}{e} e^{2H(\alpha)} - \alpha p_n + \frac{1}{2} \quad (4)$$

$$\geq \frac{1}{e} 2^{H(p) + p_n h(\alpha)} - \alpha p_n + \frac{1}{2}.$$ 

The first inequality follows taking the supremum over $\alpha$ in eq. (4), the second by substituting $\alpha = 1/2$. To justify the third, we use $2^x > 1 + x \log 2$ for $x = p_n$ obtaining

$$\frac{1}{e} 2^{H(p) + p_n h(\alpha)} - \alpha p_n \geq \frac{1}{e} 2^{H(p)} (1 + p_n \log 2) - \frac{1}{2} p_n$$

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

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

We can further refine this lemma using the techniques of [2], [5] as follows.

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

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

Proof. Given the initial decreasing $p$, we construct a sequence of probability distributions $\{Q_k\}$ 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 $Q_{k,t}$ the $t$th component of the sequence $Q_k$, we define the terms of the list $\{Q_k\}$ as follows.

We let the support of the first term coincide with $p$, i.e. $Q_0 = (p_0, p_1, \ldots, p_n, 0, 0, \ldots, 0, \ldots)$, and we define the other terms by recurrence:

$$Q_{k+1} = (Q_{k,0}, Q_{k,1}, \ldots, Q_{k,n+k-1}, (1 - \alpha) Q_{k,n+k}, \alpha Q_{k,n+k}, 0, 0, \ldots, 0, \ldots).$$

and at each step of the construction we have the inequality

$$E \left[ G (Q_k) \right] = E \left[ G (Q_{k+1}) \right] - \alpha Q_{k,n+k} \geq \frac{2^{H(Q_{k+1})}}{e} - \alpha Q_{k,n+k} \geq \frac{2^{H(Q_k)}}{e} + \frac{1}{2}.$$ 

After the first $k$ steps of the construction we find

$$E \left[ G (p) \right] = E \left[ G (Q_k) \right] - p_n \alpha \frac{1 - \alpha^k}{1 - \alpha}$$

$$= E \left[ G (Q_k) \right] + \sum_{j=0}^{k-1} (E \left[ G (Q_j) \right] - E \left[ G (Q_{j+1}) \right]).$$
where the tightest of the enumerated bounds is
\[
\mathbb{E}[G(p)] \geq \frac{1}{e^2} H(p) + \frac{1}{2} - \frac{2}{e^2} \frac{1 - \alpha}{1 - \alpha},
\]
which as we have shown increases with \( k \) up to the limit
\[
\mathbb{E}[G(p)] \geq \frac{1}{e^2} H(p) + \frac{1}{2} - \frac{2}{e^2} \frac{1 - \alpha}{1 - \alpha}
\]
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 \( \frac{1}{e^2} H(p) + \frac{1}{2} \). \( \Box \)

IV. APPLICATION TO SIDE-CHANNEL ANALYSIS

The improvements shown in previous sections can be very useful in the evaluation of side-channel attacks. In this context, Choudary and Popescu \[10\] presented a new approach, based on mathematical bounds of the guessing entropy \[11\], to bound the guessing entropy remaining after a side-channel attack for very large cryptographic keys (or other secret data). They showed that their method works for keys of up to 1024 bytes and beyond, working in constant time and memory, which none of the other methods could do. This provided a great improvement for security evaluations of cryptographic devices.

We remark here that all bounds from this paper are highly computationally scalable, because they are based on the Shannon entropy, which is additive i.e. \( H(\otimes P_i) = \sum_i H(P_i) \) for any probability distributions \( P_1, P_2, \ldots, P_n \) \[11\].

A. Evaluation of bounds

In this context of security evaluations, it is interesting to evaluate the accuracy of different bounds for the guessing entropy in different settings. In this section, we analyse the bounds derived in the preceding sections, along with those presented at CHES 2017 \[10\], using lists of probabilities obtained from the application of Template Attacks \[12\] on side-channel traces.

For easier comparison and future reference, we used the same data as in the CHES 2017 paper: A simulated dataset (MATLAB generated power consumption from the execution of the AES S-box) and a real dataset (power traces from the execution of AES in the AES hardware engine of an AVR XMEGA microcontroller).

For our analysis we have focused on 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 on the combination of two bytes – this is interesting to observe the scalability of the bounds; 3) application of the bounds on the combination of all 16 AES bytes – this represents a complete attack on the full AES key and hence is a representative scenario of a full-fledged security evaluation.

B. 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.

An interesting artifact appears when the guessing entropy decreases below two bits (\( \log(G(p)) = 1 \)), where the Massey inequality (and the ones in ISIT 2019 \[2\]) does not necessarily hold (considering for example geometric distributions with \( p_1 \geq 1/2 \)). In this case, most bounds seem to be tighter than the CHES 2017 \[10\] lower bound. Meanwhile, bounds based on Rioul’s inequality all continue to hold in this regime, owing to the fact that it does not impose preconditions on the minimal value of \( H(p) \).

C. 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 Theorem
1 provides a better (tighter) lower bound than Rioul’s lower bound.

However, in this case Theorem 2 provides numerically similar results to Theorem 1, just as the ISIT 2019 lower bound provides numerically similar results to Massey’s lower bound. These results are due to the fact that these bounds only differ pairwise in a term containing the minimum probability in the combined list and this minimum becomes zero (or almost zero) when combining two (or more) lists of probabilities in our experiments, which is just a particularity of such experiments.

D. 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. However, as soon as the guessing entropy decreases below 120 bits, the CHES 2017 lower bound becomes closer to the upper bound than the other lower bounds.

We can also confirm here that Rioul’s lower bound is a better (tighter) lower bound than Massey’s lower bound. However, in this case we observe that Theorem 1 and 2 provide numerically similar results to Rioul’s lower bound. Nevertheless, we are impressed by the scalability of such bounds, thanks to the easy computation of the Shannon entropy of product distributions.

V. Conclusion

In this paper, the asymptotically optimal Massey-like inequality is determined as an improved Rioul’s inequality by an additive constant of 1/2. Then, using the techniques of [2, 5], this inequality is further refined for finite support distributions allowing us to increase the multiplicative constant depending on the smallest probability p_0. Finally, the results are applied to the task of side-channel attack evaluation and compared to the best on the market. It is shown that under certain conditions, the Shannon entropy is in fact a precious quantifier of guessing entropy because it is computationally scalable thanks to its additivity property.

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 [6] past their original use in moment inequalities [13, 14, 15] and other derived problems such as guessing with limited (or no) memory [16].

ACKNOWLEDGMENT

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

REFERENCES

[1] J. L. Massey, “Guessing and entropy,” in Proceedings of 1994 IEEE International Symposium on Information Theory. IEEE, 1994, p. 204.
[2] P. G. Popescu and M. O. Choudary, “Refinement of Massey inequality,” in 2019 IEEE International Symposium on Information Theory (ISIT). IEEE, 2019, pp. 495–496.
[3] O. Rioul, “On guessing,” unpublished note, 11 2013.
[4] E. de Chéritsey, S. Guille, O. Rioul, and P. Plantanida, “Best information is most successful,” in 2019 International Conference on Cryptographic Hardware and Embedded Systems (CHES), 2019, pp. 49–79.
[5] A. Tănăsescu and P. G. Popescu, “Exploiting the Massey gap,” Entropy, vol. 22, no. 12, p. 1398, 2020.
[6] O. Rioul, “Variations on a theme by Massey,” arXiv preprint arXiv:2102.04200 (submitted to IEEE Transactions on Information Theory), 2021.
[7] B. Mazumdar, D. Mukhopadhyay, and I. Sengupta, “Constrained search for a class of good bijective S-boxes with improved dpa resistivity,” IEEE Transactions on Information Forensics and Security, vol. 8, no. 12, pp. 2154–2163, 2013.
[8] M. O. Choudary and M. G. Kuhn, “Efficient, portable template attacks,” IEEE Transactions on Information Forensics and Security, vol. 13, no. 2, pp. 490–501, 2017.
[9] S. Carré, S. Guilley, and O. Rioul, “Persistent fault analysis with few encryptions,” in 2020 International Workshop on Constructive Side-Channel Analysis and Secure Design (COSADE), 2020.
[10] M. O. Choudary and P. G. Popescu, “Back to Massey: Impressively fast, scalable and tight security evaluation tools,” in 2017 International Conference on Cryptographic Hardware and Embedded Systems (CHES), 2017, pp. 367–386.
[11] T. M. Cover and J. A. Thomas, Elements of information theory. John Wiley & Sons, 1999.
[12] S. Chari, J. R. Rao, and P. Rohatgi, “Template attacks,” in 2003 Cryptographic Hardware and Embedded Systems (CHES). Springer Berlin Heidelberg, 2003, pp. 13–28.
[13] E. Arikan, “An inequality on guessing and its application to sequential decoding,” IEEE Transactions on Information Theory, vol. 42, no. 1, pp. 99–105, 1996.
[14] I. Sason and S. Verdú, “Improved bounds on lossless source coding and guessing moments via Rényi measures,” IEEE Transactions on Information Theory, vol. 64, no. 6, pp. 4323–4346, 2018.
[15] S. Katuoka, “On the conditional smooth rényi entropy and its applications in guessing and source coding,” IEEE Transactions on Information Theory, vol. 66, no. 3, pp. 1674–1690, 2019.
[16] W. Huleihel, S. Salamatian, and M. Médard, “Guessing with limited memory,” in 2017 IEEE International Symposium on Information Theory (ISIT). IEEE, 2017, pp. 2253–2257.
Andrei Tănăsescu is a young researcher at University POLITEHNICA of Bucharest. His main research interests are Quantum Computing and Quantum Information Theory.

Marios O. Choudary is Senior Lecturer in Computer Science at the University Politehnica of Bucharest. His research interests include authentication and security protocols, applied cryptography and side-channel attack evaluations. He graduated from the University Politehnica of Bucharest in 2008 and then did his MPhil and PhD in Computer Science at the University of Cambridge Computer Laboratory.

Olivier Rioul is full Professor at the Department of Communication and Electronics, in the Laboratoire de Traitement et Communication de l’Information (LTCI), Télécom Paris, Institut Polytechnique de Paris, France. He graduated from École Polytechnique, Paris, France in 1987 and from École Nationale Supérieure des Télécommunications, Paris, France in 1989. He obtained his PhD degree from École Nationale Supérieure des Télécommunications, Paris, France in 1993. His research interests are in applied mathematics and include various, sometimes unconventional, applications of information theory such as inequalities in statistics, hardware security, and experimental psychology. He has been teaching information theory at various French universities for more than twenty years and has published a textbook which has become a classical French reference in the field.

Pantelimon George Popescu is Professor at the Computer Science and Engineering Department of University POLITEHNICA of Bucharest. His main fields of interest include Quantum Computing, Numerical Methods, Information Theory and Inequalities.