Maximal leakage of masked implementations using Mrs. Gerber's lemma for min-entropy

Julien Béguinot, Yi Liu, Olivier Rioul, Wei Cheng, Sylvain Guilley

Maximal Leakage of Masked Implementations Using Mrs. Gerber's Lemma for Min-Entropy

Julien Béguinot[1], Yi Liu[1], Olivier Rioul[1], Wei Cheng[1,2], and Sylvain Guilley[1,2]
[1]LTCI, Télécom Paris, Institut Polytechnique de Paris, France
[2]Secure-IC S.A.S., France
firstname.lastname@telecom-paris.fr

***Abstract*—A common countermeasure against side-channel attacks on secret key cryptographic implementations is dth-order masking, which splits each sensitive variable into $d+1$ random shares. In this paper, maximal leakage bounds on the probability of success of any side-channel attack are derived for any masking order. Maximal leakage (Sibson's information of order infinity) is evaluated between the sensitive variable and the noisy leakage, and is related to the conditional "min-entropy" (Arimoto's entropy of order infinity) of the sensitive variable given the leakage. The latter conditional entropy is then lower-bounded in terms of the conditional entropies for each share using majorization inequalities. This yields a generalization of Mrs. Gerber's lemma for min-entropy in finite Abelian groups.**

I. Introduction

When a cryptographic device is operating, any kind of physical leakage (time, power, electromagnetic emanations, etc.) can be exploited by an attacker. The attacker queries the device multiple times, and measures the corresponding leakages to infer the secret key. The security of devices against side-channel attacks has become a major concern.

To evaluate the probability of success for any side-channel attack, information-theoretic metrics turn out to be effective and have been used in many studies. Using conditional mutual information and Fano's inequality, de Chérisey et al. [6] established several universal bounds on the probability of success for a given number of queries, or equivalently, the minimum number of queries required to achieve a given level of success. This approach has been extended to conditional Sibson's α-information by Liu et al. [15]. However, both [6] and [15] were restricted to unprotected cryptographic devices.

Masking is one of the most well-established countermeasures. The main issue in this context is the fact that a direct evaluation of the information leakage requires data and computational complexities that increase rapidly with the masking order [5]. Therefore, it is important to derive bounds in terms of the individual information leakages for each share. Duc et al. [7] conjectured a general form of such bounds. Rigorous bounds were obtained in two independent recent works by Ito et al. [13] and Masure et al. [18]. Even more recently, Béguinot et al. [3] improved these results using Mrs. Gerber's lemma [14], [27] to derive sharp bounds in terms of mutual information for masking in additive groups of order 2^n.

In the case of unprotected implementations (without masking), it is shown by simulation in [15] that the probability of success of a side-channel attack is evaluated using Sibson's α-information all the more accurately as α increases. Therefore, the case of mutual information, which corresponds to $\alpha = 1$ is not optimal. This motivates the derivation of new bounds in the limiting case $\alpha = +\infty$.

The usual setup of masking countermeasures involves bitwise XOR (exclusive or) operations, which are particularly well suited to symmetric cryptographic algorithms like AES. However, modern cryptography also relies on operations performed in groups of prime order, and masking can also be multiplicative [1] and not only additive [9]. For all these reasons, there is a strong incentive to extend the previous bounds to arbitrary finite Abelian groups. This motivates the generalization of Mrs. Gerber's lemma to any such Abelian group.

Mrs. Gerber's lemma was initially derived by Wyner and Ziv [27] to lower bound the entropy of a modulo 2 addition of binary random variables in terms of the entropies of each summand. It was extended by Jog and Anatharam [14] to the case of additive groups of order 2^n, and by Hirche [10] to the case of Rényi entropy of binary variables. The general case of additive groups was only considered by Tao [23] for Shannon entropy and independent copies of two shares, in relation to sumset theory. While the original binary Mrs. Gerber's lemma was used to derive a binary version of the entropy power inequality [21], a generalization of the entropy power inequality to any prime cyclic additive group and Rényi entropy was investigated by Madiman et al. [16], but does not reduce to an explicit "Mrs. Gerber's lemma"-type inequality. Therefore, it appears that the case of min-entropy (Rényi entropy of order ∞) and additive groups of any order has not been investigated yet in our context.

Contributions

In this paper, we show that when evaluating the performance of side-channel attacks of masked implementations using conditional Sibson's α-information, the exact performance of optimal maximum likelihood attacks is attained in the limiting case $\alpha = +\infty$. This motivates the investigation of Mrs. Gerber's lemma for conditional min-entropy (Arimoto's conditional entropy of order ∞). We derive a variation of such Mrs. Gerber's lemma for any finite Abelian group and for any masking order.

The remainder of this paper is organized as follows. Section II gives some notations and preliminaries on α-informational quantities. Section III shows that the optimal evaluation of side-channel attack success by Fano's inequality is achieved in the limiting case $\alpha = +\infty$ and derives the corresponding bound in terms of the information between the sensitive variable and the leakage, which is linear in the number of queries. Section IV derives Mrs. Gerber's lemma for min-entropy, first for two summands in any finite Abelian group, then extends it to the general case of $d+1$ summands. Section V concludes and gives some perspectives.

II. Preliminaries and Notations

A. Framework and Notations

Let K be the secret key and T be a public variable (usually plaintext or ciphertext) known to the attacker. It is assumed that T is independent of K, and K is uniformly distributed over an Abelian group $\mathcal{G}$ of order M. The cryptographic algorithm operates on K and T to compute a sensitive variable X, which takes values in the same group $\mathcal{G}$ and is determined by K and T, in such a way that X is also uniformly distributed over $\mathcal{G}$.

In a masking scheme of order d, the sensitive variable X is randomly split into $d+1$ *shares* X_0, X_1, ..., X_d and cryptographic operations are performed on each share separately. Thus, $X = X_0 \oplus X_1 \oplus \cdots \oplus X_d$, where each share X_i is a uniformly distributed random variable over $\mathcal{G}$ and $\oplus$ is the group operation in $\mathcal{G}$. For this group operation, we let $\ominus g$ denote the opposite of $g \in \mathcal{G}$. A typical example is "Boolean masking", for which $\oplus \equiv \ominus$ is the bitwise XOR operation.

During computation, shares $\boldsymbol{X} = (X_0, X_1, \ldots, X_d)$ are leaking through some side channel. Noisy "traces," denoted by $\boldsymbol{Y} = (Y_0, Y_1, \ldots, Y_d)$, are measured by the attacker, where $\boldsymbol{Y}$ is the output of a memoryless side channel with input $\boldsymbol{X}$. Since masking shares are drawn uniformly and independently, both $\boldsymbol{X}$ and $\boldsymbol{Y}$ are i.i.d. sequences. The attacker measures m traces $\boldsymbol{Y}^m = (\boldsymbol{Y}_1, \boldsymbol{Y}_2, \ldots, \boldsymbol{Y}_m)$ corresponding to the i.i.d. text sequence $T^m = (T_1, T_2, \ldots, T_m)$, then exploits her knowledge of $\boldsymbol{Y}^m$ and T^m to guess the secret key $\hat{K}$. Again, since the side-channel is memoryless, both $\boldsymbol{X}^m$ and $\boldsymbol{Y}^m$ are i.i.d. sequences.

Let $\mathbb{P}_s = \mathbb{P}(K = \hat{K})$ be the probability of success of the attack upon observing T^m and $\boldsymbol{Y}^m$. In theory, maximum success is obtained by the MAP (maximum a posteriori probability) rule with success probability denoted by $\mathbb{P}_s = \mathbb{P}_s(K|\boldsymbol{Y}^m, T^m)$. The whole process is illustrated in Fig. 1.

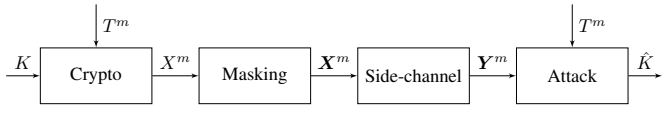

Fig. 1. Side-channel analysis as a (unintended) "communication" channel. "Crypto" can be any sensitive computation (encryption or decryption). T is a public random variable (e.g., a plain or cipher text byte).

B. Rényi's α-Entropy and Arimoto's Conditional α-Entropy

Assume that either $0 < \alpha < 1$ or $1 < \alpha < +\infty$ (the limiting values $0, 1, +\infty$ can be obtained by taking limits). We consider probability distributions P, Q with a dominating measure μ, with respect to which they follow densities denoted by the corresponding lower-case letters p, q. We follow the notations of [15] in the following

Definition 1 (Rényi α-Entropy and α-Divergence):

$$H_\alpha(P) = \tfrac{\alpha}{1-\alpha} \log \|p\|_\alpha \tag{1}$$

$$D_\alpha(P\|Q) = \tfrac{1}{\alpha-1} \log \langle p\|q\rangle_\alpha^\alpha \tag{2}$$

with the special notation:

$$\|p\|_\alpha = \Big(\int |p|^\alpha d\mu\Big)^{1/\alpha} \tag{3}$$

$$\langle p\|q\rangle_\alpha = \Big(\int p^\alpha q^{1-\alpha} d\mu\Big)^{1/\alpha}. \tag{4}$$

The usual Shannon entropy and Kullback-Leibler divergence are recovered by letting $\alpha \to 1$. The α-entropy is nonincreasing in α and achieves its *min-entropy* H_∞ at the limit $\alpha = \infty$:

Definition 2 (Min-Entropy): For a probability distribution P over a finite alphabet, the min-entropy is

$$H_\infty(P) = -\log(\max\ p). \tag{5}$$

Many different definitions of conditional α-entropy $H_\alpha(X|Y)$ were proposed in the literature. We use Arimoto's definition, which is argued to be the most promising one [8]:

Definition 3 (Arimoto's Conditional α-Entropy [2]): The conditional α-entropy of X given Y is defined as

$$H_\alpha(X|Y) = \frac{\alpha}{1-\alpha} \log \mathbb{E}_Y \|p_{X|Y}\|_\alpha. \tag{6}$$

Assuming X takes values in a finite alphabet, the conditional min-entropy can be obtained by letting $\alpha \to \infty$ in $H_\alpha(X|Y)$:

Definition 4 (Conditional Min-Entropy [24]):

$$H_\infty(X|Y) = -\log(\mathbb{E}_Y \max_x p_{X|Y}) = -\log \mathbb{P}_s(X|Y) \tag{7}$$

where $\mathbb{P}_s(X|Y)$ is the maximum average probability of success in estimating X having observed Y, by the MAP rule.

C. Sibson's α-Information and Liu et al.'s Conditional Version

Again, several different definitions of α-information $I_\alpha(X;Y)$ have been proposed, and Sibson's α-information is perhaps the most appropriate one because it satisfies several useful properties that other definitions do not [26].

Definition 5 (Sibson's α-Information [22], [26]):

$$I_\alpha(X;Y) = \min_{Q_Y} D_\alpha(P_{XY}\|P_X \times Q_Y) \tag{8}$$

$$= \tfrac{\alpha}{\alpha-1} \log \mathbb{E}_Y \langle p_{X|Y}\|p_X\rangle_\alpha. \tag{9}$$

Definition 6 (Max-Information [11, Thm. 4]): Assuming X, Y are discrete random variables, one has

$$I_\infty(X;Y) = \log \sum_y \sup_{x: p_X(x)>0} p_{Y|X}(y|x). \tag{10}$$

Max-information is also studied in [12] as *maximal leakage*.

Again, there are many different proposals for *conditional* α-information. We use the following definition which seems most appropriate in the context of side-channel analysis [15]:

Definition 7 (Conditional α-Information [15]):

$$I_\alpha(X;Y|Z) = \min_{Q_{YZ}} D_\alpha(P_{XYZ}\|P_{X|Z}Q_{YZ}) \tag{11}$$

$$= \tfrac{\alpha}{\alpha-1}\log \mathbb{E}_{YZ}\langle p_{X|YZ}\|p_{X|Z}\rangle_\alpha. \tag{12}$$

III. Fano's Equality for Order ∞: Linear Bound

A. Fano Inequality for Conditional α-Information as $\alpha \to \infty$

Using conditional α-information, Liu et al. [15] derived a universal bound on the probability of success as follows.

Theorem 1 (Generalized Fano's Inequality [15, Thm. 1]):

$$I_\alpha(K;\boldsymbol{Y}^m|T^m) \geq d_\alpha(\mathbb{P}_s(K|\boldsymbol{Y}^m,T^m)\|(\mathbb{P}_s(K))) \tag{13}$$

where $d_\alpha(p\|q)$ is the binary α-divergence:

$$d_\alpha(p\|q) = \tfrac{1}{\alpha-1}\log(p^\alpha q^{1-\alpha} + (1-p)^\alpha(1-q)^{1-\alpha}). \tag{14}$$

When $\alpha \to 1$, this bound recovers the previous bound in [6]. The simulation results in [15] show that (13) is tighter as α increases.

In this section, we prove that Fano's inequality for conditional α-information becomes an *equality* in the limiting case $\alpha = \infty$. Thus, conditional max-information can accurately characterize the probability of success.

Theorem 2 (Generalized Fano's Inequality at $\alpha = +\infty$): For a uniformly distributed secret K,

$$\begin{aligned} I_\infty(K;\boldsymbol{Y}^m|T^m) &= d_\infty(\mathbb{P}_s(K|\boldsymbol{Y}^m,T^m)\|(\mathbb{P}_s(K))) \\ &= \log(M\mathbb{P}_s) \end{aligned} \tag{15}$$

where $d_\infty(p\|q) = \lim_{\alpha\to\infty} d_\alpha(p\|q) = \log \max_{x,q(x)>0}(p(x)/q(x))$, $\mathbb{P}_s = \mathbb{P}_s(K|\boldsymbol{Y}^m,T^m)$ is the optimal probability of success, and $\mathbb{P}_s(K) = 1/M$ is the corresponding probability of success in the case of blind estimation (without any observation).

To prove this theorem, we need the explicit expression of conditional max-information.

Proposition 1 (Conditional Max-Information): Assuming X takes values in a finite alphabet, one has

$$I_\infty(X;Y|Z) = \log \mathbb{E}_Z \int_y \big(\max_{x:p_{X|Z}(x|z)>0} p_{Y|XZ}\big)\, d\mu_Y. \tag{16}$$

This result easily follows from the following Lemmas 1 and 2, which are proved in Appendices B and C respectively. In [12], conditional maximal leakage is defined as a maximum over Z, while our conditional max-information is averaged over Z—which is less than or equal to the conditional maximal leakage of [12].

Lemma 1: Given any fixed y, z, we have

$$\lim_{\alpha\to\infty}\ p_{Y|Z}\cdot\langle p_{X|YZ}\|p_{X|Z}\rangle_\alpha = \max_{x:p_{X|Z}(x|z)>0} p_{Y|XZ}. \tag{17}$$

Lemma 2:

$$\begin{aligned} &\lim_{\alpha\to\infty}\log\ \mathbb{E}_{YZ}\langle p_{X|YZ}\|p_{X|Z}\rangle_\alpha \\ &\qquad = \log \mathbb{E}_Z \int_y \lim_{\alpha\to\infty}\ p_{Y|Z}\cdot\langle p_{X|YZ}\|p_{X|Z}\rangle_\alpha. \end{aligned} \tag{18}$$

Proof of Theorem 2: Under the MAP rule, the probability of success writes

$$\begin{aligned} \mathbb{P}_s &= \mathbb{E}_{\boldsymbol{Y}^m T^m}(\max_k\ p_{K|\boldsymbol{Y}^m,T^m}) \\ &= \mathbb{E}_{T^m}\int_{\boldsymbol{y}^m}(\max_k\ p_{\boldsymbol{Y}^m|K,T^m}p_{K|T^m})d\mu_{\boldsymbol{Y}^m}. \end{aligned} \tag{19}$$

Recall K is uniformly distributed and independent from T^m. Therefore, (19) becomes

$$\mathbb{P}_s = \frac{1}{M}\cdot\mathbb{E}_{T^m}\int_{\boldsymbol{y}^m}\big(\max_k\ p_{\boldsymbol{Y}^m|K,T^m}\big)d\mu_{\boldsymbol{Y}^m}. \tag{20}$$

Combining (16) and (20) we have $I_\infty(K;\boldsymbol{Y}^m|T^m) = \log(M\mathbb{P}_s)$. Since $\mathbb{P}_s \geq 1/M$, one has $\mathbb{P}_s\cdot M \geq (1-\mathbb{P}_s)\cdot M/(M-1)$ and $d_\infty(\mathbb{P}_s(K|\boldsymbol{Y}^m,T^m)\|(\mathbb{P}_s(K))) = \log(M\mathbb{P}_s)$, which proves (15). ∎

B. Linear Bound Using Maximal Leakage $I_\infty(X;\mathbf{Y})$

Evaluating $I_\infty(K;\boldsymbol{Y}^m|T^m)$ directly turns out to be cumbersome (see Remark 1 below). Instead we use the unconditional max-information measure, i.e., maximal leakage $I_\infty(X;\mathbf{Y})$ to bound the probability of success, which is linear in the number m of measurements:

Theorem 3 (Linear Bound):

$$\log(M\mathbb{P}_s) \leq mI_\infty(X;\boldsymbol{Y}). \tag{21}$$

Proof: By Definition 6,

$$I_\infty(K,T^m;\boldsymbol{Y}^m) = \log\int_{\boldsymbol{y}^m}\max_{k,t^m}\ p_{\boldsymbol{Y}^m|K,T^m}d\mu_{\boldsymbol{Y}^m}. \tag{22}$$

Because $\max_{k,t^m}\ p_{\boldsymbol{Y}^m|K,T^m} \geq \mathbb{E}_{T^m}\ (\max_k\ p_{\boldsymbol{Y}^m|K,T^m})$, by (15) and (16) we have

$$I_\infty(K,T^m;\boldsymbol{Y}^m) \geq I_\infty(K;\boldsymbol{Y}^m|T^m) = \log(M\mathbb{P}_s). \tag{23}$$

Because $(K,T^m) \leftrightarrow X^m \leftrightarrow \boldsymbol{Y}^m$ forms a Markov chain, using the data processing inequality (DPI) for Sibson's α-information [19], [20] we have

$$I_\alpha(K,T^m;\boldsymbol{Y}^m) \leq I_\alpha(X^m;\boldsymbol{Y}^m). \tag{24}$$

Also, when T^m is not observed, each component of X^m is i.i.d., and since the side-channel is memoryless, $(X^m;\boldsymbol{Y}^m)$ is an i.i.d. sequence. It easily follows from the definition that

$$I_\alpha(X^m;\boldsymbol{Y}^m) = mI_\alpha(X;\boldsymbol{Y}). \tag{25}$$

Letting $\alpha\to\infty$ in (24) and (25) we have $I_\infty(K,T^m;\boldsymbol{Y}^m) \leq mI_\infty(X;\boldsymbol{Y})$. ∎

Remark 1: For conditional α-information we have the inequality $I_\alpha(K;\boldsymbol{Y}^m|T^m) \leq I_\alpha(X^m;\boldsymbol{Y}^m|T^m)$ similar to (24). However, one does not have an equality similar to (25) when T^m is observed.

Remark 2: This proof cannot use the result in [12, Theorem 1] because in this theorem $\boldsymbol{Y^m}$ is not on a finite alphabet. What's more, if we use Definition 1 and Theorem 1 in [12] we will have

$$I_\infty(X^m;\boldsymbol{Y}^m,T^m) \geq \log(M\cdot \mathbb{P}_s(K|\boldsymbol{Y}^m,T^m)) \tag{26}$$

but $I_\infty(X^m;\boldsymbol{Y}^m)$ is less than $I_\infty(X^m;\boldsymbol{Y}^m,T^m)$.

IV. Mrs. Gerber's Lemma for Min-Entropy in Any Finite Abelian Group

To benefit from Theorem 3 it remains to upper bound $I_\infty(X;\boldsymbol{Y})$. Since X is uniformly distributed, it is easily seen from the definition that $I_\infty(X;\boldsymbol{Y}) = \log M - H_\infty(X|\boldsymbol{Y})$. Thus, it remains to lower bound the conditional min-entropy $H_\infty(X|\boldsymbol{Y})$. This can be seen as an extension of Mrs. Gerber's lemma to min-entropy in finite additive groups.

A. Mrs. Gerber's Lemma for Two Random Variables

Wyner and Ziv [27] lower bounded the entropy of a sum of binary random variables with the entropies of each summand. This is known as Mrs. Gerber's lemma.

Theorem 4 (Mrs. Gerber's Lemma [27]): Let X_0, X_1 be two independent $\mathbb{Z}_2$-valued random variables with side information $\mathbf{Y} = (Y_0, Y_1)$ and sensitive bit $X = X_0 \oplus X_1$. Then

$$H(X|\mathbf{Y}) \geq h(h^{-1}(H(X_0|Y_0)) \star h^{-1}(H(X_1|Y_1))) \tag{27}$$

where $h(p) = -p\log p - \bar{p}\log\bar{p}$, $a\star b = a\bar{b}+\bar{a}b$ and $\bar{x} = 1-x$.

Jog and Anatharam [14] extended Mrs. Gerber's lemma to additive groups of order 2^n. Hirche [10] extended Mrs. Gerber's lemma for binary random variables to the case of Rényi entropies. In particular for min-entropy, one has equality:

Theorem 5 (Christoph Hirche [10, Lem. IV.7]): Let X_0, X_1 be two independent $\mathbb{Z}_2$-valued random variables with side information $\mathbf{Y} = (Y_0, Y_1)$ and $X = X_0 \oplus X_1$. Then

$$H_\infty(X|\mathbf{Y}) = h_\infty(h_\infty^{-1}(H_\infty(X_0|Y_0)) \star h_\infty^{-1}(H_\infty(X_1|Y_1))) \tag{28}$$

where $h_\infty(p) = -\log\max\{p,\bar{p}\}$.

In this section, Mrs. Gerber's lemma is extended for the min-entropy in any additive finite group:

Theorem 6: Let X_0, X_1 be two independent $\mathcal{G}$-valued random variables with side information $\mathbf{Y} = (Y_0, Y_1)$ and sensitive variable $X = X_0 \oplus X_1$. Then for $k = \max\{\lfloor p^{-1}\rfloor, \lfloor q^{-1}\rfloor\}$, one has the optimal bound

$$\exp(\text{–}H_\infty(X|\mathbf{Y})) \leq \begin{cases} kpq + (1-kp)(1-kq) & \text{if } \frac{1}{k+1} \leq p,q \leq \frac{1}{k} \\ \min\{p,q\} & \text{otherwise,} \end{cases} \tag{29}$$

where $p = \exp(\text{–}H_\infty(X_0|Y_0))$ and $q = \exp(\text{–}H_\infty(X_1|Y_1))$.

Remark 3: Since $kpq+(1\text{–}kp)(1\text{–}kq) = \frac{1}{k+1}+\frac{k}{k+1}((k+1)p\text{–}1)((k+1)q\text{–}1)$, $\frac{1}{k+1} \leq p,q \leq \frac{1}{k}$ implies $\frac{1}{k+1} \leq kpq+(1\text{–}kp)(1\text{–}kq) \leq \frac{1}{k}$. Thus, if both $H_\infty(X_0|Y_0)$ and $H_\infty(X_1|Y_1)$ lie in the interval $[\log k, \log(k+1)]$, then so does the corresponding bound on $H_\infty(X|\mathbf{Y})$.

Proof: We first prove the inequality in the unconditional case. The probability mass function of $X_0 \oplus X_1$ is given by the convolution with respect to $\mathcal{G}$ of the probability mass functions of X_0 and X_1. That is, for any $x \in \mathcal{G}$,

$$\mathbb{P}(X_0 \oplus X_1 = x) = \sum_{i\in\mathcal{G}} \mathbb{P}(X_0 = x\oplus i)\mathbb{P}(X_1 = \ominus i). \tag{30}$$

In particular,

$$\exp(\text{–}H_\infty(X_0 \oplus X_1)) = \max_{x\in\mathcal{G}} \sum_{i\in\mathcal{G}} \mathbb{P}(X_0 = x\oplus i)\mathbb{P}(X_1 = \ominus i). \tag{31}$$

Hence the problem reduces to upper-bound

$$\max_{x\in\mathcal{G}} \sum_{i\in\mathcal{G}} \mathbb{P}(X_0 = x\oplus i)\mathbb{P}(X_1 = \ominus i). \tag{32}$$

Since $\exp(\text{–}H_\infty(X_0 \ominus x)) = \exp(\text{–}H_\infty(X_0))$ we can assume without loss of generality that the maximum is reached for $x = 0$ and the problem reduces to the maximization of

$$\sum_{i\in\mathcal{G}} \mathbb{P}(X_0 = i)\mathbb{P}(X_1 = \ominus i). \tag{33}$$

Let $(1),\ldots,(M) \in \mathcal{G}$ be an ordering of the group elements so that $\mathbb{P}(X_0 = (1)) \geq \mathbb{P}(X_0 = (2)) \geq \ldots \geq \mathbb{P}(X_0 = (M))$. The problem is to maximize

$$\sum_{i=1}^{M} \underbrace{\mathbb{P}(X_0 = (i))}_{p_{(i)}} \underbrace{\mathbb{P}(X_1 = \ominus(i))}_{q_{(i)}}. \tag{34}$$

The min-entropy of X_1 is invariant under any permutation of its probability mass function. Furthermore, by the *rearrangement inequality* (Lemma 5 in Appendix A) a permutation of the probability mass function of X_1 maximizing the sum is such that $\mathbb{P}(X_1 = \ominus(1)) \geq \mathbb{P}(X_1 = \ominus(2)) \geq \ldots \geq \mathbb{P}(X_1 = \ominus(M))$. Finally the problem is reduced to

$$\max_{\mathbf{p},\mathbf{q}} \phi(\mathbf{p},\mathbf{q}) \triangleq \sum p_{(i)}q_{(i)} \tag{35}$$

under the constraint that $\exp(\text{–}H_\infty(X_0)) = p_{(1)} = p$ and $\exp(\text{–}H_\infty(X_1)) = q_{(1)} = q$. Moreover, h is Schur-convex in $\mathbf{p}$ when $\mathbf{q}$ is fixed and vice-versa (see Lemma 3 in Appendix A). Hence the maximum in (35) is reached for the least spread out probability mass function under the min entropy constraints. That is (Lemma 4 in Appendix A),

$$\begin{cases} (p_{(1)},\ldots,p_{(M)}) = (p,\ldots,p,1-kp,0,\ldots,0) \\ (q_{(1)},\ldots,q_{(M)}) = (q,\ldots,q,1-l\,q,0,\ldots,0) \end{cases} \tag{36}$$

where $k = \lfloor p^{-1}\rfloor$ and $l = \lfloor q^{-1}\rfloor$. Hence we obtain the bound

$$\exp(\text{–}H_\infty(X)) \leq \begin{cases} kpq + (1-kp)(1-kq) & \text{if } k = l \\ \min\{p,q\} & \text{otherwise.} \end{cases} \tag{37}$$

It remains to prove that (37) carries over to the conditional case. Note that the bound is concave in p for a fixed q and vice-versa. Indeed, let $\frac{1}{k+1} \leq q \leq \frac{1}{k}$ be fixed. Then the inequality is piecewise linear in p, equal to

$$\begin{cases} p & \text{if } p \leq \frac{1}{k+1} \\ kpq + (1-kp)(1-kq) & \text{if } \frac{1}{k+1} \leq p \leq \frac{1}{k} \\ q & \text{otherwise.} \end{cases} \tag{38}$$

The three successive slopes are 1, $k(k+1)q-k$ and 0. Since $k(k+1)q-k \in [0,1]$, these slopes are in decreasing order and the function is indeed concave. Therefore, applying Jensen's inequality (twice) proves (29). ∎

B. Extension to $d+1$ Summands

Jog and Anatharam [14] extended their generalization of Mrs. Gerber's lemma (for Shannon entropy) for random variables in group of order 2^n with two summands by repeating their inequality. In the same fashion, Theorem 6 is extended to $d+1$ summands by repeated application of Theorem 6:

Theorem 7 (Extension to $d+1$ summands): Let $p_i = \exp(-H_\infty(X_i|Y_i))$, without loss of generality assume $p_0 \leq p_1 \leq \ldots \leq p_d$. Let $k = \lfloor p_0^{-1} \rfloor$, $r = \max\{i | p_i \leq \frac{1}{k}\}$. Then $H_d = H_\infty(X|\mathbf{Y})$ is lower bounded as

$$H_d \geq -\log\left(\frac{1}{k+1} + \frac{k^r}{k+1}\prod_{i=0}^{r}((k+1)p_i - 1)\right). \quad (39)$$

Proof: See Appendix D. ∎

In the side-channel context, it is particularly interesting to characterize the behavior of the inequality in the high entropy regime in terms of maximal leakage. This corresponds to the high noise regime of Theorem 3.

Theorem 8 (Asymptotic for High Noise): Let $I_d = I_\infty(X;\mathbf{Y})$ in bits, then as $I_\infty(X_i;Y_i) \to 0$,

$$I_d \leq C_d \prod_{i=0}^{d} I_\infty(X_i;Y_i) + o\left(\prod_{i=0}^{d} I_\infty(X_i;Y_i)\right) \quad (40)$$

where $C_d = (M-1)^d(\ln 2)^d$.

Proof: See Appendix E. ∎

C. Refined Unconditioned Extension to $d+1$ Summands

In contrast to Theorem 6, Theorem 7 is not guaranteed to be optimal when $d > 1$. The inequality can be improved by exploiting the structure of the sum of multiple random variables. We derive an improved bound which is optimal for entropies in the range $[\log(k-1), \log(k)]$ provided that there is a subgroup of $\mathcal{G}$ of order k. In particular, it is optimal in the high entropy regime $[\log(M-1), \log(M)]$ (since the group itself is a subgroup of order M).

Theorem 9 (Refined extension): Let $p_i = \exp(-H_\infty(X_i))$, without loss of generality we assume $p_0 \leq p_1 \leq \ldots \leq p_d$. Let $k = \lfloor p_0^{-1} \rfloor$, $r = \max\{i | p_i \leq \frac{1}{k}\}$. Let $H_d = H_\infty(X)$,

$$H_d \geq \begin{cases} -\log\left(\frac{1}{k+1} + \frac{1}{k+1}\prod_{j=0}^{r}((k+1)p_i - 1)\right) & \text{if } r \text{ is even,} \\ -\log\left(\frac{1}{k+1} + \frac{k}{k+1}\prod_{j=0}^{r}((k+1)p_i - 1)\right) & \text{if } r \text{ is odd.} \end{cases} \quad (41)$$

Proof: See Appendix F. ∎

Contrary to Theorem 7, Theorem 9 does not apply to conditional min-entropy in general. In fact, when all the variables are fixed except one, the bound inside the logarithm is piecewise linear but discontinuous in $\frac{1}{k}$ when r is even. This discontinuity breaks the convexity of the inequality. Ensuring continuity for the desired convexity, we are led back to the expression of Theorem 7. However, under the assumption that

$$\frac{1}{M} \leq \exp(-H_\infty(X_i|Y_i = y)) \leq \frac{1}{M-1} \quad (42)$$

for all i and y, the bound of Theorem 9 inside the logarithm is linear and we do obtain a conditional inequality. Fortunately, assumption (42) makes sense in the side-channel context. In fact, a common leakage model is $Y_i = f_i(X_i) + \sigma\mathcal{N}(0,1)$ where f_i is a fixed (possibly unknown) leakage function, such as the Hamming weight or a linear combination of the bits of the variable X_i. In particular (42) holds for large enough σ (high noise regime). Then we have the following

Theorem 10 (Taylor Expansion): Assume (42) and let $I_d = I_\infty(X;\mathbf{Y})$ in bits, then as $I_\infty(X_i;Y_i) \to 0$,

$$I_d \leq C_d \prod_{j=0}^{d} I_\infty(X_i;Y_i) + o\left(\prod_{j=0}^{d} I_\infty(X_i;Y_i)\right) \quad (43)$$

where

$$C_d = \begin{cases} (\ln 2)^d & \text{if } d \text{ is even,} \\ (M-1)(\ln 2)^d & \text{if } d \text{ is odd.} \end{cases} \quad (44)$$

Proof: Taylor expansion of the exponential about 0 and of the logarithm about 1. ∎

Theorem 10 is particularly interesting because it suggests that, with respect to the *worst* case leakage distribution, masking of *odd* order d is not useful compared to masking with order $d-1$ at high noise. In practice, however, for observed leakages this phenomenon may not apply. Theorem 10 is different from Theorem 8 as the constant C_d is improved largely. Though Theorem 10 requires the high noise assumption (42) to hold.

Finally, combining Theorem 10 and Theorem 3 yields a bound on the probability of success

Corollary 1 (Bound on $\mathbb{P}_s$): For m traces, as $\mathbb{P}_s \to \frac{1}{M}$,

$$\mathbb{P}_s \leq \frac{\exp(mI_\infty(X;\mathbf{Y}))}{M} \approx \frac{1}{M} + \frac{mC_d}{M}\prod_{i=0}^{d} I_\infty(X_i;Y_i). \quad (45)$$

This is to be compared with the bound of [3, Eqn. 8]:

Proposition 2: As $\mathbb{P}_s \to \frac{1}{M}$,

$$\mathbb{P}_s \leq \frac{1}{M} + \sqrt{m}A_d\left(\prod_{i=0}^{d} I(X_i;Y_i)\right)^{\frac{1}{2}} \quad (46)$$

where $A_d = \sqrt{M-1}(2\ln 2)^{\frac{d+1}{2}}M^{-1}$.

Proof: See Appendix G. ∎

As expected both bounds decrease exponentially in d to the minimum value $\frac{1}{M}$. Although I and I_∞ are different metrics, we observe that

- the constant factor C_d/M for I_∞ in (44) is exponentially lower in d than the factor A_d for I;
- the exponential decay in d is twice higher for I_∞;
- the inequality scales better for I than for I_∞ in terms of number m of traces (since we compared both bounds for $\mathbb{P}_s \approx \frac{1}{M}$, m is not necessarily taken large).

Finally, we can contrast both bounds on a toy example. Let Y_i be uniformly distributed in $\{x \in \mathcal{G} | x \neq X_i\}$. Then it is easily seen that $I(X_i,Y_i) = I_\infty(X_i,Y_i) = \log(\frac{M}{M-1})$. In this case, the bound of this paper outperforms the bound of [3] in the high noise regime ($\mathbb{P}_s \to \frac{1}{M}$). Both bounds are compared numerically in Figs. 5 and 6 in Appendix I for $d = 1$ and 2, respectively, and $M = 256$.

V. Conclusion and Perspectives

We have shown that maximal leakage for masked implementations can be used to bound the probability of success of any side-channel attack. Maximal leakage is bounded by an efficiently computable bound based on a new variation of Mrs. Gerber's lemma for min-entropy. The bound tightness is commented with some example groups and probability mass function with figures in Appendix H.

Improving the inequality when there is no subgroup of order $k+1$ in $\mathcal{G}$ is an interesting perspective. Indeed, groups of prime order which have no subgroup except the trivial ones are of major interest for their application to masking in asymmetric cryptographic schemes (especially post-quantum schemes). Besides, it would also be of interest to check whether the parity of d does play a practical role in the efficiency of masked implementations.

REFERENCES

[1] M. Akkar and C. Giraud, "An implementation of DES and AES, secure against some attacks," in *Cryptographic Hardware and Embedded Systems - CHES 2001, Third International Workshop, Paris, France, May 14-16, 2001, Proceedings*, ser. Lecture Notes in Computer Science, Ç. K. Koç, D. Naccache, and C. Paar, Eds., vol. 2162. Springer, 2001, pp. 309–318. [Online]. Available: https://doi.org/10.1007/3-540-44709-1_26

[2] S. Arimoto, "Information measures and capacity of order α for discrete memoryless channels," in *Topics in Information Theory, Proc. 2nd Colloq. Math. Societatis János Bolyai*, A. Joux, Ed., vol. 16, 1975, pp. 41–52.

[3] J. Béguinot, W. Cheng, S. Guilley, Y. Liu, L. Masure, O. Rioul, and F.-X. Standaert, "Removing the field size loss from Duc et al.'s conjectured bound for masked encodings," *IACR Cryptol. ePrint Arch.*, pp. 1–18, 2022. [Online]. Available: https://eprint.iacr.org/2022/1738

[4] J. Béguinot, W. Cheng, S. Guilley, and O. Rioul, "Be my guess: guessing entropy vs. success rate for evaluating side-channel attacks of secure chips," in *25th Euromicro Conference on Digital System Design, DSD 2022, Maspalomas, Spain, August 31 - Sept. 2, 2022*. IEEE, 2022, pp. 496–503. [Online]. Available: https://doi.org/10.1109/DSD57027.2022.00072

[5] W. Cheng, Y. Liu, S. Guilley, and O. Rioul, "Attacking masked cryptographic implementations: information-theoretic bounds," in *2022 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2022, pp. 654–659.

[6] E. 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, pp. 49–79, 2019. [Online]. Available: https://tches.iacr.org/index.php/TCHES/article/view/7385/6557

[7] A. Duc, S. Faust, and F.-X. 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: https://doi.org/10.1007/978-3-662-46800-5_16

[8] S. Fehr and S. Berens, "On the conditional Rényi entropy," *IEEE Transactions on Information Theory*, vol. 60, pp. 6801–6810, 2014.

[9] L. Goubin and J. Patarin, "DES and differential power analysis (the "duplication" method)," in *Cryptographic Hardware and Embedded Systems, First International Workshop, CHES'99, Worcester, MA, USA, August 12-13, 1999, Proceedings*, ser. Lecture Notes in Computer Science, Ç. K. Koç and C. Paar, Eds., vol. 1717. Springer, 1999, pp. 158–172. [Online]. Available: https://doi.org/10.1007/3-540-48059-5_15

[10] C. Hirche, "Rényi bounds on information combining," in *2020 IEEE International Symposium on Information Theory (ISIT)*, 2020, pp. 2297–2302.

[11] S.-W. Ho and S. Verdú, "Convexity/concavity of Rényi entropy and α-mutual information," in *2015 IEEE International Symposium on Information Theory (ISIT)*, 2015, pp. 745–749.

[12] I. Issa, A. B. Wagner, and S. Kamath, "An operational approach to information leakage," *IEEE Transactions on Information Theory*, vol. 66, no. 3, pp. 1625–1657, 2020.

[13] A. Ito, R. Ueno, and N. Homma, "On the success rate of side-channel attacks on masked implementations: information-theoretical bounds and their practical usage," in *Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security, CCS 2022, Los Angeles, CA, USA, November 7-11, 2022*, H. Yin, A. Stavrou, C. Cremers, and E. Shi, Eds. ACM, 2022, pp. 1521–1535. [Online]. Available: https://doi.org/10.1145/3548606.3560579

[14] V. Jog and V. Anantharam, "The entropy power inequality and Mrs. Gerber's lemma for groups of order 2^n," *2013 IEEE International Symposium on Information Theory (ISIT)*, pp. 594–598, 2013.

[15] Y. Liu, W. Cheng, S. Guilley, and O. Rioul, "On conditional alpha-information and its application to side-channel analysis," in *IEEE Information Theory Workshop, ITW 2021, Kanazawa, Japan, October 17-21, 2021*. IEEE, 2021, pp. 1–6. [Online]. Available: https://doi.org/10.1109/ITW48936.2021.9611409

[16] M. Madiman, L. Wang, and J. O. Woo, "Entropy inequalities for sums in prime cyclic groups," *SIAM Journal on Discrete Mathematics*, vol. 35, no. 3, pp. 1628–1649, 2021.

[17] A. W. Marshall, I. Olkin, and B. C. Arnold, *Inequalities: Theory of Majorization and Its Applications*. Springer, 1980.

[18] L. Masure, O. Rioul, and F.-X. Standaert, "A nearly tight proof of Duc et al.'s conjectured security bound for masked implementations," *IACR Cryptol. ePrint Arch.*, p. 600, 2022. [Online]. Available: https://eprint.iacr.org/2022/600

[19] Y. Polyanskiy and S. Verdú, "Arimoto channel coding converse and Rényi divergence," in *2010 48th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, 2010, pp. 1327–1333.

[20] O. Rioul, "A primer on alpha-information theory with application to leakage in secrecy systems," in *International Conference on Geometric Science of Information*. Springer, 2021, pp. 459–467.

[21] S. Shamai and A. Wyner, "A binary analog to the entropy-power inequality," *IEEE Transactions on Information Theory*, vol. 36, no. 6, pp. 1428–1430, 1990.

[22] R. Sibson, "Information radius," *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, vol. 14, pp. 149–160, 1969.

[23] T. Tao, "Sumset and inverse sumset theory for Shannon entropy," *Combinatorics, Probability and Computing*, vol. 19, pp. 603 – 639, 2009.

[24] A. Teixeira, A. Matos, and L. Antunes, "Conditional Rényi entropies," *IEEE Transactions on Information Theory*, vol. 58, no. 7, pp. 4273–4277, 2012.

[25] T. van Erven and P. Harremos, "Rényi divergence and kullback-leibler divergence," *IEEE Transactions on Information Theory*, vol. 60, no. 7, pp. 3797–3820, 2014.

[26] S. Verdú, "α-mutual information," in *IEEE Information Theory and Applications Workshop (ITA2015)*, San Diego, USA, 2015, pp. 1–6. [Online]. Available: https://doi.org/10.1109/ITA.2015.7308959

[27] A. D. Wyner and J. Ziv, "A theorem on the entropy of certain binary sequences and applications-I," *IEEE Transactions on Information Theory*, vol. 19, pp. 769–772, 1973.

APPENDIX

A. Background on Majorization

We recall definitions and basic results of majorization theory. An extensive presentation can be found in the reference textbook [17].

Definition 8 (Statistical Ordering): If $\mathbf{p} = (p_1, \ldots, p_M)$ is a probability mass function, an arrangement $(1), (2), \ldots, (M)$ of $\mathbf{p}$ so that $p_{(1)} \geq \ldots \geq p_{(M)}$ is said to be the statistical ordering of $\mathbf{p}$. The associated cumulative mass function is noted $P_{(i)} = p_{(1)} + \ldots + p_{(i)}$ where $P_{(0)} = 0$ by convention.

Definition 9 (Majorization): Let $\mathbf{p}, \mathbf{q}$ be two probability mass functions. We say that $\mathbf{q}$ majorizes $\mathbf{p}$ and write $\mathbf{p} \preceq \mathbf{q}$ if

$$P_{(i)} \leq Q_{(i)} \qquad (i = 1, \ldots, M). \tag{47}$$

This partial order on the probability mass functions quantifies whether a distribution is more spread out than the other.

Definition 10 (Schur-Convexity): $f : \mathbf{p} \mapsto f(\mathbf{p}) \in \mathbb{R}$ is said to be Schur-convex if it is increasing with respect to majorization i.e. $\mathbf{p} \preceq \mathbf{q} \implies f(\mathbf{p}) \leq f(\mathbf{q})$.

Lemma 3 (Schur-Convex Combination): If $\alpha_1 \geq \ldots \geq \alpha_M$ then $(p_1, \ldots, p_M) \mapsto \sum_{i=1}^{M} \alpha_i p_{(i)}$ is Schur-convex.

Proof: This can be shown by an Abel transform as pointed out in [4, Remark 2].

$$\sum \alpha_i p_{(i)} = \sum \alpha_i (P_{(i)} - P_{(i-1)}) \tag{48}$$

$$= \alpha_M P_{(M)} - \alpha_1 P_{(0)} - \sum_{i=1}^{M-1} (\alpha_{i+1} - \alpha_i) P_{(i)} \tag{49}$$

$$= \alpha_M - \sum_{i=1}^{M-1} (\alpha_{i+1} - \alpha_i) P_{(i)}. \tag{50}$$

Since $\alpha_{i+1} - \alpha_i \leq 0$ the Schur-convexity follows from the definition. ∎

Lemma 4 (Majorization and min-entropy): Let $\mathbf{p}$ be a probability mass functions whose min-entropy is equal to $-\log p$ and $k = \lfloor p^{-1} \rfloor$ then

$$(p, \frac{1-p}{M-1}, \ldots, \frac{1-p}{M-1}) \preceq \mathbf{p} \preceq (p, \ldots, p, 1-kp, 0, \ldots, 0). \tag{51}$$

Lemma 5 (Rearrangement Inequality): Let $(a_1, \ldots, a_n)$, $(b_1, \ldots, b_n) \in \mathbb{R}^{+n}$ be two sequences in descending order. Then for all permutations σ of $\{1, \ldots, n\}$ it holds that

$$\sum a_i b_{n+1-i} \leq \sum a_i b_{\sigma(i)} \leq \sum a_i b_i. \tag{52}$$

Proof: See [17] for a proof using majorization. ∎

B. Proof of Lemma 1

Method 1:
By Theorem 6 of [25] we have

$$\lim_{\alpha \to \infty} \langle p_{X|YZ} \| p_{X|Z} \rangle_\alpha = \exp\big(D_\infty(P_{X|YZ} \| P_{X|Z})\big) \tag{53}$$

$$= \max_{x : p_{X|Z}(x|z) > 0} \frac{p_{X|YZ}}{p_{X|Z}}. \tag{54}$$

Because $p_{Y|Z} \cdot p_{X|YZ} / p_{X|Z} = p_{Y|XZ}$, the proof is finished.
Method 2:
We use L^∞-norm to prove this lemma.

$$\begin{aligned} p_{Y|Z} \langle p_{X|YZ} \| p_{X|Z} \rangle_\alpha &= p_{Y|Z} \ \big(\sum_{x \in \mathcal{X}} p_{X|YZ}^{\alpha} \ p_{X|Z}^{1-\alpha}\big)^{\frac{1}{\alpha}} \\ &= \big(\sum_{x \in \mathcal{X}} p_{XY|Z}^{\alpha} \ p_{X|Z}^{1-\alpha}\big)^{\frac{1}{\alpha}} = \Big(\sum_{x \in \mathcal{X}} \big(p_{XY|Z} \ p_{X|Z}^{\frac{1-\alpha}{\alpha}}\big)^{\alpha}\Big)^{\frac{1}{\alpha}} \\ &= \Big(\sum_{x \in \mathcal{X}} \big(p_{Y|XZ} \ p_{X|Z}^{\frac{1}{\alpha}}\big)^{\alpha}\Big)^{\frac{1}{\alpha}}. \end{aligned} \tag{55}$$

For any $\varepsilon > 0$, there exists a sufficiently large $\alpha > 0$ such that

$$p_{Y|XZ} - \varepsilon \leq p_{Y|XZ} \ p_{X|Z}^{\frac{1}{\alpha}} \leq p_{Y|XZ}. \tag{56}$$

Because $\mathcal{X}$ is finite, one always has a sufficiently large $\alpha > 0$ such that (56) holds for any $x \in \mathcal{X}$. By L^∞-norm we have

$$\lim_{\alpha \to \infty} \Big(\sum_{x : p_{X|Z}(x|z) > 0} \big(p_{Y|XZ} - \varepsilon\big)^{\alpha} \Big)^{\frac{1}{\alpha}} = \max_{x : p_{X|Z}(x|z) > 0} p_{Y|XZ} - \varepsilon \tag{57}$$

Since $\varepsilon > 0$ is arbitrary, combined with the squeeze theorem, the proof is finished.

C. Proof of Lemma 2

By definition we have

$$\begin{aligned} &\lim_{\alpha \to \infty} \log \mathbb{E}_{YZ} \langle p_{X|YZ} \| p_{X|Z} \rangle_\alpha \\ &= \lim_{\alpha \to \infty} \log \mathbb{E}_{YZ} \exp\big(\tfrac{\alpha-1}{\alpha} D_\alpha \langle p_{X|YZ} \| p_{X|Z} \rangle_\alpha\big). \end{aligned} \tag{58}$$

This value is bounded because $I(X;Y|Z) \leq \log M$. Since $\frac{\alpha-1}{\alpha} D_\alpha \langle p_{X|YZ} \| p_{X|Z} \rangle_\alpha$ is increasing in α, the lemma follows from the monotone convergence theorem. ∎

D. Proof of Theorem 7

We prove the inequality by induction. Theorem 6 settles the case of $d + 1 = 2$ variables. We assume it is true for all sets of at most $d + 1$ variables and show it is true for all set of at most $d + 2$ variables. Let k, r_{d+1} be the value of k, r in the theorem associated to $X_0, \ldots, X_d, X_{d+1}$. If $r_{d+1} < d + 1$ we lower bound the min-entropy of the sum $X_0 \oplus \ldots \oplus X_{d+1}$ by the entropy of $X_0 \oplus \ldots \oplus X_d$. We conclude by applying the induction hypothesis to this sum of d random variables. Else $r_{d+1} = d+1$. Since $X_0 \oplus \ldots \oplus X_d \oplus X_{d+1} = (X_0 \oplus \ldots \oplus X_d) \oplus X_{d+1}$, we apply the induction hypothesis $\mathcal{H}_d$ to $X_0, \ldots, X_d$ then we apply Theorem. 6 to X_{d+1} and $X_0 \oplus \ldots \oplus X_d$. Let $K = \exp(\text{–}H_\infty(X_0 \oplus \ldots \oplus X_{d+1} | Y_0 \ldots Y_{d+1}))$.

$$\begin{aligned} K &\overset{(a)}{\leq} 1 - k(p_{d+1} + \exp(\text{–}H_\infty(X_0 \oplus \ldots \oplus X_d | Y_0 \ldots Y_d))) \\ &\quad + k(k+1) p_{d+1} \exp(\text{–}H_\infty(X_0 \oplus \ldots \oplus X_d | Y_0 \ldots Y_d)) \end{aligned} \tag{59}$$

$$\begin{aligned} &\overset{(b)}{\leq} 1 - k(p_{d+1} + \frac{1}{k+1} + \frac{k^d}{k+1} \prod_{i=0}^{d} ((k+1)p_i - 1)) \\ &\quad + k(k+1) p_{d+1} (\frac{1}{k+1} + \frac{k^d}{k+1} \prod_{i=0}^{d} ((k+1)p_i - 1)) \end{aligned} \tag{60}$$

$$= \frac{1}{k+1} + \frac{k^{d+1}}{k+1} \prod_{i=0}^{d} ((k+1)p_i - 1))((k+1)p_{d+1} - 1) \tag{61}$$

$$= \frac{1}{k+1} + \frac{k^{d+1}}{k+1} \prod_{i=0}^{d+1} ((k+1)p_i - 1)) \tag{62}$$

where (a) holds by $\mathcal{H}_1$ and (b) holds by $\mathcal{H}_d$. As a repeated application of Theorem 6 the inequality naturally extends to the conditional case. ∎

E. Proof of Theorem 8

We upper bound $I_d = \log M - H_\infty(X|\mathbf{Y})$ using the lower bound on the min entropy. At high entropy $k = M - 1$ hence

$$\log M - I_d \geq -\log\left(\frac{1}{M} + \frac{(M-1)^d}{M} \prod_{i=0}^{d} (Mp_i - 1)\right) \tag{63}$$

where

$$p_i = \frac{\exp(I_\infty(X_i; Y_i))}{M}. \tag{64}$$

$$I_d \leq \log\left(1 + (M-1)^d \prod_{i=0}^{d} \big(\exp(I_\infty(X_i; Y_i)) - 1\big)\right) \tag{65}$$

$$= (M-1)^d (\ln 2)^d \prod_{i=0}^{d} I_\infty(X_i; Y_i) + o\left(\prod_{i=0}^{d} I_\infty(X_i; Y_i)\right) \tag{66}$$

∎

F. Proof of Theorem 9

We first prove the following usefull lemma. It intuitively tells that to minimize the min-entropy the pmf should not spread out to other values. For instance when the summed random variables are in a sub-group the value of their sum is confined in this sub-group.

Lemma 6: If X_0 and X_1 have pmfs up to permutation $(q, \ldots, q, 1-kq, 0, \ldots, 0)$ and $(p, \ldots, p, 1-kp, 0, \ldots, 0)$ then the pmf of $X_0 \oplus X_1$ is majorized by the pmf $(r, \ldots, r, 1 - kr, 0, \ldots, 0)$ where $r = p + q - (k+1)pq$. There is equality when X_0, X_1 are supported on the coset of a subgroup of $\mathcal{G}$ of order $k+1$.

Proof: The convolution involves $(k+1)^2$ strictly positive terms. Namely

$$\begin{cases} pq & k^2 \text{ times} \\ p(1-kq) & k \text{ times} \\ q(1-kp) & k \text{ times} \\ (1-kp)(1-kq) & \text{once} \end{cases}. \tag{67}$$

Further, in each mass of the results they are at most $k+1$ terms that are added and at most once an expression containing $(1-kp)$ and $(1-kq)$. Let us assume that $q \geq 1-kq$ and $p \geq 1-kp$ or $1-kq \geq q$ and $1-kp \geq p$. By rearrangement inequality (Lemma 5), $kpq + (1-kp)(1-kq)$ is the largest terms that can be obtained. The 2^{nd} to $(k+1)$-th largest terms are $(k-2)pq + p(1-kq) + q(1-kp) = p+q-(k+1)pq$. This majorizes all possible results since each term of the statistical ordering is maximized the sequence of cumulative mass function is also maximized. If $q \geq 1-kq$ and $1-kp \geq p$ or $1-kq \geq q$ and $p \geq 1-kp$ the proof is the same but the 1^{st} to k-th largest terms are $(k-2)pq + p(1-kq) + q(1-kp) = p+q-(k+1)pq$ and the $k+1$-th largest term is $kpq + (1-kp)(1-kq)$ which is the same pmf up to a permutation. The case of equality is clear. ∎

We derive the inequality of Thm. 9. The proof is composed of three steps. The first step is to prove that the inequality is achieved for pmf of the form $\mathbf{p_j} = (p_j, \ldots, p_j, 1 - k_j p_j, 0, \ldots, 0)$. The second step is to majorize the resulting convolution by induction. The final steps is to conclude the majorization argument. As in the case of two summands the problem is to maximize

$$\max_{x \in \mathcal{G}} \sum_{i_0, i_1, \ldots, i_{d-1} \in \mathcal{G}} \left(\prod_{j=0}^{d-1} \mathbb{P}(X_j = i_j)\right) \mathbb{P}(X_d = x \ominus \bigoplus_{j=0}^{d-1} i_j). \tag{68}$$

Without loss of generality, we can assume that the maximum is reached in $x = 0$, it remains to upper bound

$$\phi(\mathbf{p_0}, \ldots, \mathbf{p_d}) \triangleq \sum_{i_0, i_1, \ldots, i_{d-1} \in \mathcal{G}} \left(\prod_{j=0}^{d-1} \mathbb{P}(X_j = i_j)\right) \mathbb{P}(X_d = \ominus \bigoplus_{j=0}^{d-1} i_j). \tag{69}$$

We fix $\mathbf{p_1}, \ldots, \mathbf{p_d}$. The maximization can be written as

$$\sum_{i_0 \in \mathcal{G}} \mathbb{P}(X_0 = i_0) \alpha_{i_0} \tag{70}$$

where

$$\alpha_{i_0} = \sum_{i_1, \ldots, i_{d-1} \in \mathcal{G}} \left(\prod_{j=1}^{d-1} \mathbb{P}(X_j = i_j)\right) \mathbb{P}(X_d = \ominus \bigoplus_{j=0}^{d-1} i_j). \tag{71}$$

This is equivalent to maximize

$$\sum_{i=1}^{M} \mathbb{P}(X_0 = (i)) \alpha_{(i)} \tag{72}$$

where $(1), \ldots, (M)$ are such that $\mathbb{P}(X_0 = (1)) \geq \ldots \geq \mathbb{P}(X_0 = (M))$. By rearrangement (Lemma 5), (72) is maximum when $\alpha_{(1)} \geq \ldots \geq \alpha_{(M)}$. By lemma 3 this mapping is Schur-Convex in $\mathbf{p_0}$ hence by lemma 4 it is maximized for statistical ordering of the probability mass function of X_0 of the form

$$\mathbf{p_0} = (p_0, \ldots, p_0, 1 - k_0 p_0, 0, \ldots, 0) \tag{73}$$

where $k_0 = \lfloor p_0^{-1} \rfloor$. Equation (73) does not depend on the fixed probability mass functions of $X_1, \ldots, X_d$. By symmetry, we also obtain that the for $j = 0, \ldots, d$ the statistical ordering of the probability mass function of X_j is of the form

$$\mathbf{p_j} = (p_j, \ldots, p_j, 1 - k_j p_j, 0, \ldots, 0) \tag{74}$$

where $k_j = \lfloor p_j^{-1} \rfloor$. This concludes the first step of the proof. As previous proof we can further assume without loss of generality that k_j is constant equal to k for all j. It remains

to determine for which permutation of these probability mass function we obtain the lowest min-entropy.

Now we fix the pmf $\mathbf{p_2}, \dots, \mathbf{p_d}$. And we consider the maximization with respect to the pmf of X_0+X_1. By lemma 3, the expression is Schur-convex. Hence it is maximized for the least spread out pmf. By lemma 6, the pmf is majorized by $(r, \dots, r, 1-kr, 0, \dots, 0)$ where

$$r = p + q - (k+1)pq. \tag{75}$$

We can proceed by induction to majorize the sum of $d+1$ random variables. Let $\mathcal{H}_d$ be the induction hypothesis: The probability mass function of the sum of $d+1$ random variables is majorized by $(r, \dots, r, 1-kr, 0, \dots, 0)$ where

$$(k+1)r = 1 + (-1)^d \prod_{i=0}^{d}((k+1)p_i - 1). \tag{76}$$

The initialization $\mathcal{H}_1$ is true from (75). We assume $\mathcal{H}_j$ holds and proves $\mathcal{H}_{j+1}$ holds. Using (75) with $\mathcal{H}_j$ we obtain that the convolution is majorized by $(r, \dots, r, 1-kr, 0, \dots, 0)$ with

$$r = p_{j+1} + \frac{1}{k+1} + \frac{(-1)^j}{k+1}\prod_{i=0}^{j}((k+1)p_i - 1) \tag{77}$$

$$- (k+1)p_{j+1}\left(\frac{1}{k+1} + \frac{(-1)^j}{k+1}\prod_{i=0}^{j}((k+1)p_i - 1)\right) \tag{78}$$

$$= \frac{1}{k+1} + \frac{(-1)^j}{k+1}\prod_{i=0}^{j}((k+1)p_i - 1)(1-(k+1)p_{j+1}) \tag{79}$$

This proves $\mathcal{H}_{j+1}$ and we conclude by induction. This concludes the second step of the proof and it remains to conclude.

We proved that the probability mass function of the sum of $d+1$ random variables is majorized by $(r, \dots, r, 1-kr, 0, \dots, 0)$ where

$$(k+1)r = 1 + (-1)^d \prod_{i=0}^{d}((k+1)p_i - 1). \tag{80}$$

This shows that

$$\exp(-H_d) \leq \begin{cases} r & \text{if d is even} \\ 1-kr & \text{if d is odd} \end{cases} = \begin{cases} \frac{1}{k+1} + \frac{1}{k+1}\prod_{j=0}^{d}((k+1)p_i - 1) & (d \text{ even}) \\ \frac{1}{k+1} + \frac{k}{k+1}\prod_{j=0}^{d}((k+1)p_i - 1) & (d \text{ odd}) \end{cases}.$$

■

G. Proof of Proposition 2

Using Fano's inequality, de Chérisey et al. [6, Eqn. 11] have shown that

$$mI(X;\mathbf{Y}) \geq \log(M) - h(\mathbb{P}_s) - (1-\mathbb{P}_s)\log(M-1). \tag{81}$$

This can be explicited by computing a Taylor expansion of degree two of the binary entropy function in $\mathbb{P}_s = \frac{1}{M}$,

$$h(\mathbb{P}_s) = h(\tfrac{1}{M}) + h'(\tfrac{1}{M})(\mathbb{P}_s - \tfrac{1}{M}) + \frac{h''(\frac{1}{M})}{2}(\mathbb{P}_s - \tfrac{1}{M})^2 + o((\mathbb{P}_s - \tfrac{1}{M})^2) \tag{82}$$

$$= \log(M) - (1-\frac{1}{M})\log(M-1) + \log(M-1)(\mathbb{P}_s - \tfrac{1}{M}) - \frac{M^2\log(e)}{2(M-1)}(\mathbb{P}_s - \tfrac{1}{M})^2 + o((\mathbb{P}_s - \tfrac{1}{M})^2). \tag{83}$$

In particular (81) reduces to

$$mI(X;\mathbf{Y}) \geq \frac{M^2\log(e)}{M-1}(\mathbb{P}_s - \tfrac{1}{M})^2 + o((\mathbb{P}_s - \tfrac{1}{M})^2), \tag{84}$$

where we leveraged the following equalities

$$h(\tfrac{1}{M}) = \log(M) - (1-\frac{1}{M})\log(M-1), \tag{85}$$

$$h'(\tfrac{1}{M}) = \log(M-1) \text{ and } h''(\tfrac{1}{M}) = \frac{-M^2\log(e)}{M-1}. \tag{86}$$

In particular, (84) shows that,

$$\mathbb{P}_s \leq \frac{1}{M} + \sqrt{\frac{2\ln 2(M-1)m}{M^2} I(X,\mathbf{Y})} \tag{87}$$

$$\approx \frac{1}{M} + \sqrt{m} A_d \sqrt{\prod_{i=0}^{d} I(X_i, Y_i)} \qquad \text{(with [3, Eqn. 8])} \tag{88}$$

where

$$A_d = \frac{\sqrt{(M-1)(2\ln 2)^{d+1}}}{M}. \tag{89}$$

■

H. Discussion on the Bound Optimality

To investigate the bound tightness we compute and plot in Figs. 2, 3, and 4 the sequence $\mathbf{p}_d$ of pmf supported on a finite additive group $\mathcal{G}$ given by a fixed pmf $\mathbf{p}_0$ and the equation $\mathbf{p}_{d+1} = \mathbf{p}_d * \mathbf{p}_0$ where $*$ is the convolution with respect to the group $\mathcal{G}$. In other words, $\mathbf{p}_d$ is the pmf of the sum of $d+1$ i.i.d. $\mathcal{G}$-valued random variables with a law given by $\mathbf{p}_0$.

Figs. 2 and 4 show that the presented bound is tight in two situations:

1) When the support of the random variables is in the coset of a sub-group of order $k+1$ the inequality is tight. This is the case in Fig. 2 as $\{\bar{0}; \bar{7}\}$ is a finite sub-group of $\mathbb{Z}_{14}$ with two elements.
2) In the high entropic regime, $k = M-1$ and there is always a sub-group, the group itself. This is the case in Fig. 4.

However, when there is no finite sub-group of order $k+1$ the inequality can be strictly violated as shown by Fig. 3. Figs. 3 and 2 differs only by their group structure changed from $\mathbb{Z}_{14}$ to $\mathbb{Z}_{13}$, though the effect is huge on the actual entropy of the sum. Indeed $\{\bar{0}; \bar{7}\}$ is not the coset of a sub-group of $\mathbb{Z}_{13}$, it is even spanning the whole group. As reported

in [16] the Cauchy-Davenport inequality shows that for A, B two subsets of $\mathbb{Z}_p$ (p prime), $|A+B| \geq \min\{|A|+|B|-1, p\}$. As a consequence, the support of the sum must spread to the whole group very quickly. The investigation of this results may improve the presented inequalities.

In Fig. 4, we observe that the min-entropy does not increase visibly neither from $d = 0$ to $d = 1$ nor from $d = 2$ to $d = 3$. This supports the observation of Theorem 9 that masking with odd order might not be relevant with respect to the worst case leakages as measured by the min-entropy.

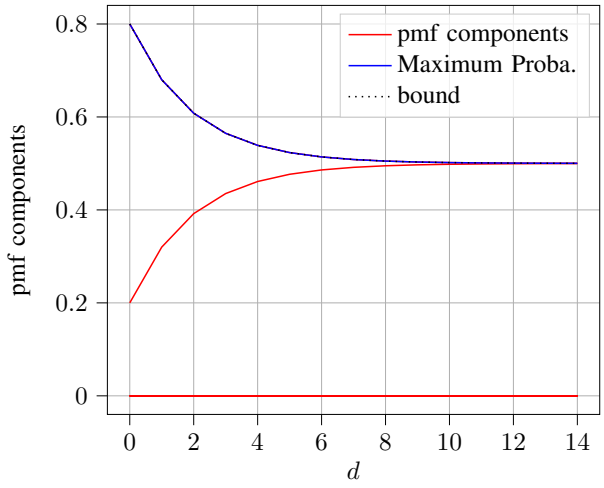

Fig. 2. Probability mass function of the sum of $d+1$ i.i.d. $\mathbb{Z}_{14}$-valued random variables with probability mass function $\mathbf{p}_0$ with $p_0 = 0.8$ and $p_7 = 0.2$.

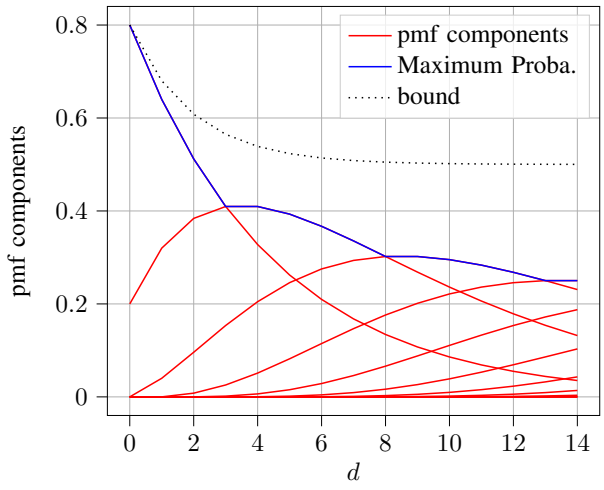

Fig. 3. Probability mass function of the sum of $d+1$ i.i.d. $\mathbb{Z}_{13}$-valued random variables with probability mass function $\mathbf{p}_0$ with $p_0 = 0.8$ and $p_7 = 0.2$.

I. Bound Comparison

Figs 5 and 6 compare both bounds for the toy example introduced. Though the bound obtained with I_∞ does not change significantly from $d = 1$ to $d = 2$. The new bound performs better for this leakage. Especially for moderate (less than 10^5) number of traces. The main limitation of this bound is that it relies on a high noise assumption.

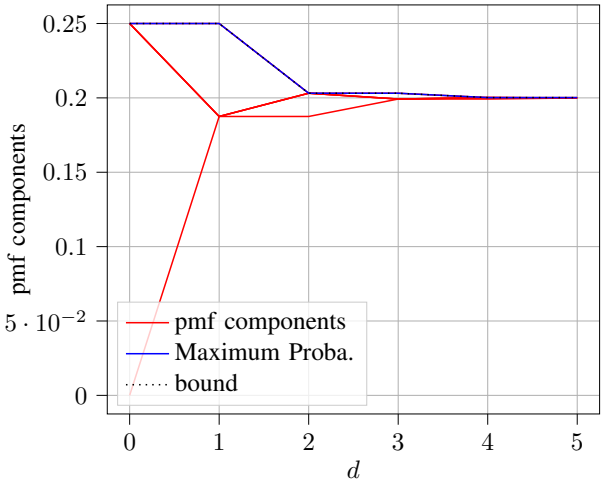

Fig. 4. Probability mass function of the sum of $d+1$ i.i.d. $\mathbb{Z}_5$-valued random variables with probability mass function $\mathbf{p}_0 = (\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, 0)$.

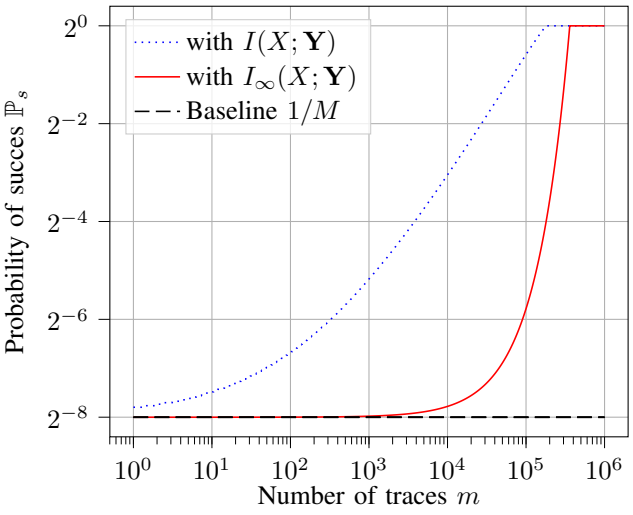

Fig. 5. Comparison of the two upper bounds (ours, Corollary 1, versus state-of-the-art, namely [3, Eqn. 8]) for $d = 1$ and $M = 256$

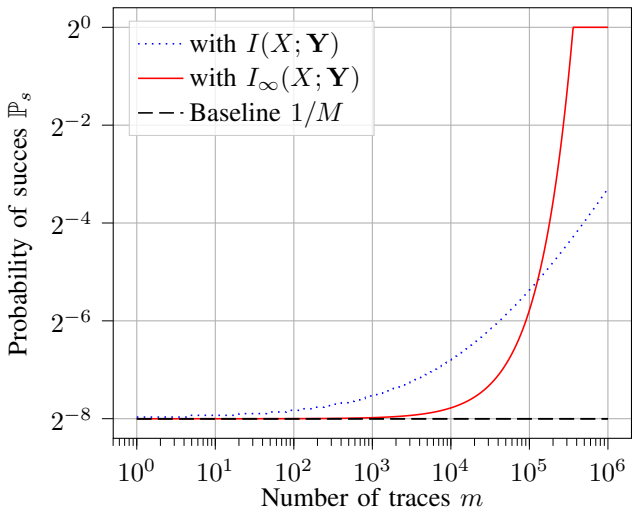

Fig. 6. Comparison of the two upper bounds (ours, Corollary 1, versus state-of-the-art, namely [3, Eqn. 8]) for $d = 2$ and $M = 256$