Cumulant Expansion of Mutual Information for Quantifying Leakage of a Protected Secret

Olivier Rioul, Wei Cheng, Sylvain Guilley

Cumulant Expansion of Mutual Information for Quantifying Leakage of a Protected Secret

Olivier Rioul*, Wei Cheng*, and Sylvain Guilley†*
*LTCI, Télécom Paris, Institut Polytechnique de Paris, Palaiseau, France, firstname.lastname@telecom-paris.fr
†Secure-IC S.A.S., Tour Montparnasse, Paris, France, sylvain.guilley@secure-ic.com

***Abstract*—The information leakage of a cryptographic implementation with a given degree of protection is evaluated in a typical situation when the signal-to-noise ratio is small. This is solved by expanding Kullback-Leibler divergence, entropy, and mutual information in terms of moments/cumulants.**

I. Introduction

Consider the following threat model in any secrecy or privacy problem where the adversary guesses a secret (cryptographic key, password, identifier, etc.), modeled as a discrete random variable X, using some observation output of some *side channel* (power consumption, electromagnetic emanation, acoustic noise, timing, etc.) modeled as a real-valued random variable Y. In side-channel applications targeting cryptographic implementations, the observation is generally made by some noisy measurement of a *sensitive variable* Z, an unknown (possibly randomized) function of the secret X which depends on the implementation. The noise is often modeled as Gaussian $N \sim \mathcal{N}(0, \sigma_N^2)$ independent of (X, Z), and the observed $Y = Z + N$ is the output of an AWGN channel. We are interested in how mutual information

$$I(X;Y) = h(Z+N) - h(Z+N \mid X) \tag{1}$$

decreases as noise power σ_N^2 increases, that is, in a typical small signal-to-noise scenario. The aim is to provide a theoretical leakage quantification as a dependency metric between secret X and attacker's observation Y. This is particularly interesting for the designer who needs to evaluate the robustness of a given implementation to side-channel attacks.

Most practical side-channel attacks actually rely on *correlation*, because it is well adapted to the Gaussian nature of the measurement noise while also being much faster than direct likelihood estimation or so-called "mutual information analysis" [1]. To protect the implementation in practice against such attacks, the cipher algorithm works with some masking scheme in such a way that leakage is perfectly balanced at all orders $k < K$:

$$\mathbb{E}(Z^k|X) = \mathbb{E}(Z^k) \quad \text{a.s.} \quad (k = 1, 2, \ldots, K-1). \tag{2}$$

Expanding powers $Y^k = (Z+N)^k$ and using the fact that N is independent of X, it follows by induction that

$$\mathbb{E}(Y^k|X) = \mathbb{E}(Y^k) \quad \text{a.s.} \quad (k = 1, 2, \ldots, K-1). \tag{3}$$

The order K is referred to as the *high-order correlation immunity* (HCI) degree by Carlet et al. [2]. It corresponds to the smallest moment of leakage that may depend on the secret. As a result, any attack from observation Y based on correlation analysis of degree $k < K$ necessarily fails; K is the minimal attack order that can possibly succeed.

HCI is now commonly adopted in quantifying security leakage in side-channel attacks. For example, the "quantitative masking strength" defined in [3] vanishes if and only if HCI> 1; similarly [4] defines "information leakage" of order 1 and 2 as being first and second-moment quantities which vanish precisely when HCI is greater than 1 and 2, respectively.

The question now becomes: How does mutual information $I(X;Y)$ capture the fact that the kth order conditional moment $m_K(Y|X=x) = \mathbb{E}(Y^K|X=x)$ depends on x when the noise increases?

Carlet et al.'s statement [2] is that $I(X;Y)$ is asymptotically $O(\sigma_N^{-2K})$ as $\sigma_N \to \infty$. This was taken as a fundamental result in the field of side-channel analysis. It was leveraged to illustrate the strength of leakage squeezing [5, Fig. 4], to compare different countermeasures [6], [7], and was extended in [8] in the case of a code-based masking implementation where countermeasures can reduce mutual information by increasing the dual distance of the code and reducing its kissing number.

Carlet et al.'s derivation [2], however, is based on Cardoso's small cumulant approximation [9, Eq. (41)] which in fact replaces Kullback-Leibler divergence by its quadratic approximation [9, Eq. (29)]. As shown in this paper, this results in a problematic expansion of mutual information [2, Eq. (6)], with erroneous coefficients. We make the appropriate corrections and find the correct asymptotic equivalent of $I(X;Y)$ up to $K = 6$. Our main result is then the following[1].

Theorem 1: Let X, Z be (discrete or continuous) real-valued random variables satisfying (2) at orders $k = 1, 2, \ldots, K-1$ but *not* at order K (i.e., with at least one value x such that $\mathbb{E}(Z^K|X=x) \neq \mathbb{E}(Z^K)$). Then if $3 \leq K \leq 6$, the following asymptotic equivalence holds as $\sigma_N \to \infty$:

$$I(X;Y) \;\sim\; \frac{\mathrm{Var}\big(\mathbb{E}(Z^K|X)\big)}{2 \cdot K! \cdot (\sigma_N^2 + \sigma_Z^2)^K} \tag{4}$$

where $\sigma_Z^2 = \mathrm{Var}(Z)$ denotes variance and $\mathrm{Var}\big(\mathbb{E}(Z^K|X)\big)$ denotes inter-class variance.

While establishing a rigorous foundation of the concept of HCI, this theorem fills the gap in proving Carlet et al.'s main

[1]Throughout we use natural logarithms so that informational quantities are expressed in *nats*.

result [2, Thm. 1], while also giving the correct asymptotic equivalent for $3 \le K \le 6$.

Higher protection orders $K > 6$ are overkill in practical implementations because the computational complexity increases at least quadratically with K. For this reason, they are currently unfeasible in embedded systems, and of theoretical interest only. We show that such high orders K actually involve additional cross-terms in the approximation of $I(X;Y)$ which make the asymptotic equivalent not as simple as in (4), but which can still be derived using the method of this paper.

Our strategy to find the the asymptotic equivalent of $I(X;Y)$ (and in particular to prove Theorem 1) is to rewrite the mutual information in terms of *non-Gaussianity* terms:

$$\begin{aligned} I(X;Y) &= h(Y^*|X) - h(Y \mid X) - \big(h(Y^*) - h(Y)\big) \\ &= D(Y\|Y^* \mid X) - D(Y\|Y^*) \end{aligned} \tag{5}$$

where Y^* is a Gaussian random variable independent of X (hence $h(Y^*|X) = h(Y^*)$) with the same first and second order moments as Y. We then go beyond the quadratic cumulant approximation of Cardoso [9, Eq. (29)] and investigate how Kullback-Leibler divergences $D(Y\|Y^*)$ and $D(Y\|Y^*|X)$ behave as σ increases, using a *Gram-Charlier* expansion [10] in terms of a sequence of "modified moments".

The remainder of the paper is organized as follows. Section II reviews a kind of Gram-Charlier expansion and derives the corresponding non-Gaussianity expansions. Section III gives the resulting expansions of mutual information and explains how to extend (4) to the problematic cases $K > 6$. Numerical validation is carried out in Section IV in a practical code-based masking scheme in AES with Hamming weight leakage model.

II. Cumulant Expansion of Non-Gaussianity

The *non-Gaussianity* of Y, defined as the Kullback-Leibler divergence $D(Y\|Y^*)$, is a nonnegative quantity which vanishes if and only if Y is Gaussian. For notational convenience write $\mu = \mu_Y = \mu_{Y^*}$ and $\sigma = \sigma_Y = \sigma_{Y^*}$. Because Y and Y^* share the same mean μ and variance σ^2, it is convenient to write their densities in the form $\frac{1}{\sigma} f(\frac{y-\mu}{\sigma})$ and $\frac{1}{\sigma} g(\frac{y-\mu}{\sigma})$, respectively, where f and g are standardized densities (in particular $g = \mathcal{N}(0,1)$). Since Kullback-Leibler divergence is invariant by invertible transformations, non-Gaussianity also writes

$$D(Y\|Y^*) = D\big(\tfrac{Y-\mu}{\sigma}\big\|\tfrac{Y^*-\mu}{\sigma}\big) = D(f\|g) = \int f \log \frac{f}{g}. \tag{6}$$

A. Density Expansion

As σ_N increases, $\sigma = \sqrt{\sigma_N^2 + \sigma_Z^2} \to \infty$ but high-order *cumulants* $\kappa_3, \kappa_4, \dots, \kappa_K$ of Y remain bounded. In fact for $k \ge 3$, $\kappa_k = \kappa_k(Y) = \kappa_k(Z) + \kappa_k(N) = \kappa_k(Z)$ are kept constant. On the other hand since Y^* is Gaussian, all its high-order cumulants are zero. This, as we show in the next Lemma, can be used to show that the Gaussian noise N dominates in $Y = Z + N$ so that f will approach the Gaussian g:

Lemma 1 (Gram-Charlier Expansion):

$$\frac{f(x)}{g(x)} = 1 + \sum_{k=3}^{K} \frac{\widetilde{m}_k}{k!\sigma^k} H_k(x) + o\big(\frac{1}{\sigma^K}\big) \tag{7}$$

where H_k is the k-th Hermite polynomial ($H_3(x) = x^3 - 3x$, $H_4(x) = x^4 - 6x^2 + 3$, $H_5(x) = x^5 - 10x^3 + 15x$, etc.) and where the "modified moments" $\widetilde{m}_k$ satisfy the recursion

$$\widetilde{m}_k = \kappa_k + \sum_{j=3}^{k-3} \binom{k-1}{j} \widetilde{m}_j \kappa_{k-j}. \tag{8}$$

The modified moments are computed exactly as the genuine moments m_k are computed from the cumulants κ_k using Smith's formula [11], except that κ_1 and κ_2 are absent. Thus $\widetilde{m}_1 = \widetilde{m}_2 = 0$, $\widetilde{m}_3 = \kappa_3$, $\widetilde{m}_4 = \kappa_4$, $\widetilde{m}_5 = \kappa_5$, $\widetilde{m}_6 = \kappa_6 + 10\kappa_3^2$, $\widetilde{m}_7 = \kappa_7 + 35\kappa_3\kappa_4$, etc. Notice that modified moments, like high-order cumulants, are bounded as $\sigma \to \infty$.

Proof: By definition of cumulants, the characteristic function $\phi_Y(t) = \mathbb{E}(e^{itY})$ of Y can be factorized as

$$\phi_Y(t) = \phi_{Y^*}(t)\, e^{\psi(t)} \tag{9}$$

where $\phi_{Y^*}(t) = e^{i\mu_Y t - \sigma_Y^2 t^2/2}$ is the characteristic function of $Y^* \sim \mathcal{N}(\mu, \sigma^2)$ and $\psi(t) = \sum_{k=3}^{K} \kappa_k \frac{(it)^k}{k!} + o(t^K)$. Taking the exponential we expand $\exp\psi(t) = 1 + \sum_{k=3}^{K} \widetilde{m}_k \frac{(it)^k}{k!} + o(t^K)$. The coefficients $\widetilde{m}_k$ can be found by Taylor's formula and Leibniz's rule: $i^k \widetilde{m}_k = (e^{\psi})^{(k)}(0) = (\psi' e^{\psi})^{(k-1)}(0) = \sum_j \binom{k-1}{j} (e^{\psi})^{(j)}(0) \psi^{(k-j)}(0)$ which simplifies to (8). Now (9) becomes

$$\phi_Y(t) = \Big(1 + \sum_{k=3}^{K} \frac{\widetilde{m}_k}{k!} (it)^k\Big) \phi_{Y^*}(t) + o(t^K)\phi_{Y^*}(t). \tag{10}$$

Taking the inverse Fourier transform gives the density of Y:

$$\frac{1}{\sigma} f\big(\frac{y-\mu}{\sigma}\big) = \Big(1 + \sum_{k=3}^{K} \frac{\widetilde{m}_k}{k!} (-\frac{\mathrm{d}}{\mathrm{d}y})^k\Big) \frac{1}{\sigma} g\big(\frac{y-\mu}{\sigma}\big) + R(y)$$

where we have used that multiplication by $(-it)$ in the Fourier domain (characteristic function) corresponds to differentiation. Now by the defining property of Hermite polynomials,

$$(-\frac{\mathrm{d}}{\mathrm{d}y})^k g\big(\frac{x-\mu}{\sigma}\big) = \frac{1}{\sigma^k} H_k\big(\frac{x-\mu}{\sigma}\big) \cdot g\big(\frac{x-\mu}{\sigma}\big).$$

The $o(t^K)$ term in (10) having at most polynomial growth at infinity, we can apply Watson's lemma [12, Chap. 2] for the remainder term $R(y)$, which gives $R(y) = o(\sigma^{-K})$ (with at most polynomial growth in y at infinity). Letting $x = \frac{x-\mu}{\sigma}$ and dividing by $g(x) > 0$ gives the announced expansion. ∎

Remark 1: Contrary to what seems to be a popular belief in the literature (see e.g., [9]), the coefficients multiplying the Hermite polynomials in the Gram-Charlier expansion (7) are not just cumulants κ_k, but "modified moments" $\widetilde{m}_k$, which differ from cumulants as soon as $k \ge 6$.

B. Divergence Expansion

Theorem 2: The expansion of divergence in power of $\frac{1}{\sigma}$ is of the form

$$D(f\|g) = \sum_{k=3}^{K} \frac{c_k}{2k!\sigma^{2k}} + o\Big(\frac{1}{\sigma^{2K}}\Big) \tag{11}$$

where $c_k = \widetilde{m}_k^2$ + other terms of the form $\alpha_m \widetilde{m}_{k_1}\widetilde{m}_{k_2}\cdots\widetilde{m}_{k_m}$ where $m \geq 3$ and $k_1 + k_2 + \cdots + k_m = 2k$.

Proof: Using (7) in the form $\frac{f}{g} = 1 + h$ where $h = \sum_{k=3}^{K} \frac{\widetilde{m}_k}{k!\sigma^k} H_k(x) + O(\sigma^{-K})$, we proceed to expand $D(f\|g) = \int g(1+h)\log(1+h)$ where $(1+h)\log(1+h) = h + \frac{h^2}{2} - \frac{h^3}{6} + \frac{h^4}{12} + \cdots + o(h^K)$. Substituting gives

$$D(f\|g) = \int gh + \frac{1}{2}\int gh^2 - \frac{1}{6}\int gh^3 + \frac{1}{12}\int gh^4 + \cdots + o\Big(\int gh^K\Big). \tag{12}$$

By the orthogonality property of Hermite polynomials

$$\int g H_k H_l = k!\, \delta_{kl}, \tag{13}$$

one has $\int gH_k = \int gH_kH_0 = 0$ $(k > 0)$ hence $\int gh = 0$. Moreover, by orthogonality, $\int gh^2 = \sum_{k=3}^{K} \big(\frac{\widetilde{m}_k}{k!\sigma^k}\big)^2 k! + o(\sigma^{-2K}) = \sum_{k=3}^{K} \frac{\widetilde{m}_k^2}{k!\sigma^{2k}} + o(\sigma^{-2K})$. Thus the quadratic part $\frac{1}{2}\int gh^2$ accounts for the $\frac{\widetilde{m}_k^2}{2k!\sigma^{2k}}$ terms $(k \geq 3)$.

The expansion of all higher-order terms $\int gh^m$ $(m \geq 3)$ involve terms of the form $\frac{\widetilde{m}_{k_1}\widetilde{m}_{k_2}\cdots\widetilde{m}_{k_m}}{\sigma^{k_1+k_2+\cdots+k_m}} \int gH_{k_1}H_{k_2}\cdots H_{k_m}$. Since each Hermite polynomial H_k has the same parity as its degree k, all such terms vanish if $k_1 + k_2 + \cdots + k_m$ is odd. Hence there remains only terms in $\frac{1}{\sigma^{2k}}$ as stated. ∎

Remark 2: The asymptotic $D(f\|g) = \frac{1}{2}\int gh^2 + o\big(\frac{1}{2}\int gh^2\big)$ was already proved in [13, Lemma 1].

C. First Few Terms in the Divergence Expansion

We can carry out the computations up to $K = 6$. The cubic and quartic terms can be evaluated at first orders using the special values [14, §6.8]: $\int gH_3^2H_4 = \frac{3!3!4!}{1!2!2!} = 216$, $\int gH_4^3 = \frac{4!4!4!}{2!2!2!} = 1728$, $\int gH_3H_4H_5 = \frac{5!4!3!}{3!2!1!} = 1440$, $\int gH_3^2H_6 = \frac{3!3!6!}{0!3!3!} = 720$, and $\int gH_3^4 = 3\frac{3!^4}{0!^43!^2} + 6\frac{3!^4}{0!^21!^22!^2} + \frac{3!^4}{1!^6} = 3348$, plus the fact that all terms in odd powers of σ are zero (since they involve integrals $\int gH_kH_lH_m = 0$ when $k + l + m$ is odd). After some calculation we obtain

$$\int gh^3 = \frac{648}{\sigma^{10}}\Big(\frac{\widetilde{m}_3}{3!}\Big)^2\frac{\widetilde{m}_4}{4!} + \frac{1728}{\sigma^{12}}\Big(\frac{\widetilde{m}_4}{4!}\Big)^3 + \frac{8640}{\sigma^{12}}\frac{\widetilde{m}_3}{3!}\frac{\widetilde{m}_4}{4!}\frac{\widetilde{m}_5}{5!} + \frac{2160}{\sigma^{12}}\Big(\frac{\widetilde{m}_3}{3!}\Big)^2\frac{\widetilde{m}_6}{6!} + O\Big(\frac{1}{\sigma^{14}}\Big)$$

and

$$\int gh^4 = \frac{3348}{\sigma^{12}}\Big(\frac{\widetilde{m}_3}{3!}\Big)^4 + O\Big(\frac{1}{\sigma^{14}}\Big).$$

Putting all pieces together and expressing modified moments in terms of cumulants, we obtain

$$\begin{aligned} D(f\|g) &= \frac{\widetilde{m}_3^2}{12\sigma^6} + \frac{\widetilde{m}_4^2}{48\sigma^8} + \frac{\widetilde{m}_5^2}{240\sigma^{10}} - \frac{\widetilde{m}_3^2\widetilde{m}_4}{8\sigma^{10}} + \frac{\widetilde{m}_6^2}{1440\sigma^{12}} \\ &\quad - \frac{\widetilde{m}_4^3}{48\sigma^{12}} - \frac{\widetilde{m}_3\widetilde{m}_4\widetilde{m}_5}{12\sigma^{12}} - \frac{\widetilde{m}_3^2\widetilde{m}_6}{72\sigma^{12}} + \frac{31\widetilde{m}_3^4}{144\sigma^{12}} + O\Big(\frac{1}{\sigma^{14}}\Big) \\ &= \frac{\kappa_3^2}{12\sigma^6} + \frac{\kappa_4^2}{48\sigma^8} - \frac{\kappa_3^2\kappa_4}{8\sigma^{10}} + \frac{\kappa_5^2}{240\sigma^{10}} \\ &\quad + \frac{7\kappa_3^4}{48\sigma^{12}} - \frac{\kappa_4^3}{48\sigma^{12}} - \frac{\kappa_3\kappa_4\kappa_5}{12\sigma^{12}} + \frac{\kappa_6^2}{1440\sigma^{12}} + O\Big(\frac{1}{\sigma^{14}}\Big). \end{aligned} \tag{14}$$

Remark 3: In order to check the validity of (14), we can recover a known expression in a different model (with entirely different assumptions and application). Instead of having $Y = Z + N$, suppose that $Y = Y_1 + Y_2 + \cdots + Y_n$ where the Y_i's are i.i.d. with mean μ, variance σ^2, and high-order cumulants κ_3, κ_4, ... The previous expansions can be used by replacing σ by $\sqrt{n}\sigma$, κ_k by $n\kappa_k$, and letting $n \to +\infty$. The *Gram-Charlier expansion*, re-ordered in powers of $\frac{1}{\sqrt{n}}$, becomes the *Edgeworth expansion*

$$\begin{aligned} \frac{f}{g} &= 1 + \frac{\kappa_3}{6\sigma^3\sqrt{n}}H_3 + \frac{\kappa_4}{24\sigma^4 n}H_4 + \frac{\kappa_3^2}{72\sigma^6 n}H_6 + \frac{\kappa_5}{120\sigma^5 n\sqrt{n}}H_5 \\ &\quad + \frac{\kappa_4\kappa_3}{144\sigma^7 n\sqrt{n}}H_7 + \frac{\kappa_3^3}{1296\sigma^9 n\sqrt{n}}H_9 + O\big(\frac{1}{n^2}\big). \end{aligned} \tag{15}$$

It is easily seen that all $O\big(\frac{1}{\sigma^{14}}\big)$ terms in (14) are then necessarily $O\big(\frac{1}{n^3}\big)$. Four terms out of the eight in (14) are also $O\big(\frac{1}{n^3}\big)$, and there remains

$$D(f\|g) = \frac{\kappa_3^2}{12n\sigma^6} + \frac{\kappa_4^2}{48n^2\sigma^8} - \frac{\kappa_3^2\kappa_4}{8n^2\sigma^{10}} + \frac{7\kappa_3^4}{48n^2\sigma^{12}} + O\Big(\frac{1}{n^3}\Big) \tag{16}$$

which is exactly the result of Comon [15, Thm 14] for his "negentropy" $D(f\|g) = h(g) - h(f) = \frac{1}{2}\log(2\pi e\sigma^2) - h(f)$.

Remark 4: The expansion (14) contrasts with Cardoso's small cumulant approximation to the Kullback-Leibler divergence [9, Eq. (41)] which in our setting would read

$$\frac{\kappa_3^2}{12\sigma^6} + \frac{\kappa_4^2}{48\sigma^8} + \frac{\kappa_5^2}{240\sigma^{10}} + \frac{\kappa_6^2}{1440\sigma^{12}} + \cdots$$

The difference with (14) is due to two facts: (a) as already noticed in Remark 1, the coefficients of the *Gram-Charlier expansion* (7) are not the cumulants κ_k for $k \geq 3$, but the modified moments $\widetilde{m}_k$, which differ from cumulants as soon as $k \geq 6$; (b) Cardoso's derivation only takes the quadratic approximation $\frac{1}{2}\int gh^2$ of divergence into account, ignoring other terms such as $\int gh^3 = O(\frac{1}{\sigma^{10}})$ which also contribute to the approximation.

While (a) and (b) have no effect for the first two terms $D(f\|g) = \frac{\kappa_3^2}{12\sigma^6} + \frac{\kappa_4^2}{48\sigma^8} + O\big(\frac{1}{\sigma^{10}}\big)$, both result in annoying higher-order cross-terms in the genuine expression (14) which do not appear in [9]. Because of this, derivations based on [9, Eq. (41)], particularly the main result of [2], become incorrect as soon as $O\big(\frac{1}{\sigma^{10}}\big)$ terms are considered.

Remark 5: Since $D(f\|g) = D(Y\|Y^*) = h(Y^*) - h(Y) = \frac{1}{2}\log(2\pi e\sigma^2) - h(Y)$ we have the following interesting expansion of (differential) entropy:

$$\begin{aligned} h(Y) &= \frac{1}{2}\log(2\pi e\sigma^2) - \frac{\kappa_3^2}{12\sigma^6} - \frac{\kappa_4^2}{48\sigma^8} + \frac{\kappa_3^2\kappa_4}{8\sigma^{10}} - \frac{\kappa_5^2}{240\sigma^{10}} \\ &\quad - \frac{7\kappa_3^4}{48\sigma^{12}} + \frac{\kappa_4^3}{48\sigma^{12}} + \frac{\kappa_3\kappa_4\kappa_5}{12\sigma^{12}} - \frac{\kappa_6^2}{1440\sigma^{12}} + O\Big(\frac{1}{\sigma^{14}}\Big). \end{aligned} \tag{17}$$

III. Cumulant Expansion of Mutual Information

A. Mutual Information Expansion

We now apply the expansion (14) to both terms $D(Y\|Y^*)$ and $D(Y\|Y^* \mid X)$ in (5). To simplify the derivation we assume that $(K-1)$th order protection (2) holds at least for the first two moments (hence $K \geq 3$): $\mu = \mu_Y = \mu_{Y|X=x}$ and $\sigma = \sigma_Y =$

$\sigma_{Y|X=x}$ for all x. We can, therefore, apply (14) for $D(Y\|Y^*)$ and $D(Y\|Y^* \mid X=x)$ for a given secret value x, and then take the expectation over X. Letting $\kappa_k(Z)=\kappa_k(Y)=\kappa_k$ and $\kappa_k(Z|X=x)=\kappa_k(Y|X=x)$ ($k\geq 3$) be the high-order cumulants of Z and $Z|X=x$, respectively, we readily obtain

$$\begin{aligned} I(X;Y) = {} & \frac{\mathbb{E}\,\kappa_3^2(Z|X)-\kappa_3^2(Z)}{12\sigma^6}+\frac{\mathbb{E}\,\kappa_4^2(Z|X)-\kappa_4^2(Z)}{48\sigma^8} \\ & -\frac{\mathbb{E}\big(\kappa_3^2(Z|X)\kappa_4(Z|X)\big)-\kappa_3^2(Z)\kappa_4(Z)}{8\sigma^{10}}+\frac{\mathbb{E}\,\kappa_5^2(Z|X)-\kappa_5^2(Z)}{240\sigma^{10}} \\ & +\frac{7\big(\mathbb{E}\,\kappa_3^4(Z|X)-\kappa_3^4(Z)\big)}{48\sigma^{12}}-\frac{\mathbb{E}\,\kappa_4^3(Z|X)-\kappa_4^3(Z)}{48\sigma^{12}} \\ & -\frac{\mathbb{E}\big(\kappa_3(Z|X)\kappa_4(Z|X)\kappa_5(Z|X)\big)-\kappa_3(Z)\kappa_4(Z)\kappa_5(Z)}{12\sigma^{12}} \\ & +\frac{\mathbb{E}\,\kappa_6^2(Z|X)-\kappa_6^2(Z)}{1440\sigma^{12}}+O\Big(\frac{1}{\sigma^{14}}\Big). \end{aligned} \tag{18}$$

Remark 6: This contrasts with the high-order expansion of mutual information in [2, Eq. (6)] which reads

$$\begin{aligned} I(X;Y) = {} & \frac{\mathbb{E}\big(\kappa_3(Z|X)-\kappa_3(Z)\big)^2}{12\sigma^6}+\frac{\mathbb{E}\big(\kappa_4(Z|X)-\kappa_4(Z)\big)^2}{48\sigma^8} \\ & +\frac{\mathbb{E}\big(\kappa_5(Z|X)-\kappa_5(Z)\big)^2}{240\sigma^{10}}+\frac{\mathbb{E}\big(\kappa_6(Z|X)-\kappa_6(Z)\big)^2}{1440\sigma^{12}}+O\Big(\frac{1}{\sigma^{14}}\Big). \end{aligned}$$

The difference with (18) is due to three facts: (a) and (b) leading to annoying cross-terms in the non-Gaussianity expansion, as explained in Remark 4; (c) terms of the form $\mathbb{E}\,\kappa_k^2(Z|X)-\kappa_k^2(Z)$ can be written as variances

$$\mathbb{E}\,\kappa_k^2(Z|X)-\kappa_k^2(Z)=\mathbb{E}\big(\kappa_k(Z|X)-\kappa_k(Z)\big)^2 \tag{19}$$

only under the condition that $\kappa_k(Z)=\mathbb{E}\,\kappa_k(Z|X)$. This condition indeed holds for $k=3,4,5$ under the above assumptions because of the well-known expressions of κ_3, κ_4, and κ_5 in terms of moments m_1, m_2, m_3, m_4, m_5, where the quantities $m_1(Z|X=x)=\mathbb{E}(Z|X=x)=\mathbb{E}(Z)=m_1(Z)$ and $m_2(Z|X=x)=\mathbb{E}(Z^2|X=x)=\mathbb{E}(Z^2)=m_2(Z)$ do not depend on $X=x$ and where $m_k(Z)=\mathbb{E}(Z^k)=\mathbb{E}\,\mathbb{E}(Z^k|X)=\mathbb{E}\,m_k(Z|X)$. However, the condition $\kappa_k(Z)=\mathbb{E}\,\kappa_k(Z|X)$ is no longer satisfied for $k=6$ because of the $-10m_3^2$ term in the expression of $\kappa_6=m_6-6m_5m_1-15m_4m_2+30m_4m_1^2-10m_3^2+120m_3m_2m_1-120m_3m_1^3+30m_2^3-270m_2^2m_1^2+360m_2m_1^4-120m_1^6$.

While (a), (b), and (c) have no effect for the first two terms $I(X;Y)=\frac{\mathbb{E}(\kappa_3(Z|X)-\kappa_3(Z))^2}{12\sigma^6}+\frac{\mathbb{E}(\kappa_4(Z|X)-\kappa_4(Z))^2}{48\sigma^8}+O\big(\frac{1}{\sigma^{10}}\big)$, they result in annoying higher-order cross-terms in the genuine expression (18) which do not appear in [2].

Proof of the main Theorem 1: The HCI condition (2) states that $m_k(Z|X)=m_k(Z)$ a.s. for $k<K$. Now from the well-known formulas expressing cumulants in terms of moments, one has $\kappa_k(Z|X)=m_k(Z|X)+$ lower-order terms in $m_1(Z|X)=m_1(Z),\ldots,m_{k-1}(Z|X)=m_{k-1}(Z)$. It follows that $\kappa_k(Z|X)=\kappa_k(Z)$ a.s. for $k<K$ while for $k=K$, we have $\kappa_K(Z|X)-m_K(Z|X)=\kappa_K(Z)-m_K(Z)$. Thus, $\kappa_K(Z|X)-\kappa_K(Z)=m_K(Z|X)-m_K(Z)$ and in particular $\mathbb{E}\,\kappa_K(Z|X)-\kappa_K(Z)=\mathbb{E}\,m_K(Z|X)-m_K(Z)=\mathbb{E}\,\mathbb{E}(Z^K|X)-\mathbb{E}(Z^K)=0$. Therefore, we can write $\mathbb{E}\,\kappa_K^2(Z|X)-\kappa_K^2(Z)=\mathrm{Var}\big(\kappa_K(Z|X)\big)=E\big(\kappa_K(Z|X)-\kappa_K(Z)\big)^2=\mathbb{E}\big(m_K(Z|X)-m_K(Z)\big)^2=\mathrm{Var}\big(m_K(Z|X)\big)=\mathrm{Var}\big(\mathbb{E}(Z^K|X)\big)$ which is nonzero since $\mathbb{E}(Z^K|X)$ is not constant a.s.

By examination of (18) when $K\leq 6$, it is easily seen that

$$\begin{aligned} I(X;Y) &= \frac{\mathbb{E}\,\kappa_K^2(Z|X)-\kappa_K^2(Z)}{2\cdot K!\cdot\sigma^{2K}}+o\Big(\frac{1}{\sigma^{2K}}\Big) \\ &= \frac{\mathrm{Var}\big(\mathbb{E}(Z^K|X)\big)}{2\cdot K!\cdot\sigma^{2K}}+o\Big(\frac{1}{\sigma^{2K}}\Big) \end{aligned} \tag{20}$$

where $\sigma^2=\sigma_Y^2=\sigma_N^2+\sigma_Z^2$. ∎

Remark 7: What makes the proof of Theorem 1 work in that in (14), all terms in $\frac{1}{\sigma^{2k}}$ ($k=3,4,5,6$) involve only cumulants of order $\leq k$.

This property, however, does not generalize to higher orders. In fact by Theorem 2, there is at least one additional term in $\frac{1}{\sigma^{14}}$ in the form $\alpha_3\widetilde{m}_8\widetilde{m}_3^2$ (since $8+3+3=14$) which will contribute to a term $\frac{\alpha_3\kappa_3^2(Z)(\mathbb{E}\,\kappa_8(Z|X)-\kappa_8(Z))}{\sigma^{14}}$ in addition of the $\frac{\mathbb{E}\,\kappa_7^2(Z|X)-\kappa_7^2(Z)}{10080\cdot\sigma^{14}}$ of (20). Assuming $\kappa_3(Z)\neq 0$, we still have $I(X;Y)=O(\sigma^{-2K})$ for $K=7$ but with a different asymptotic equivalent.

Furthermore, again assuming $\kappa_3(Z)\neq 0$, for $K=8$ the term $\frac{\alpha_3\kappa_3^2(Z)(\mathbb{E}\,\kappa_8(Z|X)-\kappa_8(Z))}{\sigma^{14}}$ still contributes to mutual information so that in the case it is no longer true that $I(X;Y)=O(\sigma^{-2K})$. We still have $I(X;Y)=O(\sigma^{-14})$ instead of $I(X;Y)=O(\sigma^{-16})$.

Incidentally, when $\kappa_3(Z)=0$ (e.g., when K's distribution is symmetric) the above mentionned annoying cross-terms disappear and Theorem 1 becomes valid not only for $K=3,4,5,6$ but also for $K=7$ and 8.

In general for higher orders, the terms $\alpha_m\widetilde{m}_{k_1}\widetilde{m}_{k_2}\cdots\widetilde{m}_{k_m}$ ($m\geq 3$, $k_i\geq 3$, $k_1+k_2+\cdots+k_m=2k$) of Theorem 2 will contribute to $I(X;Y)$ only when all k_i are necessarily $<K$. Since the maximum possible k_i is $2k-6$ (for $m=3$, the other two k_i's being equal to 3), we must have at least (2) satisfied at order $2k-5$ to ensure that $I(X;Y)=O(\sigma^{-2(k+1)})$. Therefore, for $K\geq 7$, $I(X;Y)=O(\sigma^{-2K})$ requires an HCI at least $2K-7$.

In practice, extremely high-order protection ($K=6,7,\ldots$) is unthinkable for all implementations. Hence Theorem 1 will apply to all cases of interest. In the following section we illustrate this using a code-based masked implementation.

IV. Numerical Simulations

Consider an advanced encryption standard (AES [16]) block cipher, which takes in input a plaintext of 16 bytes, and outputs a ciphertext of the same size. The attacker is able to monitor inputs and outputs, but does not know the secret key. In such a cryptographic algorithm, it is practically impossible to deduce the secret from inputs and outputs: all the security relies on the secrecy of the key, in keeping with Kerckhoffs's principle [17] (a.k.a. Shannon's maxim [18]).

Side-channel attacks consist in measuring power consumption [19] or electromagnetic (EM) waves [20] produced during the execution of the AES algorithm. As shown in Fig. 1, the

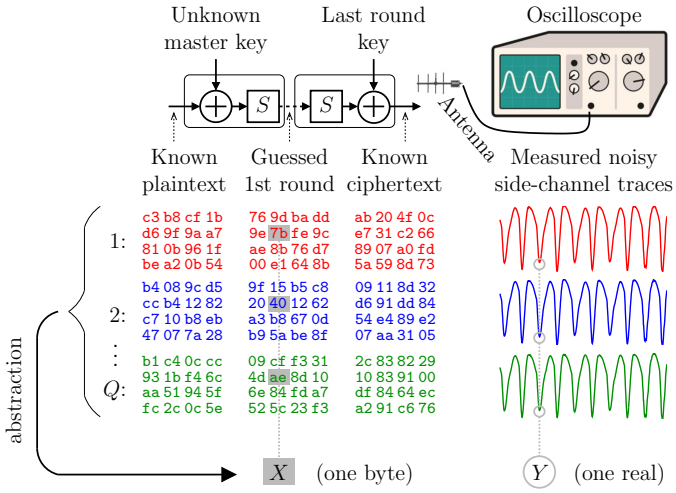

Fig. 1. Public information (plaintext and ciphertext) available to an attacker and first-round key-dependent intermediate (discrete) value, put in front of corresponding side-channel execution traces (analog) observed by the attacker.

attacker measures waveforms corresponding to the side-channel emanation of the AES computation. Such side information (repeatedly collected many times) is correlated to the secret key, and the attacker tries to exploit it in order to validate assumptions on small chunks of the key. In Fig. 1, the reference 128-bit key is `0x2b7e151628aed2a6abf7158809cf4f3c` (taken as an example in the NIST specification [16, Appendix A]) and the guessed values are that of the first round of AES, which consists in the application of AES SubBytes on the plaintext XORed with the key. The measured waveforms are time series of power or EM emanations, which depend on the plaintext (or equivalently, on the ciphertext, since encryption is symmetric). Some specific samples depend on small chunks of the plaintext/ciphertext and of the secret key, and are used by the attacker to assess hypotheses on the key.

We consider a practical case where the block cipher algorithm is protected by a masking scheme [21]. Firstly, to demonstrate Theorem 1, we target a dth-order masking scheme [8] in which the key chunk $X \in \mathbb{F}_q$ is encoded as $\left(X \oplus (\sum_{i=1}^{d} C_i \otimes M_i),\, M_1, \dots\, M_d\right)$, using an independent uniformly distributed mask $\mathbf{M} = (M_1, \dots, M_d) \in \mathbb{F}_q^d$ and a nonzero constant $\mathbf{C} \in \mathbb{F}_q^d$. Here $\oplus$ and $\otimes$ denote the addition and multiplication, respectively, in a finite field $\mathbb{F}_q = \mathbb{F}_{16}$ [2].

Hamming weight model with a Gaussian noise is a commonly used model in side-channel analysis. For instance, considering $d = 2$, the leaked sensitive variable is modeled as a $Z = w_H\left(X \oplus (C_1 \otimes M_1)\right) + w_H(M_1)$ where $w_H(\cdot)$ denotes the Hamming weight (number of nonzero bits) and $Y = Z + N$ where $N \sim \mathcal{N}(0, \sigma_N^2)$. Similarly, the Hamming weights of each share are summed together in the five-share case. As demonstrated in [8] and shown in Table I, both HCI K and $\mathrm{Var}\big(\mathbb{E}(Z^K|X)\big)$ change with different choices of $\mathbf{C}$. We can, therefore, validate Theorem 1 in multiple cases.

[2]The irreducible polynomial we used in this paper is $\alpha^4 + \alpha + 1$ for $\mathbb{F}_{16}$.

In fact, since X is taken as uniformly distributed, Hamming weights' distribution is of a binomial type, which is symmetric, hence $\kappa_3(Z) = 0$. Thus in our setting Theorem 1 should be valid at least up to $K = 8$.

TABLE I
DIFFERENT K AND $V_k = \mathrm{Var}\big(\mathbb{E}(Z^k|X)\big)$ $(k = 1, \dots, K)$ BY USING DIFFERENT $\mathbf{C}$ (IN DECIMAL REPRESENTATION).

$X, C \in \mathbb{F}_{16}$			$X \in \mathbb{F}_{16}, \mathbf{C} \in \mathbb{F}_{16}^3$			$X \in \mathbb{F}_{16}, \mathbf{C} \in \mathbb{F}_{16}^4$			
C	1	3	$\mathbf{C}$	(2,1,1)	(6,1,1)	$\mathbf{C}$	(6,1,2,4)	(6,2,4,5)	(6,7,2,2)
K	2	3	K	4	5	K	6	7	8
						$V_1 \sim V_5$	0	0	0
V_1	0	0	$V1 \sim V_3$	0	0	V_6	253.13	0	0
V_2	1	0	V_4	6.75	0	V_7	-	3100.78	0
V_3	-	2.25	V_5	-	42.19	V_8	-	-	74418.75

The numerical results of mutual information are shown in Fig. 2 in log-log scale, where slope $-K$ indicate $I(X;Y) \sim \mathrm{Cst} \cdot \sigma_N^{-2K}$. We observe the first nonzero order expansion of mutual information dominates when the noise level is high enough (e.g., when $\sigma_N^2 \geq 10$). Overall Theorem 1 gives an accurate approximation of mutual information in all practical cases.

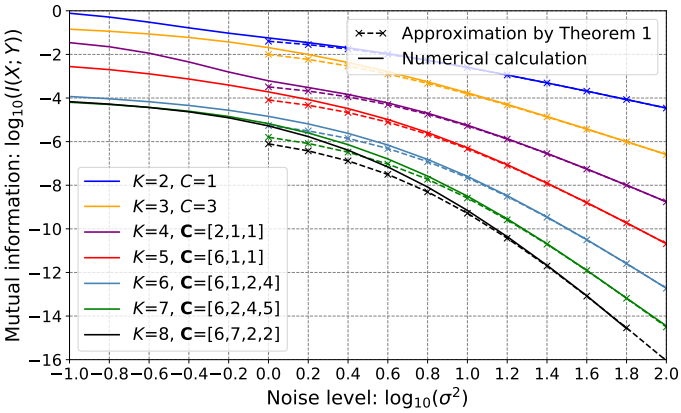

Fig. 2. Numerical validation of Theorem 1 by taking different C (or $\mathbf{C}$) therefore different $K \in \{3, 4, 5, 6, 7, 8\}$ and $\mathrm{Var}\big(\mathbb{E}(Z^K|X)\big)$.

V. CONCLUSION

In this paper, we presented a cumulant-based expansion of Kullback-Leibler divergence and mutual information with application to side-channel analysis. We fixed the mathematical issue that existed in the literature and proposed a rigorous proof for the main result in [2] in most cases of interest.

REFERENCES

[1] A. Moradi and F. Standaert, "Moments-Correlating DPA," in *Proceedings of the ACM Workshop on Theory of Implementation Security, TIS@CCS 2016 Vienna, Austria, October, 2016*, B. Bilgin, S. Nikova, and V. Rijmen, Eds. ACM, 2016, pp. 5–15. [Online]. Available: https://doi.org/10.1145/2996366.2996369

[2] C. Carlet, J.-L. Danger, S. Guilley, H. Maghrebi, and E. Prouff, "Achieving side-channel high-order correlation immunity with leakage squeezing," *J. Cryptographic Engineering*, vol. 4, no. 2, pp. 107–121, 2014.

[3] H. Eldib, C. Wang, M. Taha, and P. Schaumont, "QMS: Evaluating the Side-Channel Resistance of Masked Software from Source Code," in *Proceedings of the The 51st Annual Design Automation Conference on Design Automation Conference*, ser. DAC '14. New York, NY, USA: ACM, 2014, pp. 209:1–209:6. [Online]. Available: http://doi.acm.org/10.1145/2593069.2593193

[4] L. Zhang, A. A. Ding, Y. Fei, and P. Luo, "A Unified Metric for Quantifying Information Leakage of Cryptographic Devices Under Power Analysis Attacks," in *Advances in Cryptology - ASIACRYPT 2015 - 21st International Conference on the Theory and Application of Cryptology and Information Security, Auckland, New Zealand, November 29 - December 3, 2015, Proceedings, Part II*, ser. Lecture Notes in Computer Science, T. Iwata and J. H. Cheon, Eds., vol. 9453. Springer, 2015, pp. 338–360. [Online]. Available: https://doi.org/10.1007/978-3-662-48800-3_14

[5] C. Carlet, J. Danger, S. Guilley, and H. Maghrebi, "Leakage squeezing: Optimal implementation and security evaluation," *J. Mathematical Cryptology*, vol. 8, no. 3, pp. 249–295, 2014. [Online]. Available: http://dx.doi.org/10.1515/jmc-2012-0018

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

[7] V. Grosso and F. Standaert, "Masking Proofs Are Tight and How to Exploit it in Security Evaluations," in *Advances in Cryptology - EUROCRYPT 2018 - 37th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Tel Aviv, Israel, April 29 - May 3, 2018 Proceedings, Part II*, ser. Lecture Notes in Computer Science, J. B. Nielsen and V. Rijmen, Eds., vol. 10821. Springer, 2018, pp. 385–412. [Online]. Available: https://doi.org/10.1007/978-3-319-78375-8_13

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

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

[9] J. Cardoso, "Dependence, correlation and Gaussianity in independent component analysis," *J. Mach. Learn. Res.*, vol. 4, pp. 1177–1203, 2003. [Online]. Available: http://jmlr.org/papers/v4/cardoso03a.html

[10] A. Hald, "The early history of the cumulants and the Gram-Charlier series," *International Statistical Review*, vol. 68, no. 2, pp. 137–153, 2000.

[11] P. J. Smith, "A recursive formulation of the old problem of obtaining moments from cumulants and vice versa," *The American Statistician*, vol. 49, no. 2, pp. 217–218, 1995.

[12] P. D. Miller, *Applied asymptotic analysis*. American Mathematical Soc., 2006, vol. 75.

[13] E. Abbe and L. Zheng, "A coordinate system for Gaussian networks," *IEEE Trans. Inf. Theory*, vol. 58, no. 2, pp. 721–733, 2012. [Online]. Available: https://doi.org/10.1109/TIT.2011.2169536

[14] G. E. Andrews, R. Askey, and R. Roy, *Special Functions*. Cambridge University Press, 1999.

[15] P. Comon, "Independent component analysis, a new concept?" *Signal Process.*, vol. 36, no. 3, pp. 287–314, 1994. [Online]. Available: https://doi.org/10.1016/0165-1684(94)90029-9

[16] NIST/ITL/CSD, "Advanced Encryption Standard (AES). FIPS PUB 197," Nov 2001, http://nvlpubs.nist.gov/nistpubs/FIPS/NIST.FIPS.197.pdf (also ISO/IEC 18033-3:2010).

[17] A. Kerckhoffs, "La cryptographie militaire (2)," *Journal des sciences militaires*, vol. 9, pp. 161–191, February 1883, http://en.wikipedia.org/wiki/Kerckhoffs_law.

[18] C. E. Shannon, "Communication theory of secrecy systems," *Bell System Technical Journal*, vol. 28, pp. 656–715, octobre 1949. [Online]. Available: https://doi.org/10.1002%2Fj.1538-7305.1949.tb00928.x

[19] P. C. Kocher, J. Jaffe, and B. Jun, "Differential Power Analysis," in *Proceedings of CRYPTO'99*, ser. LNCS, vol. 1666. Springer-Verlag, 1999, pp. 388–397.

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