Formal Security Proofs via Doeblin Coefficients:
Optimal Side-channel Factorization from Noisy Leakage to Random Probing

Julien Béguinot[1], Wei Cheng[2,1], Sylvain Guilley[2,1], and Olivier Rioul[1]

[1] LTCI, Télécom Paris, Institut Polytechnique de Paris, France
[2] Secure-IC S.A.S., France

Abstract. Masking is one of the most popular countermeasures to side-channel attacks, because it can offer provable security. However, depending on the adversary's model, useful security guarantees can be hard to provide. At first, masking has been shown secure against *t-threshold probing adversaries* by Ishai *et al.* at Crypto'03. It has then been shown secure in the more generic *random probing model* by Duc *et al.* at Eurocrypt'14. Prouff and Rivain have introduced the *noisy leakage model* to capture more realistic leakage at Eurocrypt'13. Reduction from noisy leakage to random probing has been introduced by Duc *et al.* at Eurocrypt'14, and security guarantees were improved for both models by Prest *et al.* at Crypto'19, Duc *et al.* in Eurocrypt'15/J. Cryptol'19, and Masure and Standaert at Crypto'23. Unfortunately, as it turns out, we found that previous proofs in either random probing or noisy leakage models are flawed, and such flaws do not appear easy to fix.
In this work, we show that the *Doeblin coefficient* allows one to overcome these flaws. In fact, it yields optimal reductions from noisy leakage to random probing, thereby providing a correct and usable metric to properly ground security proofs. This shows the inherent inevitable cost of a reduction from the noisy leakages to the random probing model. We show that it can also be used to derive *direct* formal security proofs using the subsequence decomposition of Prouff and Rivain.

1 Introduction

1.1 Context

All cryptographic implementations leak some side information about the sensitive variables they manipulate through the so-called side-channels. These leakages can be of different natures: Timing [DKL+98], power consumption [KJJ99, KGG+18], electromagnetic [GMO01, AARR02]. The corresponding side-channel attacks can be very harmful if there is no countermeasure or if the countermeasure is not carefully implemented. One may classify countermeasures into three categories, that can be jointly implemented:

– **Key refreshing** regularly replaces the secret key by a new one, e.g., each time a given number of operations has been performed [AB00, UHIM24].

– **Hiding** equalizes the leakage, either by removing the variations caused by computation, or by creating artificial noise in the circuit. It can be achieved by physical means such as *shielding* [AARR02], *noise makers* [LBB19], *dual rail technology* [MSS09], *balancing* or adding *dummy operations* [LH20].
– **Noise amplification** leverages existing noise from the given side-channels to make their measurements harder. It can be achieved using *wire shuffling* [ISW03, CS21], *operation shuffling* [VCMKS12] or *masking* [ISW03].

Masking is one of the most effective countermeasures known so far. It is especially relevant because of its provable security [ISW03, RP10a, PR13, DDF14, DFS15, BBD+16, DFS19, BCG+23, MS23b]. Previously published security proofs for masking fall into two classes:

– **Simulation paradigm (indirect approach):** A black-box adversary is modeled by an algorithm that only accesses the public information, which corresponds to the usual cryptanalysis. By contrast, the side-channel adversary is given both public and side-channel information. If there always exists a black-box adversary whose output is indistinguishable from the side-channel adversary's output, then the implementation is considered secure.
– **Information Theoretic Paradigm (direct approach):** The implementation is considered secure if the mutual information (or some other informational metric) between the side-channel leakage and the corresponding sensitive variable is negligible, given the available public information. Equivalently, the required number of side-channel queries required to achieve a given success rate is prohibitively large for any attack.

Those two approaches have been respectively termed *indirect* and *direct* by Prest *et al.* [PGMP19]. Both approaches have their pros and cons. On the one hand, the security proof based on information theory is conceptually simpler and provides more realistic security parameters. On the other hand, the simulation-based approach is very generic and can be applied to a whole cipher at once. Indeed, as remarked in [DDF14, Footnote 4] a few pairs of plaintext/ciphertext completely reveal the key of an AES *in the information theoretic sense*. Hence block ciphers such as AES are not secure in this sense, a fortiori in the presence of side-channel information. The security of AES relies on a one-way computational assumption which cannot be taken into account in the information theoretic paradigm[3].

To prove the security of a cryptographic implementation in any of the two paradigms above, it is necessary to define the side-channel adversary's model. Some restrictions should be imposed on the adversary, since if she/he is allowed to observe all variables manipulated in the circuit, then the implementation would be trivially broken. Micali and Reyzin introduced *physically observable cryptography* [MR04], in which only computation can leak information. *Leakage resilient cryptography* [DP08, KR19] also considers memory leakage models. In

[3] Unless, for a given round in a divide-and-conquer attack, the round's output is assumed not disclosed to the attacker because it is hidden by the one-way computational assumption in the subsequent rounds.

this context, the simplest model is the *t-threshold probing model* [ISW03] in which the adversary is only allowed to probe the values of t wires within the circuit. In the more elaborated *region probing model* [GPRV22], the circuit is divided into small regions and the adversary can probe t values in each region. A more realistic model is the *random probing model* [DDF14] where the side-channels correspond to erasure channels. The most generic type of model is the *noisy leakage model* from Prouff and Rivain [PR13] where the *$\mathcal{D}$-noisy adversary* has minimal distortion $\mathcal{D}$ between the channel input and output. In this list of models, the security proof is all the more hard to establish as the model is more complex and realistic.

1.2 Contributions

In this paper, we aim at grounding security proofs of side-channel analysis countermeasures on solid mathematical foundations. This work has the following contributions.

1. We carry out a systematic mathematical study of the *complementary Doeblin coefficient* (CDC). This coefficient was originally used to study Markov chains [Doe37] and appeared in the side-channel literature as the value of ϵ in [DDF19, Eqn. 9, Proof of Lemma 4]. We show that the CDC provides the optimal reduction from a noisy leakage model to the random probing model. Since the reduction is optimal it exhibits the unavoidable loss to pay to use a security proof based on the random probing model.
2. Bounds on the success rate (SR) and guessing entropy (GE) of a side-channel attack are derived using the CDC. Such bounds holds with equality for erasure side-channels, scale well with the number of channel queries, can be applied to adaptive adversaries and are amenable to practical evaluation, e.g., in Gaussian additive noise (Hamming weight or least significant bit model, etc.).
3. A new direct security proof is presented based on the CDC and on the Prouff-Rivain subsequence decomposition. As a supplementary material, some flaws in previous direct security proofs for masking in the noisy leakage model are identified (this does not necessarily mean that the corresponding results cannot hold) and some patches or bypasses are presented. Namely, this concerns Lemma 4 (hence Theorem 3) in [PR13], Lemma 8 (hence Theorem 6, Corollary 4) in [PGMP19], and Theorem 5 in [MS23b]. Details are in Appendix E.
4. A new methodology providing indirect security proofs is presented, based on the optimal CDC reduction from the noisy leakage model to the random probing model. As a supplementary material, minor errors are also corrected from the original proof on a reduction to the t-threshold probing model of Lemma 4 in [DDF14]. As a result, the bounds derived in [DDF14, DFS15, DFS19, PGMP19] that leveraged this Lemma can be improved significantly. Again details can be found in Appendix E.

1.3 Outline

The remainder of this paper is structured as follows. Preliminaries and mathematical results on channels, leakage measures (including CDC) and figures of merit are presented in Section 2. The key property of the CDC and the resulting bounds on figures of merit are presented in Section 3, along with some theoretical expressions for concrete evaluation. Direct security proofs leveraging CDC based on Prouff-Rivain subsequence decompositions are provided in Section 4. A new methodology for the derivation of indirect proofs is shown in Section 5. Section 6 concludes.

This paper also contains supplementary material in the Appendix, which are not necessary in order to follow the main arguments of the article. Appendix A provides a formal channel definition, relating it to the notion of random function. Appendices B, C and D contain technical proofs and results. Appendix E provides a comprehensive list of flaws that we identified in the state-of-the-art papers, along with patches we devised. It should be noted that we have obtained confirmation from the various authors of the papers in which we have detected flaws in the security proof in the noisy leakage model.

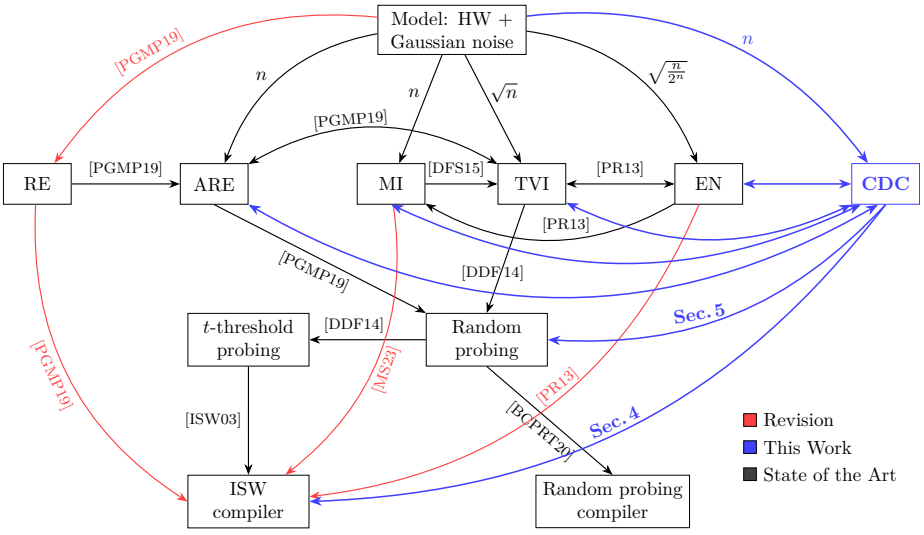


Fig. 1: Overview of formal security proofs, organized in four levels. Novelty is in blue, revisions of the state of the art is in red, and the state of the art is in black.

Figure 1 updates Fig. 1 from [PGMP19] where the black arrows indicate the state of the art, the red arrows are flaws that we identified and revised and the blue arrows correspond to our new derivations using the CDC. The figure is organized in four levels: The top level corresponds to the Hamming weight leakage model. The second level contains the main leakage measures,

corresponding to different noisy leakage adversaries. Each arrow label from the first to the second level indicates how the leakage measure scales with respect to the number n of bits, while each arrow label between two leakage models indicates the appropriate reference of the reduction from one model to another. The third layer contains the various adversarial models based on probing, and the bottom level contains the secure compilers that generate secure circuits against a given adversarial model. Each arrow label from the second or third to the fourth level indicates the appropriate reference of the corresponding security proof. The comparison with the other informational leakage measure is elaborated in Tab. 1 below.

1.4 Detailed Technical Overview

Theorem 1 provides the optimal factorization of a given channel into an erasure channel followed by another channel. As a consequence, any channel can be seen as a stochastically degraded erasure channel with the largest possible erasure probability. In particular, this implies that the optimal reduction of a side-channel adversary from the noisy leakage model (arbitrary channel) to the random probing model (erasure channel) is measured by the CDC.

The CDC equally applies to multivariate leakages. Lemma 6 shows that the CDC with multiple traces is bounded in terms of the CDC with one trace, even for an adaptive "chosen channel" adversary.

The main figures of merit (Definition 13) satisfy the data processing inequality (DPI) recalled in Lemma 8. In other words, the stochastically degraded adversary can only perform worse than the non-degraded one. As a consequence, the performance of any side-channel attacker can be bounded in terms of CDC as shown in Proposition 1. Intuitively, these bounds result from averaging two extreme cases: Either the leakage value is an erasure symbol and the figure of merit is that of a blind guess, or it is probed and the figure of merit is that of a disclosed value. This applies even for computationally bounded adversaries, allowing one to avoid complex simulation arguments [DDF14].

Lemma 7 shows that the CDC satisfies a strengthened data processing inequality which is useful in the derivation of the security proof.

Theorem 2 gives a direct security bound for ISW masked computations of an AES following the subsequence decomposition of Prouff and Rivain [PR13]. To achieve this, we derive a security lemma for each type of subsequence. For type 1 and 2 subsequences, we prove an analog of Mrs Gerber's Lemma (MGL) [WZ73] in Lemma 10 which shows that the CDC between the leakage and a masked value is upper bounded by the product of the CDCs share by share. This is expected since a sensitive value is probed if and only if all shares are probed. For type 3 subsequences, Lemma 11 provides a security bound for the cross-wise terms, in terms of the domination polynomial of the rook graph of Definition 17. The idea is that a value is probed if and only if all shares are probed at least once through cross-wise terms of one of the two shared inputs.

Theorem 4 explains how any formal security proof in the random probing model can be lifted to the noisy leakage model using the CDC. This is illustrated for the security proof of Duc et al. [DDF14].

The descriptions of the flaws appearing in previous derivations is deferred to Appendix E. To summarize, the derivations of [PGMP19] and [PR13] are invalidated because of an incorrect chain rule on probabilities. The flaw in [MS23b] is due to the fact that the bound appearing in the MGL for mutual information is separately but not jointly convex in the variables. The CDC overcomes these difficulties by allowing a direct bound which is then degraded in terms of the corresponding leakage measure through Lemma 9.

2 Mathematical Framework

In this Section, we present the mathematical framework of side-channel analysis that we use in our analysis. The notations are given in Subsection 2.1. The formal definition of a side-channel is given in Subsection 2.2. The main informational leakages measures are recalled in Subsection 2.3. Some useful properties of the complementary Doeblin coefficient (CDC) such as an adaptive single letterization and strengthened data processing inequality (DPI) are provided in Subsection 2.4. The model for a side-channel attack is described in Subsection 2.5. Finally, the figures of merit to evaluate the advantage of a side-channel adversary are introduced in Subsection 2.6.

2.1 Notations

Random variables are denoted by uppercase letters like X, Y. The corresponding set of values taken by the random variables are denoted by the corresponding calligraphic letters like $\mathcal{X}, \mathcal{Y}$. Lowercase letters denote values taken by random variables, e.g., $x \in \mathcal{X}$, $y \in \mathcal{Y}$. Bold letters denote random vectors $\mathbf{X}$ taking vector values $\mathbf{x}$. The probability distribution of X is denoted P_X; we write $X \sim P_X$.

- When X is discrete, taking values in a discrete set $\mathcal{X}$ of cardinality $|\mathcal{X}|$, its probability mass function (pmf) is noted $p_X(x) = \mathbb{P}(X = x)$;
- When X is continuous, its probability density function (pdf) is also noted $p_X(x)$ where $\mathrm{d}P_X(x) = p_X(x)\,\mathrm{d}x$.

We use the unified notation $\fint$ which is a sum in the discrete case and an integral in the continuous case. Therefore, we write $\mathbb{P}(X \in E) = \fint_{x \in E} p_X(x)$. Expectation is denoted by $\mathbb{E}_X[\cdot]$. The p-norm is noted $\|\cdot\|_p$.

- The uniform distribution on a set $\mathcal{X}$ is denoted by $\mathcal{U}(\mathcal{X})$;
- $\mathcal{B}(p)$ denotes the Bernoulli distribution with parameter p and $\mathcal{B}(n, p)$ denotes the Binomial distribution with parameters n, p. The survival function of $B \sim \mathcal{B}(n, p)$ is noted $Q_B(x, n, p) \triangleq \mathbb{P}(B > x) = \mathbb{P}(B \geqslant x + 1)$.
- $\mathcal{N}(\mu, \sigma^2)$ denotes the Gaussian distribution of mean μ and variance σ^2. The survival function of the standard Gaussian $\mathcal{N}(0, 1)$ is denoted by Q.

The joint probability distribution of (X, Y) is noted $P_{X,Y}$ with pmf or pdf $p_{X,Y}$. When X and Y are independent, $P_{X,Y} = P_X P_Y$, that is, $p_{X,Y}(x, y) =$

$p_X(x)p_Y(y)$. The conditional probability distribution of Y given X is denoted $P_{Y|X}$ where $p_{Y|X}(y|x) = \frac{p_{X,Y}(x,y)}{p_X(x)}$.

Finally, the positive (resp. negative) part of x is $x^+ \triangleq \max(0, x)$ (resp. $x^- \triangleq \max(-x, 0)$), and the complementary of $x \in [0, 1]$ is $\overline{x} \triangleq 1 - x$.

2.2 Side-Channels

A *random transformation* with input X and output Y is defined by a transitional probability distribution $P_{Y|X}$, also known as a *Markov kernel* [PW23]. For example, when X is discrete, one has $p_Y(y) = \sum_x p_X(x)p_{Y|X}(y|x)$. When X is continuous, one has $p_Y(y) = \int p_X(x)p_{Y|X}(y|x)\,\mathrm{d}x$. This random transformation is noted $X \to \boxed{P_{Y|X}} \to Y$ or $X \to Y$ for short.

In the sequel, a *side-channel* is defined as a random transformation $X \to Y$, and we shall refer to any transformation $X \to Y$ as a "channel". In the side-channel literature, it is also commonly defined as a random function $Y = F(X)$ where $F = f$ is picked at random among a set of deterministic functions f according to some probability distribution P_F. It is true, but not obvious, that the two descriptions coincide: See Appendix A for details. We say that the channel $X \to \boxed{P_{Y|X}} \to Y$ is *opaque* if Y is independent of its input X, that is, $p_{Y|X}(y|x) = p_Y(y)$ for all x and y.

Notice that any deterministic function $Y = f(X)$ can be seen as a "random" transformation where $p_{Y|X}(y|x) = \delta(y = f(x))$ (Dirac distribution). This functional channel will be denoted by $X \to \boxed{f} \to Y$. A functional channel with constant f is *opaque*. If f is the identity, the corresponding functional channel is named *identity channel*. When we write $X \to Y \to Z$ we always assume that it forms a Markov chain.

Additive masking of order $d \geqslant 0$ can be seen as a channel $X \to \boxed{\mathrm{Mask}_d} \to \mathbf{X}$, where $\mathbf{X} = (X_0, \ldots, X_d) \triangleq (R_0, \ldots, R_{d-1}, X - \sum_{i=0}^{d-1} R_i)$ where the R_i are independent and uniformly distributed $R_i \sim \mathcal{U}(\mathcal{X})$. The $d + 1$ components of $\mathbf{X}$ are called *shares* of X. By the well-known *secret sharing property*, any subset of at most d shares of X is independent of X.

An important class of channels is as follows:

Definition 1 (Erasure Channel). *The channel*

$$X \to \boxed{\mathrm{EC}_{\mathcal{E}}^{\perp}} \to Y \tag{1}$$

is said to be an erasure channel with erasure probability $\mathcal{E} \in [0, 1]$ and special erasure symbol $\perp$ if on input x, $\mathrm{EC}_{\mathcal{E}}^{\perp}$ outputs x with probability

$$\overline{\mathcal{E}} = 1 - \mathcal{E} \tag{2}$$

and the special erasure symbol $\perp$ otherwise (with probability $\mathcal{E}$). That is

$$\begin{cases} p_{Y|X}(\perp|x) = \mathcal{E} \\ p_{Y|X}(x|x) = \overline{\mathcal{E}} \end{cases} \qquad (\forall x \neq \perp) \tag{3}$$

When convenient, we also consider $\perp$ as input value, and let $p_{Y|X}(\perp|\perp) = 1$. Notice that an erasure channel with erasure probability $\mathcal{E} = 1$ is opaque.

Remark 1. The notation $\mathcal{E}$ for the erasure probability is classical in information theory. A few articles such as [DDF14, PGMP19] use the complementary $1 - \mathcal{E}$ instead. Here we follow the standard information theoretic convention.

Erasure channels satisfy useful properties that are easy to check:

Lemma 1 (Commutative Property). *Let $P_{Y|X}$ be any channel from $\mathcal{X}$ to $\mathcal{Y}$ and $P^{\perp}_{Y|X}$ its extension to the input $\perp$ by setting $P^{\perp}_{Y|X}(\perp|\perp) = 1$. Then*

$$\left(X \to \boxed{\mathrm{EC}^{\perp}_{\mathcal{E}}} \to \boxed{P^{\perp}_{Y|X}} \to Y'\right) = \left(X \to \boxed{P_{Y|X}} \to \boxed{\mathrm{EC}^{\perp}_{\mathcal{E}}} \to Y'\right). \quad (4)$$

Note that on the left-hand side the erasure channel is defined on $\mathcal{X}$ while on the right-hand side it is defined on $\mathcal{Y}$.

Proof. See Appendix B.1. □

Lemma 2 (Composition of Erasure Channels). *Let $\mathcal{E}_0, \mathcal{E}_1 \in [0, 1]$ and set $\overline{\mathcal{E}} = \overline{\mathcal{E}}_0\overline{\mathcal{E}}_1$. Then*

$$\left(X \to \boxed{\mathrm{EC}^{\perp}_{\mathcal{E}_1}} \to \boxed{\mathrm{EC}^{\perp}_{\mathcal{E}_0}} \to Y\right) = \left(X \to \boxed{\mathrm{EC}^{\perp}_{\mathcal{E}}} \to Y\right) \quad (5)$$

Proof. The output is not erased if and only if it is not erased in both channels, hence with probability $\overline{\mathcal{E}} = \overline{\mathcal{E}}_0\overline{\mathcal{E}}_1$. □

2.3 Informational Leakage Measures

There exist many noisiness metrics in the literature that quantify how noisy a channel $X \to Y$ can be. In this Subsection we list different leakage measures used in this paper.

The correlation coefficient is widely adopted in side-channel analysis for its simplicity in e.g., the associated correlation power analysis (CPA) [BCO04].

Definition 2 (Pearson's Correlation Coefficient).

$$\rho(X;Y) \triangleq \frac{\mathbb{E}_{XY}\left[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])\right]}{\sqrt{\mathbb{E}_X\left[(X - \mathbb{E}[X])^2\right]\mathbb{E}_Y\left[(Y - \mathbb{E}[Y])^2\right]}}. \quad (6)$$

The correlation coefficient is symmetric $\rho(X;Y) = \rho(Y;X)$. Note, however, that $\rho(X;Y) = 0$ does not imply that X and Y are statistically independent.

Definition 3 (Kullback-Leibler Divergence and Total Variation Distance). *Let P, Q be two probability distributions with respective pdf or pmf p, q defined over $\mathcal{X}$. The Kullback–Leibler (KL) divergence between P and Q is*

$$D_{\mathrm{KL}}(P\|Q) \triangleq \sum_{\mathcal{X}} p \log \frac{p}{q} \quad (7)$$

and the total variation distance (TV) between P and Q is

$$D_{\mathrm{TV}}(P\|Q) = \frac{1}{2}\sum\!\!\!\!\!\!\int_{\mathcal{X}} |p-q| = \frac{1}{2}\|p-q\|_1. \tag{8}$$

KL divergence is not symmetric in general $D_{\mathrm{KL}}(P\|Q) \neq D_{\mathrm{KL}}(Q\|P)$ but the total variation is symmetric $D_{\mathrm{TV}}(P\|Q) = D_{\mathrm{TV}}(Q\|P)$.

Remark 2. Total variation is known to characterize indistinguishability in the sense that no statistical test can distinguish P and Q if $D_{\mathrm{TV}}(P\|Q)$ is negligible [PW23, § 7.3]. Both D_{TV} and D_{KL} are particular instances of *f-divergences*, that satisfy a data processing inequality (see [PW23, Def. 7.1]).

Definition 4 (Mutual Information). *The* mutual information *(*MI*) is the KL divergence between the joint distribution of (X,Y) and the product of its marginals:*

$$I(X;Y) \triangleq D_{\mathrm{KL}}(p_{XY}\|p_X p_Y) = \sum\!\!\!\!\!\!\int_{\mathcal{X}\times\mathcal{Y}} p_{XY}(x,y)\log\frac{p_{X,Y}(x,y)}{p_X(x)p_Y(y)}. \tag{9}$$

MI is symmetric $I(X;Y) = I(Y;X)$. It is a measure of statistical dependence: if X and Y are statistically independent then $p_{XY} = p_X p_Y$ so that $I(X;Y) = 0$.

Definition 5 (Total Variation Information). *The* total variation information *(*TVI*) is the TV distance between the joint distribution of (X,Y) and the product of its marginals:*

$$\Delta(X;Y) \triangleq D_{\mathrm{TV}}(p_{XY}\|p_X p_Y) = \frac{1}{2}\|p_{XY} - p_X p_Y\|_1 \tag{10}$$

$$= \frac{1}{2}\sum\!\!\!\!\!\!\int_{\mathcal{X}\times\mathcal{Y}} |p_{XY}(x,y) - p_X(x)p_Y(y)|. \tag{11}$$

Note that TVI is symmetric, $\Delta(X;Y) = \Delta(Y;X)$. A negligible TVI implies that no test can distinguish p_{XY} from $p_X p_Y$, that is no test can exhibit a statistical dependence between X and Y. TVI can be seen as a particular f-information [PW23, Eqn. 7.46]. In [PGMP19] TV is referred to as Statistical Distance (SD).

Definition 6 (Maximal Leakage [IWK20]**).** *The* maximal leakage *quantifies the maximal advantage in estimating X from the side-channel information Y:*

$$\mathcal{L}(X \to Y) = \log \sum\!\!\!\!\!\!\int_{\mathcal{Y}} \sup_{x\in\mathcal{X}} p_{Y|X}(y|x). \tag{12}$$

We use an arrow instead of a semicolon in the definition of $\mathcal{L}(X \to Y)$ because it depends only on the channel $X \to Y$ and not on the input probability distribution of X. Note that maximal leakage is not symmetric $\mathcal{L}(X \to Y) \neq \mathcal{L}(Y \to X)$.

Definition 7 (Euclidean Bias [PR13]**).** *The* Euclidean norm bias (EN) *is the expected Euclidean distance between the posterior distribution $p_{X|Y}$ and its prior p_X:*

$$\beta(X;Y) \triangleq \mathbb{E}_Y \|p_{X|Y}(\cdot|Y) - p_X\|_2. \tag{13}$$

Remark 3. $\beta(X;Y)$ is similar to $\Delta(X;Y)$ where $\|\cdot\|_1$ is replaced by $\|\cdot\|_2$ inside the expectation. However, $\beta(X;Y)$ is not equal to $\|p_{XY} - p_X p_Y\|_2$ because of the square root appearing in the Euclidean norm. In particular it is not symmetric $\beta(X;Y) \neq \beta(Y;X)$. A similar quantity $\Delta_2(X;Y)$ with squared norm, related to the Rényi 2-information, is used for side-channel leakage evaluation in [LBC+23].

Definition 8 ((Average) Relative Error [PGMP19]**).**

$$\mathrm{RE}(X;Y) \triangleq \sup_{x,y} \left| \frac{p_{X|Y}(x|y)}{p_X(x)} - 1 \right| = \sup_{x,y} \left| \frac{p_{XY}(x,y)}{p_X(x)p_Y(y)} - 1 \right|. \tag{14}$$

$$\mathrm{ARE}(X;Y) \triangleq \mathbb{E}_Y \left[\sup_x \left| \frac{p_{X|Y}(x|Y)}{p_X(x)} - 1 \right| \right]. \tag{15}$$

Remark 4. While relative error is symmetric $\mathrm{RE}(X;Y) = \mathrm{RE}(Y;X)$ the average relative error is not symmetric $\mathrm{ARE}(X;Y) \neq \mathrm{ARE}(Y;X)$.

This paper focuses on another important quantity:

Definition 9 (Complementary Doeblin Coefficient [Doe37, Dob56, MS23a]**).**

$$\overline{\mathcal{E}}(X \to Y) = 1 - \sum_y \inf_x p_{Y|X}(y|x) = \sum_y \sup_x \left(p_Y(y) - p_{Y|X}(y|x) \right) \tag{16}$$

$$= \mathbb{E}_Y \left[\sup_x \left(1 - \frac{p_{Y|X}(Y|x)}{p_Y(Y)} \right) \right] = \mathbb{E}_Y \left[\sup_x \left(1 - \frac{p_{X|Y}(x|Y)}{p_X(x)} \right) \right]. \tag{17}$$

Remark 5. Doeblin's coefficient appeared implicitly in [Doe37, p. 1], later explicitly in [Sen73, Eqn. 2.6] and has been recently known as the "Doeblin ergodicity coefficient", see e.g., [CL10, Eqn. 10].

Remark 6. While the expression of CDC resembles both that of maximal leakage and ARE, it is fundamentally different. CDC is non-symmetric $\overline{\mathcal{E}}(X \to Y) \neq \overline{\mathcal{E}}(Y \to X)$. Like for maximal leakage, we use an arrow instead of a semicolon in $\overline{\mathcal{E}}(X \to Y)$ because it depends only on $X \to Y$ and not on P_X. The original Doeblin coefficient is $\mathcal{E}(X \to Y) = 1 - \overline{\mathcal{E}}(X \to Y)$.

Remark 7. Maximal leakage is a particular Sibson's α-information [Ver15, EVG22] of order $\alpha = +\infty$: $\mathcal{L}(X \to Y) = I_\infty(X;Y)$, while Mutual information can be seen as Sibson's α-information of order $\alpha = 1$: $I(X;Y) = I_1(X;Y)$. The Doeblin coefficient is the exponential of minus Sibson's α-information of order $\alpha = -\infty$ also known as maximal cost leakage: $\mathcal{E}(X \to Y) = \exp(-I_{-\infty}(X;Y))$ [EVG22].

2.4 Properties of the Complementary Doeblin Coefficient

In Lemma 2 we have seen the composition property of erasure channels sharing the same erasure symbol. What happens now if we compose two erasure channels with different erasure symbols? The following Lemma shows that even though the resulting channel is not an erasure channel, its CDC is identical:

Lemma 3 (Erasures Composition). *For the channel* $X \to \boxed{\mathrm{EC}_{\mathcal{E}_1}^{\perp_1}} \to Y \to \boxed{\mathrm{EC}_{\mathcal{E}_0}^{\perp_0}} \to Z$,

$$\overline{\mathcal{E}}(X \to Z) = \overline{\mathcal{E}}(X \to Y)\overline{\mathcal{E}}(Y \to Z). \tag{18}$$

Proof. Straightforward from the definitions (see Appendix B.4). □

Consider a channel $X \to Y$ and suppose one has a post processing $Y \to Z$ (such that $X \to Y \to Z$ is a *Markov chain* [PW23]). Intuitively, Z does not contain more information than Y about X, and we have the following:

Lemma 4 (CDC **Consistency).** *For any* $X \to Y \to Z$*, one has*

$$\mathcal{E}(X \to (Y, Z)) = \mathcal{E}(X \to Y). \tag{19}$$

Proof. Straightforward from the definitions (see Appendix B.5). □

A sensitive variable X may leak several times in a side-channel attack. For instance, the adversary may access two side-channels $X \to Y_1$ and $X \to Y_2$. What can be said about the CDC of the combined leakages? The following Lemma provides an answer.

Lemma 5 (Single Letterization). *For the multi-channel* $X \nearrow \boxed{P_{Y_1|X}} \to Y_1$, $X \searrow \boxed{P_{Y_2|X}} \to Y_2$ *denoted by* $X \to Y_1, Y_2$*, we have*

$$\mathcal{E}(X \to Y_1, Y_2) \geqslant \mathcal{E}(X \to Y_1)\mathcal{E}(X \to Y_2). \tag{20}$$

Generally with q *channels in parallel i.e.* $P_{Y_1,\ldots,Y_q|X} = \prod_{i=1}^{q} P_{Y_i|X}$ *we have*

$$\mathcal{E}(X \to Y_1, \cdots, Y_q) \geqslant \prod_{i=1}^{q} \mathcal{E}(X \to Y_i). \tag{21}$$

In terms of CDC*, equation* (21) *reformulates as*

$$\overline{\mathcal{E}}(X \to Y_1, \ldots, Y_q) \leqslant 1 - \prod_{i=1}^{q} \big(1 - \overline{\mathcal{E}}(X \to Y_i)\big) \leqslant \sum_{i=1}^{q} \overline{\mathcal{E}}(X \to Y_i). \tag{22}$$

Proof. See Appendix B.6.

In the adaptive setting, the adversary may observe Y_1 through the side-channel $X \to Y_1$ and then chose the channel $X \to Y_2$ based on his observation of Y_1. The following adaptive single letterization lemma extends Lemma 5 by showing how the CDC of the combined leakages can be derived even when the channels are chosen adaptively:

Lemma 6 (Adaptive Single Letterization). *In the adaptive setting where all channels satisfy $\overline{\mathcal{E}}(X \to Y_i) \leqslant \overline{\mathcal{E}}$, we still have*

$$\mathcal{E}(X \to Y_1, Y_2) \geqslant \mathcal{E}^2. \tag{23}$$

More generally we have $\mathcal{E}(X \to Y_1, \ldots, Y_q) \geqslant \mathcal{E}^q$.

Proof. See the Appendix B.7

2.5 Side-Channel Attack Models

We use the following terminology from [PR13, PGMP19, MS23b].

Definition 10 (δ-Noisy Channel). *A channel $X \to \boxed{P_{Y|X}} \to Y$ is said to be* δ-noisy for input X with respect to some metric $\mathcal{D}$ *if $\mathcal{D}(X;Y) \leqslant \delta$. For short, it is said to be* δ-noisy *with respect to $\mathcal{D}$ (without reference to X) when X is taken uniformly distributed $X \sim \mathcal{U}(\mathcal{X})$. $\mathcal{D}$ should be understood as a distortion measure of the channel. For instance $\mathcal{D}$ can be $\rho, I, \Delta, \mathcal{L}, \beta, \mathrm{RE}, \mathrm{ARE}$ or $\overline{\mathcal{E}}$. The lower δ, the noisier the channel.*

Definition 11 (Side-Channel Exploitability). *Consider a set of l sensitive values $(X_1, \ldots, X_l)$. A side-channel adversary obtains multiple side information $(Y_1, \ldots, Y_l)$ through the channels $\varphi_i = (X_i \to Y_i)$, $i = 1, \ldots, l$. The tuple of channels $\varphi = (\varphi_1, \ldots, \varphi_l)$ is restricted so that the adversary's ability is limited. Typically, the adversary is said to be:*

- *t-**threshold probing** [ISW03]: if φ contains at most t identity channels and opaque channels on the remaining positions;*
- *$\overline{\mathcal{E}}$-**random probing** [DDF14]: if φ is made of $\mathcal{E}$-erasure channels;*
- *δ-**noisy** [PR13, PGMP19]: if φ contains only δ-noisy channels with respect to some metric $\mathcal{D}$;*
- *(σ, f)-**additive**: if φ is made of channels of the form $X \to Y \triangleq f(X) + \sigma N$ where f is a fixed deterministic leakage function and $N \sim \mathcal{N}(0,1)$ is an independent additive Gaussian noise. Typically, f can be the Hamming weight function or the least significant bit function. When X is a bit leaking as $X \to Y = X + \sigma N$ it specializes to the leakage model of Chari et al. [CJRR99].*

A cryptographic implementation is classically modeled as a circuit Γ. There are two main paradigms about the channel inputs that are both legitimate. Either it is assumed that every wire within the circuit leaks like [DDF14]. Or it is assumed that every gate within the circuit leaks like [PR13]. In this case, the channel takes as input the operands of the gate. For unary gates both models

are exactly equivalent. The models differ whenever the gates process multiple operands. Assuming that the wires leak leads to tighter security bounds, while assuming that the gates leak seems to be closer to the physical nature of leakages. [DDF14, § 5.5] discusses more in depth the trade off between both models.

One concern in side-channel analysis is to cover adaptive adversaries $\mathcal{A}$. This term can be confusing as it is used with different meanings. To avoid any ambiguity, we make more precise the terminology with the following definitions:

Definition 12 (Adaptive Adversary Flavors). *We clarify the different notations of adaptivity in the context of side-channel attacks:*

1. *When $\mathcal{A}$ is allowed to choose sequentially the public information [MS23b] used by Γ then she/he is a* ***chosen public information adversary****. It corresponds to the usual setting of chosen plaintext or ciphertext adversary in cryptology.*
2. *When $\mathcal{A}$ is allowed to specify φ sequentially [DDF14] she/he is said to be a* ***chosen channel adversary****. This differs from chosen public information adversary; in this setting the adversary is allowed to move the position of the side-channel acquisition instruments (probes) from one query to the other.*
3. *If $\mathcal{A}$ can specify $\varphi_1, \ldots, \varphi_l$ sequentially within a query [DDF14], she/he is said to be a* ***strong chosen channel adversary****. The adversary is even allowed to move the position of its probe within a query. This last type of adaptivity is, however, unrealistic in most of practical settings.*

The activity of a side-channel adversary $\mathcal{A}$ with q queries can be viewed as a game. This game unfolds differently depending on the side-channel adaptivity and depending on the gate/wire leakage model. After side-channel collection, the adversary exploits them to distinguish the correct key K and outputs $\mathrm{out}_{\mathcal{A}}(K)$.

The complete acquisition and attack led by $\mathcal{A}$ is formalized by Alg. 1. In practice $\mathrm{out}_{\mathcal{A}}(K)$ can be a score vector sorting the key hypotheses (or parts of the key). If $\mathcal{A}$ is restricted to opaque channels she/he does not learn anything through them. In this case, $\mathcal{A}$ is said to be a black-box adversary.

2.6 Figures of Merit

When $\mathrm{out}_{\mathcal{A}}(K)$ is a key ranking the performances of the adversary are measured via three classical figures of merits: The *success rate* (SR) $\mathbb{P}_s$, the *success rate of order o* (SR_o, success rate in o-trials) $\mathbb{P}_{s,o}$ [SMY09] and the *guessing entropy* GE [Mas94]. We follow Ito *et al.* [IUH22, § 2.3] and express these metrics in terms of the a posteriori rank of the key hypothesis given the side-information.

Definition 13 (Success Rate (SR) and Guessing Entropy (GE)). *Let K be a secret random variable taking values in a finite set $\mathcal{K}$. Let Y be an arbitrary random variable representing a side-information. The* success rate *(SR) is given by the Maximum a Posteriori (MAP) rule*

$$\mathbb{P}_s(K|Y) \triangleq \mathbb{P}(\mathrm{rank}(K|Y) = 1) \tag{24}$$

Algorithm 1: Side-Channel Acquisition and Attack

Data: A number of queries q and a set of allowed side-channels.
Result: The output of the adversary $\mathcal{A}$.

Oracle $\mathcal{O}$ draws uniformly at random a secret key K.
for $i = 1, \ldots, q$ **do** `/* Sequential Acquisition of q Traces. */`
  $\mathcal{A}$ specifies a public information $\mathbf{t}_i$ and send it to $\mathcal{O}$. `/* Sequential Choice of Public Information */`
  $\mathcal{O}$ draws uniformly at random the randomness $\mathbf{R}_i$ and computes the corresponding wire values $\mathbf{X}_i$.
    `/* This is the wire leaking model. In the gate leakage model the loop is over the gates instead. */`
  for $j = 1, \ldots, l$ **do** `/* Sequential Choice of the Side-Channels */`
    $\mathcal{A}$ specifies φ_i to $\mathcal{O}$
    $\mathcal{O}$ sends back the corresponding leakage x_i from side-channel φ_i to $\mathcal{A}$ under the constraint that φ is an allowed tuple of channel.
      `/* Restriction on the type of allowed side-channels */`
return $\mathcal{A}$ outputs $\mathrm{out}_{\mathcal{A}}(K)$

The success rate of order o, SR_o *is*

$$\mathbb{P}_{s,o}(K|Y) \triangleq \mathbb{P}(\mathrm{rank}(K|Y) \leqslant o) \tag{25}$$

The guessing entropy *(*GE*) is the minimum average number of guesses of an optimal guessing strategy*

$$G(K|Y) \triangleq \mathbb{E}\{\mathrm{rank}(K|Y)\} \tag{26}$$

In general rank *is defined as a function such that for each y, $k \to \mathrm{rank}(k|y)$ is a permutation of the key space $\mathcal{K}$. In an optimal guessing strategy, for each $y \in \mathcal{Y}$, $\mathrm{rank}(k|y) \in \{1, \ldots, |\mathcal{K}|\}$ is the rank of $p_{K|Y}(k|y)$ in the list $\{p_{K|Y}(k|y)|k \in \mathcal{K}\}$ sorted in decreasing order. In case of collisions, ties are resolved randomly which does not change the statistical quantities at stake.*

Remark 8. Since $\mathbb{P}_s(K|Y) = \mathbb{P}_{s,o=1}(K|Y)$ holds, results for SR will be derived from results in terms of SR_o.

Definition 14 (Blind Guess). *When no side-information is available, the adversary performs a* blind guess *whose figures of merits are constants depending only on the a priori key-distribution. Namely*

$$\mathbb{P}_{s,o}(K) \triangleq \mathbb{P}(\mathrm{rank}(K) \leqslant o) \qquad \textit{and} \qquad G(K) \triangleq \mathbb{E}\{\mathrm{rank}(K)\} \tag{27}$$

where $\mathrm{rank}(k) \in \{1, \ldots, |\mathcal{K}|\}$ *is the rank of $p_K(k)$ in the list $\{p_K(k)|k \in \mathcal{K}\}$ sorted in decreasing order.*

The *advantage* of the adversary is quantified by $\mathbb{P}_{s,o}(K|Y) - \mathbb{P}_{s,o}(K) \geqslant 0$, $G(K) - G(K|Y) \geqslant 0$ and for statistical tests $\Delta(K;Y) \geqslant 0$. If further K is uniformly distributed then $\mathbb{P}_{s,o}(K) = \frac{o}{|\mathcal{K}|}$ and $G(K) = \frac{|\mathcal{K}|+1}{2}$.

If an adversary $\mathcal{A}$ is computationally bounded then she/he may not be able to fully exploit the side-information Y. The corresponding figures of merit are denoted by $\mathbb{P}_s^{\mathcal{A}}(K|Y)$, $\mathbb{P}_{s,o}^{\mathcal{A}}(K|Y)$ and $G^{\mathcal{A}}(K|Y)$. Obviously $\mathbb{P}_s^{\mathcal{A}}(K|Y) \leqslant \mathbb{P}_s(K|Y)$, $\mathbb{P}_{s,o}^{\mathcal{A}}(K|Y) \leqslant \mathbb{P}_{s,o}(K|Y)$ and $G^{\mathcal{A}}(K|Y) \geqslant G(K|Y)$.

3 Mathematical Key Properties of the CDC

In this section we derive the key mathematical properties of the CDC that will be useful to derive security bounds. In Subsection 3.1 we exhibit the optimal factorization of a side-channel into a stochastically degraded erasure channel. This shows that CDC is the optimal parameter in the reduction from noisy leakages to the random probing model. The word optimal refer to the reduction from noisy leakage to a random probing adversary, we do not claim however that the CDC yields an optimal bound on the success rate of a side-channel attack. In Subsection 3.2 we show how the figures of merit of a side-channel attack can be bounded using CDC leveraging the Data Processing Inequality (DPI). We show that CDC is amenable to evaluation in Subsection 3.3. Finally, we compare the CDC to the informational measure (introduced in Subsection 2.3) in Subsection 3.4.

3.1 Optimal Channel Degradation

It is known that security in the noisy leakage model can be reduced to security in the random probing model. In [DDF14, Lemma 3], security in the noisy leakage model measured by TVI is reduced to security in the random probing model. In [PGMP19, Lemma 3], security in the noisy leakage model measured by ARE is reduced to security in the random probing model. Finally, [DFS19, Theorem 3] proves security in the noisy leakage model measured by MI by upper bounding [DDF14, Lemma 3] using Pinsker's inequality [PW23, Thm 7.10].

The key property is that any channel can be seen as a stochastically degraded erasure channel. This stochastic degradation can be seen as a factorization like [DDF14, Lemma 4] of $\mathrm{Noise}(X)$ into $\mathrm{Noise}'(\varphi(X))$ where φ is an erasure channel (termed ϵ-identity function in their presentation). In this section, we derive the optimal parameter in this reduction and show it corresponds to the complementary Doeblin coefficient (CDC). This unifies previous results that can seen as a weakened version of our reduction by upper bounding the CDC.

Definition 15 (Degraded Channel). *The channel $X \to Z$ is said to be* stochastically degraded *with respect to the channel $X \to Y$ if there exists a channel $Y \to Z$ such that*

$$\left(X \to \boxed{P_{Y|X}} \to Y \to \boxed{P_{Z|Y}} \to Z\right) = \left(X \to \boxed{P_{Z|X}} \to Z\right). \tag{28}$$

Theorem 1. *Any channel $X \to \boxed{P_{Y|X}} \to Y$ is a stochastically degraded erasure channel:*

$$X \to \boxed{\mathrm{EC}_{\mathcal{E}}^{\perp}} \to X' \to \boxed{P_{Y|X'}} \to Y \tag{29}$$

with the maximum *erasure probability*

$$\mathcal{E}(X \to Y) = \sum\!\!\!\!\!\!\int_{y} \inf_{x \in \mathcal{X}} p_{Y|X}(y|x). \tag{30}$$

Proof. We provide a proof for completeness in Appendix B.3. □

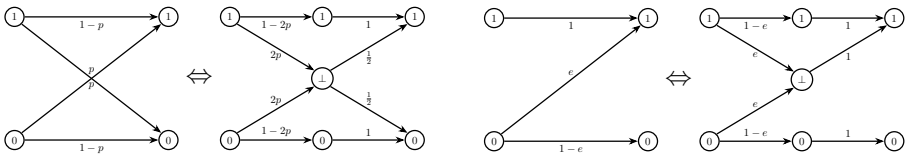
Remark 9. $\mathcal{E}(X \to Y)$ is known in the literature as the *Doeblin coefficient of ergodicity* of the channel [Doe37, Dob56, Mak20, MS23a]. In our context $\mathcal{E}(X \to Y)$ represents the erasure probability while $\overline{\mathcal{E}}(X \to Y)$ represents the probing probability. Thm 1 was proved for binary input channels in [BB11, Prop. 6.4] in the context of physical layer security and wiretap channels and in the general for network coding in [Mak20, Lemma 6] or key agreements in [GGK20, Lemma 5]. The CDC appears for the first time in the side-channel literature as the value of ϵ in [DDF19, Eqn. 9, Proof of Lemma 4].

Remark 10. Maximum erasure probability in Theorem 1 means that there exists at least one channel degradation (in the form of Equation (29)) achieving $\mathcal{E} = \mathcal{E}(X \to Y)$ and that there does not exist any channel degradation with $\mathcal{E} > \mathcal{E}(X \to Y)$. In this sense, the CDC is the optimal parameter in the reduction from noisy leakage to the random probing model.

Obviously, the erasure channel is optimally degraded into itself, that is, $\mathcal{E}(X \to \boxed{\mathrm{EC}^{\perp}_{\mathcal{E}}(X)} \to Y) = \mathcal{E}$.

Example 1 (Binary Symetric Channel (BSC)). If X is a binary random variable and $X \to Y$ a BSC with crossover probability $0 \leqslant p \leqslant \frac{1}{2}$ then $\mathcal{E}(X \to Y) = 2p$. The factorization given by Theorem 1 is shown in Figure 2a.

Example 2 (Z-Channel). If X is binary and $X \to Y$ is a Z-channel with parameter $0 \leqslant e \leqslant 1$ then $\mathcal{E}(X \to Y) = e$. The factorization given by Theorem 1 is shown in Figure 2b.



(a) BSC (b) Z-Channel

Fig. 2: Illustration of Theorem 1

Theorem 1 implies the following corollary in terms of simulatability similar to [DDF14, Lemma 3] in terms of CDC:

Corollary 1. *Any $\overline{\mathcal{E}}$-noisy adversary $\mathcal{A}$ with respect to* CDC *can be perfectly simulated by a $\overline{\mathcal{E}}$-random probing adversary $\mathcal{S}$.*

Proof. Let us assume that $\mathcal{A}$ receives the side information Y about K through the channel $K \to Y$. By assumption $\overline{\mathcal{E}}(K \to Y) = \overline{\mathcal{E}}$ so that Theorem 1 implies that $(K \to Y) = (K \to K' \to Y)$ where $K \to K'$ is an erasure channel with erasure probability $\mathcal{E}$. An adversary $\mathcal{S}$ that receives side information K' about K through $K \to K'$ is $\overline{\mathcal{E}}$-random probing. Now $\mathcal{S}$ can sample $\tilde{Y}$ by passing its observation K' through the channel $K' \to Y$. By construction $\mathcal{S}$ obtains $\tilde{Y}$ equals in law with Y. □

Strengthened Data Processing Inequality By Lemma 2, the composition of several erasure channels is an erasure channel with a larger erasure probability. In general, the composition of several channels should leak less information. This is formalized by a strengthened data processing inequality (DPI):

Lemma 7 (Strengthened Data Processing Inequality (DPI)). *For any $X \to Y \to Z$,* CDC *satisfies the following strengthened-DPI*

$$\overline{\mathcal{E}}(X \to Z) \leqslant \overline{\mathcal{E}}(X \to Y)\overline{\mathcal{E}}(Y \to Z) \tag{31}$$

which implies a preprocessing-DPI (preprocessing $X \to Y$ can only reduce leakage): $\overline{\mathcal{E}}(X \to Z) \leqslant \overline{\mathcal{E}}(Y \to Z)$, and a post-processing-DPI (post-processing $Y \to Z$ can only reduce leakage) $\overline{\mathcal{E}}(X \to Z) \leqslant \overline{\mathcal{E}}(X \to Y)$. In a nutshell, stochastic degradation reduces the value of the CDC.

Proof. By Theorem 1 we can degrade the channels $X \to Y$ and $Y \to Z$ so that the channel $X \to Z$ rewrites as

$$X \to \boxed{\mathrm{EC}^{\perp_x}_{\mathcal{E}(X\to Y)}} \to X' \to \boxed{P_{Y|X'}} \to Y \to \boxed{\mathrm{EC}^{\perp_y}_{\mathcal{E}(Y\to Z)}} \to Y' \to \boxed{P_{Z|Y'}} \to Z. \tag{32}$$

Since by Lemma 1 an erasure channel commutes with any other channel, the channel is equivalent to

$$X \to \boxed{\mathrm{EC}^{\perp_x}_{\mathcal{E}(X\to Y)}} \to X' \to \boxed{\mathrm{EC}^{\perp_y}_{\mathcal{E}(Y\to Z)}} \to X'' \to \boxed{P^{\perp_y}_{Y'|X''}} \to Y \to \boxed{P_{Z|Y'}} \to Z. \tag{33}$$

We have two different erasures symbols but by Lemma 3, the concatenated channel is stochastically degraded with respect to an erasure channel with $\overline{\mathcal{E}} = \overline{\mathcal{E}}(X \to Y)\overline{\mathcal{E}}(Y \to Z)$. So we have

$$X \to \boxed{\mathrm{EC}^{\perp}_{\mathcal{E}}} \to \tilde{X} \to \boxed{P_{\tilde{X}|X''}} \to X'' \to \boxed{P^{\perp_y}_{Y'|X''}} \to Y \to \boxed{P_{Z|Y'}} \to Z. \tag{34}$$

Now let $p_{Z|\tilde{X}} = p_{Z|Y'} \to p^{\perp_y}_{Y'|X''} \to p_{\tilde{X}|X''}$ so that

$$X \to \boxed{\mathrm{EC}^{\perp}_{\mathcal{E}}} \to \tilde{X} \to \boxed{P_{Z|\tilde{X}}} \to Z. \tag{35}$$

Since $\overline{\mathcal{E}}(X \to Z)$ is the infimum such that this factorization holds, we have $\overline{\mathcal{E}}(X \to Z) \leqslant \overline{\mathcal{E}} = \overline{\mathcal{E}}(X \to Y)\overline{\mathcal{E}}(Y \to Z)$ which concludes the proof. □

Remark 11. Makur and Singh [MS23a], with a very different proof, established a similar property of CDC in the discrete setting and interpreted it as the sub-multiplicativity of CDC.

3.2 Bounds on the Figures of Merit

An adversary tries to recover the sensitive variable X with the help of the side-information Z through the side-channel $X \to Z$ which is stochastically degraded with respect to the channel $X \to Y$. Intuitively she/he can only perform worse than an adversary that accesses Y. This intuition is formalized by a DPI data processing inequality (DPI).

Lemma 8 (Data Processing Inequality (DPI)). *Consider the channel $U \to V \to W \to X$. Then*

$$I(V;W) \geqslant I(U;X) \quad \textit{and} \qquad \Delta(V;W) \geqslant \Delta(U;X). \tag{36}$$

Consider the channel $X \to Y \to Z$, where X is valued in the finite set $\mathcal{X}$, the SR_o *and* GE *verify a post-processing DPI,*

$$\mathbb{P}_{s,o}(X|Y) \geqslant \mathbb{P}_{s,o}(X|Z) \qquad \textit{and} \qquad G(X|Y) \leqslant G(X|Z). \tag{37}$$

Proof. It is well known that the SR_o and GE verify a post-processing DPI (see for instance [BCGR22, Rio23]). As shown in [PW23, Theorem 7.16], the f-information also verifies a DPI which includes MI and TVI.

Proposition 1. *Let $\lambda_{\mathrm{SR}_o} = (1 - \mathbb{P}_{s,o}(K)), \lambda_{\mathrm{GE}} = (G(K) - 1), \lambda_{\mathrm{TVI}} = (1 - \exp(-H_2(K)))$ be three constants that only depend on the a priori secret key distribution (where H_2 is the collision entropy). The adversary's advantage for* SR, GE *and* TVI *can be bounded as follows:*

$$\begin{aligned} 0 \leqslant \mathbb{P}_{s,o}(K|Y) - \mathbb{P}_{s,o}(K) &\leqslant \overline{\mathcal{E}}(K \to Y)\lambda_{\mathrm{SR}_o}, \\ 0 \leqslant \quad G(K) - G(K|Y) \quad &\leqslant \overline{\mathcal{E}}(K \to Y)\lambda_{\mathrm{GE}}, \\ 0 \leqslant \qquad \Delta(K;Y) \qquad &\leqslant \overline{\mathcal{E}}(K \to Y)\lambda_{\mathrm{TVI}}. \end{aligned} \tag{38}$$

Proof. See Appendix B.2. □

We cannot directly deduce from Lemma 8 that $\mathbb{P}_s^{\mathcal{A}}(K|Y)$, $\mathbb{P}_{s,o}^{\mathcal{A}}(K|Y)$ and $G^{\mathcal{A}}(K|Y)$ verify a DPI. But as shown by Duc *et al.* [DDF14, Lemma 2] if the channel $K' \to Y$ can be efficiently sampled then $\mathcal{A}$ can efficiently reproduce $\tilde{Y}$ equal in distribution with Y from X'. As a consequence, under this hypothesis we can assume that $\mathbb{P}_s^{\mathcal{A}}(K|Y)$, $\mathbb{P}_{s,o}^{\mathcal{A}}(K|Y)$ and $G^{\mathcal{A}}(K|Y)$ also verify the bounds from Proposition 1. As shown by Brian *et al.* [BFO+22], a large class of noisy channels $K' \to Y$ can be indeed simulated almost for free. In fact, we do not

need to have an efficient simulation of the channel noise $K' \to Y$. Indeed, since $K \to K'$ is an erasure channel we obtain

$$\mathbb{P}_s^{\mathcal{A}}(K|Y) \overset{(a)}{\leqslant} \mathbb{P}_s(K|Y) \overset{(b)}{\leqslant} \mathbb{P}_s(K|K') \overset{(c)}{=} \mathbb{P}_s^{\mathcal{A}}(K|K') \tag{39}$$

where (a) holds because a computationally bounded adversary can only perform worse than the optimal unbounded adversary, (b) is the usual DPI and (c) is due to the fact that for an erasure channel an optimal attack is efficiently computable. For example, the attack that outputs the key when it is not erased and a random ranking otherwise is both optimal and efficient. The same derivation holds for GE (with reversed inequalities).

Discussion on the Optimality of the CDC The trade-off between a CDC-based bound and a MI based bound depends on the nature of the channel $K \to Y$ which is factorized optimally with the CDC into $K \to K' \to Y$.

- If $K \to Y$ is an erasure channel then the factorization is $K \to K' = Y$ so that using the DPI to consider K' instead of Y as a leakage can be done without any degradation of the final bound. In this case the bound using CDC is optimal and holds with equality.
- If the channel $K \to Y$ is far from an erasure channel then the channel $K' \to Y$ can be noisy. As a consequence, using the DPI to consider K' as the leakage instead of Y incurs an unavoidable loss. In this case the CDC based bound can be loose and another informational leakage measure such as MI or TVI may be more suitable to capture the noise in the side-channel. A very bad channel in this respect is the channel from $K \to Y$ where K and Y are both taking their values in $\{1, \ldots, |\mathcal{X}|\}$, $p_{Y|K}(y|k) = (|\mathcal{X}| - 1)^{-1}$ if $y \neq k$ and $p_{Y|K}(y|k) = 0$ otherwise. In this case, $\overline{\mathcal{E}}(K \to Y) = 1$ while $I(K;Y) = \log(|\mathcal{X}|/(|\mathcal{X}| - 1))$ is small.

Both of these two extreme cases are toy examples that do not occur in practice. Depending on the nature of the practical side-channel the bound based on CDC will be tight or not.

3.3 Theoretical Expressions for Concrete Evaluation

In this Subsection we show how CDC can be evaluated in a practical setting. We first show how we can derive a closed form expression for univariate functional channels perturbed by an additive Gaussian noise. (This corresponds to (σ, f)-additive adversaries.) Then we show how this allows to bound CDC in the multivariate case perturbed by an additive Gaussian noise with a given correlation matrix $\mathbf{\Sigma}$. This corresponds to the widely used setting from *template attack* (TA) [CRR02, LRP07, UKM+17]

We show that in this model CDC is suitable for concrete evaluation, even when the noise is multivariate and potentially high dimensional.

Univariate Case

Definition 16 (Radially Symmetric Decreasing). *The real-valued r.v. Z is said to be radially symmetric decreasing if $p_Z(z) = p_Z(|z|)$ and decreasing in $|z|$.*

We derive a closed-form expression for CDC when the channel is functional perturbed by a radially symmetric decreasing additive noise. This includes Gaussian, Laplacian or Cauchy distribution for example. As CDC only depends on the channel the result does not depend on the probability distribution of X. As expected the CDC tends to zero as the noise increases.

Proposition 2. *Let X be a random variable taking values in $\mathcal{X}$ and $Y = f(X) + Z$ where f is an arbitrary real-valued function and Z is a radially symmetric decreasing noise with survival function S. Let $m = \inf_{x\in\mathcal{X}} f(x)$ and $M = \sup_{x\in\mathcal{X}} f(x)$. Then*

$$\mathcal{E}(X \to Y) = 2S\left(\frac{M-m}{2}\right). \tag{40}$$

For the widely adopted linear leakage model, $\mathcal{X} = \mathbb{F}_2^n$ and $f(X) = \sum_{i=1}^n a_i X_i$ where $\mathbf{a} = (a_1, \ldots, a_n) \in \mathbb{R}^n$ and X_i denotes the i-th bit in the binary representation of X, we have $m = -\sum_i a_i^-$ and $M = \sum_i a_i^+$ so that $M - m = \sum_i |a_i| = \|\mathbf{a}\|_1$ and the expression simplifies to

$$\mathcal{E}(X \to Y) = 2S\left(\frac{\|\mathbf{a}\|_1}{2}\right). \tag{41}$$

If $Z \sim \sigma\mathcal{N}(0,1)$, the survival function S is the Marcum function Q, and

$$\overline{\mathcal{E}}(X \to Y) = 1 - 2Q\left(\frac{\|\mathbf{a}\|_1}{2\sigma}\right) \overset{\sigma\to\infty}{=} \frac{\|\mathbf{a}\|_1}{\sqrt{2\pi}}\frac{1}{\sigma} + O\left(\sigma^{-3}\right). \tag{42}$$

For the classical Hamming weight (HW) model, $\mathbf{a} = (1, \ldots, 1)$ and $\|\mathbf{a}\|_1 = n$.

Proof. See Appendix B.8. □

Remark 12. When f is constant then $m = M$, and we obtain $\mathcal{E}(X \to Y) = 1$. This is expected as in this case the channel is opaque.

Multivariate Case Let $f : x \in \mathbb{F} \mapsto f(x) \in \mathbb{R}^m$ be a multivariate leakage function and $\mathbf{Y} = f(X) + \mathcal{N}(0, \mathbf{\Sigma})$. Then $\tilde{\mathbf{Y}} \triangleq \mathbf{W}\mathbf{Y} = (\mathbf{W} \cdot f)(X) + \tilde{\mathbf{Z}}$ where $\mathbf{W}$ is a given whitening matrix (e.g. $\mathbf{W} = \mathbf{\Sigma}^{-\frac{1}{2}}$) so that $\tilde{\mathbf{Z}} = \mathbf{W}\mathbf{Z} \sim \mathcal{N}(0, \mathbf{I}_m)$. Given X, $\tilde{\mathbf{Y}}$ is a Gaussian vector whose covariance matrix is diagonal. Hence, by theorem on Gaussian vectors, the different components of $\tilde{\mathbf{Y}}$ are independent given X and Lemma 5 implies that

$$\mathcal{E}(X \to \mathbf{Y}) = \mathcal{E}(X \to \tilde{\mathbf{Y}}) \geqslant \prod_{i=1}^{m} \mathcal{E}(X \to \tilde{Y}_i). \tag{43}$$

Since every channel $X \to \tilde{Y}_i$ is univariate additive Gaussian noise, its expression is given by Proposition 2. This methodology yields a positively biased estimator of $\mathcal{E}(X \to \mathbf{Y})$ from the non-biased estimator of each $\mathcal{E}(X \to \tilde{Y}_i)$. This approach is more conservative but ensures that we do not overestimate the security parameter which would result in a false sentiment of security. This is by opposition with the perceived information (PI [RSV+11, Eqn. 3]) which is a negatively biased estimator of MI as shown in [BHM+19, IUH22].

Example 3. As an example, consider the channel $\mathbf{Y} \triangleq (f(X), f(X))^T + \mathbf{Z}$ where f is a univariate leakage function and $\mathbf{Z} \sim \mathcal{N}(0, \mathbf{\Sigma})$ with a covariance matrix $\mathbf{\Sigma} = \sigma^2 \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$ where σ is the noise standard deviation and $\rho \in [-1, 1]$ is the correlation coefficient of the noise components. Let $m = \inf_x f(x)$ and $M = \sup_x f(x)$. With a little of linear algebra we observe that $\mathbf{\Sigma} = \mathbf{P}\mathbf{D}\mathbf{P}^T$ where $\mathbf{P} = \mathbf{P}^T = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$ is orthonormal and $\mathbf{D} = \sigma^2 \begin{pmatrix} 1+\rho & 0 \\ 0 & 1-\rho \end{pmatrix}$ is diagonal. We compare two noise whitening techniques.

Karhunen–Loève transform (PCA Whitening): Let $\mathbf{W} = \mathbf{D}^{-\frac{1}{2}}\mathbf{P}^T$ be the noise whitening of Karhunen–Loève transform and $\mathbf{W}\mathbf{Z} \triangleq \tilde{\mathbf{Z}} \sim \mathcal{N}(0, \mathbf{I}_2)$,

$$\tilde{\mathbf{Y}} = \mathbf{W}\mathbf{Y} = \sqrt{\frac{2}{1+\rho}} \frac{1}{\sigma} \begin{pmatrix} f(X) \\ 0 \end{pmatrix} + \tilde{\mathbf{Z}}. \tag{44}$$

By Proposition 2, $\mathcal{E}(X \to \tilde{Y}_1) = 2Q\Big(\sqrt{\frac{2}{1+\rho}} \frac{1}{\sigma} \frac{M-m}{2}\Big)$. Furthermore, $\mathcal{E}(X \to \tilde{Y}_2) = 1$ so that Equation (43) becomes an equality here

$$\mathcal{E}(X \to \mathbf{Y}) = 2Q\Big(\sqrt{\frac{2}{1+\rho}} \frac{1}{\sigma} \frac{M-m}{2}\Big) = 1 - \frac{1}{\sigma}\sqrt{\frac{2}{1+\rho}} \frac{M-m}{\sqrt{2\pi}} + O(\sigma^{-2}). \tag{45}$$

Equation (45) is coherent:

- When $\rho = 1$, the leakage is repeated and as expected from Lemma 4 the CDC remains unchanged.
- When $\rho = 0$, we have two independent noises, and it is optimal to average the samples such that the global noise variance is halved.
- As $\rho \to -1$, $\mathcal{E}(X \to \mathbf{Y}) \to 0$ which is expected since averaging both components completely cancels out the noise and $f(X)$ is revealed.

Mahalanobis transform (ZCA Whitening): Let $\mathbf{W} = \mathbf{\Sigma}^{-\frac{1}{2}} = \mathbf{P}\mathbf{D}^{-\frac{1}{2}}\mathbf{P}^T$ be the noise whitening of Mahalanobis transform and $\mathbf{W}\mathbf{Z} \triangleq \tilde{\mathbf{Z}} \sim \mathcal{N}(0, \mathbf{I}_2)$,

$$\tilde{\mathbf{Y}} = \mathbf{\Sigma}^{-\frac{1}{2}}\mathbf{Y} = \sqrt{\frac{1}{1+\rho}} \frac{1}{\sigma} \begin{pmatrix} f(X) \\ f(X) \end{pmatrix} + \tilde{\mathbf{Z}}. \tag{46}$$

By Proposition 2, $\mathcal{E}(X \to \tilde{Y}_i) = 2Q\Big(\sqrt{\frac{1}{1+\rho}} \frac{1}{\sigma} \frac{M-m}{2}\Big)$. Equation (43) becomes

$$\mathcal{E}(X \to \mathbf{Y}) \geqslant 4Q^2\Big(\sqrt{\frac{1}{1+\rho}} \frac{1}{\sigma} \frac{M-m}{2}\Big). \tag{47}$$

Interestingly, with Mahalanobis transform we have an inequality in (43) which shows that the choice of the whitening technique can affect the bound tightness.

3.4 Comparison with the Other Informational Leakage Measures

Let X be a random variable taking values in a finite set $\mathcal{X}$ and a channel $X \to Y$. The following inequalities proved in Appendix D hold:

Lemma 9.

$$\left.\begin{array}{r} \frac{I(X;Y)}{\log|\mathcal{X}|} \leqslant \frac{I(X;Y)}{H(X)} \\ \left.\begin{array}{r} \frac{\mathrm{ARE}(X;Y)}{2\gamma_X\lambda_{\mathrm{TVI}}} \\ \\ \frac{\beta(X;Y)}{2\lambda_{\mathrm{TVI}}} \end{array}\right\} \leqslant \frac{\Delta(X;Y)}{\lambda_{\mathrm{TVI}}} \\ \\ \frac{\exp(\mathcal{L}(X\to Y))-1}{|\mathcal{X}|-1} \end{array}\right\} \leqslant \overline{\mathcal{E}}(X\to Y) \leqslant \left\{\begin{array}{l} \mathrm{ARE}(X;Y) \leqslant \mathrm{RE}(X;Y) \\ \\ \gamma_X\beta(X;Y) \\ \gamma_X\Delta(X;Y) \leqslant \gamma_X\left(\frac{I(X;Y)}{2\log e}\right)^{\frac{1}{2}} \\ \\ (|\mathcal{X}|-1)(\exp\left(\mathcal{L}(X\to Y)\right)-1) \end{array}\right. \tag{48}$$

where H is Shannon entropy, H_2 is the collision entropy, $\lambda_{\mathrm{TVI}} = 1-\exp(-H_2(X))$ and $\gamma_X \triangleq \left(\inf_{x\in\mathcal{X}} p_X(x)\right)^{-1}$. If $X \sim \mathcal{U}(\mathcal{X})$ then $\gamma_X = |\mathcal{X}|$ and $\lambda_{\mathrm{TVI}} = 1 - \frac{1}{|\mathcal{X}|}$.

Lemma 9 does not lower bound the CDC in terms of RE because it is impossible to obtain a meaningful bound. Indeed, as remarked by Masure & Standaert [MS23b] if $X \to Y$ is an erasure channel with an arbitrarily small parameter $\mathcal{E} > 0$ then $\mathrm{RE}(X;Y) = |\mathcal{X}| - 1$. As a consequence, RE cannot provide a smooth reduction from noisy leakages to the random probing model. We compare the different leakage measures in Table. 1 via three criteria:

- the ratio of their lower bound by their upper bound in Lemma 9 when $X \sim \mathcal{U}(\mathcal{X})$ (as a measure of relative looseness);
- their maximal value (which measures their normalization);
- their asymptotic values in the Hamming weight leakage model when $X \sim \mathcal{U}(\mathbb{F}_2^n)$ hence $|\mathcal{X}| = 2^n$ (which measures their performance for a typical leakage model).

This allows us to label the introductive Figure 1. Tab. 1 shows that CDC and ARE have the same asymptotic expression in the Hamming weight leakage model with high noise. While ARE is suboptimal and do not verify the properties of the CDC, it provides a tight reduction to the random probing model in this scenario. However, the range of ARE is $[0, |\mathcal{X}| - 1]$ and its relative looseness is $2(|\mathcal{X}| - 1)$ which indicates that in a sense ARE contains the field size in its definition. In any case, it remains preferable to use the CDC which provides the optimal reduction from noisy leakage to random probing.

4 Direct Proofs *via* CDC and Prouff-Rivain Subsequences

In this section we revisit the direct security proof in the noisy leakage model based on Prouff and Rivain's subsequence decomposition [PR13] to obtain a new derivation in terms of CDC. The subsequence decomposition of Prouff and Rivain is recalled in Subsection 4.1. We first prove security for subsequences of

Table 1: **T** is the ratio of the lower bound by the upper bound in Lemma 9 when $X \sim \mathcal{U}(\mathcal{X})$. **M** indicates the maximal value of the leakage measures. Finally, **H** indicates the asymptotic values of the leakage measure in the Hamming weight leakage model. We used the values of [PGMP19, Prop. 3] for ARE, EN and TVI, for MI we used [BCPZ16], for CDC we used Prop. 2. We derived the value for RE in Appendix E.2.

	$I(X;Y)$	$\Delta(X;Y)$	$\mathcal{L}(X \to Y)$	$\beta(X;Y)$	$\mathrm{RE}(X;Y)$	$\mathrm{ARE}(X;Y)$	$\overline{\mathcal{E}}(X \to Y)$
T	$\frac{\lvert\mathcal{X}\rvert \log \lvert\mathcal{X}\rvert}{\sqrt{2 \log e I(X;Y)}}$	$\lvert\mathcal{X}\rvert - 1$	$(\lvert\mathcal{X}\rvert - 1)^2$	$2(\lvert\mathcal{X}\rvert - 1)$	$+\infty$	$2(\lvert\mathcal{X}\rvert - 1)$	1
M	$\log \lvert\mathcal{X}\rvert$	$1 - \frac{1}{\lvert\mathcal{X}\rvert}$	$\log \lvert\mathcal{X}\rvert$	$\sqrt{1 - \frac{1}{\lvert\mathcal{X}\rvert}}$	$\lvert\mathcal{X}\rvert - 1$	$\lvert\mathcal{X}\rvert - 1$	1
H	$\frac{n \log e}{8} \frac{1}{\sigma^2}$	$\frac{\sqrt{n}}{2\pi\sigma}$	$\frac{n \log e}{\sqrt{2\pi}\sigma}$	$\sqrt{\frac{n}{2\pi 2^n}} \frac{1}{\sigma}$	$2^n - 1$	$\frac{n}{\sqrt{2\pi}\sigma}$	$\frac{n}{\sqrt{2\pi}\sigma}$

type 1 and 2 in Subsection 4.2. The security of type 3 and type 4 subsequences is obtained in Subsection 4.3 and Subsection 4.4 respectively. Finally, we combine the security bounds obtained for each subsequence into a security bound for the whole circuit in Subsection 4.5 and compare it with a MI based bound in Subsection 4.6.

4.1 Subsequence Decomposition

For typical block ciphers like the AES, featuring substitution boxes (denoted by Sboxes), Prouff and Rivain [PR13, § 4.2] decompose the computations in four different types of subsequences:

Type 1 $(z_i \leftarrow g(x_i))_i$ where g is a linear function (of the block cipher)
Type 2 $(x_i \leftarrow g(y_i))_i$ where g is an affine function (of Sbox evaluation)
Type 3 $(v_{i,j} \leftarrow a_i b_j)_{i,j}$ (First step of non-linear secure multiplication)
Type 4 $(t_{i,j} \leftarrow t_{i,j-1} + v_{i,j})_{i,j}$ (Last step of non-linear secure multiplication)

This decomposition has become standard to derive security proofs [PGMP19, MS23b]. Note that in this model it is classically assumed that the gates leak.

The first type of subsequences considers linear operations on a shared uniform variable. The second type of subsequence considers linear operations on a shared polynomial expression of a uniform variable. This is typically the case of linear operations within Sboxes. The third type of subsequences deals with the first part of the ISW multiplications involving the cross-product of the input shares of two (non-necessarily independent) random variables. Finally, the type 4 subsequences correspond to the compression layer of the ISW multiplication.

Flaws for Type 3 Subsequences in the State of the Art In Appendix E, we list some flaws in the preceding direct proofs in the noisy leakage model [PR13], [PGMP19], [MS23b]. While these three proofs are different in nature,

the flaws appear at a similar step: proving security for type 3 subsequences. For [PR13], [PGMP19] it is due to an incorrect derivation of the chain rule for conditional probabilities. For [MS23b] it is due to the fact that the function used in Mrs. Gerber's Lemma is convex in one variable when the others are fixed but not jointly convex. We patch part of the flaws in Appendix E using the reductions of MI, ARE and EN to the CDC presented in Subsection 3.4. In this section, X is a sensitive random variable taking values in $\mathcal{X} = \mathbb{F}_2^n$ that can be expressed as a function of the secret K and a public information (e.g., plaintext or ciphertext).

4.2 Security of Type 1 and Type 2 Subsequences

In [BCG+23, Coro. 1] Béguinot *et al.* leveraged Mrs. Gerber's Lemma (MGL) to derive security bounds for encodings in terms of MI. Masure and Standaert [MS23b, Coro. 4 & 5] showed how to exploit such MGL to prove the security of type 1 and type 2 subsequences. We now show that CDC also verifies a sort of MGL that quantifies the security for both type 1 ($f = \mathrm{id}$) and type 2 subsequences (generic f).

Lemma 10 (Mrs. Gerber's Lemma for CDC, Type 1 and Type 2 Subsequences). *Let $\mathbf{G} = (G_i)_{i=0}^d$ be a d-th order encoding of $G = g(X)$ where g is a given function. Assume that each share leaks independently through the side-channels $(G_i \to Y_i)_{i=0}^d$. Let $\mathbf{Y} \triangleq (Y_0, \ldots, Y_d)$ then, $\overline{\mathcal{E}}(X \to \mathbf{Y}) \leqslant \prod_i \overline{\mathcal{E}}(G_i \to Y_i)$.*

Proof. The key observation is that by the *secret sharing property* a masked value is probed if and only if all of its shares are probed. See Appendix B.9. □

4.3 Security of Type 3 Subsequences

Let $\mathbf{G} = (G_i)_{i=0}^d$ and $\mathbf{H} = (H_i)_{i=0}^d$ be d-th order encodings of $g(X)$ and $h(X)$ where g, h are given functions. This section proves security of type 3 subsequences involving the computations with the pairs (H_i, G_h). We need to introduce family of polynomials to express the security bound:

Definition 17 (Rook Domination Polynomial [Mer24]). *Let $(E_{i,j})_{0 \leqslant i,j \leqslant d}$ be a collection of independents events with respective probabilities $((\overline{\mathcal{E}}_{i,j})_{0 \leqslant i,j \leqslant d})$. Let*

$$\Upsilon((\overline{\mathcal{E}}_{i,j})_{0 \leqslant i,j \leqslant d}) \triangleq \mathbb{P}\left((\cap_{i=0}^d \cup_{j=0}^d E_{i,j}) \cup (\cap_{j=0}^d \cup_{i=0}^d E_{i,j})\right). \tag{49}$$

For short $\Upsilon_d(\overline{\mathcal{E}}) \triangleq \Upsilon((\overline{\mathcal{E}}_{i,j})_{0 \leqslant i,j \leqslant d})$ when for all i, j we have $\overline{\mathcal{E}}_{i,j} = \overline{\mathcal{E}}$.

In fact, Υ_d corresponds to the domination polynomial of the rook graph [Mer24]. It can be sandwiched explicitly as follows:

Proposition 3.

$$\max\{\prod_{i=0}^{d}(1-\prod_{j=0}^{d}\mathcal{E}_{i,j}),\prod_{j=0}^{d}(1-\prod_{i=0}^{d}\mathcal{E}_{i,j})\}\leqslant\Upsilon((\overline{\mathcal{E}}_{i,j})_{0\leqslant i,j\leqslant d}) \tag{50}$$

$$\leqslant\min\{\prod_{i=0}^{d}(1-\prod_{j=0}^{d}\mathcal{E}_{i,j})+\prod_{j=0}^{d}(1-\prod_{i=0}^{d}\mathcal{E}_{i,j}),1\}. \tag{51}$$

In particular when for all i,j, $\overline{\mathcal{E}}_{i,j}=\overline{\mathcal{E}}$ it yields:

$$\left(1-\mathcal{E}^{d+1}\right)^{d+1}\leqslant\Upsilon_d(\overline{\mathcal{E}})\leqslant\min\{2\left(1-\mathcal{E}^{d+1}\right)^{d+1},1\}. \tag{52}$$

Proof. By monotonicity of probability,

$$\Upsilon((\overline{\mathcal{E}}_{i,j})_{0\leqslant i,j\leqslant d})=\mathbb{P}\left((\cap_{i=0}^{d}\cup_{j=0}^{d}E_{i,j})\cup(\cap_{j=0}^{d}\cup_{i=0}^{d}E_{i,j})\right) \tag{53}$$

$$\geqslant\mathbb{P}(\cap_{i=0}^{d}\cup_{j=0}^{d}E_{i,j})=\prod_{i=0}^{d}(1-\prod_{j=0}^{d}\mathcal{E}_{i,j}). \tag{54}$$

And similarly $\Upsilon((\overline{\mathcal{E}}_{i,j})_{0\leqslant i,j\leqslant d})\geqslant\prod_{j=0}^{d}(1-\prod_{i=0}^{d}\mathcal{E}_{i,j})$. As a consequence

$$\max\{\prod_{i=0}^{d}(1-\prod_{j=0}^{d}\mathcal{E}_{i,j}),\prod_{j=0}^{d}(1-\prod_{i=0}^{d}\mathcal{E}_{i,j})\}\leqslant\Upsilon((\overline{\mathcal{E}}_{i,j})_{0\leqslant i,j\leqslant d}). \tag{55}$$

Also, by the union bound,

$$\Upsilon((\overline{\mathcal{E}}_{i,j})_{0\leqslant i,j\leqslant d})\leqslant\mathbb{P}(\cap_{i=0}^{d}\cup_{j=0}^{d}E_{i,j})+\mathbb{P}(\cap_{j=0}^{d}\cup_{i=0}^{d}E_{i,j}) \tag{56}$$

$$=\prod_{i=0}^{d}(1-\prod_{j=0}^{d}\mathcal{E}_{i,j})+\prod_{j=0}^{d}(1-\prod_{i=0}^{d}\mathcal{E}_{i,j}). \tag{57}$$

Finally, $\Upsilon((\overline{\mathcal{E}}_{i,j})_{0\leqslant i,j\leqslant d})\leqslant 1$ since it is a probability. □

Stephan Mertens [Mer24, Thm. 4] provides a recursive formula for this polynomial, which gives an efficient way to compute exhaustively the coefficients of Υ_d. The coefficients of the first polynomials Υ_d for $d\leqslant 5$ are shown in Tab. 2. Additional properties are mentioned in Appendix C.

Rationale For The Rook Domination Polynomial Within type 3 subsequences of ISW the cross-wise terms G_iH_j are computed so that each pair (G_i,H_j) leaks. After degradation into an erasure channel, for each pair (i,j) the degraded adversary $\mathcal{A}$ either probes (G_i,H_j) or receives an erasure symbol. Let $E_{i,j}$ be the event that $\mathcal{A}$ probes the pair (G_i,H_j). By the secret sharing property, the sensitive value leaks if and only if we probe each G_i or each H_j.

Let $E\triangleq(\cap_{i=0}^{d}\cup_{j=0}^{d}E_{i,j})\cup(\cap_{j=0}^{d}\cup_{i=0}^{d}E_{i,j})$. Equation (49) defines $\Upsilon\triangleq\mathbb{P}(E)$. If one represents the $E_{i,j}$ in a checkerboard, the sensitive value is probed if and

Table 2: Some explicit values of $\Upsilon_d(\overline{\mathcal{E}})$ from small values of d.

d	$\Upsilon_d(\overline{\mathcal{E}})$
0	$\overline{\mathcal{E}}$
1	$6\overline{\mathcal{E}}^2\mathcal{E}^2 + 4\overline{\mathcal{E}}^3\mathcal{E} + \overline{\mathcal{E}}^4$
2	$48\overline{\mathcal{E}}^3\mathcal{E}^6 + 117\overline{\mathcal{E}}^4\mathcal{E}^5 + 126\overline{\mathcal{E}}^5\mathcal{E}^4 + 84\overline{\mathcal{E}}^6\mathcal{E}^3 + 36\overline{\mathcal{E}}^7\mathcal{E}^2 + 9\overline{\mathcal{E}}^8\mathcal{E} + \overline{\mathcal{E}}^9$
3	$488\overline{\mathcal{E}}^4\mathcal{E}^{12} + 2640\overline{\mathcal{E}}^5\mathcal{E}^{11} + 6712\overline{\mathcal{E}}^6\mathcal{E}^{10} + 10864\overline{\mathcal{E}}^7\mathcal{E}^9 + 12726\overline{\mathcal{E}}^8\mathcal{E}^8 + 11424\overline{\mathcal{E}}^9\mathcal{E}^7$ $+8008\overline{\mathcal{E}}^{10}\mathcal{E}^6 + 4368\overline{\mathcal{E}}^{11}\mathcal{E}^5 + 1820\overline{\mathcal{E}}^{12}\mathcal{E}^4 + 560\overline{\mathcal{E}}^{13}\mathcal{E}^3 + 120\overline{\mathcal{E}}^{14}\mathcal{E}^2 + 16\overline{\mathcal{E}}^{15}\mathcal{E} + \overline{\mathcal{E}}^{16}$
4	$6130\overline{\mathcal{E}}^5\mathcal{E}^{20} + 58300\overline{\mathcal{E}}^6\mathcal{E}^{19} + 269500\overline{\mathcal{E}}^7\mathcal{E}^{18} + 808325\overline{\mathcal{E}}^8\mathcal{E}^{17} + 1778875\overline{\mathcal{E}}^9\mathcal{E}^{16}$ $+3075160\overline{\mathcal{E}}^{10}\mathcal{E}^{15} + 4349400\overline{\mathcal{E}}^{11}\mathcal{E}^{14} + 5154900\overline{\mathcal{E}}^{12}\mathcal{E}^{13} + 5186300\overline{\mathcal{E}}^{13}\mathcal{E}^{12}$ $+4454400\overline{\mathcal{E}}^{14}\mathcal{E}^{11} + 3268360\overline{\mathcal{E}}^{15}\mathcal{E}^{10} + 2042950\overline{\mathcal{E}}^{16}\mathcal{E}^9 + 1081575\overline{\mathcal{E}}^{17}\mathcal{E}^8$ $+480700\overline{\mathcal{E}}^{18}\mathcal{E}^7 + 177100\overline{\mathcal{E}}^{19}\mathcal{E}^6 + 53130\overline{\mathcal{E}}^{20}\mathcal{E}^5 + 12650\overline{\mathcal{E}}^{21}\mathcal{E}^4 + 2300\overline{\mathcal{E}}^{22}\mathcal{E}^3$ $+300\overline{\mathcal{E}}^{23}\mathcal{E}^2 + 25\overline{\mathcal{E}}^{24}\mathcal{E}^1 + \overline{\mathcal{E}}^{25}$
5	$92592\overline{\mathcal{E}}^6\mathcal{E}^{30} + 1356480\overline{\mathcal{E}}^7\mathcal{E}^{29} + 9859140\overline{\mathcal{E}}^8\mathcal{E}^{28} + 47187180\overline{\mathcal{E}}^9\mathcal{E}^{27} + 167284836\overline{\mathcal{E}}^{10}\mathcal{E}^{26}$ $+469268496\overline{\mathcal{E}}^{11}\mathcal{E}^{25} + 1086623400\overline{\mathcal{E}}^{12}\mathcal{E}^{24} + 2137381200\overline{\mathcal{E}}^{13}\mathcal{E}^{23} + 3642777000\overline{\mathcal{E}}^{14}\mathcal{E}^{22}$ $+5453014080\overline{\mathcal{E}}^{15}\mathcal{E}^{21} + 7235196885\overline{\mathcal{E}}^{16}\mathcal{E}^{20} + 8558765100\overline{\mathcal{E}}^{17}\mathcal{E}^{19} + 9057864300\overline{\mathcal{E}}^{18}\mathcal{E}^{18}$ $+8591124600\overline{\mathcal{E}}^{19}\mathcal{E}^{17} + 7305959610\overline{\mathcal{E}}^{20}\mathcal{E}^{16} + 5567447160\overline{\mathcal{E}}^{21}\mathcal{E}^{15} + 3796214400\overline{\mathcal{E}}^{22}\mathcal{E}^{14}$ $+2310778800\overline{\mathcal{E}}^{23}\mathcal{E}^{13} + 1251676800\overline{\mathcal{E}}^{24}\mathcal{E}^{12} + 600805260\overline{\mathcal{E}}^{25}\mathcal{E}^{11} + 254186856\overline{\mathcal{E}}^{26}\mathcal{E}^{10}$ $+94143280\overline{\mathcal{E}}^{27}\mathcal{E}^9 + 30260340\overline{\mathcal{E}}^{28}\mathcal{E}^8 + 8347680\overline{\mathcal{E}}^{29}\mathcal{E}^7 + 1947792\overline{\mathcal{E}}^{30}\mathcal{E}^6 + 376992\overline{\mathcal{E}}^{31}\mathcal{E}^5$ $+58905\overline{\mathcal{E}}^{32}\mathcal{E}^4 + 7140\overline{\mathcal{E}}^{33}\mathcal{E}^3 + 630\overline{\mathcal{E}}^{34}\mathcal{E}^2 + 36\overline{\mathcal{E}}^{35}\mathcal{E}^1 + \overline{\mathcal{E}}^{36}$

only if at least one event $E_{i,j}$ occurs at least once in each line $(\cap_{i=0}^{d} \cup_{j=0}^{d} E_{i,j})$ or at least once in each column $(\cap_{j=0}^{d} \cup_{i=0}^{d} E_{i,j})$. This results in Equation (49).

Considering that $E_{i,j}$ occurs when a rook is placed at coordinates (i, j) of the checkerboard, the event E is realized if and only if every position in the checkerboard is attacked (or occupied) by at least one rook. The domination polynomial of the rook graph counts the number of such configurations of m rooks. Therefore, it allows one to compute Υ_d when all $E_{i,j}$ are equiprobable.

We now show the security of type 3 subsequences using the domination polynomial of the rook graph:

Lemma 11 (Type 3 Subsequences). *Consider the channels* $((G_i, H_j) \to Y_{i,j})_{0 \leqslant i,j \leqslant d}$ *and let* $\mathbf{Y} \triangleq (Y_{i,j})_{0 \leqslant i,j \leqslant d}$. *Then one has*

$$\overline{\mathcal{E}}(X \to \mathbf{Y}) \leqslant \Upsilon((\overline{\mathcal{E}}((G_i, H_j) \to Y_{i,j}))_{0 \leqslant i,j \leqslant d}). \tag{58}$$

Proof. See Appendix B.10. □

4.4 Security of Type 4 Subsequences

Type 4 subsequences consider the compression layer in multiplication gadgets. At this stage, the sensitive variable is masked in $(d+1)^2$ shares. In the compression layer, the shares are grouped in $d+1$ groups each of size $d+1$, and are recombined to obtain a d-th order encoding of the sensitive variable. Formally, let $(V_{i,j})$ be

an encoding in $(d+1)^2$ shares of $f(X)$ where f is a given function. Let $T_{i,0} = V_{i,0}$ and $T_{i,j} = T_{i,j-1} \oplus V_{i,j}$. In particular $(T_{i,d})_{i=0}^{d}$ is a d-th order encoding of $f(X)$.

Lemma 12 (Cumulative Sum). *Consider the cumulative sum function* $h_d : (x_0, \ldots, x_d) \in \mathbb{F}^{d+1} \mapsto (x_0, x_0 + x_1, \ldots, x_0 + \ldots + x_d) \in \mathbb{F}^{d+1}$. $(V_{i,0}, \ldots, V_{i,d})$ *is a* d*-th order sharing of* $T_{i,d}$ *and since* $T_{i,j} - T_{i,j-1} = V_{i,j}$*, a channel from* $(T_{i,j-1}, V_{i,j})$ *can be seen as a channel from the pair* $(T_{i,j}, T_{i,j-1})$*. Let*

$$\mathbf{T}_i = (T_{i,0}, \ldots, T_{i,d+1}) \to \boxed{p_{\mathbf{Y}|\mathbf{T}_i} \triangleq \prod_{j=0}^{d} p_{Y_{i,j}|(T_{i,j-1},T_{i,j})}} \to \mathbf{Y}_i = (Y_{i,0}, \ldots, Y_{i,d})$$

with the convention that $T_{i,-1} = 0$*. Then the channel rewrites,*

$$T_{i,d} \to \boxed{\text{Mask}_d} \to \mathbf{V_i} = (V_{i,0}, \ldots, V_{i,d}) \to \boxed{h_d} \to \mathbf{T}_i \to \boxed{P_{\mathbf{Y}|\mathbf{T}_i}} \to \mathbf{Y}_i \quad (59)$$

and $\overline{\mathcal{E}}(T_{i,d} \to \mathbf{Y}_i) \leqslant \overline{\mathcal{E}}((T_{i,d-1}, V_{i,d}) \to Y_{i,d})$.

Proof. See Appendix B.11. □

We obtain now the following security result for type 4 subsequences:

Lemma 13 (Type 4 Subsequences). *Consider* $((T_{i,j-1}, V_{i,j}) \to Y_{i,j})_{0 \leqslant i,j \leqslant d}$ *and let* $\mathbf{Y} = (Y_{i,j})_{0 \leqslant i,j \leqslant d}$ *then,* $\overline{\mathcal{E}}(X \to \mathbf{Y}) \leqslant \prod_{i=0}^{d} \overline{\mathcal{E}}((T_{i,d-1}, V_{i,d}) \to Y_{i,d})$.

Proof. Combine Lemma 10 with Lemma 12. □

4.5 Security for the Whole Circuit

Combining Lemmas 10,11,13 for the different subsequences together with the single-letterization Lemma 6 we obtain a security guarantee for the whole circuit:

Theorem 2 (Direct Security Proof). *Consider an implementation with* n_i *subsequences of type* i *(*$i = 1, 2, 3, 4$*) and a* $\overline{\mathcal{E}}$*-noisy adversary with respect to* CDC *with* q *queries. Let* $\mathbf{Y}$ *be the vector of all corresponding side-informations acquired by the chosen channel adversary. Then one has*

$$0 \leqslant \overline{\mathcal{E}}(K \to \mathbf{Y}) \leqslant 1 - \left(\left(1 - \overline{\mathcal{E}}^{d+1}\right)^{n_1+n_2+n_4} \left(1 - \Upsilon_d(\overline{\mathcal{E}})\right)^{n_3}\right)^q \leqslant 1. \quad (60)$$

The upper bound can be weakened via the union bound *to*

$$\overline{\mathcal{E}}(K \to \mathbf{Y}) \leqslant q\left((n_1 + n_2 + n_4) + 2n_3(d+1)^{d+1}\right)\overline{\mathcal{E}}^{d+1}. \quad (61)$$

Also, using the asymptotic equivalent of the domination polynomial of the rook graph the upper bound of Equation (60) *is asymptotically equivalent to*

$$q\left(n_1 + n_2 + \left(2(d+1)^{d+1} - (d+1)!\right)n_3 + n_4\right)\overline{\mathcal{E}}^{d+1}. \quad (62)$$

Remark 13. Equation (61) is of a similar form as [MS23b, Thm. 5] but the constants n_i are not scaled by a term depending on the field size $|\mathcal{X}|$, contrary to the t_i occurring in [MS23b, Thm. 7].

Proof. See Appendix B.12. □

De Chérisey *et al.* obtained a lower bound on the minimum number of queries required by the adversary to achieve a given advantage in terms of SR in the unprotected setting with MI [dCGRP19, Thm. 2, Eqn. 4]. Béguinot *et al.* derived a tight bound for masked encoding with MI [BCG+23, Coro. 2] and maximal leakage [BLR+23, Coro. 1]. Liu *et al.* derived a tight bound for masked encoding with Sibson's α-information of order 2 [LBC+23, Thm. 2]. The combination of Theorem 2 with Proposition 1 yields a lower bound on the minimum number of queries required by the adversary to achieve a given advantage in terms of SR, GE or TVI for the entire protected computations (not only encodings).

Theorem 3 (Lower Bound on the Number of Queries). *Let*

$$\lambda(\overline{\mathcal{E}}, d) = \left(\ln\left(\left(1-\overline{\mathcal{E}}^{d+1}\right)^{n_1+n_2+n_4}\left(1-\Upsilon_d(\overline{\mathcal{E}})\right)^{n_3}\right)\right)^{-1} \tag{63}$$

$$= \left((n_1+n_2+n_4)\log\left(1-\overline{\mathcal{E}}^{d+1}\right) + n_3\log\left(1-\Upsilon_d(\overline{\mathcal{E}})\right)\right)^{-1} \tag{64}$$

$$\approx \left(\left(n_1+n_2+n_4+n_3(2(d+1)^{d+1}-(d+1)!)\right)\overline{\mathcal{E}}^{d+1}\right)^{-1}. \tag{65}$$

The number of queries to achieve $\mathbb{P}_{s,o}(K|\mathbf{Y}) = \mathbb{P}_{s,o}$, $G(K|\mathbf{Y}) = G$ *or* $\Delta(K;Y) = \Delta$ *is at least:*

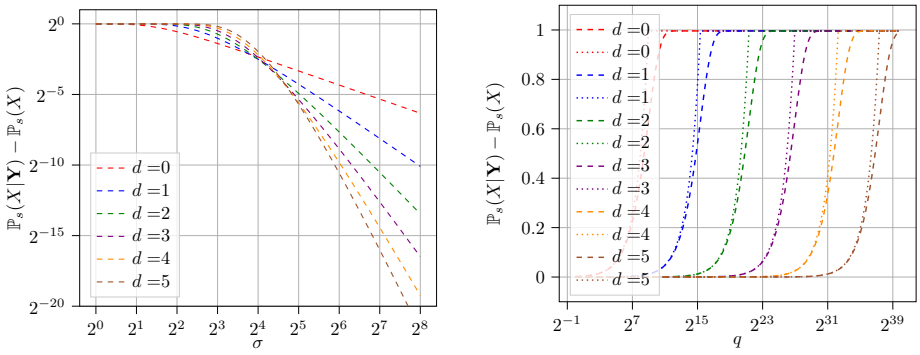
$$\begin{aligned} q_{\mathrm{sr}} &\geqslant \lambda(\overline{\mathcal{E}}, d)\ln\left((1-\mathbb{P}_{s,o})^{-1}\lambda_{\mathrm{SR}_o}\right), \\ q_{\mathrm{ge}} &\geqslant \lambda(\overline{\mathcal{E}}, d)\ln\left((G-1)^{-1}\lambda_{\mathrm{GE}}\right), \\ q_{\mathrm{tvi}} &\geqslant \lambda(\overline{\mathcal{E}}, d)\ln\left(\Delta^{-1}\lambda_{\mathrm{TVI}}\right). \end{aligned} \tag{66}$$

Theorem 3 is illustrated by Figure. 3. It shows how it behaves for a fixed number of queries and increasing level of noise in Figure 3a and how it behaves for a fixed value of the CDC and increasing number of queries q in Figure 3b. We can observe on Figure. 3a that there is an optimal masking order with respect to Theorem 3 depending on the noise level. Further, Figure. 3b shows that the bound benefits from the single letterization of Lemma 5 and Lemma 6 as it is an "S"-shaped curve. The weakened version of Lemma 6 into a linear bound corresponds to the dotted line.

4.6 Comparison With Bounds Based on Mutual Information

Theorem 2 provides bounds on all figures of merits that we compare with bounds based on MI. Let $K \to Y$ be a side-channel of a uniform key $K \sim \mathcal{U}(\mathbb{F}_2^n)$. Let $Y = \mathrm{HW}(K) + N$ where $N \sim \mathcal{N}(0, \sigma^2)$ is an additive Gaussian noise. Let $\mathbf{Y} = (Y_1, \ldots, Y_q)$ be a side-channel leakage with q queries.

No Constraint on the Noise Level A strength of Theorem 2 is that the bound on the CDC (Equation (60)) is always less than 1. In particular, our bound does not require any constraint on the noise level to apply. This is by contrast with MI based bound such as [MS23b].



(a) One query ($q = 1$) and variable σ under Hamming weight leakage model.

(b) $\overline{\mathcal{E}} = 2 \times 10^{-3}$ and increasing number of queries q.

Fig. 3: Plots of Thm. 3 for one subsequence of type 3 ($n_1 = n_2 = n_4 = 0, n_3 = 1$) and different masking order. The y-axis corresponds to the adversary's advantage in terms of success rate.

Scaling with the Number of Queries De Chérisey *et al.* [dCGRP19, Lemma 5] established an upper bound on the MI for side-channels that grows linearly with respect to q i.e., $I(K;\mathbf{Y}) \leqslant qI(K;Y)$. However, the MI $I(K;\mathbf{Y})$ is bounded above by the constant $\log|\mathcal{K}|$. Therefore, the linear bound cannot be tight for large values of q. Finding an alternative practical non-linear bound on the MI that remains bounded when q increases remains an interesting open question. However, our bound $\overline{\mathcal{E}}(K \to \mathbf{Y}) \leqslant 1 - (1 - \overline{\mathcal{E}}(K \to Y))^q \leqslant q\overline{\mathcal{E}}(K \to Y)$ of Lemma 6 for the CDC is nonlinear with respect to q and can be weakened into a linear bound using Boole's inequality. This linear approximation is tight when $\overline{\mathcal{E}} \to 0$.

Attack with One Trace (SPA) and Multiple Traces (DPA) On the one hand, Béguinot *et al.* [BLR+23, Eqn. 87] showed that as $I(K;\mathbf{Y}) \to 0$, we have

$$\mathbb{P}_s(K|\mathbf{Y}) - \mathbb{P}_s(K) \leqslant \sqrt{\frac{2(2^n-1)}{2^{2n}} \frac{qI(K;Y)}{\log e}}. \tag{67}$$

On the other hand, Proposition 1 yields

$$\mathbb{P}_s(K|\mathbf{Y}) - \mathbb{P}_s(K) \leqslant (1 - (1 - \overline{\mathcal{E}}(K \to Y))^q)\lambda_{\mathrm{SR}} \approx q\overline{\mathcal{E}}(K \to Y)\lambda_{\mathrm{SR}}. \tag{68}$$

While $\delta_{\mathrm{MI}} \propto \frac{1}{\sigma^2}$ [BCPZ16, Appendix A] and $\delta_{\mathrm{CDC}} \propto \frac{1}{\sigma}$ (Proposition 2), the square root in the MI-based security bounds (e.g., Eqn. (67)) makes both bounds on the SR advantage to be $O(\sigma^{-1})$. As a consequence, the CDC-based bound is comparable with the MI-based bound for single trace attack ($q = 1$) and at large noise.

However, in the context of a DPA where the goal is to lower bound the minimum number of queries q to achieve a given figure of merit the MI based bound Eqn. (67) will be in $O(\sigma^2)$ while the CDC based bound Eqn. (67) is only $O(\sigma)$. In the presence of masking the same observation applies to Theorem 3 and [MS23b, Theorem 7] where σ is replaced by σ^{d+1}. In conclusion CDC may not be the most suitable informational noisiness measure to capture the leakage in a DPA with many traces. However, since the factorization is optimal this leads to the important conclusion that such a loss is inherent to a reduction from noisy leakages to the random probing model. For this reason informational bounds based on MI remain an interesting tool for side-channel analysis.

5 Indirect Proofs *via* CDC With Random Probing

In this Section, we explain how existing proof in the random probing model can be combined with the CDC. We show how a security proof in the random probing model can be lifted into the noisy leakage model by improving [DDF14, Thm. 1] in Subsection 5.1. The resulting bound is shown to yield an upper bound on the side-channel adversary's advantage in Subsection 5.2. The asymptotic behavior of the security bound is analyzed in Subsection 5.3. This analysis confirms theoretically the finding from Battistello *et al.* [BCPZ16] that increasing indefinitely the masking order of ISW gadgets whose noise rate is not constant can be detrimental to security.

5.1 Lifting Security Proof in the Random Probing Model to Noisy Leakage

We first refer back the existing security proofs in the random probing model. Cassiers *et al.* [CFOS21] showed how to derive tight security bounds in the random probing model using *probe distribution table* (PDT). Belaïd *et al.* [BCP$^+$20, BRT21] also derived security proof in the random probing model based on an *expansion strategy* of small gadgets. Their improvements are based on the fact that when an adversary probes more than t wires it does not necessarily learn any information about the sensitive variable. The t-threshold probing security ensures that no subset of at most t wires leak information. However, some subsets of more than t wires does not leak information either. By carefully determining the subsets of leaking wires and the subset of non-leaking wires the so-called *probability of simulation failure* can be reduced.

Our goal is to show how we can lift a proof in the random probing model to the noisy leakage model using CDC. To do so we show how to improve [DDF14, Thm. 1]. In the same way the above-mentioned proof in the random probing model can be lifted to the noisy leakage model using the CDC.

In this section, the circuit Γ is decomposed in $|\Gamma|$ regions (gadgets) whose numbers of wires is specified by the sequence (l_i). We assume that Γ is secure in the region probing model, i.e. any set of at most t (probed) wires in each region of the circuit is independent with the secret key.

Theorem 4 (Indirect Security Proof). *Let $\mathcal{A}$ be a $\overline{\mathcal{E}}$-noisy adversary with respect to* CDC *with q queries. Let $\mathbf{Y}$ be the vector of all corresponding side-information acquired by the chosen channel adversary. Then one has*

$$\mathcal{E}(K \to \mathbf{Y}) \leqslant \mathrm{fail}(t, (l_i), \overline{\mathcal{E}}, q) \triangleq 1 - \prod_{i=1}^{|\Gamma|} \left(1 - Q_B(t, l_i, \overline{\mathcal{E}})\right)^q \leqslant q \sum_{i=1}^{|\Gamma|} Q_B(t, l_i, \overline{\mathcal{E}}). \tag{69}$$

Remark 14. Here the wires leak while the gates leak in Section 4.

Proof. See Appendix B.13. □

Theorem 4 is very generic, it is "agnostic" to the countermeasure implemented to achieve security against t-threshold probing adversary in the region probing model. Masure & Standaert observed in [MS23b, Tab. 1] that [DDF14, Thm. 1] does not provide incentive to noisier leakage. Theorem 4 does not suffer from this weakness. In particular, the adversary's advantage vanishes as the noise level increases. Theorem 4 depends on five parameters:

1. The noise level as quantified by $\overline{\mathcal{E}}$.
2. The security order of the gadgets as quantified by t.
3. The "temporal attack surface" quantified by the number of queries q.
4. The attack surface of the adversary within a gadget as measured by the number l of wires potentially probed within the gadgets.
5. The attack surface of the adversary on the whole circuit as measured by $|\Gamma|$.

5.2 Bounds on SCA Advantage

If a cryptographic algorithm is insecure against a black-box adversary then it is also insecure against a side-channel adversary. For this reason we are interested in the advantage of the side-channel adversary compared to its black-box counterpart. In Proposition 4 we upper bound this advantage, contrary to the usual bound for side-channel attacks it depends on the computational assumption made on the adversary through the term $\mathbb{P}^{\mathrm{BB}}_{s,o}(q)$.

Proposition 4. *Let $\mathcal{A}$ be a $\overline{\mathcal{E}}$-noisy with respect to* CDC *adversary with q queries. Let $\mathbb{P}^{\mathrm{BB}}_{s,o}(q)$ and $\mathbb{P}^{\mathrm{SCA}}_{s,o}(q, \overline{\mathcal{E}})$ be the respective SR_o of the best black box and side-channel adversary with q queries. Then,*

$$\begin{cases} \mathbb{P}^{\mathrm{SCA}}_{s,o}(q, \overline{\mathcal{E}}) - \mathbb{P}^{\mathrm{BB}}_{s,o}(q) & \leqslant \left(1 - \mathbb{P}^{\mathrm{BB}}_{s,o}(q)\right) \ \mathrm{fail}(t, (l_i), \overline{\mathcal{E}}, q) \\ \mathrm{GE}^{\mathrm{BB}}(q) - \mathrm{GE}^{\mathrm{SCA}}(q, \overline{\mathcal{E}}) & \leqslant \left(\mathrm{GE}^{\mathrm{BB}}(q) - 1\right) \mathrm{fail}(t, (l_i), \overline{\mathcal{E}}, q). \end{cases} \tag{70}$$

Proof. This is a direct consequence of Theorem 4 and Proposition 1 applied to a (possibly computationally bounded) adversary. □

5.3 Properties of the Failure Probability Function

We analyze how our bound depends on the order of an implemented countermeasure. Consider a masking countermeasure of order d, and for illustrative purposes assume that $t = d$ (which would obtained for the implementation of [RP10b] separated by leak-free refresh). Also assume that the l_i associated with the multiplication gadgets grow quadratically with respect to d (e.g., ISW). We use the shorthand notations $l(d) = \max_i l_i(d)$, $\mathrm{fail}(d, \overline{\mathcal{E}}) = \mathrm{fail}(t(d), (l_i(d)), \overline{\mathcal{E}}, 1)$ and $\mathrm{fail}(d, \overline{\mathcal{E}}, q) = \mathrm{fail}(t(d), (l_i(d)), \overline{\mathcal{E}}, q)$. We can define an optimal masking order with respect to Theorem 4:

$$d^*(\overline{\mathcal{E}}) \triangleq \arg\min_{d \in \mathbb{N}} \mathrm{fail}(d, \overline{\mathcal{E}}, q). \tag{71}$$

Since $\mathrm{fail}(d, \overline{\mathcal{E}}, q) = 1 - \left(1 - \mathrm{fail}(d, \overline{\mathcal{E}})\right)^q$ we have $d^*(\overline{\mathcal{E}}) = \arg\min_{d \in \mathbb{N}} \mathrm{fail}(d, \overline{\mathcal{E}})$ so that the optimal masking order is independent of the number of queries.

It is not true in general that $\mathrm{fail}(d, \overline{\mathcal{E}})$ is a decreasing function of the masking order d. Though we show that it essentially holds in the limits of high noise.

Proposition 5. *For all $d_1 < d_2$ there exist a noise threshold $\overline{\mathcal{E}}_0$ such that for all noise level $\overline{\mathcal{E}} \leqslant \overline{\mathcal{E}}_0$, $\mathrm{fail}(d_1, \overline{\mathcal{E}}) > \mathrm{fail}(d_2, \overline{\mathcal{E}})$, which indicates that $d^*(\overline{\mathcal{E}}) \xrightarrow{\overline{\mathcal{E}} \to 0} \infty$.*

Proof. See Appendix B.14. □

Proposition 5 means that while the bound is not decreasing with respect to the masking order d there always exists a noise level for which masking at higher order is more interesting.

Proposition 6. *If $l(d)/t(d) \xrightarrow{d \to \infty} \infty$ then $\mathrm{fail}(d, \overline{\mathcal{E}}) \xrightarrow{d \to \infty} 1$. Therefore, there exist a finite optimal masking order with respect to the noise level $d^*(\overline{\mathcal{E}}) < \infty$ and $\mathrm{fail}(d, \overline{\mathcal{E}})$ cannot be reduced further than $\mathrm{fail}(d^*, \overline{\mathcal{E}}) > 0$. This in turn implies that the adversary's advantage cannot be made arbitrarily small.*

Proof. See Appendix B.15. □

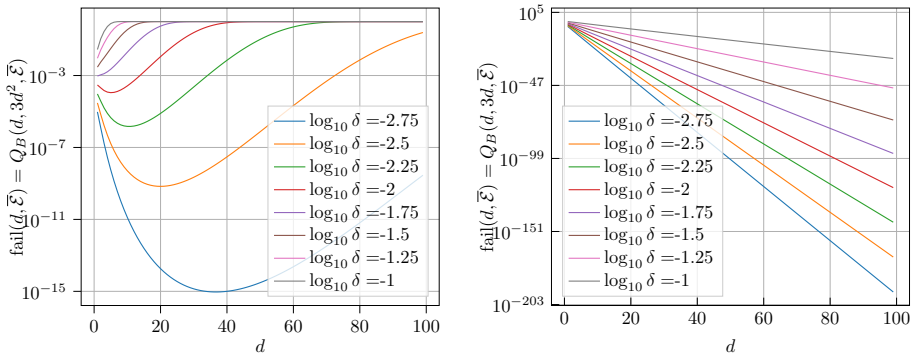
Proposition 6 means that for a fixed level of noise increasing indefinitely the masking order is detrimental to the security bounds. As a consequence, there exists a finite optimal masking order with respect to the noise level.

Proposition 7. *If $\overline{\mathcal{E}} < \overline{\mathcal{E}}_0 \leqslant \frac{t}{l}$ then $Q_B(t, l, \overline{\mathcal{E}}) \leqslant \exp\left(-l d_{\mathrm{KL}}\left(t/l \middle\| \overline{\mathcal{E}}\right)\right) \leqslant \exp\left(-l d_{\mathrm{KL}}\left(\overline{\mathcal{E}}_0 \middle\| \overline{\mathcal{E}}\right)\right)$ where $d_{\mathrm{KL}}(p \| q) \triangleq D_{\mathrm{KL}}(\mathcal{B}(p) \| \mathcal{B}(q)) = p \log \frac{p}{q} + \bar{p} \log \frac{\bar{p}}{\bar{q}}$. In this case $d^*(\overline{\mathcal{E}}) = \infty$.*

Proof. We apply Chernoff-Hoeffding bound [Hoe94]. □

Proposition 7 shows that when the gadget size grows at most linearly with respect to the security parameter t then masking provides an exponential gain with respect to the gadget size. The coefficient in the exponential is lower bounded by a binary divergence between a *protection rate* $\frac{t}{l}$ and a *leakage rate* $\overline{\mathcal{E}}$.

The original expression of [DDF14, Thm. 1] can be misinterpreted as it seems that the probability of error converges to 0 when d increases. But when the gadget size l grows quadratically with respect to d then the maximum value of d tolerated in the proof is upper bounded by a function inversely proportional to the noise level. Hence, it is not true to say that the advantage of the adversary decreases exponentially with respect to d. This confirms the observations from [BCPZ16] where Battistello *et al.* observed that the noise is expected to decrease linearly with the number of shares and that this assumption is not met in practice. Figure 4a derived from Theorem 4 shows that for quadratic gadget and a fixed level of noise, the advantage of the adversary is either increasing or decreasing and then increasing with respect to the masking order depending on the noise level. For linear gadget, Figure 4b shows that masking does provide an exponential gain with respect to the adversaries advantage. Note that this is a weakness of the ISW gadget whose noise rate is not constant and not a weakness of the bound. This emphasizes the importance to obtain gadget with improved noise rate [CS19]. In particular quasi-linear masking scheme [GJR18, GPRV22, CDGT24] and Toom-Cook based gagdets [Pla22] are promising approaches.



(a) Quadratic gadget (multiplication) with $t = d$ and $l = 3d^2$.

(b) Linear gadget (linear operation) with $t = d$ and $l = 3d$.

Fig. 4: Evolution of Theorem 4 for different masking order d for quadratic and linear gadget.

Remark 15. Proposition 6 may appear as contradictory with [DFS19, Coro. 2]. It turns out that [DFS19, Coro. 2] is incorrectly derived from [DFS19, Thm. 3]. See the Appendix E.6 for more details.

6 Conclusion

We showed how the complementary Doeblin coefficient (CDC) can be used to reduce optimally a noisy adversary to a random probing adversary. This allows us to exhibit the unavoidable inherent cost of a reduction from noisy leakages to the random probing model. We derived a set of properties of the CDC which makes it a sound leakage measure that is easy to manipulate and showed that it is amenable to evaluation in a multivariate setting.

The CDC yields security bounds for all figures of merits that scale well with the number of side-channel queries (single letterization property). As a byproduct we also lower bounded the minimum number of queries to achieve a given figure of merit.

Furthermore, security bounds in terms of CDC are easily derived using the Prouff-Rivain subsequence decomposition or can be naturally combined with existing security proofs in the random probing model for any type of countermeasures. We analyzed the asymptotic behavior of the obtained bounds in terms of countermeasure order and confirmed the existence of an optimal masking order with respect to the security bounds.

Overall, we believe that these contributions are essential to ground the security of masked implementations in the noisy leakage model on solid foundations. As perspectives, we would like to obtain direct security proof of code-based masking implementation using CDC leveraging a new appropriate subsequence decomposition. Investigating formal security proof of masking combined with shuffling [ABG+22] using CDC could also be relevant as a way to reduce the physical noise requirement to obtain relevant security parameters.

Acknowledgements

We are grateful to Stephan Mertens that provided us an efficient way to compute the coefficients of the rook domination polynomials appearing in Type 3 subsequence. We also thank him for sharing us the explicit values of the coefficients of the domination polynomials up to $d = 20$. We are also grateful to Loïc Masure, François-Xavier Standaert, Matthieu Rivain and Thomas Prest for our very insightful discussions. We would also like to thank the anonymous reviewers for their in-depth reviews of the article and their extremely valuable suggestions. Secure-IC acknowledges partial funding from the European Union's Horizon Europe research and innovation programme through the ALLEGRO project, under grant agreement No. 101070009. Julien Béguinot is a PhD candidate funded by Institut Mines-Télécom through the *Futur & Ruptures* program. Wei Cheng is also partially supported by National Key R&D Program of China No. 2022YFB3103800.

References

AARR02. Dakshi Agrawal, Bruce Archambeault, Josyula R. Rao, and Pankaj Rohatgi. The EM Side-Channel(s). In *CHES*, volume 2523 of *LNCS*, pages 29–45. Springer, 2002.

AB00. Michel Abdalla and Mihir Bellare. Increasing the lifetime of a key: a comparative analysis of the security of re-keying techniques. In *International Conference on the Theory and Application of Cryptology and Information Security*, pages 546–559. Springer, 2000.

ABG+22. Melissa Azouaoui, Olivier Bronchain, Vincent Grosso, Kostas Papagiannopoulos, and François-Xavier Standaert. Bitslice masking and improved shuffling: How and when to mix them in software? *IACR Trans. Cryptogr. Hardw. Embed. Syst.*, 2022(2):140–165, 2022.

BB11. Matthieu R. Bloch and João Barros. *Physical-Layer Security: From Information Theory to Security Engineering*. Cambridge University Press, 2011.

BBD+16. Gilles Barthe, Sonia Belaïd, François Dupressoir, Pierre-Alain Fouque, Benjamin Grégoire, Pierre-Yves Strub, and Rébecca Zucchini. Strong non-interference and type-directed higher-order masking. In Edgar R. Weippl, Stefan Katzenbeisser, Christopher Kruegel, Andrew C. Myers, and Shai Halevi, editors, *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, Vienna, Austria, October 24-28, 2016*, pages 116–129. ACM, 2016.

BCG+23. Julien Béguinot, Wei Cheng, Sylvain Guilley, Yi Liu, Loïc Masure, Olivier Rioul, and François-Xavier Standaert. Removing the Field Size Loss from Duc et al.'s Conjectured Bound for Masked Encodings. In *Constructive Side-Channel Analysis and Secure Design: 14th International Workshop, COSADE 2023, Munich, Germany, April 3–4, 2023, Proceedings*, pages 86–104. Springer, 2023.

BCGR22. Julien Béguinot, Wei Cheng, Sylvain Guilley, and Olivier 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*, pages 496–503. IEEE, 2022.

BCO04. Éric Brier, Christophe Clavier, and Francis Olivier. Correlation power analysis with a leakage model. In Marc Joye and Jean-Jacques Quisquater, editors, *Cryptographic Hardware and Embedded Systems - CHES 2004: 6th International Workshop Cambridge, MA, USA, August 11-13, 2004. Proceedings*, volume 3156 of *Lecture Notes in Computer Science*, pages 16–29. Springer, 2004.

BCP+20. Sonia Belaïd, Jean-Sébastien Coron, Emmanuel Prouff, Matthieu Rivain, and Abdul Rahman Taleb. Random probing security: Verification, composition, expansion and new constructions. In Daniele Micciancio and Thomas Ristenpart, editors, *Advances in Cryptology - CRYPTO 2020 - 40th Annual International Cryptology Conference, CRYPTO 2020, Santa Barbara, CA, USA, August 17-21, 2020, Proceedings, Part I*, volume 12170 of *Lecture Notes in Computer Science*, pages 339–368. Springer, 2020.

BCPZ16. Alberto Battistello, Jean-Sébastien Coron, Emmanuel Prouff, and Rina Zeitoun. Horizontal Side-Channel Attacks and Countermeasures on the ISW Masking Scheme. In Benedikt Gierlichs and Axel Y. Poschmann,

editors, *Cryptographic Hardware and Embedded Systems - CHES 2016 - 18th International Conference, Santa Barbara, CA, USA, August 17-19, 2016, Proceedings*, volume 9813 of *Lecture Notes in Computer Science*, pages 23–39. Springer, 2016.

BFO+22. Gianluca Brian, Antonio Faonio, Maciej Obremski, João Ribeiro, Mark Simkin, Maciej Skórski, and Daniele Venturi. The mother of all leakages: How to simulate noisy leakages via bounded leakage (almost) for free. *IEEE Trans. Inf. Theory*, 68(12):8197–8227, 2022.

BHM+19. Olivier Bronchain, Julien M. Hendrickx, Clément Massart, Alex Olshevsky, and François-Xavier Standaert. Leakage certification revisited: Bounding model errors in side-channel security evaluations. In Alexandra Boldyreva and Daniele Micciancio, editors, *Advances in Cryptology - CRYPTO 2019 - 39th Annual International Cryptology Conference, Santa Barbara, CA, USA, August 18-22, 2019, Proceedings, Part I*, volume 11692 of *Lecture Notes in Computer Science*, pages 713–737. Springer, 2019.

BLR+23. Julien Béguinot, Yi Liu, Olivier Rioul, Wei Cheng, and Sylvain Guilley. Maximal leakage of masked implementations using Mrs. Gerber's lemma for min-entropy. *CoRR*, abs/2305.06276, 2023.

BRT21. Sonia Belaïd, Matthieu Rivain, and Abdul Rahman Taleb. On the power of expansion: More efficient constructions in the random probing model. In Anne Canteaut and François-Xavier Standaert, editors, *Advances in Cryptology - EUROCRYPT 2021 - 40th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Zagreb, Croatia, October 17-21, 2021, Proceedings, Part II*, volume 12697 of *Lecture Notes in Computer Science*, pages 313–343. Springer, 2021.

CDGT24. Claude Carlet, Abderrahman Daif, Sylvain Guilley, and Cédric Tavernier. Quasi-linear masking against SCA and FIA, with cost amortization. *IACR Trans. Cryptogr. Hardw. Embed. Syst.*, 2024(1):398–432, 2024.

CFOS21. Gaëtan Cassiers, Sebastian Faust, Maximilian Orlt, and François-Xavier Standaert. Towards tight random probing security. In Tal Malkin and Chris Peikert, editors, *Advances in Cryptology - CRYPTO 2021 - 41st Annual International Cryptology Conference, CRYPTO 2021, Virtual Event, August 16-20, 2021, Proceedings, Part III*, volume 12827 of *Lecture Notes in Computer Science*, pages 185–214. Springer, 2021.

CJRR99. Suresh Chari, Charanjit S. Jutla, Josyula R. Rao, and Pankaj Rohatgi. Towards Sound Approaches to Counteract Power-Analysis Attacks. In Wiener [Wie99], pages 398–412.

CL10. Stephen Chestnut and Manuel E. Lladser. Occupancy distributions in Markov chains via Doeblin's ergodicity coefficient. In *DMTCS Proceedings vol. AM, 21st International Meeting on Probabilistic, Combinatorial, and Asymptotic Methods in the Analysis of Algorithms (AofA'10)*, 2010.

CRR02. Suresh Chari, Josyula R. Rao, and Pankaj Rohatgi. Template attacks. In Burton S. Kaliski, Jr., Çetin Kaya Koç, and Christof Paar, editors, *Cryptographic Hardware and Embedded Systems - CHES 2002, 4th International Workshop, Redwood Shores, CA, USA, August 13-15, 2002, Revised Papers*, volume 2523 of *Lecture Notes in Computer Science*, pages 13–28. Springer, 2002.

CS19. Gaëtan Cassiers and François-Xavier Standaert. Towards globally optimized masking: From low randomness to low noise rate or probe isolating multiplications with reduced randomness and security against horizontal

attacks. *IACR Trans. Cryptogr. Hardw. Embed. Syst.*, 2019(2):162–198, 2019.

CS21. Jean-Sébastien Coron and Lorenzo Spignoli. Secure wire shuffling in the probing model. In Tal Malkin and Chris Peikert, editors, *Advances in Cryptology - CRYPTO 2021 - 41st Annual International Cryptology Conference, CRYPTO 2021, Virtual Event, August 16-20, 2021, Proceedings, Part III*, volume 12827 of *Lecture Notes in Computer Science*, pages 215–244. Springer, 2021.

CT06. Thomas M. Cover and Joy A. Thomas. *Elements of information theory (2nd edition)*. Wiley, 2006. ISBN: 978-0471-24195-9.

dCGRP19. Éloi de Chérisey, Sylvain Guilley, Olivier Rioul, and Pablo Piantanida. Best Information is Most Successful — Mutual Information and Success Rate in Side-Channel Analysis. *IACR Trans. Cryptogr. Hardw. Embed. Syst.*, 2019(2):49–79, 2019.

DDF14. Alexandre Duc, Stefan Dziembowski, and Sebastian Faust. Unifying Leakage Models: From Probing Attacks to Noisy Leakage. In Phong Q. Nguyen and Elisabeth Oswald, editors, *Advances in Cryptology - EUROCRYPT 2014 - 33rd Annual International Conference on the Theory and Applications of Cryptographic Techniques, Copenhagen, Denmark, May 11-15, 2014. Proceedings*, volume 8441 of *Lecture Notes in Computer Science*, pages 423–440. Springer, 2014.

DDF19. Alexandre Duc, Stefan Dziembowski, and Sebastian Faust. Unifying leakage models: From probing attacks to noisy leakage. *J. Cryptol.*, 32(1):151–177, 2019.

DFS15. Alexandre Duc, Sebastian Faust, and François-Xavier Standaert. Making Masking Security Proofs Concrete - Or How to Evaluate the Security of Any Leaking Device. In Elisabeth Oswald and Marc Fischlin, editors, *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*, volume 9056 of *Lecture Notes in Computer Science*, pages 401–429. Springer, 2015.

DFS19. Alexandre Duc, Sebastian Faust, and François-Xavier Standaert. Making Masking Security Proofs Concrete (Or How to Evaluate the Security of Any Leaking Device), Extended Version. *J. Cryptol.*, 32(4):1263–1297, 2019.

DKL+98. Jean-François Dhem, François Koeune, Philippe-Alexandre Leroux, Patrick Mestré, Jean-Jacques Quisquater, and Jean-Louis Willems. A Practical Implementation of the Timing Attack. In Jean-Jacques Quisquater and Bruce Schneier, editors, *CARDIS*, volume 1820 of *Lecture Notes in Computer Science*, pages 167–182. Springer, 1998.

Dob56. Roland L Dobrushin. Central limit theorem for nonstationary Markov chains. i. *Theory of Probability & Its Applications*, 1(1):65–80, 1956.

Doe37. Wolfgang Doeblin. Le cas discontinu des probabilités en chaîne. *Publ. Fac. Sci. Univ. Masaryk (Brno), 236:1–13*, 1937. munispace.muni.cz/library/catalog/book/1837.

DP08. Stefan Dziembowski and Krzysztof Pietrzak. Leakage-resilient cryptography. In *49th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2008, October 25-28, 2008, Philadelphia, PA, USA*, pages 293–302. IEEE Computer Society, 2008.

EVG22. Amedeo Roberto Esposito, Adrien Vandenbroucque, and Michael Gastpar. On Sibson's α-mutual information. In *IEEE International Symposium on Information Theory, ISIT 2022, Espoo, Finland, June 26 - July 1, 2022*, pages 2904–2909. IEEE, 2022.

FGG97. Nir Friedman, Dan Geiger, and Moisés Goldszmidt. Bayesian network classifiers. *Mach. Learn.*, 29(2-3):131–163, 1997.

GGK20. Amin Gohari, Onur Günlü, and Gerhard Kramer. Coding for positive rate in the source model key agreement problem. *IEEE Trans. Inf. Theory*, 66(10):6303–6323, 2020.

GJR18. Dahmun Goudarzi, Antoine Joux, and Matthieu Rivain. How to securely compute with noisy leakage in quasilinear complexity. In Thomas Peyrin and Steven D. Galbraith, editors, *Advances in Cryptology - ASIACRYPT 2018 - 24th International Conference on the Theory and Application of Cryptology and Information Security, Brisbane, QLD, Australia, December 2-6, 2018, Proceedings, Part II*, volume 11273 of *Lecture Notes in Computer Science*, pages 547–574. Springer, 2018.

GMO01. Karine Gandolfi, Christophe Mourtel, and Francis Olivier. Electromagnetic Analysis: Concrete Results. In *CHES*, volume 2162 of *LNCS*, pages 251–261. Springer, May 14-16 2001. Paris, France.

GPRV22. Dahmun Goudarzi, Thomas Prest, Matthieu Rivain, and Damien Vergnaud. Probing security through input-output separation and revisited quasilinear masking. *IACR Cryptol. ePrint Arch.*, page 45, 2022.

Hoe94. Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *The collected works of Wassily Hoeffding*, pages 409–426, 1994.

ISW03. Yuval Ishai, Amit Sahai, and David Wagner. Private Circuits: Securing Hardware against Probing Attacks. In *CRYPTO*, volume 2729 of *Lecture Notes in Computer Science*, pages 463–481. Springer, August 17–21 2003. Santa Barbara, California, USA.

IUH22. Akira Ito, Rei Ueno, and Naofumi Homma. Perceived Information Revisited New Metrics to Evaluate Success Rate of Side-Channel Attacks. *IACR Trans. Cryptogr. Hardw. Embed. Syst.*, 2022(4):228–254, 2022.

IWK20. Ibrahim Issa, Aaron B. Wagner, and Sudeep Kamath. An Operational Approach to Information Leakage. *IEEE Trans. Inf. Theory*, 66(3):1625–1657, 2020.

KGG+18. Paul Kocher, Daniel Genkin, Daniel Gruss, Werner Haas, Mike Hamburg, Moritz Lipp, Stefan Mangard, Thomas Prescher, Michael Schwarz, and Yuval Yarom. Spectre Attacks: Exploiting Speculative Execution. *CoRR*, abs/1801.01203, 2018.

KJJ99. Paul C. Kocher, Joshua Jaffe, and Benjamin Jun. Differential Power Analysis. In Wiener [Wie99], pages 388–397.

KR19. Yael Tauman Kalai and Leonid Reyzin. A survey of leakage-resilient cryptography. In Oded Goldreich, editor, *Providing Sound Foundations for Cryptography: On the Work of Shafi Goldwasser and Silvio Micali*, pages 727–794. ACM, 2019.

LBB19. Jacqueline Lagasse, Christopher Bartoli, and Wayne P. Burleson. Combining clock and voltage noise countermeasures against power side-channel analysis. In *30th IEEE International Conference on Application-specific Systems, Architectures and Processors, ASAP 2019, New York, NY, USA, July 15-17, 2019*, pages 214–217. IEEE, 2019.

LBC+23. Yi Liu, Julien Béguinot, Wei Cheng, Sylvain Guilley, Loïc Masure, Olivier Rioul, and François-Xavier Standaert. Improved alpha-information bounds for higher-order masked cryptographic implementations. In *IEEE Information Theory Workshop, ITW 2023, Saint-Malo, France, April 23-28, 2023*, pages 81–86. IEEE, 2023.

LH20. JongHyeok Lee and Dong-Guk Han. Security analysis on dummy based side-channel countermeasures - case study: AES with dummy and shuffling. *Appl. Soft Comput.*, 93:106352, 2020.

LRP07. Kerstin Lemke-Rust and Christof Paar. Gaussian Mixture Models for Higher-Order Side Channel Analysis. In Pascal Paillier and Ingrid Verbauwhede, editors, *CHES*, volume 4727 of *LNCS*, pages 14–27. Springer, 2007.

Mak20. Anuran Makur. Coding theorems for noisy permutation channels. *IEEE Trans. Inf. Theory*, 66(11):6723–6748, 2020.

Mas94. James L. Massey. Guessing and entropy. In *Proceedings of 1994 IEEE International Symposium on Information Theory*, pages 204–, Jun 1994.

Mer24. Stephan Mertens. Domination polynomial of the rook graph. *Journal of Integer Sequences, Vol. 27, Article 24.3.7* and *arXiv preprint arXiv:2401.00716*, 2024.

MR04. Silvio Micali and Leonid Reyzin. Physically observable cryptography. In *TCC*, volume 2951 of *Lecture Notes in Computer Science*, pages 278–296. Springer, February 19-21 2004. Cambridge, MA, USA.

MS23a. Anuran Makur and Japneet Singh. Doeblin coefficients and related measures. *arXiv preprint arXiv:2309.08475*, 2023.

MS23b. Loïc Masure and François-Xavier Standaert. Prouff and Rivain's Formal Security Proof of Masking, Revisited - Tight Bounds in the Noisy Leakage Model. In Helena Handschuh and Anna Lysyanskaya, editors, *Advances in Cryptology - CRYPTO 2023 - 43rd Annual International Cryptology Conference, CRYPTO 2023, Santa Barbara, CA, USA, August 20-24, 2023, Proceedings, Part III*, volume 14083 of *Lecture Notes in Computer Science*, pages 343–376. Springer, 2023.

MSS09. Amir Moradi, Mohammad Taghi Manzuri Shalmani, and Mahmoud Salmasizadeh. Dual-rail transition logic: A logic style for counteracting power analysis attacks. *Computers & Electrical Engineering*, 35(2):359–369, 2009.

PGMP19. Thomas Prest, Dahmun Goudarzi, Ange Martinelli, and Alain Passelègue. Unifying Leakage Models on a Rényi Day. In Alexandra Boldyreva and Daniele Micciancio, editors, *Advances in Cryptology - CRYPTO 2019 - 39th Annual International Cryptology Conference, Santa Barbara, CA, USA, August 18-22, 2019, Proceedings, Part I*, volume 11692 of *Lecture Notes in Computer Science*, pages 683–712. Springer, 2019.

Pla22. Maxime Plançon. Exploiting algebraic structures in probing security. *IACR Cryptol. ePrint Arch.*, page 1540, 2022.

PR13. Emmanuel Prouff and Matthieu Rivain. Masking against Side-Channel Attacks: A Formal Security Proof. In Thomas Johansson and Phong Q. Nguyen, editors, *Advances in Cryptology - EUROCRYPT 2013, 32nd Annual International Conference on the Theory and Applications of Cryptographic Techniques, Athens, Greece, May 26-30, 2013. Proceedings*, volume 7881 of *Lecture Notes in Computer Science*, pages 142–159. Springer, 2013.

PW23. Yury Polyanskiy and Yihong Wu. *Information Theory, From Coding to Learning (1. ed.)*. Cambridge University Press, 2023.

Rio23. Olivier Rioul. The Interplay between Error, Total Variation, Alpha-Entropy and Guessing: Fano and Pinsker Direct and Reverse Inequalities. *Entropy*, 25(7):978, 2023.

Riv22. Matthieu Rivain. *On the Provable Security of Cryptographic Implementations*. PhD thesis, Habilitation thesis. Personal website, 2022.

RP10a. Matthieu Rivain and Emmanuel Prouff. Provably Secure Higher-Order Masking of AES. *IACR Cryptology ePrint Archive*, 2010:441, 2010. Extended version of [RP10b].

RP10b. Matthieu Rivain and Emmanuel Prouff. Provably Secure Higher-Order Masking of AES. In Stefan Mangard and François-Xavier Standaert, editors, *CHES*, volume 6225 of *LNCS*, pages 413–427. Springer, 2010.

RSV+11. Mathieu Renauld, François-Xavier Standaert, Nicolas Veyrat-Charvillon, Dina Kamel, and Denis Flandre. A Formal Study of Power Variability Issues and Side-Channel Attacks for Nanoscale Devices. In Kenneth G. Paterson, editor, *Advances in Cryptology - EUROCRYPT 2011 - 30th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Tallinn, Estonia, May 15-19, 2011. Proceedings*, volume 6632 of *Lecture Notes in Computer Science*, pages 109–128. Springer, 2011.

Sen73. Eugene Seneta. On the historical development of the theory of finite inhomogeneous Markov chains. *Mathematical Proceedings of the Cambridge Philosophical Society*, 74(3):507–513, 1973.

SF04. Nadav Shulman and Meir Feder. The Uniform Distribution as a Universal Prior. *IEEE Trans. Inf. Theory*, 50(6):1356–1362, 2004.

SMY09. François-Xavier Standaert, Tal Malkin, and Moti Yung. A Unified Framework for the Analysis of Side-Channel Key Recovery Attacks. In *EUROCRYPT*, volume 5479 of *LNCS*, pages 443–461. Springer, April 26-30 2009. Cologne, Germany.

UHIM24. Rei Ueno, Naofumi Homma, Akiko Inoue, and Kazuhiko Minematsu. Fallen Sanctuary: A Higher-Order and Leakage-Resilient Rekeying Scheme. *IACR Trans. Cryptogr. Hardw. Embed. Syst.*, 2024(1):264–308, 2024.

UKM+17. Thomas Unterluggauer, Thomas Korak, Stefan Mangard, Robert Schilling, Luca Benini, Frank K. Gürkaynak, and Michael Muehlberghuber. Leakage Bounds for Gaussian Side Channels. In Thomas Eisenbarth and Yannick Teglia, editors, *Smart Card Research and Advanced Applications - 16th International Conference, CARDIS 2017, Lugano, Switzerland, November 13-15, 2017, Revised Selected Papers*, volume 10728 of *Lecture Notes in Computer Science*, pages 88–104. Springer, 2017.

VCMKS12. Nicolas Veyrat-Charvillon, Marcel Medwed, Stéphanie Kerckhof, and François-Xavier Standaert. Shuffling against Side-Channel Attacks: A Comprehensive Study with Cautionary Note. In Xiaoyun Wang and Kazue Sako, editors, *ASIACRYPT*, volume 7658 of *Lecture Notes in Computer Science*, pages 740–757. Springer, 2012.

Ver15. Sergio Verdú. α-mutual information. In *2015 Information Theory and Applications Workshop, ITA 2015, San Diego, CA, USA, February 1-6, 2015*, pages 1–6. IEEE, 2015.

vW74. Heinrich von Weizsäcker. Zur Gleichwertigkeit zweier Arten der Randomisierung. *manuscripta mathematica*, 11:91–94, 1974.

Wie99. Michael J. Wiener, editor. *Advances in Cryptology - CRYPTO '99, 19th Annual International Cryptology Conference, Santa Barbara, California,*

USA, August 15-19, 1999, Proceedings, volume 1666 of *Lecture Notes in Computer Science*. Springer, 1999.

WZ73. Aaron D. Wyner and Jacob Ziv. A theorem on the entropy of certain binary sequences and applications-i. *IEEE Trans. Inf. Theory*, 19(6):769–772, 1973.

— SUPPLEMENTARY MATERIAL —

A Two Equivalent Descriptions of a Channel

In this section, we provide a proof sketch that any random transformation $X \to \boxed{P_{Y|X}} \to Y$ can be seen as a random function $Y = F(X)$ where F is distributed according to a distribution P_F ($F \sim P_F$).

On one hand, the random function description $Y = F(X)$ is obviously a particular case of a channel $X \to \boxed{P_{Y|X}} \to Y$ where for every input x and every event $E \subset \mathcal{Y}$,

$$P_{Y|X}(Y \in E | X = x) = \mathbb{P}(F(x) \in E). \tag{72}$$

and the probability on the right-hand side is taken over $F \sim P_F$.

Conversely, any channel $X \to \boxed{P_{Y|X}} \to Y$ can be seen as a random function $Y = F(X)$ for a suitably defined $F \sim P_F$. This fact was proved rigorously in [vW74] and can be seen as follows. For each fixed input x define the random variable $F(x)$ by the probability distribution

$$\mathbb{P}(F(x) \in E) = P_{Y|X}(Y \in E | X = x) \tag{73}$$

for any event $E \subset \mathcal{Y}$. This defines a stochastic process $F = \{F(x) | x \in \mathcal{X}\}$ indexed by $\mathcal{X}$ that follows some probability distribution P_F. It can be shown [vW74] that F is the desired random function such that $X \to \boxed{P_{Y|X}} \to Y = F(X)$.

B Technical Proofs

B.1 Proof of Lemma 1

Proof. We verify that we obtain the same transition probability. Let $x \in \mathcal{X}$ and $y' \in \mathcal{Y}$. First on the left-hand side:

$$\begin{cases} p_{Y'|X}(\perp|x) &= \sum\!\!\!\!\!\!\int_{\mathcal{X}\cup\{\perp\}} p_{X'|X}(x'|x) p^{\perp}_{Y|X}(\perp|x') = \bar{\mathcal{E}} p^{\perp}_{Y|X}(\perp|x) + \mathcal{E} p^{\perp}_{Y|X}(\perp|\perp) = \mathcal{E} \\ p_{Y'|X}(y'|x) &= \bar{\mathcal{E}} p^{\perp}_{Y|X}(y'|x) + \mathcal{E} p^{\perp}_{Y|X}(y'|\perp) = \bar{\mathcal{E}} p_{Y|X}(y'|x). \end{cases} \tag{74}$$

On the right-hand side:

$$\begin{cases} p_{Y'|X}(\perp|x) &= \sum\!\!\!\!\!\!\int_{\mathcal{Y}} p_{Y|X}(y|x) p_{Y'|Y}(\perp|y) = \sum\!\!\!\!\!\!\int_{\mathcal{Y}} p_{Y|X}(y|x) \mathcal{E} = \mathcal{E} \\ p_{Y'|X}(y'|x) &= \sum\!\!\!\!\!\!\int_{\mathcal{Y}} p_{Y|X}(y|x) p_{Y'|Y}(y'|y) = \bar{\mathcal{E}} p_{Y|X}(y'|x). \end{cases} \tag{75}$$

□

B.2 Proof of Proposition 1

Proof. This is a consequence of a DPI. By Theorem 1, we obtain

$$K \to \boxed{\mathrm{EC}^{\perp}_{\mathcal{E}(K\to Y)}} \to K' \to Y. \tag{76}$$

Hence, by DPI on the SR (Lemma 8), we have

$$\mathbb{P}_{s,o}(K|Y) \leqslant \mathbb{P}_{s,o}(K|K') = \overline{\mathcal{E}}(K \to Y) + \mathcal{E}(K \to Y)\mathbb{P}_{s,o}(K). \tag{77}$$

Further by DPI on the GE (Lemma 8), we have

$$G(K|Y) \geqslant G(K|K') = \overline{\mathcal{E}}(K \to Y) + \mathcal{E}(K \to Y)G(K). \tag{78}$$

Finally, by DPI on the TVI (Lemma 8), we have

$$\Delta(K|Y) \leqslant \Delta(K|K') = \overline{\mathcal{E}}(K \to Y) \sum_{k\in\mathcal{K}} p_K(k)(1 - p_K(k)) \tag{79}$$

$$= \overline{\mathcal{E}}(K \to Y) \left(1 - \exp\left(-H_2(K)\right)\right). \tag{80}$$

□

B.3 Proof of Theorem 1

Proof. Consider a channel $P_{Y|X}$ with a given Doeblin coefficient $\mathcal{E}$. We show that there exists a channel $P_{Y|X'}$ such that the channel $P_{Y|X}$ factorizes into $P_{X'|X}P_{Y|X'}$ where $P_{X'|X}$ is an erasure channel with parameter $\mathcal{E}$. Let

$$\begin{cases} p_{Y|X'}(y|\perp) = \mathcal{E}^{-1} \inf\limits_{x\in\mathcal{X}} p_{Y|X}(y|x) \\ p_{Y|X'}(y|x) = \overline{\mathcal{E}}^{-1}(p_{Y|X}(y|x) - \inf\limits_{x\in\mathcal{X}} p_{Y|X}(y|x)) \end{cases}. \tag{81}$$

This is a well defined channel since all terms are nonnegative and $\fint_{\mathcal{Y}} p_{Y|X'}(y|\perp) = \fint_{\mathcal{Y}} p_{Y|X'}(y|x) = 1$. This gives the desired factorization since $\mathcal{E}\mathcal{E}^{-1} \inf\limits_{x\in\mathcal{X}} p_{Y|X}(y|x) + \overline{\mathcal{E}\mathcal{E}}^{-1}(p_{Y|X}(y|x) - \inf\limits_{x\in\mathcal{X}} p_{Y|X}(y|x)) = p_{Y|X}(y|x)$.

Conversely, assume that there exists a channel $P_{Y|X'}$ such that the channel $P_{Y|X}$ factorizes into $P_{X|X'}P_{Y|X'}$ where $P_{X|X'}$ is an erasure channel with parameter $\mathcal{E}$. Then for any pair x, y we have $p_{Y|X}(y|x) = \overline{\mathcal{E}}p_{Y|X'}(y|x) + \mathcal{E}p_{Y|X'}(y|\perp) \geqslant \mathcal{E}p_{Y|X'}(y|\perp)$. Since this is true for all x, this is also true by taking the infimum over x: $\inf_x p_{Y|X}(y|x) \geqslant \mathcal{E}P_{Y|X}(y|\perp)$. Summing over y and noting that $\fint_{y\in\mathcal{Y}} P_{Y|X}(y|\perp) = 1$ gives $\fint_{y\in\mathcal{Y}} \inf_{x\in\mathcal{X}} p_{Y|X}(y|x) \geqslant \mathcal{E}$. □

B.4 Proof of Lemma 3

Proof. The composition can be computed explicitly

$$\begin{cases} p_{Z|X}(z|x) = 0 & \text{if } x \neq z \\ p_{Z|X}(z|x) = \overline{\mathcal{E}}_0\overline{\mathcal{E}}_1 & \text{if } x = z \\ p_{Z|X}(\perp_0|x) = \mathcal{E}_0 & \text{for all } x \\ p_{Z|X}(\perp_1|x) = \overline{\mathcal{E}}_0\mathcal{E}_1 & \text{for all } x \end{cases} \tag{82}$$

Hence the result by computing equation (30), $\mathcal{E} = 0 + \mathcal{E}_0 + \overline{\mathcal{E}}_0\mathcal{E}_1 = 1 - \overline{\mathcal{E}}_0\overline{\mathcal{E}}_1$. □

B.5 Proof of Lemma 4

Proof.

$$\begin{aligned} \bigoint\!\!\!\bigoint_{\mathcal{Y}\times\mathcal{Z}} \inf_{x\in\mathcal{X}} p(yz|x) &= \bigoint\!\!\!\bigoint_{\mathcal{Y}\times\mathcal{Z}} \inf_{x\in\mathcal{X}} p_{Y|X}(y|x)p_{Z|XY}(z|xy) && (83)\\ &= \bigoint\!\!\!\bigoint_{\mathcal{Y}\times\mathcal{Z}} \inf_{x\in\mathcal{X}} p_{Y|X}(y|x)p_{Z|Y}(z|y) && (84)\\ &= \bigoint\!\!\!\bigoint_{\mathcal{Y}\times\mathcal{Z}} p_{Z|Y}(z|y) \inf_{x\in\mathcal{X}} p_{Y|X}(y|x) && (85)\\ &= \bigoint\!\!\!\bigoint_{\mathcal{Y}} \left(\bigoint\!\!\!\bigoint_{\mathcal{Z}} p_{Z|Y}(z|y)\right) \inf_{x\in\mathcal{X}} p_{Y|X}(y|x) && (86)\\ &= \bigoint\!\!\!\bigoint_{\mathcal{Y}} \inf_{x\in\mathcal{X}} p_{Y|X}(y|x). && (87) \end{aligned}$$

□

B.6 Proof of Lemma 5

Remark 16. Very recently Makur and Singh [MS23a] established the same property of CDC in the discrete setting (with transition matrices) for Bayesian networks [FGG97].

Proof.

$$\begin{aligned} \bigoint\!\!\!\bigoint_{\mathcal{Y}_1\times\mathcal{Y}_2} \inf_x p_{Y_1Y_2|X}(y_1y_2|x) &= \bigoint\!\!\!\bigoint_{\mathcal{Y}_1\times\mathcal{Y}_2} \inf_x p_{Y_1|X}(y_1|x)p_{Y_2|XY_1}(y_2|xy_1) && (88)\\ &= \bigoint\!\!\!\bigoint_{\mathcal{Y}_1\times\mathcal{Y}_2} \inf_x p_{Y_1|X}(y_1|x)p_{Y_2|X}(y_2|x) && (89)\\ &\geqslant \bigoint\!\!\!\bigoint_{\mathcal{Y}_1\times\mathcal{Y}_2} \inf_x p_{Y_1|X}(y_1|x) \inf_x p_{Y_2|X}(y_2|x) && (90)\\ &= \bigoint\!\!\!\bigoint_{\mathcal{Y}_1} \inf_x p_{Y_1|X}(y_1|x) \bigoint\!\!\!\bigoint_{\mathcal{Y}_2} \inf_x p_{Y_2|X}(y_2|x). && (91) \end{aligned}$$

The result with q channels follows by induction.

The inequality is a consequence of the union bound. Indeed, if $0 \leqslant x_1, \ldots, x_q \leqslant 1$ we can see each as a probability of a given event E_i, $x_i = \mathbb{P}(E_i)$ and $1 - \prod_{i=1}^q (1 - x_i) = \mathbb{P}(\cup_{i=1}^q E_i) \leqslant \sum_{i=1}^q \mathbb{P}(E_i) = \sum_{i=1}^q x_i$. □

B.7 Proof of Lemma 6

Proof. The key point is that for any observation y_1, the channel $X \to \boxed{P_{Y_2|X,Y_1=y_1}} \to Y_2$ is $\overline{\mathcal{E}}$-noisy with respect to CDC. This exactly means that for all y_1, $\oint_{\mathcal{Y}_2} \inf_x p(y_2|xy_1) = \mathcal{E}_{y_1}(X \to Y_2) \geqslant \mathcal{E}$.

$$\oint_{\mathcal{Y}_1 \times \mathcal{Y}_2^x} \inf p_{Y_1Y_2|X}(y_1y_2|x) = \oint_{\mathcal{Y}_1 \times \mathcal{Y}_2} \inf_x p_{Y_1|X}(y_1|x) p_{Y_2|XY_1}(y_2|xy_1) \tag{92}$$

$$\geqslant \oint_{\mathcal{Y}_1 \times \mathcal{Y}_2} \inf_x p_{Y_1|X}(y_1|x) \inf_x p(y_2|xy_1) \tag{93}$$

$$= \oint_{\mathcal{Y}_1} \left(\left(\oint_{\mathcal{Y}_2} \inf_x p_{Y_2|XY_1}(y_2|xy_1) \right) \inf_x p_{Y_1|X}(y_1|x) \right) \tag{94}$$

$$\geqslant \oint_{\mathcal{Y}_1} \left(\mathcal{E} \inf_x p_{Y_1|X}(y_1|x) \right) \tag{95}$$

$$\geqslant \mathcal{E}^2. \tag{96}$$

The result with q channels follows by induction. □

B.8 Proof of Proposition 2

Proof. Since the channel noise is additive $p_{Y|X}(y|x) = p_Z(y - f(x))$. As the noise is radially symmetric $p_Z(y - f(x)) = p_Z(|y - f(x)|)$. Since the noise is radially symmetric decreasing $\inf_x p_{Y|X}(y|x) = p_Z(\sup_x |y - f(x)|)$. If $y \geqslant \frac{m+M}{2}$ then $\sup_x |y - f(x)| = y - m$ else $\sup_x |y - f(x)| = M - y$. Hence the result by splitting the integral over $\mathcal{Y}$ into two parts,

$$\mathcal{E}(X \to Y) = \int_{-\infty}^{\frac{m+M}{2}} p_Z(M - y)\, \mathrm{d}y + \int_{\frac{m+M}{2}}^{\infty} p_Z(y - m)\, \mathrm{d}y = 2 \int_{\frac{M-m}{2}}^{\infty} p_Z(u)\, \mathrm{d}u. \tag{97}$$

Since $Q(x) = \frac{1}{2} - \frac{x}{\sqrt{2\pi}} + O(x^3)$, we also obtain the announced approximation for high noise and additive Gaussian noise. □

B.9 Proof of Lemma 10

Proof. Consider the channel

$$X \to \boxed{f} \to G \to \boxed{\mathrm{Mask}_d} \to \mathbf{G} \to \boxed{\prod_{i=0}^{d} p_{Y_i|G_i}} \to \mathbf{Y}. \tag{98}$$

Since by the strengthened DPI (Lemma 7), $\overline{\mathcal{E}}(X \to \mathbf{Y}) \leqslant \overline{\mathcal{E}}(X \to G)\overline{\mathcal{E}}(G \to \mathbf{Y}) \leqslant \overline{\mathcal{E}}(G \to \mathbf{Y})$. We can limit ourselves to the channel $G \to \mathbf{Y}$. By Theorem 1,

$G_i \to \boxed{P_{Y_i|G_i}} \to Y_i$ is a stochastically degraded erasure channel with erasure probability $\mathcal{E}_i = \mathcal{E}(G_i \to Y_i)$. Consequently, the channel rewrites

$$G_i \to \boxed{\mathrm{EC}_{\mathcal{E}_i}^{\perp_i}} \to G_i' \to \boxed{P_{Y_i|G_i'}} \to Y_i. \tag{99}$$

So that the channel is

$$G \to \boxed{\mathrm{Mask}_d} \to \boxed{\prod_{i=0}^{d} \mathrm{EC}_{\mathcal{E}_i}^{\perp_i}} \to \mathbf{Y}' \to \boxed{\prod_{i=0}^{d} p_{Y_i|G_i}} \to \mathbf{Y}. \tag{100}$$

Hence, the channel $G \to \mathbf{Y}$ is stochastically degraded with respect to the channel $G \to \mathbf{Y}'$. By the strengthened DPI (Lemma 7) we obtain

$$\overline{\mathcal{E}}(X \to \mathbf{Y}) \leqslant \overline{\mathcal{E}}(G \to \mathbf{Y}) \leqslant \overline{\mathcal{E}}(G \to \mathbf{Y}'). \tag{101}$$

Recall that by definition

$$\overline{\mathcal{E}}(G \to \mathbf{Y}') = \mathbb{E}_{Y_0', \dots, Y_d'} \left[\sup_{g \in f(\mathcal{X})} \left(1 - \frac{p(g|Y_0', \dots, Y_d')}{p(g)} \right) \right]. \tag{102}$$

If there exists an index i such that $y_i' = \perp_i$ then by *secret sharing property* $p(g|y_0', \dots, y_d') = p(g)$ so that

$$\sup_{g \in f(\mathcal{X})} \left(1 - \frac{p(g|y_0', \dots, y_d')}{p(g)} \right) = 0. \tag{103}$$

Otherwise, for all $i \in \{0, \dots, d\}$, $y_i' \neq \perp_i$ and

$$\begin{cases} p(g|y_0', \dots, y_d') = 1 & \text{if } g = y_0' + \ldots + y_d' \\ p(g|y_0', \dots, y_d') = 0 & \text{if } g \neq y_0' + \ldots + y_d' \end{cases} \tag{104}$$

so that

$$\sup_{g \in f(\mathcal{X})} \left(1 - \frac{p(g|y_0', \dots, y_d')}{p(g)} \right) = \sup(1, 1 - \frac{1}{p(y_0' + \ldots + y_d')}) = 1. \tag{105}$$

As a consequence,

$$\overline{\mathcal{E}}(G \to \mathbf{Y}') = \mathbb{E}_{Y_0', \dots, Y_d'} \left[\mathbb{1}_{Y_0' \neq \perp_0, \dots, Y_d' \neq \perp_d} \right] = \mathbb{P}\left(Y_0' \neq \perp_0, \dots, Y_d' \neq \perp_d \right) = \prod_{i=0}^{d} \overline{\mathcal{E}}_i. \tag{106}$$

□

B.10 Proof of Lemma 11

Proof. By Theorem 1, $(G_i, H_j) \to Y_{i,j}$ is stochastically degraded with respect to an erasure channel with erasure probability $\mathcal{E}_{i,j} \triangleq \mathcal{E}((G_i, H_j) \to Y_{i,j})$. Hence, we can write

$$(G_i, H_j) \to \boxed{\mathrm{EC}_{\mathcal{E}_{i,j}}^{\perp_{i,j}}} \to (G'_i, H'_j) \to Y_{i,j}. \tag{107}$$

As a consequence the channel $X \to \mathbf{Y}$ is stochastically degraded with respect to the channel $X \to ((G'_i, H'_j))_{0 \leqslant i,j \leqslant d}$. By the strengthened DPI (Lemma 7) we have $\mathcal{E}(X \to \mathbf{Y}) \geqslant \mathcal{E}(X \to ((G'_i, H'_j))_{0 \leqslant i,j \leqslant d})$. It only remains to show that $\overline{\mathcal{E}}(X \to ((G'_i, H'_j))_{0 \leqslant i,j \leqslant d}) = \Upsilon((\overline{\mathcal{E}}_{i,j})_{0 \leqslant i,j \leqslant d})$.

By the *secret sharing property* we have $p(x|((g'_i, h'_j))_{i,j}) = p(x)$ whenever $\exists i, \forall j, (g'_i, h'_j) = \perp_{i,j}$ and $\exists j, \forall i, (g'_i, h'_j) = \perp_{i,j}$. In this case

$$\sup_{x \in \mathcal{X}} \left(1 - \frac{p(x|((g'_i, h'_j))_{i,j})}{p(x)}\right) = 0. \tag{108}$$

Otherwise, when $\forall i, \exists j, (g'_i, h'_j) \neq \perp_{i,j}$ or $\forall j, \exists i, (g'_i, h'_j) \neq \perp_{i,j}$, as in the previous proof,

$$\sup_{x \in \mathcal{X}} \left(1 - \frac{p(x|((g'_i, h'_j))_{i,j})}{p(x)}\right) = 1. \tag{109}$$

Let $E_{i,j}$ be the event $(G'_i, H'_j) \neq \perp_{i,j}$. Then the $E_{i,j}$ are mutually independents events with probabilities $\overline{\mathcal{E}}_{i,j}$. Further we have

$$\sup_{x \in \mathcal{X}} \left(1 - \frac{p(x|((G'_i, H'_j))_{i,j})}{p(x)}\right) = 1 \tag{110}$$

on the event

$$E = (\cap_{i=0}^{d} \cup_{j=0}^{d} E_{i,j}) \cup (\cap_{j=0}^{d} \cup_{i=0}^{d} E_{i,j}) \tag{111}$$

and 0 otherwise. As a consequence,

$$\overline{\mathcal{E}}(X \to ((G'_i, H'_j))_{0 \leqslant i,j \leqslant d}) = \mathbb{E}\left[\mathbb{1}_E\right] = \mathbb{P}(E). \tag{112}$$

□

B.11 Proof of Lemma 12

Proof. By Theorem 1 the channel $(T_{i,j}, T_{i,j-1}) \to Y_{i,j}$ rewrites

$$(T_{i,j}, T_{i,j-1}) \to \boxed{\mathrm{EC}_{\mathcal{E}_{i,j}}^{\perp_i}} \to Y'_{i,j} \to Y_{i,j} \tag{113}$$

where $\mathcal{E}_{i,j} \triangleq \mathcal{E}((T_{i,j}, T_{i,j-1}) \to Y_{i,j})$. Hence, $T_{i,d} \to \mathbf{Y}_i$ is stochastically degraded with respect to

$$T_{i,d} \to \mathbf{T}_i \to \boxed{\prod_{j=0}^{d} \mathrm{EC}_{\mathcal{E}_j}^{\perp_j}} \to \mathbf{Y}'_i. \tag{114}$$

Thus, by the strengthened DPI (Lemma 7) $\mathcal{E}(T_{i,d} \to \mathbf{Y}_i) \geqslant \mathcal{E}(T_{i,d} \to \mathbf{Y}'_i)$. It remains to show that $\mathcal{E}(T_{i,d} \to \mathbf{Y}'_i) = \mathcal{E}((T_{i,d}, T_{i,d-1}) \to Y_{i,d})$. By *secret sharing property* if $y'_{i,d} = \perp$ then $p(t|y'_{i,1}, \ldots, y'_{i,d}) = p(t)$. Hence,

$$\sup_t \left(1 - \frac{p(t|y'_{i,1}, \ldots, y'_{i,d})}{p(t)}\right) = 0 \tag{115}$$

in this case. Otherwise, $y_{i,d} = (t_{i,d}, t_{i,d-1})$ so that $p(t|y'_{i,1}, \ldots, y'_{i,d} = (t_{i,d}, t_{i,d-1})) = \mathbb{1}_{t=t_{i,d}}$. Hence,

$$\sup_t \left(1 - \frac{p(t|y'_{i,1}, \ldots, y'_{i,d})}{p(t)}\right) = 1 \tag{116}$$

in this case. Finally,

$$\overline{\mathcal{E}}(T_{i,d} \to \mathbf{Y}'_i) = \mathbb{E}_{\mathbf{Y}'_i}[\mathbb{1}_{Y'_{i,d} \neq \perp}] = \mathbb{P}(Y'_{i,d} \neq \perp) = \overline{\mathcal{E}}((T_{i,d}, T_{i,d-1}) \to Y_{i,d}). \tag{117}$$

This concludes the proof since $\overline{\mathcal{E}}((T_{i,d-1}, V_{i,d}) \to Y_{i,d}) = \overline{\mathcal{E}}((T_{i,d}, T_{i,d-1}) \to Y_{i,d})$. □

B.12 Proof of Theorem 2

Proof. Each sensitive variable in the different subsequences is a function f of the secret key K and public information T i.e $X = f(K, T)$. By Lemma 7 we have $\overline{\mathcal{E}}(K \to Y) \leqslant \overline{\mathcal{E}}(X \to Y)$. Since the CDC does not depend on the input distribution and by Lemma 6 carries over adaptively chosen channel we can recombine the different leakages from the different queries and subsequences. Since the subsequences are separated by leakage-free refreshing and that the channel noise is independent of one query to another we can use Lemma 6 to combine the different leakages of each subsequence and query. Further, the Lemmas 10,11,13 provide the bounds for each subsequence. The asymptotic equivalence is a consequence of Proposition 14. □

B.13 Proof of Theorem 4

Proof. Let $\mathbf{Y} = (Y_{j,i})$ be the vector of all leakages, where $Y_{j,i}$ corresponds to the leakage of the l_i wires of the i-th gadget for the j-th trace. By Lemma 6 we obtain for an adaptive chosen channel adversary that

$$\overline{\mathcal{E}}(K \to \mathbf{Y}) \leqslant 1 - \prod_{j=1}^{q} \prod_{i=1}^{|\Gamma|} 1 - \overline{\mathcal{E}}(K \to Y_{j,i}). \tag{118}$$

It remains to show that $\overline{\mathcal{E}}(K \to Y_{j,i}) \leqslant Q_B(t, l_i, \overline{\mathcal{E}})$ for all (j, i). Let $X_1, \ldots, X_{l_i}$ be the l_i considered wires and $L_1, \ldots, L_{l_i}$ be the corresponding leakage for the j-th trace $Y_{j,i} = (L_1, \ldots, L_{l_i})$. The channel $K \to Y_{j,i}$ is factorized by Theorem 1 to

$$K \to (X_1, ..., X_{l_i}) \to (X'_1, \ldots, X'_{l_i}) \to Y_{j,i} \tag{119}$$

where $X_i \to X'_i$ is an erasure channel with erasure probability $\mathcal{E}$. By the DPI (Lemma 7), $\overline{\mathcal{E}}(K \to Y_{j,i}) \leqslant \overline{\mathcal{E}}(K \to (X'_1, \dots, X'_{l_i}))$.

Now $\sup_k 1 - \frac{p(k|x'_1,\dots,x'_{l_i})}{p(k)} = 0$ provided that there exists at most t non-erased values in the tuple $(x'_1, \dots, x'_{l_i})$. Otherwise, $\sup_k 1 - \frac{p(k|x'_1,\dots,x'_{l_i})}{p(k)} \leqslant 1$. Hence $\overline{\mathcal{E}}(K \to (X'_1, \dots, X'_{l_i})) \leqslant \mathbb{E}_{X'_1,\dots,X'_{l_i}} \mathbb{1}_E = \mathbb{P}(E)$ where E is the event that there is at least t non-erased values in the tuple $(X'_1, \dots, X'_{l_i})$. Since the number of non-erased values follows a binomial distribution with parameters l_i and $\overline{\mathcal{E}}$ we obtain that $\mathbb{P}(E) = Q_B(t, l_i, \overline{\mathcal{E}})$ which concludes the proof. □

B.14 Proof of Proposition 5

Proof. As $\overline{\mathcal{E}} \to 0$ we have

$$\mathrm{fail}(d_1, \overline{\mathcal{E}}) \sim \left(\sum_i \binom{l_i(d_1)}{t(d_1)+1} \right) \overline{\mathcal{E}}^{d_1} = C_{d_1} \overline{\mathcal{E}}^{d_1} \tag{120}$$

and similarly for d_2. As a consequence,

$$\frac{\mathrm{fail}(d_2, \overline{\mathcal{E}})}{\mathrm{fail}(d_1, \overline{\mathcal{E}})} \sim \frac{C_{d_2}}{C_{d_1}} \overline{\mathcal{E}}^{d_2 - d_1}. \tag{121}$$

Since $d_2 > d_1$ the ratio in the equivalent becomes strictly inferior to 1 whenever $\overline{\mathcal{E}} < \left(\frac{C_{d_1}}{C_{d_2}}\right)^{\frac{1}{d_2-d_1}}$. Finally, $\mathrm{fail}(d_1, \overline{\mathcal{E}}) > \mathrm{fail}(d_2, \overline{\mathcal{E}})$ if $\overline{\mathcal{E}}$ is small enough. □

B.15 Proof of Proposition 6

Proof. We assume $\overline{\mathcal{E}} < 1$. We have

$$Q_B(t(d), l(d), \overline{\mathcal{E}}) = 1 - Q_B(l(d) - t(d) + 1, l(d), \overline{\mathcal{E}}) \tag{122}$$

Since $\frac{l(d)}{t(d)} \overset{d\to\infty}{\longrightarrow} \infty$, from a certain rank $\overline{\mathcal{E}} l(d) < l(d) - t(d) + 1$, and we can apply Chernoff-Hoeffding [Hoe94] bound

$$1 \geqslant Q_B(t(d), l(d), \overline{\mathcal{E}}) \geqslant 1 - \exp\left(-l(d) d_{\mathrm{KL}} \left(\frac{l(d) - t(d) + 1}{l(d)} \middle\| \overline{\mathcal{E}} \right) \right) \tag{123}$$

where $d_{\mathrm{KL}}(p\|q) \triangleq D_{\mathrm{KL}}(\mathcal{B}(p)\|\mathcal{B}(q)) = p \log \frac{p}{q} + \bar{p} \log \frac{\bar{p}}{\bar{q}}$.

We have $d_{\mathrm{KL}} \left(\frac{l(d)-t(d)+1}{l(d)} \middle\| \overline{\mathcal{E}} \right) \to d_{\mathrm{KL}}(1\|\overline{\mathcal{E}}) > 0$ and $l(d) \to \infty$ so by sandwiching theorem $1 - Q_B(t(d), l(d), \overline{\mathcal{E}})$ converges exponentially fast to 0. Then

$$0 \leqslant \prod_i \left(1 - Q_B(t(d), l_i(d), \overline{\mathcal{E}})\right) \leqslant \left(1 - Q_B(t(d), l(d), \overline{\mathcal{E}})\right). \tag{124}$$

As a consequence, $\mathrm{fail}(d, \overline{\mathcal{E}}) \overset{d\to\infty}{\longrightarrow} 1$. □

C Technical Properties of the Rook Domination Polynomial

Lemma 14 (Properties of Υ_d). *Υ_d verifies the following properties:*

- *Lower and upper bounds:*

$$\left(1-\mathcal{E}^{d+1}\right)^{d+1} \leqslant \Upsilon_d(\overline{\mathcal{E}}) \leqslant \min\{2\left(1-\mathcal{E}^{d+1}\right)^{d+1}, 1\}. \tag{125}$$

- *Υ_d is a polynomial function of degree $(d+1)^2$. The multiplicity of zero as a root of Υ_d is $d+1$. Namely we can write:*

$$\Upsilon_d(\overline{\mathcal{E}}) = \sum_{j=d+1}^{(d+1)^2} a_j \overline{\mathcal{E}}^j \mathcal{E}^{(d+1)^2-j}. \tag{126}$$

- *For all $i \in \{0,\dots,(d+1)^2\}$, $a_i \in \mathbb{N}$ and*

$$\begin{cases} a_{d+1} = 2(d+1)^{d+1} - (d+1)! & \\ 0 \leqslant a_k \leqslant \binom{(d+1)^2}{k} - \binom{d^2}{k} & d+2 \leqslant k \leqslant d^2 \\ a_k = \binom{(d+1)^2}{k} & k \geqslant d^2+1. \end{cases} \tag{127}$$

  In particular, $\sum_{j=d+1}^{(d+1)^2} a_i < 2^{(d+1)^2} - 2^{d^2}$.

Proof. See Stephan Mertens's article [Mer24] on the domination polynomial of the rook graph. □

D Comparison with Other Noisiness Metrics

Proof. The lower bound is obtained via the DPI. Namely, we degrade the channel $X \to Y$ as

$$X \to \boxed{\mathrm{EC}_{\mathcal{E}}^{\perp}} \to X' \to Y \tag{128}$$

where $\mathcal{E} = \mathcal{E}(X \to Y)$ Then by DPI Lemma 8 we obtain

$$I(X;X') \geqslant I(X;Y). \tag{129}$$

Since $I(X;X') = H(X) - H(X|X') = H(X) - \mathcal{E}H(X) - \bar{\mathcal{E}}0 = \bar{\mathcal{E}}H(X)$ we obtain

$$\bar{\mathcal{E}} \geqslant \frac{I(X;Y)}{H(X)} \geqslant \frac{I(X;Y)}{\log|\mathcal{X}|}. \tag{130}$$

For TVI we proceed similarly to proposition 1. By DPI Lemma 8, $\Delta(X;X') \geqslant \Delta(X;Y)$ and

$$\Delta(X;X') = \sum_{\mathcal{X}} p_{X'}(x)(1-p_X(x)) \tag{131}$$

$$= \bar{\mathcal{E}} \sum_{\mathcal{X}} p_X(x)(1-p_X(x)) \tag{132}$$

$$= \bar{\mathcal{E}}\big(1 - \sum_{\mathcal{X}} p_X(x)^2\big) \tag{133}$$

$$= \bar{\mathcal{E}}\big(1 - \exp(-H_2(X))\big) \tag{134}$$

$$\leqslant \bar{\mathcal{E}}\big(1 - \frac{1}{|\mathcal{X}|}\big). \tag{135}$$

We obtain

$$\bar{\mathcal{E}} \geqslant \frac{\Delta(X;Y)}{1-\exp(-H_2(X))} \geqslant \frac{\Delta(X;Y)|\mathcal{X}|}{|\mathcal{X}|-1}. \tag{136}$$

For maximal leakage we can also leverage the DPI,

$$\log(1 + (|\mathcal{X}|-1)\bar{\mathcal{E}}(X \to Y)) = \mathcal{L}(X \to X') \geqslant \mathcal{L}(X \to Y). \tag{137}$$

By Jensen's inequality, as in Rivain's HDR [Riv22], we obtain that $\beta(X;Y) \leqslant 2\Delta(X;Y)$. Also,

$$\mathrm{ARE}(X;Y) = \mathbb{E}_y \sup_x \left| \frac{p_{X|Y}(x|y)}{p_X(x)} - 1 \right| \tag{138}$$

$$= \sum\!\!\!\!\!\!\int_{y\in\mathcal{Y}} \sup_{x\in\mathcal{X}} \frac{1}{p_X(x)} \left| p_{X,Y}(x,y) - p_X(x)p_Y(y) \right| \tag{139}$$

$$\leqslant 2\gamma_X \Delta(X;Y). \tag{140}$$

We have

$$\overline{\mathcal{E}}(X \to Y) = 1 - \sum\!\!\!\!\!\!\int_y \inf_x p(y|x) = \sum\!\!\!\!\!\!\int_y \sup_x p(y) - p(y|x) = \mathbb{E}_y \left[\sup_x \left(1 - \frac{p(y|x)}{p(y)} \right) \right]. \tag{141}$$

Hence similarly to [PGMP19, Proposition 1] we obtain,

$$\overline{\mathcal{E}}(X \to Y) \leqslant \mathbb{E}_y \left[\sup_x \left| 1 - \frac{p(y|x)}{p(y)} \right| \right] = ARE(X;Y) \leqslant RE(X;Y). \tag{142}$$

For TVI we simplify the derivations from [DDF14, Appendix C, Eqn. 14] and use Pinsker's inequality as [DFS19] to obtain a bound for MI.

$$\overline{\mathcal{E}}(X \to Y) \leqslant \mathbb{E}_y \left[\sup_x \left(1 - \frac{p(y|x)}{p(y)} \right)^+ \right] \tag{143}$$

$$\leqslant \mathbb{E}_y \left[\sum_x \left(1 - \frac{p(y|x)}{p(y)} \right)^+ \right] \tag{144}$$

$$= \mathbb{E}_y \left[\sum_x \frac{1}{p(x)} \left(p(x) - p(x|y) \right)^+ \right] \tag{145}$$

$$\leqslant \frac{1}{\inf_{x \in \mathcal{X}} p_X(x)} \Delta(X;Y) \tag{146}$$

$$\leqslant \frac{1}{\inf_{x \in \mathcal{X}} p_X(x)} \sqrt{\frac{I(X;Y)}{2 \log e}}. \tag{147}$$

Finally by adding positive term we obtain the bound for EN,

$$\overline{\mathcal{E}}(X \to Y) \leqslant \mathbb{E}_y \left[\sqrt{\sum_x \left(1 - \frac{p(y|x)}{p(y)} \right)^2} \right] \tag{148}$$

$$= \mathbb{E}_y \left[\sqrt{\sum_x \frac{1}{p_X(x)^2} \left(p(x) - p(x|y) \right)^2} \right] \tag{149}$$

$$\leqslant \frac{1}{\inf_{x \in \mathcal{X}} p_X(x)} \beta(X;Y). \tag{150}$$

The upper bound in terms of maximal leakage is obtained as follows, let x_y be a value of x achieving $\inf_x p(y|x)$ then

$$\inf_x p(y|x) = \sum_x p(y|x) - \sum_{x \neq x_y} p(y|x) \tag{151}$$

$$\geqslant \sum_x p(y|x) - (|\mathcal{X}| - 1) \sup_x p(y|x). \tag{152}$$

Summing over y yields the announced result. □

E Extensive State of the Art: Flaws and Patches

E.1 "Masking against Side-Channel Attacks: a Formal Security Proof"

In their seminal work Prouff and Rivain [PR13] introduced the noisy leakage model and provided for the first time a direct proof of security for masking in

this model. The proof of the technical lemma used in [PR13] are available in Rivain HDR [Riv22]. Though, we identified a critical flaw in [PR13, Lemma 4] used to derive [PR13, Thm. 3] and the main theorem. Recall [PR13, Lemma 4]:

> *Let A, B be two uniform and mutually independent random variables defined over $\mathbb{F}$. Let $L = f(A, B)$ be a leakage corresponding to the side-channel $f : (a, b) \in \mathbb{F}^2 \mapsto l \in \mathcal{L}$. Then for every $a, b \in \mathbb{F}$ and $l \in \mathcal{L}$ we have*
>
> $$\beta(A; f(A, b) = l) \leqslant |\mathbb{F}|\beta((A, B); L = l). \tag{153}$$

Lemma 15 (Counter Example to Lemma 4 in [PR13]). *If $\mathbb{F} = \mathbb{F}_2$ and $L = f(A, B) = A \cdot B$ then*

$$|\mathbb{F}|\beta((A, B); L = 0) = \frac{1}{\sqrt{3}} < \frac{1}{\sqrt{2}} = \beta(A; f(A, 1) = 0) \tag{154}$$

which violates strictly the statement of [PR13, Lemma 4]

Proof. If $f(A, 1) = 0$ then $A = 0$ and $\beta^2(A; f(A, 1) = 0) = \frac{1}{2}$. If $L = 0$ then there are three equilikely cases $(A, B) \in \{(0, 0); (0, 1); (1, 0)\}$ hence $|\mathbb{F}|^2\beta^2((A, B); L = 0) = 4 * \left(3 * (\frac{1}{3} - \frac{1}{4})^2 + (\frac{1}{4})^2\right) = \frac{1}{3}$. □

The error comes from a flawed statement of the chain rule for conditional probabilities in the third equation in the proof of the lemma. Indeed, it is not true that $p(a|bl)p(b) = p(ab|l)$ since $p(b|l) \neq p(b)$ in general. We provided a counter-example to [Lemma 4, [PR13]]. As a consequence, the proof of Theorem 3 in [PR13] recalled below is also flawed.

> *Let A and B be two random variables uniformly distributed over some finite set $\mathcal{X}$. Let d be a positive integer, and let $(A_i)_i$ and $(B_j)_j$ be two dth-order encoding of A and B respectively. Let $\mathcal{E}$ be a real number such that $\mathcal{E} \leqslant \frac{\alpha}{(d+1)|\mathcal{X}|^2}$ for some $\alpha \in (0, 1]$ and let $(f_{i,j})_{i,j}$ be noisy leakage functions defined over $\mathcal{X} \times \mathcal{X}$ and belonging to $\mathcal{N}(\mathcal{E})$. We have:*
>
> $$\beta((A, B); (f_{i,j}(A_i, B_j))_{i,j}) \leqslant 2|\mathcal{X}|^{\frac{3d+2}{2}}(\lambda(\mathcal{E}, d)\mathcal{E})^{d+1} \tag{155}$$
>
> *where $\lambda(\mathcal{E}, d) = \inf_{\alpha \in [(d+1)|\mathcal{X}|^2\mathcal{E}, 1]} \frac{e^\alpha - 1}{\alpha} d + \frac{e^\alpha - 1}{\alpha} + e^\alpha$.*

We propose the following patch to their main theorem based on our derivations. A weakened version of Theorem 2 using Lemma 9 yields:

Patch 1

$$\begin{aligned}
\frac{1}{2}\left(\frac{|\mathcal{X}|}{|\mathcal{X}| - 1}\right)\beta(X; \mathbf{Y}) &\leqslant \overline{\mathcal{E}}(X \to \mathbf{Y}) && (156)\\
&\leqslant 1 - \left(\left(1 - \overline{\mathcal{E}}^{d+1}\right)^{n_1+n_2+n_4}\left(1 - \Upsilon_d(\overline{\mathcal{E}})\right)^{n_3}\right)^q && (157)\\
&\leqslant 1 - \left(\left(1 - (|\mathcal{X}|\beta)^{d+1}\right)^{n_1+n_2+n_4}\left(1 - \Upsilon_d(|\mathcal{X}|^2\beta)\right)^{n_3}\right)^q && (158)\\
&\leqslant q\left((n_1 + n_2 + n_4) + 2n_3(d+1)^{d+1}|\mathcal{X}|^{d+1}\right)|\mathcal{X}|^{d+1}\beta^{d+1}. && (159)
\end{aligned}$$

E.2 "Unifying Leakage Model on a Rényi Day"

The proof of [PGMP19, Lemma 8] is flawed, which also affects [PGMP19, Thm. 6, Corollary 4] and the main result. Recall Lemma 8 from [PGMP19]:

> *Let* A, B be two independent random variable over $\mathbb{F}$ with $|\mathbb{F}|$. Let $f : (a, b) \in \mathbb{F}^2 \mapsto l \in \mathcal{L}$ be a side-channel for (A, B) i.e., $L = f(A, B)$. Then for every $b \in \mathbb{F}$,
>
> $$\mathrm{RE}(A|f(A, b)) \leqslant \mathrm{RE}((A, B)|L). \tag{160}$$

The flaw is similar as in Lemma 4 of [PR13]. The chain rule of probability is used wrongly since $p(b|l) \neq p(b)$. As a consequence, [PGMP19, Thm. 6] recalled below is, unfortunately, incorrect.

> *Let A,B be two uniform random variables over a finite field* $\mathcal{X}$*, d a positive integer, and* (A_i)*,*(B_i) *be* $d + 1$ *additive sharing of A and B respectively. Let* $\delta \in \mathbb{R}$ *such that* $\delta \leqslant \frac{1}{2d+1}$*, and* $(f_{i,j})_{i,j}$ *be a family of randomized and mutually independent functions such that each* $f_{i,j}$ *is* δ_{RE}*-noisy with respect to* RE*. We have:*
>
> $$\mathrm{RE}((A, B)|(f_{i,j}(A_i, B_j))_{i,j}) \leqslant 3\left(\frac{(d + 1)\delta_{\mathrm{RE}}}{1 - d\delta_{\mathrm{RE}}}\right)^{d+1}. \tag{161}$$

We revise the asymptotic evaluation of RE in [PGMP19, Prop. 3]:

Patch 2 *Let* $\mathcal{X} = \mathbb{F}_2^n$ *and consider* $X \sim \mathcal{U}(\mathcal{X})$*. Let* $Y = w_H(X) + Z$ *where is an independent additive Gaussian noise* $Z \sim \sigma\mathcal{N}(0, 1)$*. Then*

$$\mathrm{RE}(X; Y) = 2^n - 1 = |\mathcal{X}| - 1. \tag{162}$$

Proof. First $p_Y(y) \sim \frac{1}{2^n}\binom{n}{n}\varphi_\sigma(y - n)$ as $y \to \infty$. Second for $x = (1, \ldots, 1)$ we have $p_{Y|X}(y|x) = \varphi_\sigma(y - n)$. Hence, $\frac{p_{Y|X}(y|x)}{p_Y(y)} \to 2^n$ as $y \to \infty$. Hence, by sandwich theorem $2^n - 1 \geqslant \sup_x \left|\frac{p_{Y|X}(y)}{p_Y(y)} - 1\right| \geqslant \frac{p_{Y|X}(y|(1,\ldots,1))}{p_Y(y)} - 1$ tends to $2^n - 1$ as $y \to \infty$. This in turn implies that $\mathrm{RE}(X; Y) = \sup_y \sup_x \left|\frac{p_{Y|X}(y)}{p_Y(y)} - 1\right| = 2^n - 1 = |\mathcal{X}| - 1$. □

A potential fix would be to redefine the RE to remove the worst y that are extremely unlikely.

Definition 18 (Alternative Definition of RE**).** *We relax the definition of* RE *to allow it to be large on a set whose probability is less than* $\mathcal{E}$*,*

$$\mathrm{RE}_{\mathcal{E}}(X; Y) \triangleq \inf_{\substack{A \subseteq \mathcal{Y} \\ \mathbb{P}(Y \in A^c) \leqslant \mathcal{E}}} \sup_{y \in A} \sup_{x \in \mathcal{X}} |\mathrm{PMI}(x; y) - 1|. \tag{163}$$

Though, this patch would require to re-derive all results in terms of this new definition. We suggest the following patch in terms of ARE weakening Theorem 2 with Lemma 9:

Patch 3 (Patch Based on CDC**)**

$$\frac{1}{2(|\mathcal{X}|-1)}\mathrm{ARE}(X;\mathbf{Y}) \leqslant \overline{\mathcal{E}}(X \to \mathbf{Y}) \tag{164}$$

$$\leqslant 1 - \left(\left(1-\overline{\mathcal{E}}^{d+1}\right)^{n_1+n_2+n_4}\left(1-\Upsilon_d(\overline{\mathcal{E}})\right)^{n_3}\right)^q \tag{165}$$

$$\leqslant 1 - \left(\left(1-\delta_{\mathrm{ARE}}^{d+1}\right)^{n_1+n_2+n_4}\left(1-\Upsilon_d(\delta_{\mathrm{ARE}})\right)^{n_3}\right)^q. \tag{166}$$

E.3 "Prouff & Rivain Formal Security Proof of Masking, Revisited"

The proof of [MS23b, Thm. 5] is flawed, which affects the main result [MS23b, Thm. 7]. In the last step of the proof the bound is averaged over the values of $\mathbf{b}$ in equation (27). However, the i-th coordinates depends on all the $b_0, \ldots, b_d$ and no only on b_i. Indeed, it is A_i that is fixed on this coordinate. Hence, the average on the i-th coordinate should be over $b_0, \ldots, b_d$ and not only on b_i. But the inequality cannot be obtained this way because the bound provided by Mrs. Gerber lemma is convex in one variable when the others are fixed, but is not jointly convex. Recall [MS23b, Thm. 5],

> *Let A, B be two independent and uniform random variables over a finite field $\mathbb{F}$. Let $(A_i)_{0\leqslant i\leqslant d}, (B_i)_{0\leqslant i\leqslant d}$ be d-encodings of A and B respectively. Let $L_{i,j}$ be the side-channel for the cross term A_i, B_j i.e., we observe $L_{i,j}(A_i, B_j)$. Let $\mathbf{L} = (L_{i,j})_{0\leqslant i,j\leqslant d}$. Further it is assumed that*
>
> $$I((A_i, B_j); L_{i,j}(A_i, B_j)) \leqslant \delta_{i,j} \leqslant \delta_{\mathrm{MI}} \leqslant 1. \tag{167}$$
>
> *Then,*
>
> $$I((A,B);\mathbf{L}) \leqslant f_{MGL}(\sum_j \delta_{0,j}, \ldots, \sum_j \delta_{d,j}) + f_{MGL}(\sum_i \delta_{i,0}, \ldots, \sum_i \delta_{i,d}) \tag{168}$$

We suggest the following patch based on the CDC. A weakened version of Theorem 2 using Lemma 9 yields:

Patch 4 (Patch Based on CDC**)**

$$\frac{I(X;\mathbf{Y})}{\log|\mathcal{X}|} \leqslant \overline{\mathcal{E}}(X \to \mathbf{Y}) \tag{169}$$

$$\leqslant 1 - \left(\left(1-\overline{\mathcal{E}}^{d+1}\right)^{n_1+n_2+n_4}\left(1-\Upsilon_d\left(\overline{\mathcal{E}}\right)\right)^{n_3}\right)^q \tag{170}$$

$$\leqslant 1 - \left(\left(1-\left(\delta_{\mathrm{CDC}}^A\right)^{d+1}\right)^{n_1+n_2+n_4}\left(1-\Upsilon_d\left(\delta_{\mathrm{CDC}}^B\right)\right)^{n_3}\right)^q \tag{171}$$

where

$$\begin{cases} \delta_{\mathrm{CDC}}^A = \min\left(1, |\mathcal{X}|\left(\frac{\delta_{\mathrm{MI}}}{2\log e}\right)^{\frac{1}{2}}\right) \\ \delta_{\mathrm{CDC}}^B = \min\left(1, |\mathcal{X}|^2\left(\frac{\delta_{\mathrm{MI}}}{2\log e}\right)^{\frac{1}{2}}\right). \end{cases} \tag{172}$$

The choice of the uniform prior in the definition of a noisy adversary is arbitrary but makes sense for most cryptographic sensitive variables. However, replacing the uniform prior by the maximum over all input distribution to obtain the analog of a communication "capacity" [4] [CT06] is also meaningful.

Definition 19 (Channel Capacity). *The* capacity *of the channel* $X \to \boxed{P_{Y|X}} \to Y$ *is*

$$C(X \to Y) \triangleq \sup_{p_X} I(X;Y). \tag{173}$$

As pointed out by by Masure & Standaert [MS23b] we can bound capacity (for MI) by a function of the δ noisiness using results on the universality of the uniform prior [SF04]. This roughly leads to an overhead of $O(|\mathcal{X}|)$ in the worst case though. Note that like CDC the channel capacity is a property of the channel (hence the notation with the arrow). This provides another patch for [MS23b, Thm. 5].

Patch 5 (Patch Based on Capacity) *Let* A, B *be two independent and uniform random variables over a finite field* $\mathbb{F}$*. Let* $(A_i)_{0\leqslant i\leqslant d}, (B_i)_{0\leqslant i\leqslant d}$ *be* d*-encodings of* A *and* B *respectively. Consider the side-channels* $((A_i, B_j) \to Y_{i,j})_{0\leqslant i,j\leqslant d}$*. Further, assume that the adversary is a* δ_C *capacity noisy adversary (i.e.,* $C((A_i, B_j) \to Y_{i,j}) \leqslant \delta_C$ *for all* i, j*). Then,*

$$I((A, B); \mathbf{L}) \leqslant 2 f_{\mathrm{MGL}}((d+1)\delta_C, \ldots, (d+1)\delta_C). \tag{174}$$

Proof. We upper-bound every term within the MGL function in [MS23b, Eqn. 27] by $(d+1)\delta_C$. We do not need to take the average anymore since we upper-bounded [MS23b, Eqn. 27] uniformly in $\mathbf{b}$. □

Also, [MS23b, Coro. 2] can be slightly improved:

Patch 6 (Slightly Improved Corollary 2) *Let* Y *be uniformly distributed over* $\mathbb{F}$*. Let* $Y \to Y^k \to L$ *be a* δ_{MI}*-noisy with respect to* MI *channel. Then,*

$$I(Y; L) \leqslant \mathrm{GCD}(k, |\mathcal{X}| - 1)\delta_{\mathrm{MI}} \tag{175}$$

where we removed the $\frac{|\mathcal{X}|-1}{|\mathcal{X}|}$ *overhead factor from the original lemma.*

Proof. Indeed, for $s = y^k$, $s \neq 0$,

$$\mathbb{P}(Y^k = s | Y \neq 0) = \frac{\mathrm{GCD}(k, |\mathcal{X}| - 1)}{|\mathcal{X}| - 1}. \tag{176}$$

Hence,

$$\begin{aligned}
\mathbb{P}(Y^k = s) &= \mathbb{P}(Y \neq 0)\mathbb{P}(Y^k = s | Y \neq 0) && (177)\\
&= \frac{|\mathcal{X}| - 1}{|\mathcal{X}|} \frac{\mathrm{GCD}(k, |\mathcal{X}| - 1)}{|\mathcal{X}| - 1} && (178)\\
&= \frac{\mathrm{GCD}(k, |\mathcal{X}| - 1)}{|\mathcal{X}|}. && (179)
\end{aligned}$$

We can conclude using the original proof from [MS23b, Coro. 2]. □

[4] This term is usually reserved for communication problems.

Now we can state a revisited [MS23b, Thm. 7]:

Proposition 8. *Consider a δ_C-noisy adversary with respect to capacity on the same setting of [MS23b, Thm. 7].*

$$I(Y, \mathbf{L}) \leqslant t_3 2 f_{\mathrm{MGL}}((d+1)\delta_C, \ldots, (d+1)\delta_C) + t_{1,2,4} f_{\mathrm{MGL}}(\delta_C, \ldots, \delta_C) \tag{180}$$

where

$$t_3 = \sum_{p,q \in \mathcal{M}} \varphi(p, q, |\mathcal{Y}|), \qquad t_{1,2,4} = \sum_{p,q \in \mathcal{M}} \xi(p, q, |\mathcal{Y}|) + \sum_{k \in \mathcal{S}} \psi(k, |\mathcal{Y}|) \tag{181}$$

and

$$\begin{cases} \varphi(p, q, |\mathcal{Y}|) = |\mathcal{Y}| \min\{\mathrm{GCD}(p, |\mathcal{Y}| - 1), \mathrm{GCD}(q, |\mathcal{Y}| - 1)\} \\ \xi(p, q, |\mathcal{Y}|) = \mathrm{GCD}(p + q, |\mathcal{Y}| - 1) \\ \psi(k, |\mathcal{Y}|) = \mathrm{GCD}(k, |\mathcal{Y}| - 1). \end{cases} \tag{182}$$

E.4 "Making Masking Security Proof Concrete"

We revise several claims from [DFS19]. Recall [DFS19, Thm. 2]:

> Let d be the number of shares used for a key encoding, m be the number of measurements, and $I(Y_i; \mathbf{L}_{Y_i}) \leqslant \delta_{\mathrm{MI}}$ for some $\delta_{\mathrm{MI}} \leqslant \frac{2}{|\mathcal{X}|^2}$. Then, if we refresh the encoding in a leak-free manner between each measurement, the probability of success of a key recovery adversary under independent leakage is:
>
> $$\mathbb{P}_s \leqslant 1 - \left(1 - \left(|\mathcal{X}| \sqrt{\frac{\delta_{\mathrm{MI}}}{2}}\right)^d\right)^m. \tag{183}$$

Equation. (183) is slightly incorrect because when the side-information is "erased" the adversary still guesses the secret with probability $\frac{1}{|\mathcal{X}|}$. Also, implicitly the unit of the MI is the nats here. We prefer to normalize the inequality by $\log e$ so that the result holds in all basis and to make explicit that $\frac{\delta_{\mathrm{MI}}}{loge}$ has no physical dimension:

Patch 7

$$\mathbb{P}_s \leqslant 1 - \left(1 - \left(|\mathcal{X}| \sqrt{\frac{\delta_{\mathrm{MI}}}{2 \log e}}\right)^d\right)^m \left(1 - \frac{1}{|\mathcal{X}|}\right). \tag{184}$$

Proof. As in [DFS19] the adversary is reduced to a random probing adversary with probing probability $\overline{\mathcal{E}} = |\mathcal{X}| \sqrt{\frac{\delta_{\mathrm{MI}}}{2 \log e}}$. The adversary probes all d shares with probability $\overline{\mathcal{E}}^d$. This does not happen with probability $1 - \overline{\mathcal{E}}^d$. This does

not happen for all m queries with probability $\left(1-\overline{\mathcal{E}}^d\right)^m$. In this case the SR is $\frac{1}{|\mathcal{X}|}$ otherwise it is 1 so that

$$\mathbb{P}_s \leqslant \left(1-\overline{\mathcal{E}}^d\right)^m \frac{1}{|\mathcal{X}|} + 1 - \left(1-\overline{\mathcal{E}}^d\right)^m = 1 - \left(1-\overline{\mathcal{E}}^d\right)^m \left(1-\frac{1}{|\mathcal{X}|}\right). \quad (185)$$

□

[DFS19, Coro. 1] is revised accordingly:

Patch 8 *In the same setting as [DFS19, Coro. 1] we have:*

$$0 \leqslant \mathbb{P}_s - \frac{1}{|\mathcal{X}|} \leqslant m\left(1-\frac{1}{|\mathcal{X}|}\right)\exp\left(-\alpha d\right) \quad (186)$$

where

$$\alpha = -\log\left(|\mathcal{X}|\sqrt{\frac{\delta_{\mathrm{MI}}}{2\log e}}\right) > 0. \quad (187)$$

E.5 "Unifying Leakage Models: from Probing Attacks to Noisy Leakage"

Tightness Recall Theorem 1 from [DDF14].

> *Let Γ be an arbitrary stateful arithmetic circuit over some field $\mathbb{F}$. Let Γ' be the circuit derived from Γ using the ISW compiler. Then Γ' is a $(\delta, |\Gamma|\exp(-\frac{d}{12}))$-noise-resilient implementation of Γ where*
>
> $$\delta \triangleq ((28d+16)|\mathbb{F}|)^{-1} = O((d|\mathbb{F}|)^{-1}). \quad (188)$$

As pointed out by Masure & Standaert in [MS23b], [DDF14, Thm. 1] is limited in the sense that the security bound does not depend on the noise level. The bound only depends on the masking order. Further, the required number of shares to obtain a negligible probability of error is prohibitive. Indeed, the bound on the adversary's advantage is

$$|\Gamma|\exp\left(-\frac{d}{12}\right). \quad (189)$$

Since the advantage is always upper bounded by 1 this bound is non-trivial only if $|\Gamma|e^{-\frac{-d}{12}} \leqslant 1$ which happens when $d \geqslant \lceil 12\log|\Gamma|\rceil$. If we consider a very minimalistic implementation with only $|\Gamma| = 2$ gadgets then we already need the unpractical masking order $d \geqslant 9$. This non-tightness essentially comes from the use of concentration inequality. It is applied in a non-asymptotic context to the number of wires l in a gadget. [DDF14] is improved by removing this concentration inequality.

Random Probing to t-Threshold Probing. [DDF14, Lemma 4] shows that the security in the random probing model can be reduced to security in the t-threshold probing model using a Markov style argument. The original proof comports two minor flaws. First the simulator presented in [DDF14, Appendix E] is not $2\overline{\mathcal{E}}l - 1$ threshold probing but $\lceil 2\overline{\mathcal{E}}l - 1 \rceil$ threshold probing. Further it is not true that [DDF14, Equation 23] holds, we clarify why in our proof. We revise [DDF14, Lemma 4] as follows:

Patch 9 *Let $\mathcal{A}$ be an $\overline{\mathcal{E}}$-random-probing adversary on $\mathcal{X}^l$. Then for all $t \in \mathbb{N}$, there exists a t-threshold probing adversary $\mathcal{S}$ on $\mathcal{X}^l$ such that for every $(x_1, \ldots, x_l) \in \mathcal{X}^l$ we have,*

$$\Delta(\mathrm{out}_{\mathcal{A}}(x_1, \ldots, x_l); \mathrm{out}_{\mathcal{S}}(x_1, \ldots, x_l)) \leqslant Q_B(t, l, \overline{\mathcal{E}}) \tag{190}$$

where $Q_B(\cdot, l, \overline{\mathcal{E}})$ is the survival function of the binomial distribution with parameters $(l, \overline{\mathcal{E}})$.

Proof. We follow the proof-line of [DDF14]. Let $\mathbf{x} \triangleq (x_1, \ldots, x_l) \in \mathcal{X}^l$. As explained in [DDF14] we can assume without loss of generality that the adversary outputs all the information that he obtains i.e.,

$$\mathrm{out}_{\mathcal{A}}(\mathbf{x}) = (\varphi_1(x_1), \ldots, \varphi_l(x_l)) \tag{191}$$

where $\varphi_1, \ldots, \varphi_l$ are erasure channel with probability of erasure $1 - \overline{\mathcal{E}}$. We show that $\mathcal{A}$ can be simulated by a t-threshold probing adversary $\mathcal{S}$. First, $\mathcal{S}$ draw l iid Bernoulli random variables $B_1, \ldots, B_l$ with parameter $\overline{\mathcal{E}}$. Let $S_l = B_1 + \cdots + B_l$, $S_l \sim \mathcal{B}(l, \overline{\mathcal{E}})$.

- If $S_l \leqslant t$ then $\mathcal{S}$ queries the values x_i such that $B_i = 1$. Then $\mathcal{S}$ outputs the vector $(y_1, \ldots, y_l)$ where $y_i = x_i$ if $B_i = 1$ and $y_i = \perp$ if $B_i = 0$.
- Else $S_l > l$ then $\mathcal{S}$ fails and outputs the error vector $(\perp, \ldots, \perp)$.

Let $N_{\overline{\mathcal{E}}}$ be the random variable that counts the number of values that $\mathcal{A}$ receives without erasures (i.e., $l - N_{\overline{\mathcal{E}}}$ is the number of erasures that $\mathcal{A}$ receives). The distribution of $\mathrm{out}_{\mathcal{A}}(\mathbf{x})$ given the event $(N_{\overline{\mathcal{E}}} \leqslant t)$ is equal to the distribution of $\mathrm{out}_{\mathcal{S}}(\mathbf{x})$ given the event $(S_l \leqslant t)$[5]. Let $\Delta \triangleq \Delta(\mathrm{out}_{\mathcal{A}}(\mathbf{x}); \mathrm{out}_{\mathcal{S}}(\mathbf{x}))$ and $y \in \mathcal{Y} \triangleq (\mathcal{X} \cup \{\perp\})^l$. By the law of total probability $\mathbb{P}(\mathrm{out}_{\mathcal{A}}(\mathbf{x}) = y) = \mathbb{P}(N_{\overline{\mathcal{E}}} \leqslant t)\mathbb{P}(\mathrm{out}_{\mathcal{A}}(\mathbf{x}) = y | N_{\overline{\mathcal{E}}} \leqslant t) + \mathbb{P}(N_{\overline{\mathcal{E}}} > t)\mathbb{P}(\mathrm{out}_{\mathcal{A}}(\mathbf{x}) = y | N_{\overline{\mathcal{E}}} > t)$, and $\mathbb{P}(\mathrm{out}_{\mathcal{S}}(\mathbf{x}) = y) = \mathbb{P}(S_l \leqslant t)\mathbb{P}(\mathrm{out}_{\mathcal{S}}(\mathbf{x}) = y | S_l \leqslant t) + \mathbb{P}(S_l > t)\mathbb{P}(\mathrm{out}_{\mathcal{S}}(\mathbf{x}) = y | S_l > t)$. Further,

$$\begin{cases} \mathbb{P}(S_l \leqslant t) = \mathbb{P}(N_{\overline{\mathcal{E}}} \leqslant t) \\ \mathbb{P}(S_l > t) = \mathbb{P}(N_{\overline{\mathcal{E}}} > t) = Q_B(t, l, \overline{\mathcal{E}}) \\ \mathbb{P}(\mathrm{out}_{\mathcal{S}}(\mathbf{x}) = y | S_l \leqslant t) = \mathbb{P}(\mathrm{out}_{\mathcal{A}}(\mathbf{x}) = y | N_{\overline{\mathcal{E}}} \leqslant t) \end{cases} \tag{192}$$

[5] However, it is not true to say that conditionally to $S_l < t$ then $\mathrm{out}_{\mathcal{S}}(\mathbf{x})$ is equal in distribution with $\mathrm{out}_{\mathcal{A}}(\mathbf{x})$ as claimed in [DDF14]. Indeed $\mathrm{out}_{\mathcal{A}}(\mathbf{x})$ is independent from the event $S_l < t$ hence its distribution is the same when conditioned on this event. On the one hand it is possible that $S_l < t$ but $\mathrm{out}_{\mathcal{A}}(\mathbf{x})$ outputs all the values $(x_1, \ldots, x_l)$ without any erasure. On the other hand when conditioned on the event $S_l < t$, $\mathrm{out}_{\mathcal{S}}(\mathbf{x})$ will produce at least $l - t$ erasures.

so part of the terms cancels, and we obtain $\mathbb{P}(\mathrm{out}_{\mathcal{A}}(\mathbf{x}) = y) - \mathbb{P}(\mathrm{out}_{\mathcal{S}}(\mathbf{x}) = y) = \mathbb{P}(S_l > t)\left(\mathbb{P}(\mathrm{out}_{\mathcal{A}}(\mathbf{x}) = y | N_{\overline{\mathcal{E}}} > t) - \mathbb{P}(\mathrm{out}_{\mathcal{S}}(\mathbf{x}) = y | S_l > t)\right)$.

Now let U be a random variable that follows the distribution of $\mathrm{out}_{\mathcal{A}}(\mathbf{x})$ given $N_{\overline{\mathcal{E}}} > t$ and V be a random variable that follows the distribution of $\mathrm{out}_{\mathcal{S}}(\mathbf{x})$ given $S_l > t$. We can conclude the proof since

$$\begin{aligned}
\Delta &= \frac{1}{2}\sum_{y\in\mathcal{Y}} \left|\mathbb{P}(\mathrm{out}_{\mathcal{A}}(\mathbf{x}) = y) - \mathbb{P}(\mathrm{out}_{\mathcal{S}}(\mathbf{x}) = y)\right| && (193)\\
&= \frac{1}{2}\sum_{y\in\mathcal{Y}} \left|Q_B(t,l,\overline{\mathcal{E}})\left(\mathbb{P}(\mathrm{out}_{\mathcal{A}}(\mathbf{x}) = y | N_{\overline{\mathcal{E}}} > t) - \mathbb{P}(\mathrm{out}_{\mathcal{S}}(\mathbf{x}) = y | S_l > t)\right)\right| && (194)\\
&= Q_B(t,l,\overline{\mathcal{E}})\frac{1}{2}\sum_{y\in\mathcal{Y}} \left|\mathbb{P}(\mathrm{out}_{\mathcal{A}}(\mathbf{x}) = y | N_{\overline{\mathcal{E}}} > t) - \mathbb{P}(\mathrm{out}_{\mathcal{S}}(\mathbf{x}) = y | S_l > t)\right| && (195)\\
&= Q_B(t,l,\overline{\mathcal{E}})\Delta(U;V) && (196)\\
&\leqslant Q_B(t,l,\overline{\mathcal{E}}). && (197)
\end{aligned}$$

□

[DDF14, Lemma 4] is used in [DFS19, Thm. 3] and [PGMP19, Lemma 5] so our patch directly improves their corresponding results.

E.6 Optimal Masking Order in Other Bounds

Proposition 6 may appear as contradictory with [DFS19, Coro. 2]. It turns out that [DFS19, Coro. 2] is incorrectly derived from [DFS19, Thm. 3]. Namely, inverting equation (13) from [DFS19, Thm. 3] yields

$$d \leqslant d^*_{\mathrm{DFS}} \triangleq \left\lfloor \frac{1 - 16|\mathcal{X}|\sqrt{\frac{\delta_{\mathrm{MI}}}{2\log e}}}{28|\mathcal{X}|\sqrt{\frac{\delta_{\mathrm{MI}}}{2\log e}}} \right\rfloor = \left\lfloor \frac{1}{28|\mathcal{X}|}\left(\frac{\delta_{\mathrm{MI}}}{2\log e}\right)^{-\frac{1}{2}} - \frac{4}{7} \right\rfloor \tag{198}$$

while the inequality is in the opposite direction in [DFS19, Coro. 2]. It follows that equation (15) in [DFS19, Coro. 2] does not hold. This corollary may lead to the misleading conclusion that provided that the masking order is high enough then the adversary's advantage can be made arbitrarily small. The patched inequality however indicates that the bound derived in [DFS19, Thm. 3] is only valid up to a given maximal masking order d^*_{DFS}. In [DDF14, Thm. 1], it is required that $\delta^{-1}_{\mathrm{TVI}} \geqslant (28d + 16)|\mathcal{X}|$ which in turn implies for a fixed noise level and because d is an integer that

$$d \leqslant d^*_{\mathrm{DDF}} \triangleq \left\lfloor \frac{\delta^{-1}_{\mathrm{TVI}}}{28|\mathcal{X}|} - \frac{4}{7} \right\rfloor. \tag{199}$$

The security bound provided in [DDF14, Thm. 1] being decreasing with respect to the masking order d, d^*_{DDF} is the optimal masking order predicted by the

inequality. The advantage of the adversary cannot be reduced further than

$$|\Gamma|\exp\left(-\frac{d^*_{\mathrm{DDF}}}{12}\right) = |\Gamma|\exp\left(-\frac{1}{12}\left\lfloor\frac{\delta^{-1}_{\mathrm{TVI}}}{28|\mathcal{X}|} - \frac{4}{7}\right\rfloor\right) \tag{200}$$

with this security proof. Once again this is in line with Proposition 6 and the observation of Battistello *et al.* [BCPZ16], from a certain rank as d increases the bound on the adversary's advantage increases in ISW gadgets whose noise rate is not constant.

In [PR13, Coro. 2] it is required that $d \leqslant \left\lfloor \delta^{-1}_{\mathrm{EN}}|\mathcal{X}|^{-2}\right\rfloor$ which implies the existence of a finite optimal masking order $d^*_{\mathrm{EN}} \leqslant \left\lfloor \delta^{-1}_{\mathrm{EN}}|\mathcal{X}|^{-2}\right\rfloor$ with respect to the bound. Also in [PGMP19, Thm. 6] it is required that $\delta^{-1}_{\mathrm{RE}} \geqslant 2d+1$ i.e., the derivation holds for $d \leqslant \left\lfloor \frac{1}{2}(\delta^{-1}_{\mathrm{RE}} - 1)\right\rfloor$. Also the security bound presented in [PGMP19] also has an optimal masking order $d^*_{\mathrm{RE}} \leqslant \left\lfloor \frac{1}{2}(\delta^{-1}_{\mathrm{RE}} - 1)\right\rfloor$. In [MS23b] it is required that $\delta^{-1}_{\mathrm{MI}} \geqslant d+1$ which in turn implies that there is an optimal masking order $d^*_{\mathrm{MI}} \leqslant \lfloor \delta^{-1}_{\mathrm{MI}}\rfloor$. For these three bounds a term of the form $((d+1)\delta)^{d+1}$ appear which is maximized for $d+1 \in \{\lfloor(\delta e)^{-1}\rfloor, \lceil(\delta e)^{-1}\rceil\}$. This indicates that in these bounds it is sub-optimal with respect to the security bound to maximize the masking order.

In the direct security proof (Theorem 2) the bottleneck is associated to the type 3 subsequences. While it is complicated to derive analytically the minimal value of $\Upsilon_d(\overline{\mathcal{E}})$ with respect to d we know that $(1-\mathcal{E}^{d+1})^{d+1} \leqslant \Upsilon_d(\overline{\mathcal{E}}) \leqslant 1$. Since $(d+1)\ln(1-\mathcal{E}^{d+1}) \sim -d\mathcal{E}^{d+1} \to 0$, $(1-\mathcal{E}^{d+1})^{d+1} \to 1$ and by the sandwich theorem we obtain that $\Upsilon_d(\overline{\mathcal{E}}) \to 1$ as $d \to +\infty$. This implies the existence of an optimal masking order $d^*(\mathcal{E})$ with respect to the direct security proof that can be computed numerically.