Security Assessment Against Side-Channel Attacks : Insights from an Information-Theoretic Perspective

Yi Liu

INSTITUT POLYTECHNIQUE DE PARIS

NNT : 2023IPPAT033

Thèse de doctorat

Security Assessment Against Side-Channel Attacks: Insights from an Information-Theoretic Perspective

Thèse de doctorat de l'Institut Polytechnique de Paris
préparée à Télécom Paris

École doctorale n°626 École doctorale de l'Institut Polytechnique de Paris (ED IP Paris)
Spécialité de doctorat : Réseaux, Informations et Communications

Thèse présentée et soutenue à Palaiseau, le 15 novembre 2023, par

YI LIU

Composition du Jury :

Michael Gastpar Professeur, EPFL (LINX)	Président & Rapporteur
Sihem Mesnager Professeure, Paris 8 (Département de Mathématiques)	Examinatrice
Alexander Barg Professeur, University of Maryland (CSM Department)	Examinateur
François-Xavier Standaert Professeur, UCLouvain (ICTEAM)	Examinateur
Vincent Grosso Chargé de recherche, CNRS (Laboratoire Hubert Curien)	Examinateur
Sheng Yang Professeur, Central Supelec (L2S)	Examinateur
Olivier Rioul Professeur, Télécom Paris (COMELEC)	Directeur de thèse
Patrick Solé Directeur de recherche, CNRS (Institut de Maths de Marseille)	Co-directeur de thèse

Acknowledgements

Time flies, and three years have passed in the blink of an eye. At this moment, as I stand on the brink of graduation, looking back at the path I've traveled, I am suddenly reminded of an ancient Chinese poem: "Yearning to speak yet holding back, instead remarking on the cool and pleasant autumn."

If it weren't for the care, encouragement, support, and love from the people around me, I might have found it difficult to persist to this day and earn this PhD degree. Therefore, I dedicate this thesis, and this sincere acknowledgment, to everyone. It belongs not just to one person or a few people, but to all of us who have shared those beautiful moments together.

The people I am most grateful for are my three supervisors: Olivier, Patrick, and Sylvain. When I first came to France three years ago, I found everything unfamiliar and novel, and naturally, there were many things I had to adapt to. Fortunately, my supervisors were very enthusiastic and kind, which helped me to adapt and integrate more quickly into this environment. They took me to delicious French restaurants and sometimes we would have Chinese food together before meetings. These touched me deeply, but due to my limited English proficiency, I never had the chance to express this deep appreciation. When I first started my PhD, my abilities were lacking in many areas: my English communication was not smooth, I didn't know how to write a good research paper, and I found it difficult to orally present my thoughts and research findings. These shortcomings brought many difficulties to my doctoral studies, but my advisors were very patient and provided me with a lot of useful advice. I often felt guilty, thinking that the gap between my level and that of my supervisors meant they had to spend more time and effort in guiding me. Any progress I have made in any academic direction over these three years is inseparable from their teaching by word and example. In them, I saw the wonderful qualities that a good professor should possess: a passion for academia, very intelligent and good at thinking, serious and responsible, honest and kind. I am fortunate to have them as my doctoral supervisors; they are undoubtedly the best academic role models for me. Especially Olivier, as my main supervisor, has invested a lot of time and effort in guiding me. He has given me many valuable suggestions and is someone I deeply admire in academia. I sincerely thank my supervisors and hope they are always successful, healthy, and happy. I will always remember these three years of life in France and the beautiful memories we shared.

I also want to thank my family, partner, and friends. They have always supported and encouraged me consistently, making me feel the happiness and warmth of being surrounded by love. I thank Wei for his help when I was just starting and feeling lost. His profound knowledge benefited me greatly, whether in our discussions during meetings or

the exchanges on the way home afterwards. Thanks to Julien, who is very smart and kind, with high intelligence and emotional quotients, like a little sun bringing warmth to those around him. And I am thankful for meeting Xiaolin in the last year of my Ph.D. In my heart, he is a perfect business partner, as he is not only capable but always full of encouragement. We have already created a website together, and I look forward to our future projects going further.

Thanks to all the friends I met in France, Yinghao, Ziyu, Di, Yihan, Xiang, Jiayi, Yibo, Yue, Jingtian, Peter, etc., and my old friends, Lijia, Shuang, Chenfeng, Xinghua, Xiaoyu, Yueping, etc. Thanks to my love Shihao. He took care of me meticulously when I was sick, and his encouraging companionship supported me when I was down. He is my serendipity in France, and the romance that warms my life. Thanks to my dearest Nan, since we met in 2010, we have accompanied each other ovor 13 years. Since that year, I've had a sister and never felt lonely again. Thanks to all my family, I love you forever.

Thanks to the scholarship provided by CSC, and to my school, Telecom Paris.

Words are too short to express deep feelings, and I don't intend to list everyone's names and our beautiful memories here. But I want to say, one day I will leave, and the world will no longer remember a person named LIU Yi who earned a Ph.D. However, the beautiful times we have experienced together will always remain in the river of history, becoming the most beautiful star in the star-filled sky.

Publications

[1] W. Cheng, Y. Liu, S. Guilley and O. Rioul, "Towards finding best linear codes for side-channel protections," in Proc. 10th International Workshop on Security Proofs for Embedded Systems (PROOFS 2021), Beijing, China, Sept. 17, 2021, EPiC Series in Computing, vol. 87, 2022, pp. 83-99.

[2] Y. Liu, W. Cheng, S. Guilley, and O. Rioul, "On conditional alpha-information and its application in side-channel analysis," in Proc. 2021 IEEE Information Theory Workshop (ITW 2021), Kanazawa, Japan, Oct. 17-21, 2021.

[3] P. Solé, Y. Liu, W. Cheng, S. Guilley, and O. Rioul, "Linear programming bounds on the kissing number of q-ary codes," in Proc. 2021 IEEE Information Theory Workshop (ITW2021), Kanazawa, Japan, Oct. 17-21, 2021.

[4] Y. Liu, W. Cheng, S. Guilley, and O. Rioul, "Evaluation of side-channel attacks using alpha-information," 18th International Workshop on Cryptographic Architectures Embedded in Logic Devices (CryptArchi 2022), Porquerolles, France, May 29-June 1st, 2022.

[5] W. Cheng, O. Rioul, Y. Liu, J. Béguinot, and S. Guilley, "Side-channel information leakage of code-based masked implementations," 17th Canadian Workshop on Information Theory (CWIT 2022), Ottawa, Ontario, Canada, June 5-8, 2022.

[6] W. Cheng, Y. Liu, S. Guilley, and O. Rioul, "Attacking masked cryptographic implementations: Information-theoretic bounds," in Proc. 2022 IEEE International Symposium on Information Theory (ISIT 2022), Espoo, Finland, June 26-July 1, 2022.

[7] W. Cheng, Y. Liu, S. Guilley, and O. Rioul, "Towards finding best linear codes for side-channel protections (extended version)," Journal of Cryptographic Engineering, online 15 Nov. 2022.

[8] J. Béguinot, W. Cheng, S. Guilley, Y. Liu, L. Masure, O. Rioul, and F.-X. Standaert, "Removing the field size loss from Duc et al.'s conjectured security bound for masked

encodings," in Proc. 14th International Workshop on Constructive Side-Channel Analysis and Secure Design (COSADE 2023), Munich, Germany, Apr. 3-4, 2023. Lecture Notes on Computer Science, Vol. 13979, pp. 86–104, Springer Nature, 2023.

[9] J. Béguinot, O. Rioul, S. Guilley, W. Cheng, and Y. Liu, "Side-Channel Security. How Much Are You Secure? Mrs. Gerber's Lemma and Majorization," Poster, Colloque de l'Institut Mines-Télécom, Risques & Cyber : Sécurité et Résilience, Palaiseau, France, Apr. 13, 2023.

[10] Y. Liu, J. Béguinot, W. Cheng, S. Guilley, L. Masure, O. Rioul, and F.-X. Standaert, "Improved alpha-information bounds for higher-order masked cryptographic implementations," IEEE Information Theory Workshop (ITW 2023), Saint Malo, France, Apr. 23-28, 2023.

[11] O. Rioul, J. Béguinot, Y. Liu, W. Cheng, and S. Guilley, "A generic information-theoretic framework for evaluating the side-channel security of masked implementations," 19th International Workshop on Cryptographic Architectures Embedded in Logic Devices (CryptArchi 2023), Cantabria, Spain, June 11-14, 2023.

[12] J. Béguinot, Y. Liu, O. Rioul, W. Cheng, and S. Guilley, "Maximal leakage of masked implementations using Mrs. Gerber's lemma for min-entropy," in Proc. IEEE International Symposium on Information Theory (ISIT 2023), Taipei, Taiwan, June 25-30, 2023.

[13] Y. Liu, W. Cheng, O. Rioul, S. Guilley, and P. Solé, "Kissing number of codes: A survey," Chapter, VIASM, Springer Lecture Notes in Mathematics, to appear, 2023.

Contents

Abstract (English)

Context

In today's world, cryptographic devices are almost omnipresent, and the necessity to guarantee their security has been increasing. When a cryptographic device is operating, any kind of unintended leakage (time, power, electromagnetic, etc.) can be exploited by an attacker. By querying the device multiple times, measuring the corresponding leakages, and correlating them with internal sensitive values, the attacker is able to guess the secret key.

Such side-channel attacks (SCAs) pose a threat to cryptographic devices. Evaluating the security of cryptographic devices against SCAs is important for both the industrial and academic sectors, and information-theoretic metrics turn out to be effective tools. To thwart SCAs, a well-established countermeasure is masking. The issues of how to achieve more effective masking, as well as how to evaluate the security of masked implementations, have become topics of widespread concern.

My main goals are to **quantify side-channel leakage**, **evaluate the security of** (both unprotected and masked) **cryptographic devices against SCAs**, and pursue strategies to **construct more effective masking codes**.

Structure of the Manuscript

The main contributions can be divided into three chapters:

The first chapter (Chapter 2) is about the construction of masking codes:

- recent research on code-based masking indicates that the protection becomes more efficient when the dual distance of the masking code is large, and the kissing number of the dual of the masking code is low. Motivated by this, we employ linear programming to derive bounds for the kissing number of q-ary linear codes with a given minimum distance;

- we show the effect of code-based masking is related to the weight enumeration of the dual of the masking code. We first present mathematical tools to study those weight enumerators, and then provide an efficient method to search for good codes, based on a lexicographic sorting of the weight enumerators from lowest to highest degrees.

The second chapter (Chapter 3) evaluates the side-channel leakage of (both unprotected and masked) cryptographic implementations:

- we proposed a conditional version of Sibson's alpha-information by a simple closed-form "log-expectation" expression. This definition satisfies important properties such as consistency, uniform expansion, and data processing inequalities. Based on this definition we evaluate the side-channel leakage of unprotected devices;

- we investigate how a code-based masked implementation leaks in an information-theoretic setting, and establish that the mutual information between the sensitive variable and the leakage decreases as the measurement noise variance increases, with an exponent equal to the dual distance of the masking code.

The final chapter (Chapter 4) is about the security analysis of masked cryptographic devices against side-channel attacks. We use alpha information measures to analyze the implementation of Boolean masking and Arithmetic masking, and derive the upper bound of the probability of success. More precisely, the divergence between the probability of success of an ML attack and a blind guess is bounded by alpha information between the sensitive variable and the side-channel leakage. Further, when $\alpha = 2$, this alpha information can be upper-bounded by a function of the leakage information quantity of each masking share.

Meanwhile, these bounds also provide lower bounds on the minimum number of queries required to achieve a given success rate, which improves the most advanced bound currently available. An important issue, resolved in this part, is the removal of the loss factor due to the masking field size.

Lastly, we discussed other metrics used in similar evaluations.

Résume (Français)

Contexte

À l'heure actuelle, les dispositifs cryptographiques jouent un rôle presque incontournable dans notre monde moderne, et la garantie de leur sécurité représente une préoccupation de plus en plus prégnante. Lorsqu'un dispositif cryptographique est opérationnel, toute fuite non intentionnelle, qu'elle soit liée au temps, à la puissance, à l'électromagnétisme, etc., devient une potentielle vulnérabilité exploitée par des attaquants. En interrogeant le dispositif de manière répétée, en mesurant les fuites correspondantes, et en les corrélant avec des valeurs sensibles internes, un attaquant peut parvenir à deviner la clé secrète utilisée.

Les attaques par canaux auxiliaires (SCAs), telles que décrites, posent une menace significative pour les dispositifs cryptographiques. En conséquence, l'évaluation de la sécurité de ces dispositifs face aux SCAs devient une préoccupation majeure, pour laquelle les métriques informationnelles se révèlent être des outils efficaces. Afin de contrer les SCAs, le masquage émerge comme une contre-mesure bien établie. Cependant, la recherche actuelle se concentre sur l'optimisation du masquage pour le rendre plus efficace, tout en évaluant la sécurité des implémentations masquées, constituant ainsi un domaine de recherche majeur.

Mes principaux objectifs sont de **quantifier les fuites par canaux auxiliaires**, d'**évaluer la sécurité des dispositifs cryptographiques** (non protégés et masqués) **contre les SCAs**, et de créer des méthodes permettant de **construire des codes de masquage plus efficaces**.

Structure du manuscrit

Les principales contributions peuvent être divisées en trois chapitres :

Le premier chapitre (Chapitre 2) concerne la construction de codes de masquage :

Les recherches les plus récentes dans le domaine du masquage basé sur le code indiquent clairement que l'efficacité de la protection atteint son maximum lorsque la distance duale du code de masquage est substantielle et que le nombre de contacts du code dual du code de masquage est maintenu à un niveau bas. Dans cette perspective, nous procédons à la démonstration rigoureuse des bornes relatives au nombre de contacts des codes linéaires q-aires, en tenant compte d'une distance minimale prédéfinie. Cette démonstration repose

sur l'utilisation de techniques avancées de programmation, nous permettant ainsi de définir avec précision ces bornes pour garantir une sécurité optimale dans les implémentations de masquage. En suivant cette approche, nous contribuons à l'évolution continue des méthodologies de protection des dispositifs cryptographiques, en particulier ceux qui reposent sur des codes linéaires q-aires, pour assurer une défense robuste contre les attaques potentielles.

Nous montrons que l'effet du masquage basé sur les codes est lié au polynome énumérateur des poids du code dual du code de masquage. Nous présentons d'abord des outils mathématiques pour étudier ce polynome énumérateur des poids, puis nous proposons une méthode efficace pour rechercher de bons codes, basée sur un tri lexicographique des polynomes énumérateurs des poids du degré le plus bas au plus élevé.

Le deuxième chapitre (Chapitre 3) évalue les fuites par canaux auxiliaires des implémentations cryptographiques (non protégées et masquées) :

Nous proposons une version conditionnelle de l'alpha-information de Sibson avec une formule explicite simple en « log-moyenne ».

Cette définition satisfait des propriétés importantes telles que la consitence, développement uniforme et les inégalités de traitement des données. Avec cette définition, nous évaluons les fuites par canaux auxiliaires des dispositifs non protégés .

Nous étudions comment une implémentation masquée basée sur les codes fuit dans un cadre informationnel et établissons que l'information mutuelle entre les variables sensibles et les fuites observées diminue lorsque la variance du bruit de mesure augmente, avec un exposant égal à la distance minimale du code duale du code de masquage.

Le dernier chapitre (Chapitre 4) concerne l'analyse de sécurité des dispositifs cryptographiques masqués contre les attaques par canaux auxiliaires. Nous utilisons des mesures d'information alpha pour analyser l'implémentation du masquage booléen et du masquage arithmétique, et bornons supérieurement la probabilité de succès du meilleur adversaire. Plus précisément, la divergence entre la probabilité de succès d'une attaque par maximum de vraisemblance et une supposition à l'aveugle est limitée par l'alpha information généralisée entre la variable sensible et la fuite du canal auxiliaire. De plus, lorsque $\alpha = 2$, cette information généralisée peut être bornée supérieurement par une fonction de l'information de fuite de chaque part de masquage.

En même temps, ces limites fournissent également des bornes inférieures sur le nombre minimum de requêtes nécessaires pour atteindre un taux de succès donné, ce qui améliore les meilleurs bornes de l'état de l'art. Un problème important, résolu dans cette partie, est de retirer une constante multiplicative égal à la taille du corps consideré des bornes obtenues.

CHAPTER 1

Background, Notations, and Research Status

Contents

1.1 Side-Channel Analysis

In contemporary society, the stature and function of information have witnessed a marked ascendance, and it has become a crucial strategic resource for societal development. Information technology is transforming the way people live and work, and the information industry has contributed significantly to economic growth. Concurrently, the safeguarding of information has emerged as an escalating concern of paramount importance for numerous individuals and entities.

The objectives of information security include: *confidentiality*, *integrity*, *authentication*, and *non-repudiation*. As the cornerstone of information security, cryptography provides protection for various data and communications, ensuring that only the appropriate recipients can access and understand this data.

Development of Modern Cryptography

From the ancient Caesar cipher, the Enigma machine during World War II, to the widely used and technically mature public-private key cryptographic systems in modern society, cryptography has undergone a long evolution. The rise of modern cryptography can be attributed to the combination of computer science and mathematics. In the 1970s, Diffie and Hellman first introduced the concept of public-key cryptography [DH22], marking a revolutionary breakthrough in modern cryptography. Soon after, the RSA (Rivest–Shamir–Adleman) public-key encryption algorithm was invented, offering a method for data encryption and digital signatures [RSA78].

For many years, the research on symmetric encryption (such as the data encryption standard (DES) [Cop94], the advanced encryption standard (AES) [RD01]) and asymmetric encryption (like the RSA, the elliptic curve cryptography (ECC) [Kob87]) became mainstream. Meanwhile, cryptography was also applied to various other security tasks, such as hash functions, digital signatures, blind signatures, and zero-knowledge proofs.

These cryptographic algorithms play a fundamental role in safeguarding data by providing secure communication and storage mechanisms. Without cryptography, many modern technologies and services, like cloud storage, online shopping, and email, could not operate securely.

Origin of Side-Channel Attack

With the advancement of cryptography, attackers began to look for new methods beyond the pure mathematical domain to challenge cryptographic systems. This led to the origin of side-channel attacks. Side-channel attacks do not target the encryption algorithms directly but focus on the "by-products" or externally leaked information of the actual implementations.

The birth of side-channel attacks can be traced back to 1996 when a significant development was published by Paul Kocher. He demonstrated that the duration required to execute cryptographic operations could inadvertently divulge information regarding the encryption key [Koc96]. This innovative approach elucidated that attackers might derive secrets merely by observing the time of execution. Such attacks, which exploit the runtime of algorithms, are denominated as Timing Attacks.

In the ensuing years, Kocher, in collaboration with his colleagues, introduced two distinct methodologies of power attacks: simple power analysis (SPA) and differential power analysis (DPA) [KJJ99]. The SPA, a category of side-channel attack, meticulously examines the power consumption of a cryptographic device during its operation. Intrinsically, varying operations within a cryptographic algorithm may manifest disparate power consumption patterns. For instance, certain operations might exhibit distinct power patterns for the binary "0" compared to "1". A thorough examination of these variances in power patterns enables attackers to deduce intermediate values or, in some instances, the entire

key utilized in the cryptographic procedure. A quintessential example would be if a device's power trace distinctly delineates the variations between squaring and multiplication operations, enabling attackers to discern specific key bits in particular algorithms.

In contrast to the relatively straightforward SPA, DPA embodies a more intricate mechanism. It exploits variations in power consumption patterns during cryptographic devices' operation, enabling attackers to statistically analyze these patterns and potentially deduce secret keys or other sensitive internal information without having to breach the algorithm directly. This vulnerability means that even if a cryptographic algorithm is mathematically secure, its physical implementation can still be susceptible to covert attacks, undermining the overall security of the system.

From then on, side-channel attacks, as a novel and effective form of attack, have gained attention from both the scientific and industrial communities, leading to a surge of research on the topic.

1.1.1 Current Research on Side-Channel Analysis

Side-channel attack (SCA) is a powerful class of attacks that exploit information leaked unintentionally during the execution of cryptographic algorithms. Unlike traditional attacks that attempt to break the mathematical foundations of encryption, side-channel attacks leverage various physical leaks, such as timing information [DKL+00], power consumption [CCD00], or electromagnetic leaks [GMO01], to gain insights into the sensitive information used in cryptographic operations.

Research concerning side-channel analysis can be broadly categorized into the following three domains:

- Identify potent side-channel attack techniques. Such studies encompass: probing further exploitable types of side-channel information (like a acoustic side-channel [SSYA19]); investigating more powerful side-channel analysis strategies (see [LBM15] for a machine learning approach); tailoring and executing side-channel attacks specific to various cryptographic devices (see [XBZ12], studying the feasibility of inferring a user's tap inputs to a smartphone with its integrated motion sensors), and so forth.

- Propose robust countermeasures against side-channel attacks. This line of research includes: hardware-based countermeasures (like the reduction of leaked electromagnetic radiation or power consumption through physical shielding, see [GMOP15]), software-based countermeasures (such as masking [Mes00, GM11], jamming the side-channel with noise); and protocol-level countermeasures, like frequently altering keys to limit the number of samples an attacker can gather in a short span of time, among others.

- Ascertain effective side-channel analysis methods. Such investigations aim to utilize the tools of mathematics and information theory to theoretically validate issues associated with side-channel analysis, see [SMY09, dCGRP19a]. This might involve quantitative analysis of the amount of information leaked in side-channel analysis, providing theoretical proof of the security of side-channel countermeasures, exploring parameters that influence the efficacy of countermeasures, and so on.

The content of this thesis predominantly pertains to the latter two categories: it encompasses research on masking—a widely-adopted countermeasure—as well as quantitative analyses of information leakage in side-channel attacks, and investigations into the security robustness of cryptographic devices against such attacks.

For designers and manufacturers of cryptographic devices, there is an inherent aspiration to ensure these devices are resilient against all potential side-channel attacks. **This thesis predominantly adopts a defender-centric perspective, conducting security assessments against the theoretically optimal attack strategy, specifically the ML (maximum likelihood) attack**.

The performance of ML attack surpasses all types of side-channel attacks encountered in practice because, by definition, the ML attack seeks to identify the most probable key based on observed measurements and the known (or assumed) statistical distribution of these measurements. Essentially, it exploits the entire statistical structure of the side-channel, a feat that's nearly unattainable in real-world attacks. If a system can withstand an ML attack, it is anticipated to fend off other less optimal side-channel assaults.

1.1.2 Framework of Side-Channel Attacks

A comprehensive evaluation is inextricably linked to appropriate modeling. In the realm of side-channel analysis, various models have been proposed and deliberated upon across a multitude of research endeavors, see [SMY09],[HRG14a],[dCGRP19a]. The side-channel model presented in this thesis follows from these previous work.

Notations

Upper case letters, like X, are used to denote random variables. The set of all possible values of X is represented in calligraphic letters like $\mathcal{X}$. The probability distribution of X is denoted as $P_X(x)$, the subscript X will be omitted when the context is clear, denoted as $P(x)$.

Let $P_X(x)$, $Q_X(x)$ be two probability distributions of X that possess a dominating measure $\mu(x)$ such that $P_X(x) \ll \mu(x)$ and $Q_X(x) \ll \mu(x)$; the corresponding lower-case letters $p_X(x)$ and $q_X(x)$ are densities of $P_X(x), Q_X(x)$ with respect to $\mu(x)$. When X is a discrete random variable, $\mu(x)$ is usually taken as the counting measure; when X is a continuous variable, $\mu(x)$ is usually taken as the Lebesgue measure.

Theoretical Model

In this thesis, the secret key is denoted as K and assumed to be uniformly distributed over $\mathcal{K}$; let $N = |\mathcal{K}|$ be the size of $\mathcal{K}$. The public variable T is known to the attacker and independent of the secret K; it can be plaintext or ciphertext, depending on whether the algorithm used is encryption or decryption. Normally one assumes T is uniformly distributed over $\mathcal{T}$.

When a device is operating, the cryptographic algorithm operates on K and T to compute a sensitive variable $X = f(K, T)$, where f is supposed to be a deterministic function. During this process, side-channel information about X will inevitably leak, which can include variations in algorithm execution time, voltage fluctuations, and so on. The

attacker measures such side-channel leakages, which are also referred to as "traces" and denoted as Y. After measuring sufficient amount of traces, the attacker will perform statistical analysis on them and attempt to recover the secret K used in the cryptographic algorithm. The guessed key is denoted as $\hat{K}$. To improve the accuracy of the guess, the attacker performs m measurements, each with corresponding T_i, X_i, and Y_i, $i = 1, 2, \ldots, m$. The m-element vector is denoted as $T^m = (T_1, T_2, \ldots, T_m)$.

The whole procedure can be seen as the following communication model in Fig. 1.1.

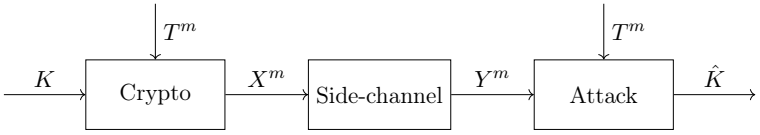


Figure 1.1: Side-channel seen as a communication channel (without masking).

The attacker exploits his knowledge of T^m and Y^m to estimate the secret $\hat{K}$ by using the MAP (maximum a posteriori) rule.[1] Obviously, the attack is successful if $\hat{K} = K$. The (maximal) probability of success is denoted as

$$\mathbb{P}_s = \mathbb{P}(K = \hat{K}|Y^m, T^m) = \mathbb{E}_{Y^m T^m} \max_k p_{K|Y^m T^m}(k|Y^m = y^m, T^m = t^m), \tag{1.1}$$

which is always no less than $\frac{1}{N}$ (better than blind guess). It maximizes the success probability $\mathbb{P}(K = \hat{K})$ because

$$\begin{aligned} \mathbb{P}(K = \hat{K}) &= \mathbb{E}_{Y^m T^m}\big(\mathbb{P}(\hat{K} = K|Y^m, T^m)\big) && (1.2)\\ &= \mathbb{E}_{Y^m T^m}\big(\sum_k p(k|Y^m, T^m)\mathbb{P}(\hat{K} = k|Y^m, T^m)\big) && (1.3)\\ &\leq \mathbb{E}_{Y^m T^m}\big(\max_k p(k|Y^m, T^m)\big) && (1.4) \end{aligned}$$

where (1.3) holds because $K - (Y^m, T^m) - \hat{K}$ is a Markov chian; (1.4) with equality if $\mathbb{P}(\hat{K} = k|Y^m, T^m) = 1$ for some x achieving $\max_k p(k|Y^m, T^m)$.

Since ML is equivalent to MAP when the distribution of K is uniform, the probability of success for the ML attack is also $\mathbb{P}_s$.

1.2 Code-Based Masking

Countermeasures against side-channel attacks involve a range of techniques aimed at mitigating the risk of sensitive information leakage through unintended side-channels. These countermeasures include algorithmic approaches like hiding and masking, physical measures such as shielding against electromagnetic emissions, protocol level countermeasures, etc. **Among them, masking is one of the most widely used countermeasures against side-channel attacks.** It involves introducing random values, known as masks, into the intermediate calculations of a cryptographic algorithm.

[1] Maximum success $\mathbb{P}_s(K|Y^m, T^m) = \mathbb{E}_{Y^m T^m}\big(\max_k p(k|Y^m, T^m)\big)$ attained with $\hat{K} = \hat{k}(Y^m, T^m) = \arg\max_k p(k|Y^m, T^m)$.

Masking Scheme

The masking scheme operates by splitting the sensitive variable into several shares, such that any t shares of them are independent of the sensitive variable. This certain threshold t is called the *security order* of the masking scheme. These shares will be processed by cryptographic algorithm independently, ensuring that the masked values are used in all intermediate computations. At the end of the computation, the masked shares are combined to obtain the final result. During this process, some side-channel leakage depending on each share are leaked and measured by the attacker. The whole procedure is shown in Fig. 1.2:

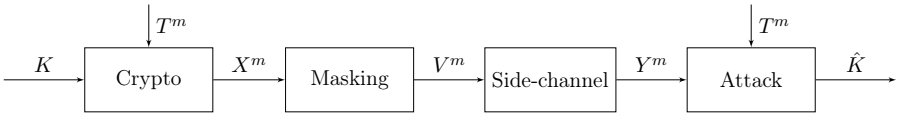


Figure 1.2: Side-channel seen as a communication channel (with masking).

As in previous model, K, T, X represent the secret key, the public variable and sensitive imformation respectively. X is a deterministic function of K and T. After masking, X is splitted into $d+1$ shares: $X_0, X_1, \ldots, X_d$. The vector composed of all shares is denoted as $V = (X_0, X_1, \ldots, X_d)$. The side-channel leakage $Y = (Y_0, Y_1, \ldots, Y_d)$ is a vector of each share's leakage. Again, the attacker performs m measurements to guess the secret key, the probability of success for the ML attack is denoted as $\mathbb{P}_s$.

Definition 1.2.1 (Arithmetic Masking)**.** Let $\mathbb{F}_q$ be a finite field with q elements. For an arithmetic masking scheme over $\mathbb{F}_q$, a sensitive variable X is masked into $V = (X_0, X_1, \ldots, X_d)$ which satisfies

$$X = X_0 + X_1 + \cdots + X_d \tag{1.5}$$

where $X, X_0, \ldots, X_d \in \mathbb{F}_q$ and "+" is the addition operation in $\mathbb{F}_q$.

1.2.1 Rationale of Code-Based Masking

Code-based masking is a class of efficient and secure countermeasures. It includes Boolean masking (BM) [CJRR99], inner product masking (IPM) [BFG+17], direct sum masking (DSM) [BCC+14, CG18, PGS+17], etc. To the best of our knowledge, generalized code-based masking [WMCS20] is the most generic masking scheme that unifies all above-mentioned schemes.

Let $\mathbb{F}_q$ be a finite field with q elements, and k, s, d be positive integers with $k+s \leq d+1$. Consider the sensitive information X as a row vector comprising k components, which is uniformly distributed across $\mathbb{F}_q^k$. The mask M is a uniform random variable over $\mathbb{F}_q^s$, it introduces randomness in the masking scheme.

To split the sensitive information into $d+1$ shares, X is multiplied by a matrix $\mathcal{G}_\mathcal{C} \in \mathbb{F}_q^{k\times(d+1)}$ and M is multiplied by a matrix $\mathcal{G}_\mathcal{D} \in \mathbb{F}_q^{s\times(d+1)}$. The generalized code-based

masking is modeled by

$$V = X\mathcal{G}_{\mathcal{C}} + M\mathcal{G}_{\mathcal{D}}, \tag{1.6}$$

where $\mathcal{G}_{\mathcal{C}}$ and $\mathcal{G}_{\mathcal{D}}$ are dictated by the masking scheme, possessing a rank of k and s respectively. Therefore the masked value $V \in \mathbb{F}_q^{d+1}$ is a row vector with $d+1$ components.

The row spaces of $\mathcal{G}_{\mathcal{C}}$ and $\mathcal{G}_{\mathcal{D}}$ are denoted as $\mathcal{V}_{\mathcal{G}_{\mathcal{C}}} = \mathcal{C}$ and $\mathcal{V}_{\mathcal{G}_{\mathcal{D}}} = \mathcal{D}$, which are two linear codes with paremeters $[d+1, k]$ and $[d+1, s]$ respectively ($d+1$ is the length of the linear codes, k and s are their dimensions). Thus $\mathcal{G}_{\mathcal{C}}$ and $\mathcal{G}_{\mathcal{D}}$ are also called generator matrices of linear codes $\mathcal{C}$ and $\mathcal{D}$. In code-based masking, we always assume $\mathcal{C} \cap \mathcal{D} = \{0\}$.

Direct Sum Masking

If constraint $\mathcal{C} + \mathcal{D} = \mathbb{F}_q^{d+1}$ is applied in (1.6), then $k + s = d + 1$, and the system becomes direct sum masking [BCC+14].

Inner Product Masking

Inner product masking [BFG+17] can be seen as a special case of direct sum masking, which has $k = 1$, $s = d$, and

$$\mathcal{G}_{\mathcal{C}} = \begin{pmatrix} 1 & 0_d \end{pmatrix}$$
$$\mathcal{G}_{\mathcal{D}} = \begin{pmatrix} \alpha^T & I_d \end{pmatrix}$$

where 0_d denotes an all-zero row vector of length d, α^T denotes the transpose of a row vector where $\alpha = (\alpha_1, \alpha_2, \cdots, \alpha_d) \in \mathbb{F}_q^d$, and I_d is the identity matrix of order d.

Boolean Masking

Boolean masking [CJRR99] seems to be the simplest family of code-based masking schemes:

Definition 1.2.2 (Boolean Masking)**.** For Boolean Masking scheme, $k = 1$, $s = d$, and two generator matrices are

$$\mathcal{G}_{\mathcal{C}} = \begin{pmatrix} 1 & 0_d \end{pmatrix}$$
$$\mathcal{G}_{\mathcal{D}} = \begin{pmatrix} 1_d^T & I_d \end{pmatrix}$$

where 0_d denotes an all-zero row vector, 1_d^T denotes the transpose of an all-one row vector and I_d is the identity matrix of order d. Since $M = (M_1, \ldots, M_d)$ is uniformly distributed over $\mathbb{F}_q^d$, the masked value

$$V = (X + \sum_{i=1}^{d} M_i, M_1, \ldots, M_d) \tag{1.7}$$

is uniformly distributed over $\mathbb{F}_q^{d+1}$. Therefore, the attacker needs to know all $d+1$ shares to recover the sensitive variable X.

Denote V as $(X_0, X_1, \ldots, X_d)$. If $X, M_1, M_2, \ldots, M_d$ are binary sequences with the same length, and $+$ is the XOR operation $\oplus$, then a Boolean masking scheme satisfies

$$X = X_0 \oplus X_1 \oplus \cdots \oplus X_d. \tag{1.8}$$

1.2.2 Comparison with Wiretap Channel Model

The wiretap channel is a communication channel characterized by the presence of an eavesdropper, which poses a security threat to the transmitted information. This section discusses its differences and connections with side-channel attack scenario.

Wiretap Channel

The wiretap channel is first proposed by Wyner [Wyn75] and then considered by Csiszár and Körner [CK78] in a more general way. The model is shown in Fig. 1.3:

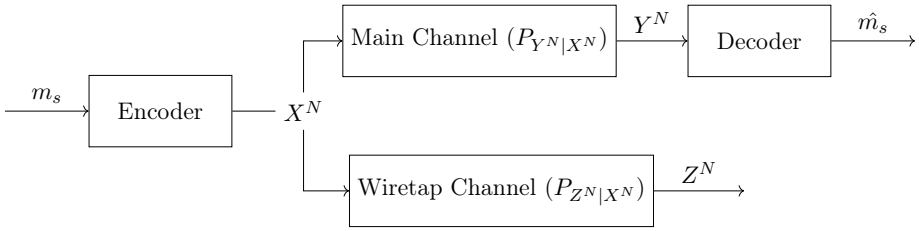


Figure 1.3: A Wiretap Channel Model.

where m_s represents the message, X^N represents the corresponding N-bit binary codeword. After m_s is encoded into X^N, it is transmitted through a main channel with probability distribution $P_{Y^N|X^N}$. Upon receipt of Y^N, the decoder makes an estimate $\hat{m}_s$ of the message. Meanwhile, the intruder wiretapped the codeword via a wiretap channel with probability distribution $P_{Z^N|X^N}$. Thus X^N is also the input of this wiretap channel, and the output is denoted as Z^N. Intuitively, the whole system is considered secure if the mutual information between m_s and Z^N is small enough (for instance, $I(m_s; Z^N) \to 0$ as $N \to \infty$).

Wiretap Channel of Type II

In 1984, Ozarow and Wyner proposed the wiretap channel of type II [OW84]. Let $\mathcal{S} \subset \{1, 2, 3, \ldots, N\}$ be a μ-element set, where μ is an integer with $1 \leq \mu < N$. Let Z_i be the i-th element of vector Z^N $(i = 1, \ldots, N)$, which is calculated from

$$Z_i = \begin{cases} X_i, & i \in \mathcal{S} \\ ?, & i \notin \mathcal{S} \end{cases} \tag{1.9}$$

Then the wiretap channel model of type II is:

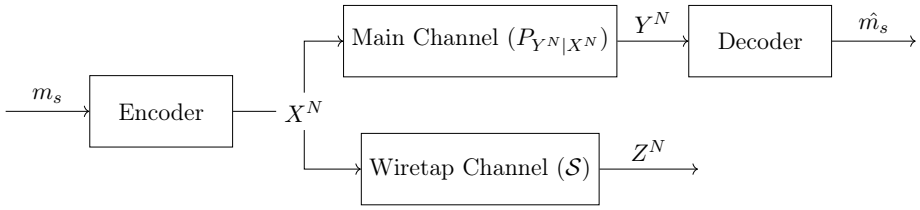


Figure 1.4: A Wiretap Channel Model of Type II.

Example 1.2.1. Assume the chosen subset of an intruder is $\mathcal{S} = [1, 5, 7, 14, 15]$ and the input of wiretap channel is a 16-bit codeword $X^N = [0\ 1\ 0\ 1\ 1\ 1\ 0\ 0\ 1\ 0\ 1\ 1\ 0\ 1\ 1\ 1]$. Then the observed value is

$$Z^N = [0\ ?\ ?\ ?\ 1\ ?\ 0\ ?\ ?\ ?\ ?\ ?\ ?\ 1\ 1\ ?]. \tag{1.10}$$

The Connections with Side-Channel Attacks

In a sense, the masking setting under the probing model[2] can be viewed as a special case of a wiretap channel of Type II, where an eavesdropper has access to a given number of bits. Assuming that the information symbols of m_s are elements in the finite field $\mathbb{F}_q$. In Fig. 1.4, the procudure can be seen as a masking scheme: the message $m \in \mathbb{F}_q^k$ is encoded as $X \in \mathbb{F}_q^n$ by using a masking code. The elements in set $\mathcal{S}$ can be considered as the positions of probed bits choosed by the side-channel attacker (see Section 3.2 for detail).

The Differences from Side-Channel Attacks

However, the classical wiretap channel scenario (Fig. 1.3) differs from the side-channel attack in an important respect: Fig. 1.1 is not a genuine communication channel, where the secret key K should always be kept as secret and is never to be transmitted reliably to any destination (there is no legitimate receiver). The attacker queries the cryptographic devices several times to guess the static secret key K. Even though the public variable T varies, the same key is used for every encryption.

On the other hand, in a wiretap channel, messages need to be reliably encoded, transmitted, and decoded in the presence of the eavesdropper. Typically, each transmitted message varies, and what concerns us is the amount of leaked information during a single information transmission. If we draw an analogy with side-channel attacks, it is like asking the attacker to guess the secret key based on a single trace, which is quite challenging for non-profiling attacks.

[2]Under the probing model, a κ-dimensional side-channel attack refers to the attacker observing the sensitive values at κ positions. See Definition 3.2.1 for more detail.

1.3 Information Measures

As the ancient saying goes, "To do a good job, one must first sharpen one's tools." Suitable metrics can significantly improve the accuracy of quantitative analysis. Information-theoretic metrics have demonstrated their effectiveness in many works on side-channel security assessment. **This section reviews several information theory concepts and their properties**, which will be instrumental in forming the content of the subsequent chapters.

Information theory measures are mathematical tools commonly used to quantify uncertainty (randomness), the amount of information, and differences between various distributions, among other things. Traditional information measures encompass Shannon entropy, Kullback-Leibler divergence, Fano's mutual information, etc. These definitions are extended to Rényi information theory measures, as they can be obtained from generalized α-information theoretic measures by letting $\alpha \to 1$.

Every metric has its own applicable usage scenarios. When dealing with a particular research question, choosing the most suitable metrics for the context can significantly improve the effectiveness of the analysis. Generally, information-theoretic metrics can be defined from two distinct perspectives: operational significance and axiomatic characterization. The latter facet aids in the metric selection process by allowing us to choose metrics that exhibit specific properties relevant to addressing research questions. For example, the characterization theorem defines Shannon entropy based on several "natural" properties like expansibility, symmetry, additivity, etc. If all these properties are necessary in a research scenario, then Shannon entropy is the only possible and good choice, as these axioms uniquely determine this metric.

1.3.1 Entropy

Entropy is primarily linked with a state of randomness, disorder, or uncertainty. This term and concept find application in various fields, ranging from classical thermodynamics, where it was initially identified, to the microscopic depiction of nature in statistical physics, and the principles of information theory.

The concept of information entropy was initially introduced by Claude Shannon in 1948, which is widely known as *Shannon entropy*. Given a random variable X with a probability distribution P_X and a dominating measure μ such that $P_X \ll \mu$, the Shannon entropy of X is

Definition 1.3.1 (Shannon Entropy)**.**

$$H(X) = -\int_{\mathcal{X}} p_X(x) \log p_X(x)\ d\mu(x) \tag{1.11}$$

When μ is a counting measure we obtain the classical definition of Shannon entropy for discrete random variables:

$$H(X) = -\sum_{x \in \mathcal{X}} p_X(x) \log p_X(x). \tag{1.12}$$

When μ is the Lebesgue measure, $p_X(x)$ is the probability density function, we obtain the integral form of (1.12):

$$h(X) = -\int_{\mathcal{X}} p_X(x) \log p_X(x)\ dx, \tag{1.13}$$

which is called the differential entropy.

Besides Shannon entropy, various other information entropy measures have been suggested in literature, including *Hartley entropy*, *collision entropy*, *min-entropy*:

Definition 1.3.2 (Hartley Entropy)**.** Suppose the domain of X has finite volume with respect to the measure μ. The Hartley entropy of X is

$$H_0(X) = \log |\mathcal{X}|. \tag{1.14}$$

where $|\mathcal{X}|$ represents the volume of X with respect to μ. When $\mathcal{X}$ has finite elements and μ is a counting measure, $|\mathcal{X}|$ is the number of elements in $\mathcal{X}$.

Definition 1.3.3 (Collision Entropy)**.**

$$H_2(X) = -\log \int_{\mathcal{X}} p_X^2(x)\ d\mu(x). \tag{1.15}$$

Definition 1.3.4 (Min-Entropy)**.**

$$H_\infty(X) = -\log\big(\operatorname*{ess\,sup}_{x\in\mathcal{X}} p_X(x)\big). \tag{1.16}$$

The aforementioned definitions can be unified under a generalized concept known as α-entropy, also commonly referred to as Rényi entropy. In 1961, Rényi proposed several postulates that can characterize Shannon entropy and α-entropy [Rén61]. The closed-form expression of Rényi entropy is

Definition 1.3.5 (Rényi Entropy (α-entropy))**.** Assume that either $0 < \alpha < 1$ or $1 < \alpha < +\infty$ (the particular values $0, 1, +\infty$ being obtained by taking limits). The *α-entropy* of a random variable X with a probability distribution P_X is defined as

$$H_\alpha(X) = H_\alpha(P_X) = \tfrac{\alpha}{1-\alpha} \log \|p_X\|_\alpha \tag{1.17}$$

where we have used the special notation

$$\|p_X\|_\alpha = \big(\int_{\mathcal{X}} p_X^\alpha(x)\ d\mu(x)\big)^{1/\alpha}. \tag{1.18}$$

When μ is a counting measure we obtain the classical definition of Rényi entropy for discrete random variables; when $\mu(x)$ is the Lebesgue measure we obtain the corresponding definitions for continuous variables.

Letting α tends to $0, 1, 2, \infty$, one obtains Hartley entropy, Shannon entropy, collision entropy and min-entropy, respectively. One has the following inequality:

$$H_0(X) \geq H(X) \geq H_2(X) \geq H_\infty(X). \tag{1.19}$$

That is because

Proposition 1.3.1. $H_\alpha(X)$ is non-increasing in α. For any $\alpha < \beta$ in $[0, +\infty]$, $H_\alpha(X) \geq H_\beta(X)$ with equality if and only if X is uniformly distributed.

Proof. When $1 < \alpha < \beta$, we have

$$\begin{aligned} H_\alpha(X) &= \frac{1}{1-\alpha} \log\Big(\int_{\mathcal{X}} p_X(x)^\alpha \, d\mu(x)\Big) \\ &= -\log\Big(\int_{\mathcal{X}} p_X(x) \cdot p_X(x)^{\alpha-1} \, d\mu(x)\Big)^{\frac{1}{\alpha-1}} \\ &= -\log\Big(\mathbb{E}_X\big(p_X(x)^{\beta-1}\big)^{\frac{\alpha-1}{\beta-1}}\Big)^{\frac{1}{\alpha-1}} \\ &\geq -\log\Big(\mathbb{E}_X\big(p_X(x)^{\beta-1}\big)\Big)^{\frac{1}{\alpha-1}\cdot\frac{\alpha-1}{\beta-1}} = H_\beta(X) \end{aligned}$$

where the last inequality comes from Jensen's inequality ($\frac{\alpha-1}{\beta-1} < 1$ when $1 < \alpha < \beta$). For other range of α and β, the inequality can be proved in a similar way using Jensen's inequality. □

1.3.2 Divergence and its Conditional Version

In information theory, divergence is a measure that quantifies the difference or "distance" between two probability distributions. The *Kullback-Leiber divergence (K-L divergence)*, also called *relative entropy*, is a widely used statistical distance.

Definition 1.3.6 (K-L Divergence)**.** Given two probability distributions P_X and Q_X defined over the same sample space $\mathcal{X}$, the divergence from P_X to Q_X is defined as

$$D(P_X \| Q_X) = \int_{\mathcal{X}} p_X(x) \log \frac{p_X(x)}{q_X(x)} \, d\mu(x). \tag{1.20}$$

For discrete variable X with a counting measure μ, it becomes

$$D(P_X \| Q_X) = \sum_{x \in \mathcal{X}} p_X(x) \log \frac{p_X(x)}{q_X(x)}. \tag{1.21}$$

When μ is the Lebesgue measure, $p_X(x), q_X(x)$ are the probability density functions, one obtains the integral form of (1.21):

$$D(P_X\|Q_X) = \int_{\mathcal{X}} p_X(x) \log \frac{p_X(x)}{q_X(x)}\, dx, \tag{1.22}$$

which is compatible with the discrete form (1.21).

1.3.2.1 Alpha-Divergence

The *Rényi divergence* (also called α-*divergence*) is a well-known generalization of K-L divergence:

Definition 1.3.7 (α-Divergence)**.** Assume that either $0 < \alpha < 1$ or $1 < \alpha < +\infty$ (the particular values $0, 1, +\infty$ being obtained by taking limits).

The α-*divergence* of P_X from Q_X are defined as

$$D_\alpha(P_X\|Q_X) = \tfrac{1}{\alpha-1} \log\langle p_X\|q_X\rangle_\alpha^\alpha \tag{1.23}$$

where we have used the special notation:

$$\langle p_X\|q_X\rangle_\alpha = \Big(\int_{\mathcal{X}} p_X(x)^\alpha q_X(x)^{1-\alpha}\, d\mu(x)\Big)^{1/\alpha} \tag{1.24}$$

with the same convention for μ as in Definition 1.3.5.

It is easy to verify

$$\lim_{\alpha\to 1} D_\alpha(P_X\|Q_X) = D(P_X\|Q_X). \tag{1.25}$$

A link between α-entropy and α-divergence is the following *uniform expansion property* (UEP). Let $U \sim \mathcal{U}(N)$ be uniformly distributed over a set of finite μ-measure N. (In the discrete case U simply takes N equiprobable values.) Let P be a distribution over the same sample space. Since $u \equiv \frac{1}{N}$ we have $\langle p\|u\rangle_\alpha = N^{\frac{\alpha-1}{\alpha}}\|p\|_\alpha$, hence

Property 1.3.1 (UEP of α-Divergence[vEH14])**.**

$$D_\alpha(P\|U) = H_\alpha(U) - H_\alpha(P) = \log N - H_\alpha(P). \tag{1.26}$$

Another important property is the *data processing inequality* (DPI). A random transformation given by a conditional distribution $P_{Y|X}$ is noted $P_X \to \boxed{P_{Y|X}} \to P_Y$ if a random variable $X \sim P_X$ is input and the output distribution P_Y satisfies

$$p_Y(y) = \int_{\mathcal{X}} p_{Y|X}(y|x) p_X(x)\, d\mu(x).$$

Similarly for $Q_X \to \boxed{P_{Y|X}} \to Q_Y$ one has

$$q_Y(y) = \int_{\mathcal{X}} p_{Y|X}(y|x) q_X(x)\, d\mu(x).$$

Property 1.3.2 (DPI for α-Divergence [PV10, Rio21])**.** Any transformation can only reduce α-divergence: $D_\alpha(P_X\|Q_X) \geq D_\alpha(P_Y\|Q_Y)$.

1.3.2.2 Conditional Alpha-Divergence

The definition of α-divergence has been extended to *conditional* versions.

Definition 1.3.8 (Conditional α-Divergence [Ver15])**.** The conditional α-divergence is defined as

$$D_\alpha(P_{Y|X}\|Q_{Y|X}|P_X) = D_\alpha(P_{Y|X}P_X\|Q_{Y|X}P_X). \tag{1.27}$$

This definition is consistent with the unconditional one:

Property 1.3.3 (Consistency of Conditional α-Divergence w.r.t. α-Divergence)**.** If $X \equiv 0$ then

$$D_\alpha(P_{Y|X}\|Q_{Y|X}|P_X) = D_\alpha(P_Y\|Q_Y). \tag{1.28}$$

Here following Shannon [Sha53] we have noted $X \equiv 0$ for any random variable independent of everything else considered (e.g., a constant variable).

In Definition 1.3.8 we remark that the *expectation* over the conditioned variable is only taken *inside the logarithm* in the α-divergence's expression:

$$D_\alpha(P_{Y|X}\|Q_{Y|X}|P_X) = \tfrac{1}{\alpha-1}\log \mathbb{E}_X\langle p_{Y|X}\|q_{Y|X}\rangle_\alpha^\alpha. \tag{1.29}$$

1.3.3 Conditional Entropy

Conditional entropy, often denoted $H(X|Y)$, is a measure that quantifies the amount of uncertainty remaining in a random variable X given the observation of another random variable Y.

Definition 1.3.9 (Conditional Entropy)**.** Given random variables X and Y with probability distributions P_X and P_Y respectively. Let P_{XY} be the joint probability distribution. The conditional Shannon entropy is defined as

$$H(X|Y) = -\int_{\mathcal{XY}} p_{XY}(x,y)\log\frac{p_{XY}(x,y)}{p_Y(y)}\,d\mu(x,y) \tag{1.30}$$

Conditional entropy is widely used in many research areas. In side-channel analysis, attackers always guess the most likely key based on their observations of measured side-channel leakages. Therefore, the conditional version of information-theoretic measures is very commonly used in such scenarios.

After Rényi entropy was proposed, several studies have been conducted to seek an appropriate definition of conditional α-entropy. Intuitively, a reasonable notion of conditional α-entropy may satisfy several natural properties, such as *non-negativity* (the value

of uncertainty should be non-negative, as the uncertainty of a completely certain variable is 0) and *conditioning reduces entropy* (having more information can decrease the level of uncertainty):

- *non-negativity*: $H_\alpha(X|Y) \geq 0$ (at least for discrete random variable X);
- *conditioning reduces entropy (CRE)*: $H_\alpha(X|Y) \leq H_\alpha(X)$, where equality holds if and only if X is independent of Y;

In addition, as a generalization of conditional (Shannon) entropy, conditional α-entropy is expected to be consistent with classical information measures as α tends to 1. So it might be a good idea to draw inspiration from different expressions of conditional entropy as follows [FB14]:

$$H(X|Y) = \mathbb{E}_Y H(X|Y=y) \tag{1.31}$$
$$= H(XY) - H(Y) \tag{1.32}$$

Besides these, when X is a uniform distribution over a finite alphabet, i.e., $X \sim \mathcal{U}(|\mathcal{X}|)$, the conditional entropy can be expressed by K-L divergence in the following way [TH18]:

$$H(X|Y) = \log|\mathcal{X}| - D(P_{XY}\|P_X \times P_Y) \tag{1.33}$$
$$= \log|\mathcal{X}| - \min_{Q_Y} D(P_{XY}\|P_X \times Q_Y) \tag{1.34}$$

Drawing from (1.31), (1.32), (1.33), and (1.34) respectively, we have different proposals for conditional α-entropy:

(1) $\tilde{H}^1_\alpha(X|Y) = \mathbb{E}_Y H_\alpha(X|Y=y)$.

(2) $\tilde{H}^2_\alpha(X|Y) = H_\alpha(XY) - H_\alpha(Y)$.

(3) $\tilde{H}^3_\alpha(X|Y) = \frac{1}{1-\alpha}\log \mathbb{E}_Y \|P_{X|Y}\|^\alpha_\alpha$.

(4) $\tilde{H}^4_\alpha(X|Y) = \frac{\alpha}{1-\alpha}\log \mathbb{E}_Y \|P_{X|Y}\|_\alpha$.

The first two suggested notions are taken from [Cac97, equation (2.15)] and [JA04, section 2.2] [3]. These two proposals are argued that violate the property *conditioning reduces entropy* [FB14]. The last two definitions are proposed by Hayashi [Hay11] and Arimoto [Ari75] respectively. Both of them satisfy *non-negativity* (for discrete random variable) and *conditioning reduces entropy*, the main difference is the latter one satisfies the (variation of) chain rule [FB14], i.e.

$$H_\alpha(X|Y) + H_0(Y) \geq H_\alpha(XY) \tag{1.35}$$

where $H_0(Y)$ is the Hartley entropy of Y, while Hayashi's definition does not (see [FB14, Example (4) and Theorem 3]).

In this thesis we use Arimoto's definition $\tilde{H}^4_\alpha(X|Y)$, and denote it as $H_\alpha(X|Y)$ in the rest of this article. In addition to the properties mentioned above, Arimoto's definition also satisfies other useful properties: *consistency, uniform expansion property (UEP) and data processing inequality (DPI)*. Consistency is obvious from the definition:

[3] The second definition was proposed again by Golshani et al. in 2009, see [GPY09, equation (2.10)].

Property 1.3.4 (Consistency of Conditional α-Entropy w.r.t. α-Entropy)**.** If $Y \equiv 0$ then

$$H_\alpha(X|Y) = H_\alpha(X). \tag{1.36}$$

Property 1.3.5 (Uniform Expansion Property)**.** If $U \sim \mathcal{U}(N)$ is uniform and independent of X, then

$$D_\alpha(P_{Y|X}\|U|P_X) = H_\alpha(U) - H_\alpha(Y|X) = \log N - H_\alpha(Y|X). \tag{1.37}$$

It is because $\langle p_{X|Y}\|u\rangle_\alpha = N^{\frac{\alpha-1}{\alpha}}\|p_{X|Y}\|_\alpha$.

Property 1.3.6 (Data Processing Inequality [FB14, Rio21])**.** If $X-Y-Z$ forms a Markov chain, then

$$H_\alpha(X|Y) \le H_\alpha(X|Z). \tag{1.38}$$

In particular for $Z \equiv 0$, it yields the *conditioning reduces α-entropy (CRE)*:

$$H_\alpha(X|Y) \le H_\alpha(X|0) = H_\alpha(X). \tag{1.39}$$

1.3.4 Generalized Mutual Information

The mutual information (MI) of two random variables is a measure of the mutual dependency between the two variables, quantifying the "amount of information" gained about one random variable through observing the other.

Definition 1.3.10 (Mutual Information)**.** Given two random variables X and Y with joint probability distribution P_{XY} and corresponding distributions, the mutual information between X and Y is

$$I(X;Y) = \int_{\mathcal{XY}} p_{XY}(x,y) \log \frac{p_{XY}(x,y)}{p_X(x)p_Y(y)}\, d\mu(x,y). \tag{1.40}$$

This quantity was first defined and analyzed by Claude Shannon, although he did not refer to it as "mutual information". The term was later coined by Robert Fano, so sometimes it is called "Fano's mutual information".

Following the proposal of general information measures like Rényi entropy and divergence, numerous studies have sought an apt generalization of mutual information. Different definitions of α-information $I_\alpha(X;Y)$ were proposed in the literature.

Intuitively, a reasonable definition should be consistent with Fano's mutual information as $\alpha \to 1$, and possibly also satisfy the following useful properties:

- *independence*: $I_\alpha(X;Y) \ge 0$ with equality if and only if X and Y are independent;

- *data post-processing inequality (post-DPI)*: if $X - Y - Z$ forms a Markov chain, then post-processing cannot increase the information, i.e., $I_\alpha(X;Z) \leq I_\alpha(X;Y)$;
- *data pre-processing inequality (pre-DPI)*: if $X - Y - Z$ forms a Markov chain, then pre-processing cannot increase the information, i.e., $I_\alpha(X;Z) \leq I_\alpha(Y;Z)$;
- *monotonicity*: $I_\alpha(X;Y)$ is nondecreasing as α increases;
- *closed-form expression* amenable to efficient numerical estimation.

1.3.4.1 Comparison of Different Proposals

At least four definitions are defined in the literature (for discrete random variables):

$$\text{(1)}\ I_\alpha^A(X;Y) = H_\alpha(X) - H_\alpha(X|Y),$$

$$\text{(2)}\ I_\alpha^C(X;Y) = \min_{Q_Y} \mathbb{E}_X(D_\alpha(P_{Y|X}\|Q_Y)),$$

$$\text{(3)}\ I_\alpha^R(X;Y) = D_\alpha(P_{XY}\|P_X \times P_Y) = \tfrac{1}{\alpha-1}\log \mathbb{E}_Y\langle p_{X|Y}\|p_X\rangle_\alpha^\alpha,$$

$$\text{(4)}\ I_\alpha(X;Y) = \min_{Q_Y} D_\alpha(P_{XY}\|P_X \times Q_Y) = \tfrac{\alpha}{\alpha-1}\log \mathbb{E}_Y\langle p_{X|Y}\|p_X\rangle_\alpha,$$

which somehow parallel the corresponding ones for conditional entropy.

Arimoto's Definition

The first definition was proposed by Arimoto [Ari75]. It is easily seen to satisfy both the *independence* and *post-DPI* property because of the DPI property of Arimoto's conditional entropy. However, it does not satisfy *monotonicity* because sometimes it can be decreasing in α (as $I_\alpha^A(X;X) = H_\alpha(X)$ is decreasing in α).

Csiszár's Definition

The second definition is from Csiszár [Csi95]. It does not seem to admit a closed-form expression, and the minimization is hard to solve analytically even in simple examples [Ver15]. However, one can prove *monotonicity* and the *independence* property, based on the corresponding properties of the α-divergence.

Rényi's α-Mutual Information

The third definition requires no minimization and appears in [TH18, equation (50)]. We call it *Rényi's α-mutual information* because it is a natural definition from Rényi's divergence, just as in the classical case $\alpha = 1$. Also, it is *mutual* in the sense that $I_\alpha^R(X;Y) = I_\alpha^R(Y;X)$. From the nonnegativity of α-divergence: $D_\alpha(P\|Q) \geq 0$ with equality if and only if $P = Q$, it is easily seen that $I_\alpha^R(X;Y)$ satisfies the *independence* property. From the monotonicity property of α-divergence, it also satisfies *monotonicity*. One can also check *post-DPI* and *pre-DPI* properties, by same reasoning line as in the proof of [LCGR21, Property 8], replacing Q_Y, Q_Z by P_Y, P_Z, respectively.

Sibson's α-Information

Finally, the fourth definition is due to Sibson [Sib69] (see also [Ver15]). In contrast to Rényi α-mutual information, symmetry does not hold in general: $I_\alpha(X;Y) \neq I_\alpha(Y;X)$. However, it is known to satisfy the independence property, monotonicity, and the pre and post-DPI [PV10] (see also [Rio21]). See Table 1.1 for a summary of all properties.

Table 1.1: Summary of properties for various definitions of α-information.

Def.	Independence	Post-DPI	Pre-DPI	Monotonicity	Closed-form
I_α^A	yes	yes	—	no	yes
I_α^C	yes	—	—	yes	no
I_α^R	yes	yes	yes	yes	yes
I_α	yes	yes	yes	yes	yes

Sibson's α-information is perhaps the preferred generalization of Fano's mutual information and has found various applications [PV10, Ver15, TH18, Rio21, EWG21, EGI21].

In this thesis, we employ both Rényi's α-mutual information and Sibson's α-information as information-theoretic tools for security evaluation. The remainder of this subsection provides further details of these two definitions.

1.3.4.2 Sibson's α-Information

The closed-form expression of Sibson's α-information is

Definition 1.3.11 (Sibson's α-Information)**.**

$$I_\alpha(X;Y) = \tfrac{\alpha}{\alpha-1} \log \mathbb{E}_Y \langle p_{X|Y} \| p_X \rangle_\alpha, \tag{1.41}$$

where we have used a special notation:

$$\langle p \| q \rangle_\alpha = \Big(\int p^\alpha q^{1-\alpha} d\mu \Big)^{1/\alpha}. \tag{1.42}$$

In addition to the properties listed in Table 1.1, Sibson's α-information also satisfies *uniform expansion property (UEP)* and *Sibson's identity*. As in the case of the conditional α-entropy, since $\langle p_{U|Y} \| u \rangle_\alpha = N^{\frac{\alpha-1}{\alpha}} \| p_{U|Y} \|_\alpha$, we have the following

Property 1.3.7 (UEP for α-Information [vEH14, Rio21])**.** If $U \sim \mathcal{U}(N)$ is uniformly distributed, then

$$I_\alpha(U;Y) = H_\alpha(U) - H_\alpha(U|Y) = \log N - H_\alpha(U|Y). \tag{1.43}$$

An important property of α-information is *Sibson's identity*. It is straightforward to compute

$$\langle p_{XY} \| p_X q_Y \rangle_\alpha^\alpha = \iint p_Y^\alpha p_{X|Y}^\alpha p_X^{1-\alpha} q_Y^{1-\alpha} d\mu \tag{1.44}$$

$$= \left\langle p_Y \langle p_{X|Y} \| p_X \rangle_\alpha \,\|\, q_Y \right\rangle_\alpha^\alpha. \tag{1.45}$$

Defining the (suitably normalized) distribution $q_Y^* = p_Y \langle p_{X|Y} \| p_X \rangle_\alpha / \mathbb{E}_Y \langle p_{X|Y} \| p_X \rangle_\alpha$, substituting and taking the logarithm gives the following

Proposition 1.3.2 (Sibson's identity [Sib69, Ver15])**.** One has

$$D_\alpha(P_{XY} \| P_X Q_Y) = D_\alpha(Q_Y^* \| Q_Y) + I_\alpha(X;Y), \tag{1.46}$$

hence the following alternate minimizing definition:

$$I_\alpha(X;Y) = \min_{Q_Y} D_\alpha(P_{XY} \| P_X Q_Y). \tag{1.47}$$

1.3.4.3 Rényi's Alpha Mutual Information

Definition 1.3.12 (Rényi's α-Mutual Information)**.**

$$I_\alpha^R(X;Y) = D_\alpha(P_{XY} \| P_X \times P_Y) = \tfrac{1}{\alpha - 1} \log \mathbb{E}_Y \langle p_{X|Y} \| p_X \rangle_\alpha^\alpha \tag{1.48}$$

Since

$$\min_{Q_Y} D_\alpha(P_{XY} \| P_X \times Q_Y) \leq D_\alpha(P_{XY} \| P_X \times P_Y),$$

Sibson's α-information is always less than or equal to *Rényi's α-mutual information*:

$$I_\alpha(X;Y) \leq I_\alpha^R(X;Y). \tag{1.49}$$

Remark 1.3.1. Rényi's α-mutual information does not satisfy the *uniform expansion property* when considering Arimoto's conditional entropy, as

$$I_\alpha^R(U;Y) \geq I_\alpha(U;Y) = \log N - H_\alpha(U|Y). \tag{1.50}$$

However, one has

$$I_\alpha^R(U;Y) = \log N - \tilde{H}_\alpha^3(U|Y) \tag{1.51}$$

as shown in (1.33), where $\tilde{H}_\alpha^3(U|Y)$ is Hayashi's proposal for conditional α-entropy.

1.3.5 Generalized Fano Inequality

Assume X is discrete and estimated from Y using the MAP rule, with (maximal) probability of success $\mathbb{P}_s = \mathbb{P}_s(X|Y) = \mathbb{E}_Y \sup_x p_{X|Y}(x|y = Y)$. The guessed X is denoted as $\hat{X}$. Also let $\mathbb{P}_s(X) = \sup_x p_X(x)$ be the probability of success when guessing X without even knowing Y. As $X - Y - \hat{X}$ is Markov, using data processing inequality for Sibson's α-information and α-divergence, we have the following

Lemma 1.3.1 (Rioul's Generalized Fano Inequality [Rio21, Thm. 1])**.**

$$I_\alpha(X;Y) \geq I_\alpha(X;\hat{X}) \geq d_\alpha\big(\mathbb{P}_s(X|Y)\|\mathbb{P}_s(X)\big) \tag{1.52}$$

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

$$d_\alpha(p\|q) = \tfrac{1}{\alpha-1}\log\big(p^\alpha q^{1-\alpha} + (1-p)^\alpha(1-q)^{1-\alpha}\big). \tag{1.53}$$

Classical Fano's inequality can be obtained by letting $\alpha \to 1$.

1.4 Security Assessment Against Side-Channel Attack

Side-channel attacks have become a concern for cryptographic systems due to device's inherent tendency to leak information, which leverages information leaked during the execution of cryptographic algorithms. To ensure resilience against these threats, a comprehensive security assessment strategy is indispensable. This subsection reviews the methods used for evaluating potential vulnerabilities and understanding the efficacy of protective measures, particularly in masked implementations.

1.4.1 Leakage Evaluation and Security Analysis

In [SMY09], the authors advocate for the evaluation of implementations by leveraging both information theoretic measures (including conditional entropy, mutual information) and security indicators (such as success rates or guessing entropy). They clarify the individual relevance of mutual information and the probability of success, but approach these metrics separately. In [DSVC14], the authors introduce a variety of metrics that assist in quantifying the side-channel leakage of cryptographic chips.

In [dCGRP19b, dCGRP19a], Chérisey et al. use classical information-theoretic tools (such as mutual information, K-L divergence) to establish some universal inequalities between the probability of success of a side-channel attack and the minimum number of queries to reach a given success rate. Such inequalities are 'universal' in the sense that they can apply to any type of attack and depend only on the leakage model. Leveraging Fano's inequality, the authors manage to forge a connection between mutual information and the probability of success. All of these works utilized the classical information-theoretic tools. As we will see in the subsequent sections of this paper, by applying generalized information theory for the analysis of side-channel security, tighter bounds can be obtained.

1.4.2 Evaluation of Masked Implementations

Masking serves as a preemptive countermeasure, deliberately obscuring the relationship between secret data and discernible leakages. By doing so, it becomes increasingly challenging for attackers to extrapolate the concealed data. Evaluating the efficacy and robustness of such masked implementations, in both practical and theoretical settings, is paramount.

As we mentioned before, the rationale of the masking scheme is splitting the sensitive variable into several shares and processing each share independently. Thus the side-channel leakage in masked implementations are vectors composed of each share's leakage.

In [CJRR99], Chari et al. prove a lower bound on the number of measurements required to carry out statistical attacks on devices: This bound increases as the number of shares increases. In their work, the physical characteristics of the targeted devices are assumed to satisfy several properties, and the security assessment focuses on leaking shares independently of any computation. In 2013, Prouff et al. [PR13] consider a more general leakage model: the only computation leaks information model. They derive a security bound for masked block cipher implementations using a measure called *bias* (related to statistical distance). Their main theorem shows that given the noisy leakages on its shares, the bias of sensitive variable decreases exponentially with the security order. Unlike [CJRR99], Prouff et al. do not provide a bound on the number of measurement, but focus on demonstrating that the information gained by observing the noisy leakage of one execution is negligible.

Subsequently, in [DFS15], Duc et al. derive a lower limit on the minimum number m of queries required to achieve a given probability of success $\mathbb{P}_s$. This lower bound is later improved in [MRS23] and [IUH22]. Even though the subsequent two papers have made considerable advancements to the results presented in Subsequently, in [DFS15], the lower bounds they established still tend to loosen when dealing with a large size of the finite field. All three papers conducted a security analysis of masked implementation using mutual information.

1.4.3 The Parameters Affecting Masking Performance

Masking performance can be influenced by various parameters, both intrinsic and extrinsic to the cryptographic system. Critical factors include: masking generation mechanisms, operational environment, and algorithmic complexity. This thesis focuses on masking generation mechanisms, more specifically, the linear codes used in the code-based masking.

It has been proven that the effectiveness of code-based masking is influenced by at least two parameters of the dual of the masking code: the *minimum distance* and the *kissing number* of the dual of masking code.

Before delving into the pertinent literature on this topic, we revisit several known definitions of linear codes.

Definition 1.4.1 (Linear code parameters [MS77])**.** A linear code $\mathcal{C}$ is a set of vectors, called codewords, which form a vector space over some finite field $\mathbb{F}_q$. The parameters of the linear code $\mathcal{C}$ is a triple $[n, k, \delta]$, where n is the code length, k is its dimension, and δ is its minimum distance (also denoted as $\delta_{\mathcal{C}}$). When δ is not known, $\mathcal{C}$ is referred to as an $[n, k]$ linear code.

Definition 1.4.2 (Hamming Weight [MS77, Chap 1, §3])**.** The *Hamming weight*, or simply the weight, of a vector $x = (x_1, \cdots, x_n)$ is the number of nonzero x_i. It is denoted as $w_H(x)$.

Definition 1.4.3 (Weight Distribution [MS77, Chap 2, §1])**.** Let $\mathcal{C}$ be an $[n, k, \delta]$ linear code and A_i be the number of codewords of Hamming weight i: $A_i = |\{x \in \mathcal{C} \mid w_H(x) = i\}|$. The sequence $A_0, A_1, \ldots A_n$ is called the *weight distribution* of $\mathcal{C}$. Obviously $A_0 = 1$, $A_1 = \cdots = A_{\delta-1} = 0$.

Definition 1.4.4 (Kissing Number)**.** The kissing number of a linear code $\mathcal{C}$ is A_δ, the number of nonzero codewords of minimum weight δ.

Definition 1.4.5 (Dual code [MS77] and dual distance)**.** The *dual code* of a code $\mathcal{C}$ is the linear code consisting of the set of all vectors orthogonal to all codewords of $\mathcal{C}$, denoted as $\mathcal{C}^\perp$. The *dual distance* $\delta_{\mathcal{C}^\perp}$ of the code $\mathcal{C}$ is the minimum distance of $\mathcal{C}^\perp$.

As stated in Section 1.2.1, code-based masking employs two linear codes $\mathcal{C}$ and $\mathcal{D}$ in the following way:

$$V = X\mathcal{G}_\mathcal{C} + M\mathcal{G}_\mathcal{D}, \tag{1.54}$$

where $\mathcal{G}_\mathcal{C}$ and $\mathcal{G}_\mathcal{D}$ are generator matrices of $\mathcal{C}$ and $\mathcal{D}$ (thus the row spaces of $\mathcal{G}_\mathcal{C}$ and $\mathcal{G}_\mathcal{D}$ are $\mathcal{C}$ and $\mathcal{D}$), respectively.

At present, the side-channel security order of code-based masking has been linked to the minimum distance and the kissing number of $\mathcal{D}^\perp$.

Dual Distance

The side-channel security order of code-based masking has been linked to $\delta_{\mathcal{D}^\perp}$, the dual distance of masking code $\mathcal{D}$. In [PGS+17], the authors prove the (bit-probing) security of DSM is equal to $\delta_{\mathcal{D}^\perp}$; For IPM, [CG18] establishes a connection between $\delta_{\mathcal{D}^\perp}$, the mutual information between sensitive variable and leakage, and success rate. Namely, the security of the masking scheme is increasing as $\delta_{\mathcal{D}^\perp}$ increases.

Kissing Number

In [CGC+21b, CGC+21a], the authors prove the impact of code-based masking depends on two properties of $\mathcal{D}^\perp$ (the dual of the masking code): its minimum distance δ and the kissing number A_δ. By evaluating the mutual information between the sensitive variable and leakage, they prove the random masking is all the more secure as δ is large and the kissing number A_δ is small.

To conclude, the security of code-based masking is related to the distance and kissing number of the dual of the masking code. Given a masking code with a specific length and dimension, the larger its dual distance, the more secure the corresponding masking becomes. If two masking codes have the same distance, then the one with a smaller kissing number of its dual code is even more secure.

CHAPTER 2

Construction of Masking Codes

This chapter consists of two sections, the content of the first section has been published in "Linear Programming Bounds on the Kissing Number of q-ary Codes" [SLC+21], the second section has been published in "Towards Finding Best Linear Codes for Side-Channel Protections" [CLGR22].

Contents

Abstract

The first section uses linear programming to derive upper and lower bounds on the "kissing number" A_δ of any q-ary linear code $\mathcal{C}$ with distance distribution frequencies A_i, in terms of the given parameters $[n, k, \delta]$.

The second section attempts to address the constructive selection of optimal codes tailored for code-based masking when the device leaks information in the Hamming weight leakage model. We show that the problem is related to the weight enumeration of the extended dual of the masking code. We provide an efficient method to search for good codes, based on a lexicographic sorting of the weight enumeration polynomial from lowest to highest degrees.

2.1 Linear Programming Bounds on the Kissing Number of q-ary Codes

The *kissing number* A_δ of a linear code is the number of nonzero codewords of minimum weight δ. As we discussed in Section 1.4.3, when parameters $n, k, \delta_{\mathcal{D}^\perp}$ of the dual of the masking code are given, the corresponding masking is all the more secure as the kissing number of $\mathcal{D}^\perp$ is small [CGC+21b]. Therefore, the bounds on the kissing number give us a clue about limits on the security improvements achieved by the related code-based masking.

In this section, we derive several bounds on the kissing number of q-ary codes by using the linear programming method, including numerical and explicit bounds. For a given minimum distance δ, the kissing number can vary significantly as shown in Figure 2.1 and Figure 2.2. Therefore, the problem we consider is: Given the length, dimension, and minimum distance of a code, how to bound the kissing number above and below?

Building on MacWilliams formula of q-ary linear codes for Hamming weight enumerators (see [MS77, Chap. 5, Eq. (47)]), we derive bounds on kissing number by linear programming. This approach can be exploited numerically, using the linear programming solver of `Magma` [BCP97], or analytically via the polynomial method of [MS77, Chap. 17, Th. 20]. As shown in Tables 2.1 and 2.2, for binary codes, the numerical method is more precise, while the polynomial method is useful to create insightful bounds with an explicit analytical expression.

The more general problem of bounding arbitrary weight frequencies is studied using similar techniques in [ABL01]. However, the results in [ABL01] are mostly asymptotic: it gives non-explicit asymptotic bounds on all weight frequencies. In Ashikhmin et al.'s work [ABV01], they investigated the existence of codes whose kissing number satisfying an asymptotic lower bound. In the present paper we have strived to derive explicitly possibility bounds for any q-ary linear codes with given parameters $[n, k, \delta]$.

2.1.1 Background

Let $\mathcal{C}$ be a $[n, k, \delta]$ linear code over the finite field $\mathbb{F}_q$, with length n, dimension k, minimum distance δ, and the weight distribution $A_0, A_1, \ldots A_n$, where $A_i = |\{x \in \mathcal{C} \mid w_H(x) = i\}|$, $i = 0, \ldots, n$. Its dual code is denoted as $\mathcal{C}^\perp$, the weight distribution of $\mathcal{C}^\perp$ is denoted as $A'_0, A'_1, \ldots A'_n$. By definition, $A_0 = 1$, $A_j = 0$ for any $0 < j < \delta$, then

$$q^k = 1 + A_\delta + \sum_{j=\delta+1}^{n} A_j. \tag{2.1}$$

Before moving to linear programming bounds, let's review the definition of the Krawtchouk Polynomial.

Definition 2.1.1 (Krawtchouk Polynomial [MS77, Chap 5, §7])**.** For any prime power q

and positive integer n, define the Krawtchouk polynomial

$$P_k(x;n) = P_k(x) = \sum_{j=0}^{k} (-1)^j (q-1)^{k-j} \binom{x}{j} \binom{n-x}{k-j}, \tag{2.2}$$

where $k = 0, 1, \ldots, n$.

Example 2.1.1. The Krawtchouk polynomials for $k = 0, 1, 2$ are:

$$P_0(x) = 1, \tag{2.3}$$

$$P_1(x) = (q-1) - qx, \tag{2.4}$$

$$P_2(x) = \frac{1}{2}\Big(q^2x^2 + q(2n + q - 2nq - 2)x + 2n(n-1)(q-1)^2\Big). \tag{2.5}$$

In particular, when $q = 2$, they become:

$$P_0(x) = 1, \tag{2.6}$$

$$P_1(x) = n - 2x, \tag{2.7}$$

$$P_2(x) = 2x^2 - 2nx + \frac{n(n-1)}{2}. \tag{2.8}$$

See [MS77, Chap 5, §7] for background on these polynomials.

2.1.2 Linear Programming Bounds

For $[n, k]$ linear codes over $\mathbb{F}_q$, by MacWilliams formula for q-ary codes [MS77, Chap. 5, Eq. (47)] we have

$$q^k \sum_{i=0}^{n} A'_i x^{n-i} y^i = \sum_{i=0}^{n} A_i (x + (q-1)y)^{n-i} (x-y)^i, \tag{2.9}$$

which yields

$$q^k A'_i = \sum_{j=0}^{n} A_j P_i(j) \tag{2.10}$$

for all $i = 0, 1, \ldots, n$.

Linear programming leads to the following theorem concerning a lower bound on the kissing number.

Theorem 2.1.1 (Lower Bound on the Kissing Number)**.** *If $\mathcal{C}$ is an q-ary $[n, k, \delta]$ linear code then $A_\delta \geq q^k - 1 - \lfloor L \rfloor$, where L denotes the maximum of $\sum_{j=\delta+1}^{n} A_j$ subject to the $2n - \delta$ constraints*

$$-P_i(0) - (q^k - 1)P_i(\delta) \leq \sum_{j=\delta+1}^{n} A_j (P_i(j) - P_i(\delta)) \tag{2.11}$$

for $i = 1, 2, \ldots, n$, and $A_j \geq 0$ for $j = \delta + 1, \delta + 2, \ldots, n$.

Proof. By definition of A'_i, we have $A'_i \geq 0$ for $i = 1, 2, \ldots, n$ which, from (2.10), reads $P_i(0) + A_\delta P_i(\delta) + \sum_{j=\delta+1}^{n} A_j P_i(j) \geq 0$. Substituting $A_\delta = q^k - 1 - \sum_{j=\delta+1}^{n} A_j$ gives (2.11). The Theorem is proved by using (2.1) again. □

We have a similar result for upper bounds.

Theorem 2.1.2 (Upper Bound on the Kissing Number)**.** *If $\mathcal{C}$ is an $[n, k, \delta]$ q-ary code then $A_\delta \leq q^k - 1 - \lceil S \rceil$ where S denotes the minimum of $\sum_{j=\delta+1}^{n} A_j$ under the same constraints as above.*

Proof. The proof is similar as Theorem 2.1.1, so it is omitted. □

Consider the n inequality constraints (2.11)

$$-P_i(0) - (q^k - 1)P_i(\delta) \leq \sum_{j=\delta+1}^{n} A_j(P_i(j) - P_i(\delta))$$

for $i = 1, 2, \ldots, n$, along with the $n-\delta$ constraints $A_j \geq 0$ for $j = \delta+1, \delta+2, \cdots, n$. In this mathematical program, the A_j's are considered as rational variables if linear programming is used, or integral variables if integer programming is intended. Both approaches can be implemented in `Magma` [BCP97].

The calculation result of the linear programming method is presented in Figure 2.1 and Figure 2.2. Here we focus on binary codes, and take different rates $R = \frac{k}{n}$ as different examples ($R \approx \frac{1}{2}$ and $R \approx \frac{1}{3}$), with δ being the best known for given parameters $[n, k]$.

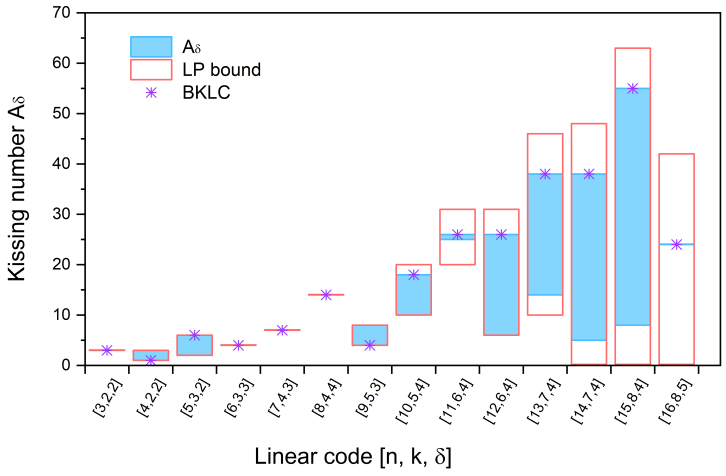


Figure 2.1: Linear programming bounds on the kissing number for $R \approx 1/2$. Bounds are tight for $n = 3, 4, 5, 6, 7, 8, 9$.

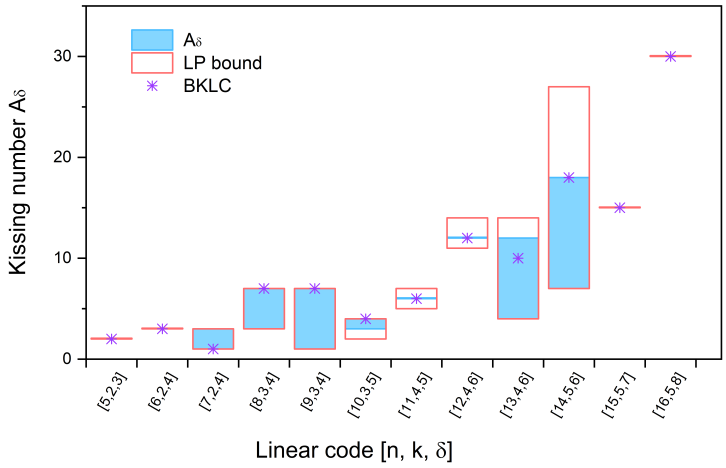


Figure 2.2: Linear programming bounds on the kissing number for $R \approx 1/3$. Bounds are tight for $n = 5, 6, 7, 8, 9, 15, 16$.

The LP bounds are represented for n ranging from 3 to 16. We omit the cases when $k = 1$ because they are trivial situations with only two codewords. For some choices $[3, 2, 2]$, $[6, 3, 3]$, $[7, 4, 3]$, $[8, 4, 4]$, $[5, 2, 3]$, $[6, 2, 4]$, $[15, 5, 7]$ and $[16, 5, 8]$, the lower and upper bounds agree and the kissing number is necessarily unique.

However, in general, the lower and upper bounds do not agree, and it is possible to find actual codes with different kissing numbers between those bounds, as represented in light blue color in Figure 2.1 and Figure 2.2. The research has been carried out by randomly selecting linear codes of parameters $[n, k, \delta]$ and the range displayed in blue correspond to actually discovered codes amongst the ones we explored. Our search could not be exhaustive so that there might exist codes with lower or higher kissing numbers.

Some exceptions are when:

- $[n, k, \delta] = [8, 4, 4]$ and $[16, 8, 5]$, as those are unique codes (extended Hamming code [MS77] and shortened QR code [MS77]). The uniqueness of the latter is proven in [BH01].

- $[n, k, \delta] \in \{[3, 2, 2], [6, 3, 3], [7, 4, 3], [5, 2, 3], [6, 2, 4], [11, 4, 5], [12, 4, 6], [15, 5, 7], [16, 5, 8]\}$, as the room between lower and upper bounds is limited.

We also superimposed in Figure 2.1 and Figure 2.2 the special case of `Magma` [BCP97] Best Known Linear Code (BKLC). The function BKLC(n, δ) returns a code with the largest known dimension, for a given length and minimum distance, consistently with Grassl database [Gra07], which favors codes obtained by some algebraic construction. On several occasions, especially for rate 1/2 codes, the kissing number of BKLC is relatively high, hence `Magma` [BCP97] is not adapted to applications requiring a small kissing number.

2.1.3 Explicit Bounds Using the Polynomial Method

The following identity is a polynomial way of expressing the duality of LP.

Lemma 2.1.3 (Polynomial Method[Del72, Eq.(18)]). *Let $\beta(x) \in \mathbb{Q}[x]$ denote a polynomial with Krawtchouk expansion*

$$\beta(x) = \sum_{j=0}^{n} \beta_j P_j(x). \tag{2.12}$$

The following identity holds

$$\sum_{i=0}^{n} \beta(i) A_i = q^k \sum_{j=0}^{n} \beta_j A'_j. \tag{2.13}$$

Proof. Immediate by (2.10), upon swapping the order of summation. □

2.1.3.1 Lower Bounds

Using Lemma 2.1.3 we have the following theorem. This theorem can also be obtained by setting appropriate parameters in [ABL01, Thm 1].

Theorem 2.1.4 (Lower Bound[ABL01]). *Let $\beta(x) \in \mathbb{Q}[x]$ satisfying*

$$\begin{aligned} \beta_j &\geq 0, \forall j = 0, 1, \ldots, n, && (2.14)\\ \beta(x) &\leq 0, \forall x \in (\delta, n], && (2.15)\\ \beta(\delta) &> 0, && (2.16)\\ q^k \beta_0 &> \beta(0). && (2.17) \end{aligned}$$

Then we have the lower bound

$$A_\delta \geq \frac{q^k \beta_0 - \beta(0)}{\beta(\delta)}. \tag{2.18}$$

Proof. By Lemma 2.1.3 we have

$$\beta(0) + A_\delta \beta(\delta) + \sum_{i=\delta+1}^{n} \beta(i) A_i \geq q^k \beta_0 A'_0 = q^k \beta_0,$$

implying

$$\beta(0) + A_\delta \beta(\delta) \geq q^k \beta_0,$$

where (2.19) and (2.19) hold due to the specific assumptions made about the polynomial $\beta(x)$, namely (2.14) and (2.15). □

The main result of this paragraph are the following corollaries. First, we consider the case of β linear.

Corollary 2.1.4.1. *If* $\delta = \lceil (n-1)(q-1)/q \rceil$, *then*

$$A_\delta \geq \frac{q^k - nq + n - 1}{(n-\delta)q - n + 1}. \tag{2.19}$$

Proof. Take $\beta(x) = nq - n + 1 - qx$. Because $P_0 = 1$ and $P_1(x) = (q-1)n - qx$ (see Example. 2.1.1), the coefficients of the Krawtchouk expansion are $\beta_0 = \beta_1 = 1$. Thus $\beta(0) = nq - n + 1$, $\beta(\delta) = nq - n + 1 - q\delta$, and $\beta(x) \leq 0$ when $x \geq \frac{nq-n+1}{q}$. So in order to satisfy $\beta(x) \leq 0$ for any integer $x \in (\delta, n]$, we must have $\delta + 1 \geq \frac{nq-n+1}{q}$. Combine with $\beta(\delta) > 0$ we have $(q-1)(n-1) \leq q\delta < (q-1)(n-1) + q$. Plugging this data into Theorem 2.1.4, the result follows. □

Next, we consider the scenario where β is a quadratic polynomial.

Corollary 2.1.4.2. *If* $q\delta > nq - n - 2q + 1$ *then*

$$A_\delta \geq \frac{q^{k-2}n(n - qn + q\delta + 2q - 1) - n\delta - n}{n - \delta}. \tag{2.20}$$

Proof. Assume $\beta = 1 + \beta_1 P_1(x) + \beta_2 P_2(x)$, where

$$P_2(x) = \frac{q^2}{2}x^2 + \frac{q(q - 2nq + 2n - 2)}{2}x + (q-1)^2\binom{n}{2}.$$

To ensure the negativity of β for $x \in (\delta, n]$, the simplest is to assume $\beta(\delta + 1) = \beta(n) = 0$. This gives a system of two equations in β_1, β_2. The solution according to Wolfram alpha [LLC] is

$$\beta_1 = \frac{nq - 2n - q\delta - 2q + 2}{n(n - qn + q\delta + 2q - 1)}, \quad \beta_2 = \frac{-2}{n(n - qn + q\delta + 2q - 1)}.$$

This yields

$$\beta(\delta) = \frac{q^2(n-\delta)}{n(n - qn + q\delta + 2q - 1)}, \quad \beta(0) = \frac{q^2(\delta + 1)}{(n - nq + q\delta + 2q - 1)}.$$

Then the result follows by Theorem 2.1.4. □

Example 2.1.2. Consider the binary code $\mathcal{C} = RM(1, m)$, with parameters $k = m + 1$, and $\delta = 2^{m-1}$. It is well-known that $\mathcal{C}$ is a two-weight code with $A_0 = A_{2^m} = 1$, and $A_{2^{m-1}} = 2^{m+1} - 2$. Since $2\delta - n + 3 > 0$, using Corollary 2.1.4.2 we have $A_\delta \geq 2^{m+1} - 2$. So $RM(1, m)$ meets the lower bound.

If more constraints are imposed on the linear code $\mathcal{C}$, using the same method as in 2.1.4.2, we can obtain the following result:

Corollary 2.1.4.3. *If $\mathcal{C}$ is a binary code and all weights of $\mathcal{C}$ lie in the range $[\delta, n-\delta]$, with distance $\delta < \frac{n}{2}$ and $(n-2\delta-1)^2 < n+1$, then*

$$A_\delta \geq \frac{2^{k-2}(n^2-4n\delta-3n)+(2^k+1)\delta(\delta+1)}{(2\delta-n)} - \delta - 1. \tag{2.21}$$

Proof. Because all weights of $\mathcal{C}$ lie in the range $[\delta, n-\delta]$, for a quadratic β, to ensure its negativity on the weights it is enough to assume $\beta(\delta+1) = \beta(n-\delta) = 0$. This gives a system of two equations in β_1, β_2, if we write $\beta = 1 + \beta_1 P_1(x) + \beta_2 P_2(x)$. The solution according to Wolfram alpha [LLC] is

$$\beta_1 = \beta_2 = \frac{2}{n+1-(n-2\delta-1)^2}.$$

This yields

$$\beta(\delta) = \frac{-4n+8\delta}{n^2-4n\delta-3n+4\delta^2+4\delta}, \quad \beta(0) = \frac{4(\delta^2+\delta-n\delta-n)}{n^2-4n\delta-3n+4\delta^2+4\delta}.$$

Then the result follows by Theorem 2.1.4. □

2.1.3.2 Upper Bounds

Like Theorem 2.1.4, the following theorem can also be obtained by setting appropriate parameters in [ABL01, Thm 1].

Theorem 2.1.5 (Upper Bound[ABL01]). *Let $\beta(x) \in \mathbb{Q}[x]$ satisfying*

$$\beta_j \leq 0, \forall j = 1, \ldots, n, \tag{2.22}$$

$$\beta(x) \geq 0, \forall x \in (\delta, n], \tag{2.23}$$

$$\beta(\delta) > 0, \tag{2.24}$$

$$q^k \beta_0 > \beta(0). \tag{2.25}$$

Then we have the upper bound

$$A_\delta \leq \frac{q^k\beta_0 - \beta(0)}{\beta(\delta)}. \tag{2.26}$$

The proof is analogous to that of Theorem 2.1.4 and is omitted.

The main results of this paragraph are the following corollaries. First, we consider the case of β linear.

Corollary 2.1.5.1. *If $n - nq + 1 + q\delta > 0$, then*

$$A_\delta \leq \frac{q^k + nq - n - 1}{n - nq + 1 + q\delta}. \tag{2.27}$$

Proof. Take $\beta(x) = n - nq + 1 + qx$, where $P_1(x) = (q-1)n - qx$. By construction $\beta_0 = 1$, and $\beta_1 = -1$. Note that $\beta(0) = n - nq + 1$, and $\beta(\delta) = n - nq + 1 + q\delta$. We see that $\beta(x) > 0$, for x an integer $> \frac{n-nq+1}{q}$. Plugging this data into Theorem 2.1.5, the result follows. □

Next, we once again consider the scenario where β is a quadratic polynomial.

Corollary 2.1.5.2. *If $\delta < \frac{(q-1)n+1}{q}$, then*

$$A_\delta \leq \frac{q^{k-2}n(qn - n - q\delta + 1) + n(\delta - 1)}{n - \delta}. \tag{2.28}$$

Proof. Assume $\beta = 1 - \beta_1 P_1(x) - \beta_2 P_2(x)$, with β_1, $\beta_2 > 0$. To ensure the positivity of β for $x \in (\delta, n]$ the simplest is to assume $\beta(\delta - 1) = \beta(n) = 0$. This gives a system of two equations in β_1, β_2. The solution according to `Magma` [BCP97] is

$$\beta_1 = \frac{2n + q\delta - 2 - nq}{qn^2 - n^2 - qn\delta + n}, \beta_2 = \frac{2}{qn^2 - n^2 - qn\delta + n}.$$

This yields

$$\beta(\delta) = \frac{q^2(n - \delta)}{qn^2 - n^2 - qn\delta + n}, \quad \beta(0) = \frac{q^2(1 - \delta)}{qn - n - q\delta + 1}.$$

The result follows by Theorem 2.1.5. □

Example 2.1.3. Still consider the binary code $\mathcal{C} = RM(1, m)$, where $n = 2^m$, $k = m + 1$, and $\delta = 2^{m-1}$. Using Corollary 2.1.5.2, we have $A_\delta \leq 2^{m+1} - 2$. From Corollary 2.1.4.2 we know $A_\delta \geq 2^{m+1} - 2$. So $A_\delta = 2^{m+1} - 2$. Because $A_0 = 1$, it proved that $RM(1, m)$ is a two-weight code. $RM(1, m)$ is the only code we know that satisfies the upper bound and the lower bound at the same time.

If more constraints are imposed on the linear code $\mathcal{C}$, we can derive the following result:

Corollary 2.1.5.3. *If $\mathcal{C}$ is a binary code and all weights of $\mathcal{C}$ lie in the range $[\delta, n - \delta]$, with $n - 2\delta > 0$ and $(n - 2\delta + 2)^2 > n$, then*

$$A_\delta \leq \frac{2^{k-2}\big((n - 2\delta + 2)^2 - n\big) + (\delta - 1)(n + 1 - \delta)}{n + 1 - 2\delta}. \tag{2.29}$$

Proof. For a quadratic β, of concavity $\cap$, to ensure its positivity on the weights it is enough to assume $\beta(\delta - 1) = \beta(n - \delta + 1) = 0$.

This gives a system of two equations in β_1, β_2, if we write $\beta = 1 - \beta_1 P_1(x) - \beta_2 P_2(x)$. The solution according to `Magma` [BCP97] is

$$\beta_1 = 0, \quad \beta_2 = \frac{2}{(n - 2\delta + 2)^2 - n}.$$

This yields

$$\beta(0) = \frac{4(\delta - 1)(\delta - n - 1)}{(n - 2\delta + 2)^2 - n}, \quad \beta(\delta) = \frac{4(1 - 2\delta + n)}{(n - 2\delta + 2)^2 - n}.$$

Then the result follows by Theorem 2.1.5. □

Table 2.1 and Table 2.2 contain the bounds for binary codes.

Table 2.1: Upper/Lower Bounds for some Linear Codes

Binary code	Lower bound of A_δ		Upper bound of A_δ	
$[n, k, \delta]$	Poly. method	LP bound	LP bound	Poly. method
$[8, 3, 4]$	2	3	7	10
$[8, 4, 4]$	14	14	14	14
$[9, 3, 4]$	−2	1	7	12
$[9, 4, 4]$	6	6	14	19
$[10, 3, 5]$	0	2	4	12
$[10, 4, 4]$	−2	12	15	25
$[11, 4, 5]$	4	5	7	22
$[12, 4, 6]$	10	11	14	18
$[13, 4, 6]$	2	4	14	24
$[14, 4, 7]$	8	8	8	20
$[14, 5, 6]$	2	7	27	50
$[15, 4, 8]$	15	15	15	15
$[15, 5, 7]$	15	15	15	41
$[16, 4, 8]$	6	7	15	22
$[16, 5, 8]$	30	30	30	30

Table 2.1 shows that the LP bound is more precise in general than the polynomial method. The interest of the latter resides in producing intuitive bounds with a closed formula.

Table 2.2 shows the LP bounds for n ranging from 17 to 32. It is a supplement to the results in Figure 2.1 and Figure 2.2. Because it is difficult to calculate all possible values of A_δ, we did not compare the bounds with the range of A_δ as Figure 2.1 and Figure 2.2. As we can see from Table 2.2, for some values $[22, 11, 7]$, $[23, 12, 7]$ and $[24, 12, 8]$, the lower and upper bounds agree, which means the LP bounds must be tight at these values. It also shows that when n is large, the lower bounds for some values may be trivial (smaller than 1), while the upper bounds are much smaller than the trivial bounds ($A_\delta \leq 2^k - 1$).

Table 2.2: LP Bounds for some Linear Codes

Binary codes ($R \approx \frac{1}{2}$)			Binary codes ($R \approx \frac{1}{3}$)		
$[n, k, \delta]$	lower bound	upper bound	$[n, k, \delta]$	lower bound	upper bound
$[17, 9, 5]$	17	50	$[17, 6, 7]$	12	23
$[18, 9, 6]$	69	142	$[18, 6, 8]$	32	50
$[19, 10, 5]$	-14	72	$[19, 6, 8]$	12	51
$[20, 10, 6]$	40	209	$[20, 7, 8]$	29	83
$[21, 11, 6]$	56	282	$[21, 7, 8]$	9	83
$[22, 11, 7]$	176	176	$[22, 7, 8]$	-3	88
$[23, 12, 7]$	253	253	$[23, 8, 8]$	-2	143
$[24, 12, 8]$	759	759	$[24, 8, 8]$	-12	163
$[25, 13, 6]$	-23	526	$[25, 8, 9]$	-29	64
$[26, 13, 7]$	-67	295	$[26, 9, 9]$	-43	100
$[27, 14, 7]$	-33	353	$[27, 9, 10]$	31	247
$[28, 14, 8]$	295	1138	$[28, 9, 10]$	-4	259
$[29, 15, 7]$	-182	509	$[29, 10, 10]$	-5	396
$[30, 15, 8]$	105	1724	$[30, 10, 11]$	-14	178
$[31, 16, 8]$	168	1985	$[31, 10, 12]$	149	442
$[32, 16, 8]$	-36	2274	$[32, 11, 12]$	298	639

2.1.4 Applications in Code-based Masking

Recall that the kissing number is one of the two factors that determine the concrete side-channel security level in the code-based masking [CGC+21b] because the mutual information that measures the informativeness of leakage is proportional to the kissing number. In this respect, Theorem 2.1.4 and 2.1.5 enable us to bound the security gains induced by the corresponding code-based masking. In particular, given the code parameters $[n, k, \delta]$, these two theorems indicate the best and the worst cases of codes that can be achieved in practice.

Taking the code $[8, 4, 4]$ in Table 2.1 as an example, it is unique and has been proven to be the best case in the code-based masking with two shares over $\mathbb{F}_{2^4}$, given the variance of Gaussian noise is greater than 1.0 [CG18]. In this case, both lower bound and upper bound coincide in 14. Another example is the linear codes with parameters $[12, 4, 6]$, which are the optimal choices in three share cases over $\mathbb{F}_{2^4}$ [CGC+21b]. In the latter case, the lower and upper LP bounds are 11 and 14, respectively, where the BKLC code in `Magma` [BCP97] gives $A_\delta = 12$. It is worth mentioning that $A_\delta = 12$ is unique for all $[12, 4, 6]$ linear codes which is verified by exhaustive code search, although there are several non-equivalent classes.

In general, algebraic codes owing to their large automorphism group have a large kissing number. Conversely, the application of code-based masking favors codes with low kissing number, which are less studied and certainly deserve more attention.

2.2 Finding Best Linear Codes for Side-Channel Protections

Let K, T, X represent the secret key, the public variable, and sensitive variable respectively. The following model has been introduced in Section. 1.2:

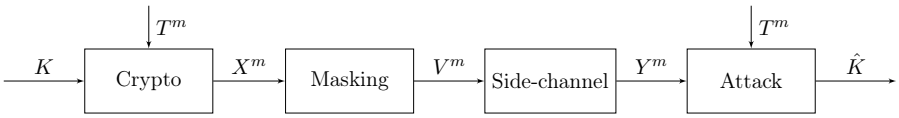


Figure 2.3: Side-channel seen as a communication channel (with masking).

In this section, the number of shares is denoted as n instead of $d+1$. This is because we have plenty discussions about the masking code in this section, and using n as the notation aligns with the commonly used code parameter $[n, k]$.

To protect the key-dependent sensitive variable, X is masked into n shares: $V = (X_0, X_1, \ldots, X_{n-1})$. The side-channel leakage is denoted as $Y = (Y_0, Y_1, \ldots, Y_{n-1})$ with each Y_i deponding on X_i, $i = 0, 1, \ldots, n-1$. The attacker queries the cryptographic device m times and obtains corresponding traces Y^m. Finally, the attacker use Y^m and T^m to compute the guessed key $\hat{K}$.

For the code-based masking, as stated in Section 1.2.1, randomness M is introduced in masking procedure and both X and M are multiplied by generator matrices of the masking codes. Let $\mathbb{F}_q$ be a finite field where q is a power of 2, and k, s be integers satisfying $k + s \leq n$. One has

$$V = X\mathcal{G}_{\mathcal{C}} + M\mathcal{G}_{\mathcal{D}}, \tag{2.30}$$

where $X \sim \mathcal{U}(\mathbb{F}_q^k)$, $M \sim \mathcal{U}(\mathbb{F}_q^s)$, $\mathcal{G}_{\mathcal{C}} \in \mathbb{F}_q^{k\times n}$ has rank k and $\mathcal{G}_{\mathcal{D}} \in \mathbb{F}_q^{s\times n}$ has rank s. The row space of $\mathcal{G}_{\mathcal{C}}$ and $\mathcal{G}_{\mathcal{D}}$ are denoted as $\mathcal{V}_{\mathcal{G}_{\mathcal{C}}} = \mathcal{C}$ and $\mathcal{V}_{\mathcal{G}_{\mathcal{D}}} = \mathcal{D}$, they are two linear codes with paremeters $[n, k]$ and $[n, s]$ respectively. We assume $\mathcal{C} \cap \mathcal{D} = \{0\}$.

It follows that from the perspective of side-channel resistance, the word-level security is only captured by the minimum distance of $\mathcal{D}^\perp$ [CG18, PGS+17]. By contrast, the bit-level security of a code-based masking is related to both the minimum distance and the kissing number of $\mathcal{D}^\perp$ [CGC+21b, CGC+21a] under the Hamming weight leakage model.

Rather than searching from all possible candidates as in [CGC+21a], we aim at constructing optimal linear codes for generalized code-based masking (GCM) by an efficient algorithm. This is an open problem to the best of our knowledge. In [CGC+21b], the authors demonstrate a good code (for masking countermeasure) should have a large minimum distance and a low kissing number. However, as we can see from the definition of weight distribution (Definition 1.4.3), the kissing number of a code is only one coefficient of the weight distribution polynomial. As we demonstrate in the sequel, the entire weight distribution is to be considered to assess the side-channel resistance of a code-based masking.

2.2.1 Background

Definition 2.2.1 (Linear Code over Specific Finite Field)**.** An $[n, k, \delta]$ linear code over the finite field $\mathbb{F}_{2^\ell}$ is denoted by $[n, k, \delta]_{2^\ell}$ to refer to the field on which the code is defined.

Definition 2.2.2 (Subfield extension of a code [MS77])**.** The subfield representation of $x \in \mathbb{F}_{2^\ell}$ is its vector of coordinates $[x] \in \mathbb{F}_2^\ell$, which depends on the choice of the basis of $\mathbb{F}_{2^\ell}$ over $\mathbb{F}_2$. The subfield extension $[\mathcal{C}]$ is the set of all vectors obtained from the codewords of $\mathcal{C}$ by taking the subfield representation of every component.

Considering a generator matrix of a linear code $\mathcal{C}$ of size $k \times n$ in $\mathbb{F}_{2^\ell}$, the generator matrix of the extended code $[\mathcal{C}]$ has a size of $k\ell \times n\ell$ in $\mathbb{F}_2$.

Definition 2.2.3 (Prefix-based lexicographical order of sequences)**.** Let (A_i) and (A'_i) $(0 \leq i \leq n)$ be two sequences of integers of length n. The sequence (A_i) is (strictly) *smaller* than the sequence (A'_i) if $A_0 < A'_0$, or if there exists $1 \leq j \leq n$, such that $A_i = A'_i$ for all $0 \leq i < j$, and $A_j < A'_j$.

Definition 2.2.4 (Best weight distribution)**.** A linear code $\mathcal{C}$ is said to be *better* than a linear code $\mathcal{C}'$ if its weight distribution is (prefix-based) *smaller* than that of $\mathcal{C}'$. A code has the *best* weight distribution if it is better than any other linear code.

Thus, to obtain the best weight distribution, we apply the following three principles:

1. maximize the minimum distance δ (recall that $\delta = \min\{i \neq 0; A_i > 0\}$);
2. (in case of a tie) minimize the kissing number A_δ;
3. (in case of a tie) minimize the following coefficients A_i, $i > \delta$ in lexicographical order.

Regarding the first principle, it is feasible to construct a maximum distance separable (MDS) code which maximizes the minimum distance. We have the following Delsarte's lemma for the dual of an MDS code.

Lemma 2.2.1 (Dual of an MDS code [Del75])**.** *The dual of an MDS code is also an MDS code.*

Corollary 2.2.1.1. *The dual distance of a linear MDS code of parameters* $[n, k]_{2^\ell}$ *is* $\delta = k + 1$.

Proof of the corollary. The dual distance of a linear MDS code is equal to the minimum distance of the dual of the code, which has parameters $[n, n-k]_{2^\ell}$. By Lemma 2.2.1, it is MDS. Therefore, the Singleton bound [Sin64] is tight and we have that $n-(n-k)+1=\delta$. Hence $\delta = k+1$. □

With these notations, we present the following Algorithm 1, provide a conceptual process for finding the best masking code for GCM. In particular, the difference comparing with [CGC+21b, CGC+21a] lies in line 4, which indicates the better code in case of a tie in A_i for $\delta \leq i \leq n$.

Input : Masking order t (at word level over $\mathbb{F}_{2^\ell}$)

Output : Codes for GCM over $\mathbb{F}_{2^\ell}$

1 Construct an MDS code $\mathcal{D}$:$[n, n-k]_{2^\ell}$ with $\delta_{\mathcal{D}^\perp} = t+1$ // Use Corollary 2.2.1.1, $\delta_{\mathcal{D}^\perp} = n-k+1$

2 Apply subfield extension on $\mathcal{D}$ to get $[\mathcal{D}]$ // Use Definition 2.2.2

3 Compute the dual code $[\mathcal{D}]^\perp$ // Use Definition 1.4.5

4 **if** $[\mathcal{D}]^\perp$ *has the best weight distribution* **then** // Use Definition 2.2.4

5 | **return** $\mathcal{D}$

6 **else**

7 | **goto** Line 1

Algorithm 1: Conceptual process for finding the best masking code for GCM.

2.2.2 Orthogonal Bases and Subfield Representations

In a code-based masking scheme, the side-channel security order at bit level is related to the weight distribution of the codes in the subfield representation [CGC+21b, CGC+21a]. Particularly, given a code $\mathcal{D}$ in (2.30) defined over $\mathbb{F}_{2^\ell}$, we wish to evaluate the weight distribution of the dual extended code $[\mathcal{D}]^\perp$, and the natural question is to assess whether this is equivalent to evaluate the weight distribution of extended dual code $[\mathcal{D}^\perp]$. However, as shown in Figure 2.4, the commutative relationship does not hold in general because depending on the choice of basis of $\mathbb{F}_{2^\ell}$ over $\mathbb{F}_2$, the two codes $[\mathcal{D}]^\perp$ and $[\mathcal{D}^\perp]$ are not always equivalent to each other.

As it turns out, the commutative relationship will hold true if the basis used in subfield representation is a *trace-orthogonal* basis. Therefore, we first show how to construct trace-orthogonal bases and then investigate the subfield extension of the code.

2.2.2.1 Construction of Trace-Orthogonal Bases

Let $\ell > 1$ and $\mathbb{F}_{2^\ell}$ be the extension field of $\mathbb{F}_2$. By the Frobenius conjugacy property, the trace function $\mathrm{tr} : \mathbb{F}_{2^\ell} \to \mathbb{F}_2$, defined as $\mathrm{tr}(x) = \sum_{i=0}^{\ell-1} x^{2^i}$, is linear.

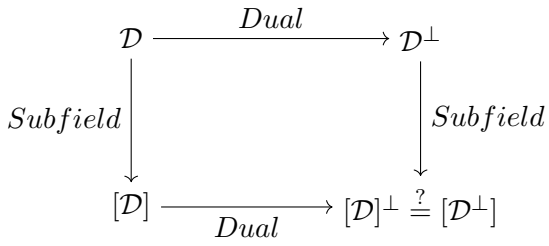


Figure 2.4: Commutative connection between sub-field representation and duality.

The (trace-)orthogonality and orthonormality is defined as follows.

Definition 2.2.5. Elements a_1, a_2 in $\mathbb{F}_{2^\ell}$ are *orthogonal* if $\mathrm{tr}(a_1 a_2) = 0$. A basis $\{a_1, a_2, \ldots, a_\ell\}$ of $\mathbb{F}_{2^\ell}$ over $\mathbb{F}_2$ is *orthonormal* if $\mathrm{tr}(a_i^2) = \mathrm{tr}(a_i) = 1$ and $\mathrm{tr}(a_i a_j) = 0$ for all $i \neq j$.

Notice that, as mentioned in [SL80], we have the following result:

Lemma 2.2.2. *A (trace-)orthogonal basis in $\mathbb{F}_{2^\ell}$ is always orthonormal.*

Proof. Let a_i be elements in a basis, where $i \in \{1, \ldots, \ell\}$. We need to show that it satisfies $\mathrm{tr}(a_i) = 1$.

The trace takes values in $\mathbb{F}_2$, which consists in two elements, namely 0 and 1. Therefore, it must be proven that $\mathrm{tr}(a_i) \neq 0$. This means that a_i is not self-orthogonal, since $\mathrm{tr}(a_i^2) = \mathrm{tr}(a_i)^2 = \mathrm{tr}(a_i)$ in $\mathbb{F}_2$.

Let us reason by the absurd. Assume that a_i is self-orthogonal. Then, not only a_i is orthogonal to all vectors a_j $(j \neq i)$, but also to itself. Therefore, it belongs to the dual of the space vector E generated by the basis $\{a_1, a_2, \ldots, a_\ell\}$. Notice that E is the universe code, hence its dual is the singleton $\{0\}$. Consequently $a_i = 0$, which contradicts the fact that a_i is a basis vector. □

Remark 2.2.1. Incidentally, we notice that the condition (36) in [Lem75, §5, p. 182] is superfluous, since already implied by condition (37).

By [LN97, Note 3, page 75] (which points to the original paper [Lem75]), we know that an orthonormal basis always exists. Although [Lem75] gives a formal construction meant to provide the existence result, the resulting implementation is double-exponential in 2^ℓ, which is far too complex to implement in practice.

In this paper, we consider instead a fast, but probabilistic, trace-orthogonal basis construction given by Algorithm 2. For $\ell = 8$, it works most of the time in one iteration (e.g., about 70.20% over 2000 times of randomly running Algorithm 2). Examples are provided below.

Input : $\ell \geq 1$, the extension order of $\mathbb{F}_2$, and α, a primitive element of $\mathbb{F}_{2^\ell}$

Output : An orthonormal basis of $\mathbb{F}_{2^\ell}$

```
1  (b_1, ..., b_ℓ) ← (0, ..., 0)                          // Basis, initialized with 0s
2  for i ∈ {1, ..., ℓ} do                                  // Find the ith element of the orthonormal basis
3      for a ∈ (F_{2^ℓ})* do                               // Candidate next vector in the basis (chosen randomly)
4          if tr(a) = 1 then                               // Test for tr(a^2) = tr(a)^2 ≠ 0 (only element ≠ 0 is 1 in F_2)
5              is_orthogonal ← true
6              for j ∈ {1, ..., i − 1} do
7                  if tr(a b_j) ≠ 0 then                   // Test whether a and b_j are orthogonal
8                      is_orthogonal ← false
9              if is_orthogonal then
10                 b_i ← a
11 return (b_1, ..., b_ℓ)
```

Algorithm 2: Randomized construction of an orthonormal basis in $\mathbb{F}_{2^\ell}$.

Remark 2.2.2. Strictly speaking, Algorithm 2 does not necessarily converge with a basis of full rank. We observed that depending on the scanning order of field elements at line 3, the algorithm can succeed or fail to return a basis. Therefore, we introduced a randomization at this line, and repeated the algorithm until it returns a (full rank) basis.

In viewing of Definition 2.2.5, the elements in a basis must satisfy $\mathrm{tr}(a_i) \neq 0$. Therefore, we can improve Algorithm 2 by removing zero-trace elements with a preliminary check of all traces. The new procedure is shown in Algorithm 3.

Table 2.3 presents the comparison on efficiency between Algorithms 2 and 3. The performance metric is the execution time, measured on a server which runs the `Magma` [BCP97] system. This clearly shows the advantage of using Algorithm 3 when the order of the finite field is large. For instance, when $\ell = 16$, Algorithm 3 have a speedup by a factor of 5 compared to Algorithm 2.

We shall use the following two examples of trace-orthogonal bases throughout the rest of this paper:

- $\mathcal{B}_0 = \{\alpha^{252}, \alpha^{156}, \alpha^{122}, \alpha^{203}, \alpha^{5}, \alpha^{126}, \alpha^{71}, \alpha^{65}\}$,
- $\mathcal{B}_1 = \{\alpha^{121}, \alpha^{252}, \alpha^{202}, \alpha^{20}, \alpha^{242}, \alpha^{15}, \alpha^{126}, \alpha^{44}\}$.

where α is the first primitive element in the finite field $\mathbb{F}_{2^8}$. Note that the irreducible polynomial used in this paper is: $g(X) = X^8 + X^4 + X^3 + X^2 + 1$. Moreover, we also investigate the default basis used in `Magma` [BCP97], which is a non-orthogonal basis:

- $\mathcal{B}_2 = \{1, \alpha^1, \alpha^2, \alpha^3, \alpha^4, \alpha^5, \alpha^6, \alpha^7\}$.

Input : ℓ, the extension order, and α, a primitive element of $\mathbb{F}_{2^\ell}$

Output : An orthonormal basis of $\mathbb{F}_{2^\ell}$

```
1  list ← {}
2  for i ∈ {1, ..., 2^ℓ − 1} do                    // Check the trace of elements in F*_{2^ℓ}
3      if tr(α^i) = 1 then
4          list ← list ∪ {i}                        // Put the power in list if trace equals 1
5  B ← {α^list[1]}                                  // Create a set with one element
6  start ← 2                                        // Set the start position of searching (can be changed)
7  while #B ≠ ℓ do
8      n ← start
9      for k ∈ {2, ..., ℓ} do                       // Find the kth element of the orthonormal basis
10         for s ∈ {n + 1, ..., #list} do
11             is_orthogonal ← true
12             for j ∈ {1, ..., k − 1} do           // Test whether the candidate is orthogonal with elements in B
13                 a ← B[j] · α^list[s]
14                 if tr(a) ≠ 0 then
15                     is_orthogonal ← false
16             if is_orthogonal then
17                 B ← B ∪ a
18                 n ← s
19         if #B < k then                           // Start again if we cannot find next base
20             break;
21     start ← start + 1                            // Change a start position (if we do not get enough basis)
22 return B
```

Algorithm 3: The improved construction of orthonormal bases in $\mathbb{F}_{2^\ell}$.

Table 2.3: The comparison on efficiency of two algorithms for constructing trace-orthogonal bases. Note that with our `Magma` [BCP97] server is with Intel Xeon CPU@2.0GHz, 4 processors (only one is used), and with 16GB Memory.

ℓ		4	8	12	16	20	24
Run time (sec)	Alg. 2	0.0001	0.0038	0.1150	1.5034	36.0350	1146.1685
	Alg. 3	0.0001	0.0019	0.0334	0.3065	4.7267	267.7467

2.2.2.2 Subfield Representation and Duality of Codes

We therefore specify the representation in Definition 2.2.2 by showing how to transform an element over $\mathbb{F}_{2^\ell}$ into $\mathbb{F}_2$. The subfield representation $[a]$ of a field element a is defined as follows.

Definition 2.2.6. Let $b = (b_1, \ldots, b_\ell)$ an orthonormal basis of $\mathbb{F}_{2^\ell}$. The subfield representation of $a \in \mathbb{F}_{2^\ell}$ is $[a] = (\mathrm{tr}(ab_1), \ldots, \mathrm{tr}(ab_\ell))$.

The subfield representation code $[\mathcal{D}]$ can be seen a concatenated code (as per Forney [For65]) with $\mathcal{D}$ of parameters $[n, k]_{2^\ell}$ as the outer code, and the universal $[\ell, \ell, 1]_2$ as the inner code. As a consequence, the side-channel security at bit level and word (ℓ-bit string) level are related by the subfield representation as follows: The security order at word level is the dual distance of the code in $\mathbb{F}_{2^\ell}$, whereas the security order at bit level is the dual distance of the subfield representation in $\mathbb{F}_2$.

A nice feature of trace-orthonormal bases is that duality and subfield representation commute:

Theorem 2.2.3. *Let $\mathcal{D}$ be a linear code. Then under a trace-orthogonal basis, we have:*

$$[\mathcal{D}]^\perp = [\mathcal{D}^\perp]. \tag{2.31}$$

Said equivalently, the duality and the sub-field representation form a commutative diagram:

$$\begin{array}{ccc} \mathcal{D} & \xrightarrow{Dual} & \mathcal{D}^\perp \\ {\scriptstyle Subfield}\downarrow & & \downarrow{\scriptstyle Subfield} \\ [\mathcal{D}] & \xrightarrow{Dual} & [\mathcal{D}]^\perp = [\mathcal{D}^\perp] \end{array}$$

Proof. Given $x, y \in \mathbb{F}_{2^\ell}^n$ and their subfield representations are $[x], [y] \in \mathbb{F}_2^{n\ell}$, respectively. Then the inner product $\langle x|y\rangle = 0$ implies that $0 = \mathrm{tr}(\langle x|y\rangle) = \sum_i \mathrm{tr}(x_i y_i) = \sum_i \sum_j [x_i]_j [y_i]_j = \langle [x]|[y]\rangle$ where the third equality holds because of the property of the trace-orthogonal basis. Therefore, we obtain $[\mathcal{D}^\perp] \subseteq [\mathcal{D}]^\perp$.

Inversely, two linear codes $[\mathcal{D}^\perp]$ and $[\mathcal{D}]^\perp$ are subspaces of $\mathbb{F}_2^{n\ell}$ that have the same length $2^{n\ell}$ and dimension $2^{(n-k)\ell}$, implying the same number of codewords in both codes. As a consequence, we have $[\mathcal{D}^\perp] = [\mathcal{D}]^\perp$. □

As a straightforward consequence of Theorem 2.2.3, the order of two transformations in lines 2 and 3 of Algorithm 1 is interchangeable. Therefore, the selection of the best codes can be achieved from the code $\mathcal{D}$ to the dual code $\mathcal{D}^\perp$ and then to the subfield extension $[\mathcal{D}^\perp]$. Section 2.2.2.3 illustrates the gain in terms of speed of this method.

Remark 2.2.3. We notice that the resulting distances are not the same depending on:

- which basis is used,
- the code itself.

We provide several examples of properties of codes $\mathcal{D}^\perp$ of parameters $[5, 3]_{256}$ (for $\ell = 8$). The `Magma` [BCP97] scripts are given in Appendix of [CLGR22]). The difference between the tables are the bases:

- $\mathcal{B}_0$ is used in Table 2.4,
- $\mathcal{B}_1$ is used in Table 2.5.

2.2.2.3 Optimized code research

We notice that the Subfield extension operation is "one-way". Namely, it is easy to extend a code from $\mathbb{F}_{2^\ell}$ to $\mathbb{F}_2^\ell$ (see Magma `SubfieldRepresentationCode` command), but the inverse operation is not trivial. Moreover, all codes of parameters $[n\ell, k\ell]_2$ cannot be interpreted as codes $[n, k]_{2^\ell}$. On the contrary, taking the dual of a linear code is invertible, and even involutive, as $(\mathcal{C}^\perp)^\perp = \mathcal{C}$.

Thus, leveraging trace-orthogonal bases, one can simplify the search for good codes by trading Alg. 4 (which is a realization of Alg. 1) by Alg. 5.

2.2.3 Characterizing Side-Channel Security by Weight Distribution

Mutual information (MI) is commonly used in tasks related to measuring side-channel leakage as an information-theoretic metric. Essentially, MI measures the statistical dependencies between the key-dependent variables and the leakage, which considers the full distributions of corresponding variables. Since the weight distribution determines how weights of codewords in a linear code are distributed, it therefore determines the leakage distribution of the masked variable from a coding-theoretic perspective [CGC+21b].

In view of this, we have the following conjecture.

Conjecture 2.2.1. MI between the sensitive variable and side-channel leakage depends on the weight distributions of the corresponding codes in the code-based masking.

Table 2.4: The dual distances for two seeds when drawing random code $\mathcal{D}$, using $\mathcal{B}_0$ of $\mathbb{F}_{256}$.

`SetSeed(0)`		`SetSeed(1)`	
$\delta_{\mathcal{D}^\perp}$	$\delta_{[\mathcal{D}]^\perp}$	$\delta_{\mathcal{D}^\perp}$	$\delta_{[\mathcal{D}]^\perp}$
4	8	4	6
3	6	4	7
4	8	4	6
4	6	4	6
4	8	4	8
4	7	4	8
4	7	4	8
4	7	4	8
4	8	4	7
4	7	4	8

Table 2.5: The dual distances for two seeds when drawing random code $\mathcal{D}$, using $\mathcal{B}_1$ of $\mathbb{F}_{256}$.

`SetSeed(0)`		`SetSeed(1)`	
$\delta_{\mathcal{D}^\perp}$	$\delta_{[\mathcal{D}]^\perp}$	$\delta_{\mathcal{D}^\perp}$	$\delta_{[\mathcal{D}]^\perp}$
4	8	4	7
3	6	4	7
4	7	4	7
4	7	4	8
4	8	4	7
4	7	4	7
4	7	4	8
4	6	4	7
4	7	4	7
4	7	4	8

Input : Number of iterations N

Output : Best found GCM code over $\mathbb{F}_{2^\ell}$

1 $w \leftarrow (2^n, 0, \ldots, 0)$ // Worst case for a weight enumeration polynomial

2 $\mathcal{D}_{\text{best}} \leftarrow \text{RandomCode}[n, k]_{2^\ell}$

3 **for** $i \in \{1, \ldots, N\}$ **do**

4 Select a random code $\mathcal{D}$

5 **if** *enumerationPolynomial($[\mathcal{D}]^\perp$) is better than w* **then**

6 $w \leftarrow \text{enumerationPolynomial}([\mathcal{D}]^\perp)$

7 $\mathcal{D}_{\text{best}} \leftarrow \mathcal{D}$

8 **return** $\mathcal{D}_{\text{best}}$

Algorithm 4: Bounded search for an efficient code

Input : Number of iterations N

Output : Best found GCM code over $\mathbb{F}_{2^\ell}$

$w \leftarrow (2^n, 0, \ldots, 0)$ // Worst case for a weight enumeration polynomial

$\mathcal{D}_{\text{best}} \leftarrow \text{RandomCode}[n, k]_{2^\ell}$

for $i \in \{1, \ldots, N\}$ **do**

  Select a random code $\mathcal{D}'$

  if *enumerationPolynomial($[\mathcal{D}']$) is better than* w **then**

    $w \leftarrow \text{enumerationPolynomial}([\mathcal{D}'])$ // No computation of dual code for all candidates

    $\mathcal{D}_{\text{best}} \leftarrow \mathcal{D}^\perp$

return $\mathcal{D}_{\text{best}}$

Algorithm 5: Optimized (compared to Alg. 4) bounded search for an efficient code

It is well-known that for a code of dual distance δ, any tuple of $\delta - 1$ coordinates is uniformly distributed, and some tuples of δ coordinates are linearly dependent [MS77, Theorem 10]. Therefore, the side-channel security order under probing model is $\delta - 1$, and an attack of order δ, corresponding to codewords of Hamming weight equal to δ, brings some mutual information that depends on $\sigma^{-2\delta}$, where σ^2 is the variance of the AWGN channel that characterized the leakage model. Moreover, since not all codewords have the same Hamming weight δ, other codewords of weights $> \delta$ should bring more information when considering mutual information as an information-theoretic metric.

Said differently, the mutual information is related to $\sum_{i=0}^{n\ell} \frac{A_i}{\sigma^{2i}}$, or more precisely (removing the useless 1 constant arising from $i = 0$), it is related to:

$$\sum_{i=\delta}^{n\ell} \frac{A_i}{\sigma^{2i}}, \tag{2.32}$$

where $n\ell$ is the length of the extended code over $\mathbb{F}_2$ and A_i is the number of codewords of weight i (in the dual of the code employed to mask the information). Hence the lexicographical order of the A_i to compare the amount of leakage is indeed associated with the masking code.

2.2.3.1 Connecting with Attacks

When evaluating with side-channel attacks, particularly in the optimal multivariate attacks (using higher-order optimal distinguishers) [BGHR14], the weight distribution also plays a significant role. More precisely, we have the following conjecture.

Conjecture 2.2.2. The success rate of optimal multivariate attack is determined by the weight distributions of the corresponding codes in the code-based masking.

2.2.4 Numerical Results

In the following, we consider a typical case of GCM by setting the generator matrices of the two codes $\mathcal{C}$ and $\mathcal{D}$ as follows:

$$\mathcal{G}_{\mathcal{C}} = \begin{pmatrix} 1 & 0 \end{pmatrix}, \tag{2.33}$$

$$\mathcal{G}_{\mathcal{D}} = \begin{pmatrix} \alpha_1 & \alpha_2 \end{pmatrix} = \begin{pmatrix} \alpha^i & \alpha^j \end{pmatrix}. \tag{2.34}$$

Clearly, the code $\mathcal{D}$ is an MDS code of parameters $[2, 1, 2]$. Considering equivalent linear codes over $\mathbb{F}_{2^8}$, we can fix $\alpha^j = 1$ in $\mathcal{G}_{\mathcal{D}}$. Hence there are only 254 candidates for the second element in $\mathcal{G}_{\mathcal{D}}$, corresponding to 254 linear codes.

As a common setting in side-channel analysis, we take the Hamming weight leakage model with the Gaussian noise. The setup is shown in Figure 2.3 in a communication channel viewpoint. Considering different bases, we launch an information-theoretic evaluation on all linear codes under different noise levels. The results are shown in Figure 2.5, 2.6 and 2.7 for the three bases, respectively. In particular, we add Figure 2.5(a) for the purpose of comparison, which illustrates the effectiveness of our lexicographical order based sorting of all codes.

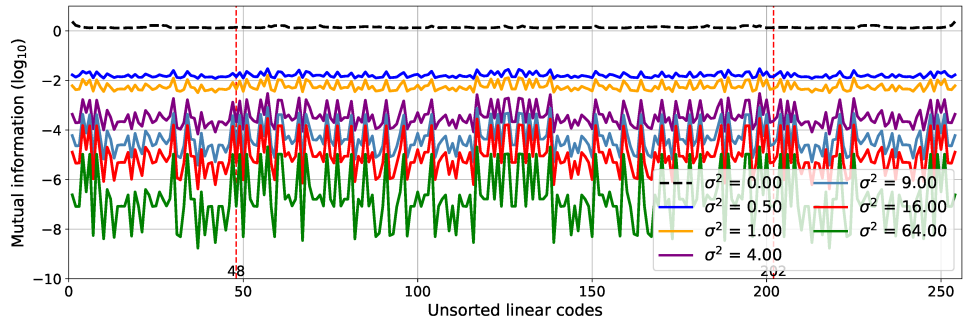
Note that the two vertical red dashed lines are for indicating the different dual distances $\delta_{\mathcal{D}^\perp} \in \{2, 3, 4\}$. For instance in Figure 2.5(b), the first vertical line marked 48 means there are 48 linear codes with $\delta_{\mathcal{D}^\perp} = 4$, and $202 - 48 = 154$ linear codes with $\delta_{\mathcal{D}^\perp} = 3$, and remaining 52 linear codes with $\delta_{\mathcal{D}^\perp} = 2$.

An interesting observation from Figure 2.5, 2.6 and 2.7 is, the bases have a significant impact on the distribution of linear codes. The mutual information increases (in most cases, except for some local minima) with the code lexicographic order on their weight enumeration polynomial. This justifies Conjecture 2.2.1. However, the number of exceptions (local minima) decreases when the noise increases, and the curves become indeed strictly increasing. Particularly, the first basis $\mathcal{B}_0$ gives the best weight distribution among the three bases, which will be investigated further in the next subsection.

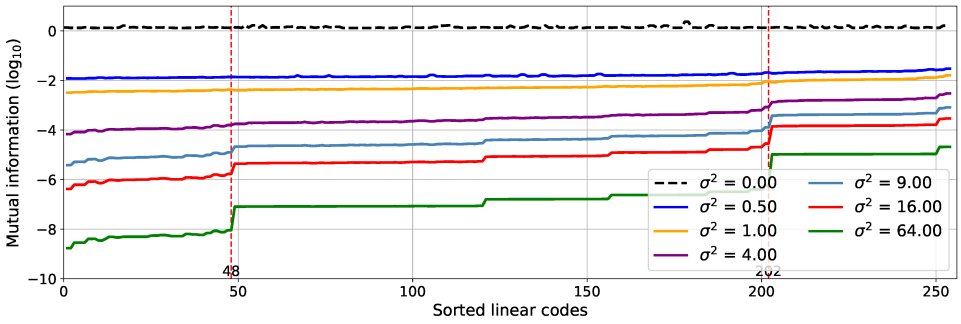
2.2.5 Classifying Linear Codes

In order to find the best weight distributions under different bases, we classify linear codes as in Table 2.6. Specifically, in Table 2.6, we first show the distribution of the minimum distance of all 254 linear codes under the three bases, and then present the best weight distribution in the last column. The takeaway point for the three bases is that the basis has a significant impact on the distribution of the minimum distances. Under condition of the prefix-based lexicographical order of weight distribution (Definition 2.2.3), we focus on the number of codes with the minimum distance equal to 4, resulting that $\mathcal{B}_2$ gives more codes with $\delta = 4$ (among the three cases). On the contrary, the first basis $\mathcal{B}_0$ gives the best weight distribution among all three bases where $A_4 = 2$.

Secondly, we randomly generate 1,000,000 linear codes over $\mathbb{F}_2$ by fixing $n = 16$ and $k = 8$ for comparison. The distribution of the minimum distances are listed in the fourth row of Table 2.6. One interesting observation is that this random approach gives a better weight distribution than all three bases over $\mathbb{F}_{2^8}$.



(a) Linear codes without sorting.



(b) Sorted linear codes in the lexicographical order.

Figure 2.5: Information-theoretic evaluation of all 254 candidates under the trace-orthogonal basis $\mathcal{B}_0$.

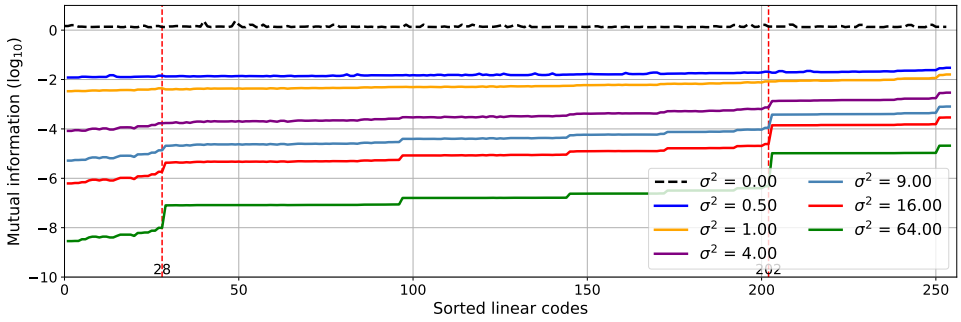


Figure 2.6: Information-theoretic evaluation of all 254 candidates under the trace-orthogonal basis $\mathcal{B}_1$ sorted in the lexicographical order.

However, all above cases do not recover the best known linear code (BKLC) given $n = 16$ and $k = 8$. We know that there is a unique linear code with parameters $[16, 8, 5]$, which has the minimum distance equal to 5 [CGC+21b]. Among all linear codes over $\mathbb{F}_2$, this BKLC code gives us the best weight distribution according to our lexicographical sorting, since it has $A_4 = 0$ while $A_4 > 0$ for other cases. From a perspective of side-channel analysis, this BKLC code provides us a masking code with the bit-level security order

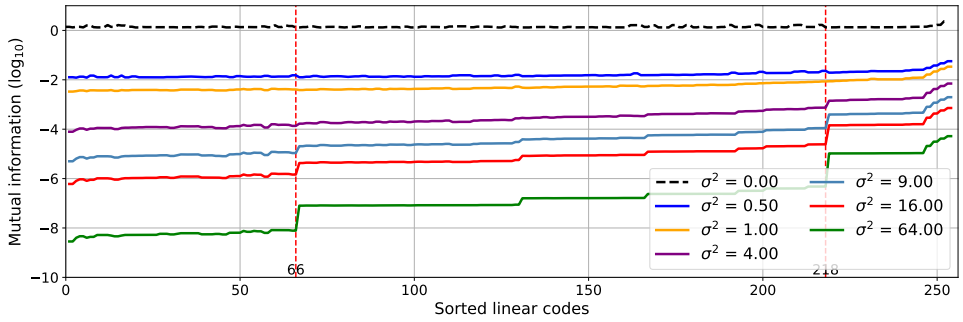


Figure 2.7: Information-theoretic evaluation of all 254 candidates under the default basis $\mathcal{B}_2$ sorted in the lexicographical order.

Table 2.6: Classifying linear codes under different bases. Note that the float number in parenthesis is the ratio between the number of codes in a class and the total number of candidates.

	Subfield	Number of linear codes with different δ					Best weight distribution
		$\#\{\delta=1\}$	$\#\{\delta=2\}$	$\#\{\delta=3\}$	$\#\{\delta=4\}$	$\#\{\delta=5\}$	
$\mathcal{B}_0$	$\mathbb{F}_{2^8} \to \mathbb{F}_2$	0	52 (0.2047)	154 (0.6063)	48 (0.1890)	0	[**1, 0, 0, 0, 2**, 22, 40, 44, 45, 40, 32, 20, 8, 2, 0, 0, 0]
$\mathcal{B}_1$	$\mathbb{F}_{2^8} \to \mathbb{F}_2$	0	52 (0.2047)	174 (0.6850)	28 (0.1102)	0	[**1, 0, 0, 0, 3**, 21, 38, 46, 45, 40, 34, 18, 7, 3, 0, 0, 0]
$\mathcal{B}_2$	$\mathbb{F}_{2^8} \to \mathbb{F}_2$	0	36 (0.1417)	152 (0.5984)	66 (0.2598)	0	[**1, 0, 0, 0, 4**, 22, 35, 42, 47, 46, 36, 14, 4, 4, 1, 0, 0]
Random codes	$\mathbb{F}_2$	60688 (0.0607)	357539 (0.3575)	528070 (0.5281)	53703 (0.0537)	0	[**1, 0, 0, 0, 1**, 23, 42, 42, 45, 40, 30, 22, 9, 1, 0, 0, 0]
BKLC	$\mathbb{F}_2$	0	0	0	0	1	[**1, 0, 0, 0, 0**, 24, 44, 40, 45, 40, 28, 24, 10, 0, 0, 0, 0]

$t = \delta_{\mathcal{D}^\perp} - 1 = 4$, that is higher than all other linear codes. Unfortunately, this code cannot be constructed by the subfield extension approach from $\mathbb{F}_{2^8}$ to $\mathbb{F}_2$ (e.g., by using bases like $\mathcal{B}_i$ for $i \in \{0, 1, 2\}$). This is also the reason why the direct sum masking can be better than the inner product masking in the sense of side-channel resistance [CG18, CGC+21b].

Evaluation of the best weight distributions under different bases. In Table 2.6, we present five best cases of the weight distribution. In order to have a fair comparison, we launch an information-theoretic evaluation by using mutual information. The results are depicted in Figure 2.8.

As shown in Figure 2.8, the main observation is that our lexicographical order-based sorting still works when comparing linear codes extended by using different bases. Note that for the best weight distribution under $\mathcal{B}_1$ and $\mathcal{B}_2$, the curve for $\mathcal{B}_1$ is slightly higher than that of $\mathcal{B}_2$. The reason is that other elements (e.g., $A_{\delta+1}, A_{\delta+2}$, etc.) in the weight

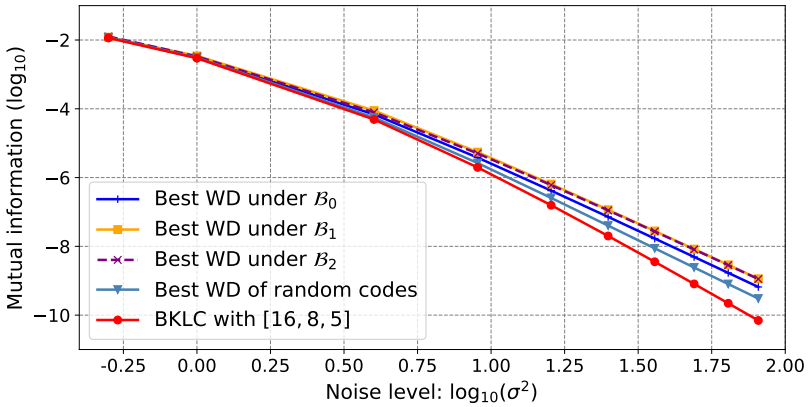


Figure 2.8: Information-theoretic evaluation of the best weight distributions (WD) under different bases as shown in Table 2.6.

distribution under $\mathcal{B}_1$ have more impact on mutual information.

CHAPTER 3

Side-Channel Leakage Evaluation

This chapter consists of two sections, the results in the first section have been published in "On Conditional Alpha-Information and its Application to Side-Channel Analysis" [LCGR21] and part of "Maximal Leakage of Masked Implementations Using Mrs. Gerber's Lemma for Min-Entropy" [BLR+23]; the second section has been published in "Side-Channel Information Leakage of Code-Based Masked Implementations" [CRL+22].

Contents

Abstract

The first section evaluates the side-channel leakage of a cryptographic device without countermeasures. To assess the leakage more precisely, we introduce a novel metric called conditional Sibson's α-information, and derive its corresponding Fano's inequality: The equality holds when $\alpha = \infty$. Based on the simulation results, this novel metric quantify the leakage more accurately than mutual information when $\alpha > 1$.

The second section evaluates the Side-Channel Leakage of Code-Based Masked Implementations using mutual information between sensitive variable and the leakage. It is proved that exploitable information necessitates the attacker to deploy a number of probes that is at least equivalent to the dual distance of the masking code, irrespective of their specific placements.

3.1 Side-Channel Leakage of Unprotected Implementations

In Section 1.1.2, the following side-channel model has been introduced:

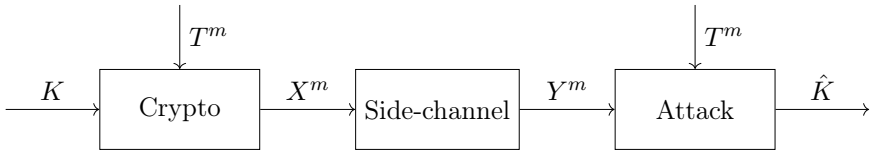


Figure 3.1: Side-channel seen as a communication channel (without masking).

where K is the secret key with $|\mathcal{K}| = N$, T is the public variable, X is sensitive variable determined by K and T, Y is the side-channel leakage, and m is the number of measurements. For conciseness, we use $\boldsymbol{X}$ to represent an m-dimensional vector $X^m = (X_1, X_2, \ldots, X_m)$, and similarly, $\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_m)$ and $\boldsymbol{T} = (T_1, T_2, \ldots, T_m)$.

The guessed key $\hat{K}$ is estimated from $\boldsymbol{Y}, \boldsymbol{T}$ using the MAP rule, with (maximal) probability of success

$$\mathbb{P}_s = \mathbb{P}(K = \hat{K}|\boldsymbol{Y}, \boldsymbol{T}) = \mathbb{E}_{\boldsymbol{YT}} \sup_k p_{K|\boldsymbol{Y},\boldsymbol{T}}(k|\boldsymbol{y} = \boldsymbol{Y}, \boldsymbol{t} = \boldsymbol{T}). \tag{3.1}$$

In [dCGRP19b, dCGRP19a], Chérisey et al. use classical information-theoretic tools (such as mutual information, conditional entropy) to evaluate the side-channel leakage of a cryptographic device. Leveraging Fano's inequality, the authors manage to forge a connection between mutual information and the probability of success $\mathbb{P}_s$. To be more specific, they derived the following theorem:

Theorem 3.1.1 (Fano Inequality for Conditional MI [dCGRP19a, Lemma 2])**.**

$$I(\boldsymbol{X}; \boldsymbol{Y}|\boldsymbol{T}) \geq d\big(\mathbb{P}_s \| \tfrac{1}{N}\big) \tag{3.2}$$

where $\frac{1}{N}$ *means a binary distribution whose probabilities are* $\frac{1}{N}$ *and* $\frac{N-1}{N}$*;* $d(p\|q)$ *denotes binary divergence:*

$$d(p\|q) = p \log \frac{p}{q} + (1-p) \log \frac{1-p}{1-q}. \tag{3.3}$$

In (3.2), $d(\mathbb{P}_s\|\frac{1}{N})$ is an increasing function of $\mathbb{P}_s$ when $\mathbb{P}_s \geq \frac{1}{N}$ ($\mathbb{P}_s$ falls within this range, as the (maximal) probability of success invariably exceeds that of blind guessing). Therefore the mutual information between the input and the output of the side-channel $I(\boldsymbol{X}; \boldsymbol{Y}|\boldsymbol{T})$, sets an upper limit for the success rate of the attack. Using Monte Carlo simulation, one can evaluate the value of $I(\boldsymbol{X}; \boldsymbol{Y}|\boldsymbol{T})$.

However, as shown in Figure 3 of [dCGRP19a], the inequality (3.2) has some room for improvement. Considering that the choice of metric may be one of the causes for this gap, we attempt to employ more general tools: α-information theoretic measures. More specifically, we aim at extending the approach of [dCGRP19b, dCGRP19a] to α-information quantities depending on a parameter α. For that we need the following ingredients that were crucial in the derivation steps of [dCGRP19b, dCGRP19a]:

- a *closed-form expression* of conditional mutual information, amenable to efficient numerical estimation;
- a *data processing inequality* of conditional mutual information over a "conditional" Markov chain for a given plain or cypher text T (known to the attacker);
- a *expansion property* of conditional mutual information, i.e., its decomposition as a difference between conditional entropies, valid at least when the secret is assumed *uniformly distributed*;
- a *Fano inequality* which yields a lower bound on mutual information that depends on the probability of success (or equivalently on the probability of error).

Our aim, therefore, is to establish all of these properties for a suitably defined conditional Rényi version of mutual information of order $\alpha > 0$.

The rest of this section is organized as follows. Section 3.1.1 proposes a natural definition of conditional α-information satisfying the required properties and Section 3.1.2 makes a detailed comparison to previous proposals. Section 3.1.3 presents the main result applied to side-channel analysis, which is then validated using simulations in Section 3.1.4 .

3.1.1 Conditional Alpha-Information

As a natural continuation of the generalized definitions in Section 1.3, we define the conditional α-information with a "log-expectation" closed-form expression, obtained by taking the expectation over the conditional variable inside the logarithm in the expression of Sibson's (unconditional) α-information (1.41):

Definition 3.1.1 (Conditional α-Information, Closed-Form Definition)**.**

$$\begin{aligned} I_\alpha(X;Y|Z) &= \tfrac{\alpha}{\alpha-1} \log \mathbb{E}_Z \mathbb{E}_{Y|Z} \langle p_{X|YZ} \| p_{X|Z} \rangle_\alpha \\ &= \tfrac{\alpha}{\alpha-1} \log \mathbb{E}_{YZ} \langle p_{X|YZ} \| p_{X|Z} \rangle_\alpha. \end{aligned} \tag{3.4}$$

where

$$\langle p \| q \rangle_\alpha = \Big(\int p^\alpha q^{1-\alpha} d\mu \Big)^{1/\alpha}. \tag{3.5}$$

The notations follow from Section 1.3. To the best of our knowledge, this definition has not been considered elsewhere.

3.1.1.1 Basic Properties

Our definition enjoys three important properties: consistency, uniform expansion property (UEP) and data processing inequality (DPI).

Property 3.1.1 (Consistency of Conditional α-Information w.r.t. α-Information)**.** If Z is independent of (X, Y) then $I_\alpha(X;Y|Z) = I_\alpha(X;Y)$.

Proof. Obvious from the definitions. □

Property 3.1.2 (UEP for Conditional α-Information)**.** If $U \sim \mathcal{U}(N)$ is uniformly distributed independent of Z, then

$$I_\alpha(U;Y|Z) = H_\alpha(U) - H_\alpha(U|YZ) = \log N - H_\alpha(U|YZ). \tag{3.6}$$

Proof. Similarly as for the preceding UEPs, we have $\langle p_{U|YZ} \| u \rangle_\alpha = N^{\frac{\alpha-1}{\alpha}} \| p_{U|YZ} \|_\alpha$. Averaging over (Y, Z) and taking the logarithm gives the announced formula. □

We say that a sequence of random variables forms a *conditional Markov chain* given some random variable T if it is Markov for any given $T = t$.

Property 3.1.3 (DPI for Conditional α-Information)**.** If $W - X - Y - Z$ forms a conditional Markov chain given T, then

$$I_\alpha(X;Y|T) \geq I_\alpha(W;Z|T). \tag{3.7}$$

Proof. By DPI of Sibson's α-information, $I_\alpha(X;Y|T=t) \geq I_\alpha(W;Z|T=t)$ for any t. From closed-form expression of Sibson's α-information this gives $\langle p_{X|Y,T=t} \| p_{X|T=t} \rangle_\alpha \geq \langle p_{W|Z,T=t} \| p_{W|T=t} \rangle_\alpha$ for $\alpha > 1$ and the opposite inequality for $0 < \alpha < 1$. This in turn from Definition 3.1.1 gives the announced inequality for any α. □

3.1.1.2 Conditional Sibson's Identity

Proposition 3.1.1 (Conditional Sibson's Identity)**.** One has

$$D_\alpha(P_{XYZ} \| P_{X|Z} Q_{YZ}) = D_\alpha(Q^*_{YZ} \| Q_{YZ}) + I_\alpha(X;Y|Z), \tag{3.8}$$

hence the following alternate minimizing definition:

$$I_\alpha(X;Y|Z) = \min_{Q_{YZ}} D_\alpha(P_{XYZ} \| P_{X|Z} Q_{YZ}) \tag{3.9}$$

Proof. Similarly as in the case of α-information, it is straightforward to compute

$$\langle p_{XYZ} \| p_{X|Z} q_{YZ} \rangle_\alpha^\alpha = \iiint p_{YZ}^\alpha p_{X|YZ}^\alpha p_{X|Z}^{1-\alpha} q_{YZ}^{1-\alpha} d\mu \tag{3.10}$$

$$= \left\langle p_{YZ} \langle p_{X|YZ} \| p_{X|Z} \rangle_\alpha \,\|\, q_{YZ} \right\rangle_\alpha^\alpha \tag{3.11}$$

Defining the (suitably normalized) distribution

$$q^*_{YZ} = \frac{p_{YZ}\langle p_{X|YZ}\|p_{X|Z}\rangle_\alpha}{\mathbb{E}_{YZ}\langle p_{X|YZ}\|p_{X|Z}\rangle_\alpha}, \tag{3.12}$$

substituting and taking the logarithm gives the announced identity. □

3.1.1.3 Conditional Max-Information

Proposition 3.1.2 (Conditional Max-Information)**.** Assuming X, Y take values in finite alphabets, one has

$$I_\infty(X;Y|Z) = \log \mathbb{E}_Z \int_y \big(\max_{x:p_{X|Z}(x|z)>0} p_{Y|XZ}\big)\, d\mu_Y. \tag{3.13}$$

This result easily follows from the following Lemmas 3.1.2 and 3.1.3.

Lemma 3.1.2. *Assume X takes values in finite alphabets. Given any fixed y, z, we have*

$$\lim_{\alpha\to\infty}\ p_{Y|Z}\cdot\langle p_{X|YZ}\|p_{X|Z}\rangle_\alpha = \max_{x:p_{X|Z}(x|z)>0} p_{Y|XZ}. \tag{3.14}$$

Proof. **[Method 1]**: By Theorem 6 of [vEH14] we have

$$\lim_{\alpha\to\infty}\langle p_{X|YZ}\|p_{X|Z}\rangle_\alpha = \exp\big(D_\infty(P_{X|YZ}\|P_{X|Z})\big) = \max_{x:p_{X|Z}(x|z)>0}\frac{p_{X|YZ}}{p_{X|Z}}. \tag{3.15}$$

Because $p_{Y|Z}\cdot p_{X|YZ}/p_{X|Z} = p_{Y|XZ}$, the proof is finished.

[Method 2]: We use L^∞-norm to prove this lemma.

$$\begin{aligned} p_{Y|Z}\langle p_{X|YZ}\|p_{X|Z}\rangle_\alpha &= p_{Y|Z}\ \Big(\sum_{x\in\mathcal{X}} p^\alpha_{X|YZ}\ p^{1-\alpha}_{X|Z}\Big)^{\frac{1}{\alpha}} \\ &= \Big(\sum_{x\in\mathcal{X}} p^\alpha_{XY|Z}\ p^{1-\alpha}_{X|Z}\Big)^{\frac{1}{\alpha}} = \Big(\sum_{x\in\mathcal{X}}\big(p_{XY|Z}\ p^{\frac{1-\alpha}{\alpha}}_{X|Z}\big)^\alpha\Big)^{\frac{1}{\alpha}} \\ &= \Big(\sum_{x\in\mathcal{X}}\big(p_{Y|XZ}\ p^{\frac{1}{\alpha}}_{X|Z}\big)^\alpha\Big)^{\frac{1}{\alpha}}. \end{aligned} \tag{3.16}$$

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

$$p_{Y|XZ} - \varepsilon \le p_{Y|XZ}\ p^{\frac{1}{\alpha}}_{X|Z} \le p_{Y|XZ}. \tag{3.17}$$

Because $\mathcal{X}$ is finite, one always has a sufficiently large $\alpha > 0$ such that (3.17) holds for any $x \in \mathcal{X}$. By L^∞-norm we have

$$\lim_{\alpha\to\infty}\Big(\sum_{x:p_{X|Z}(x|z)>0}\big(p_{Y|XZ}-\varepsilon\big)^\alpha\Big)^{\frac{1}{\alpha}} = \max_{x:p_{X|Z}(x|z)>0} p_{Y|XZ} - \varepsilon \tag{3.18}$$

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

Lemma 3.1.3. *Assume Y takes values in finite alphabets. One has*

$$\lim_{\alpha\to\infty} \log\ \mathbb{E}_{YZ}\langle p_{X|YZ}\|p_{X|Z}\rangle_\alpha = \log \mathbb{E}_Z \int_y \lim_{\alpha\to\infty}\ p_{Y|Z}\cdot\langle p_{X|YZ}\|p_{X|Z}\rangle_\alpha\ d\mu.$$

Proof. By definition we have

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

where $\log \mathbb{E}_{YZ}\langle p_{X|YZ}\|p_{X|Z}\rangle_\alpha$ is bounded because $\mathbb{E}_{YZ}\langle p_{X|YZ}\|p_{X|Z}\rangle_\alpha \le |\mathcal{Y}|$ when $\alpha > 1$.

Because $\frac{\alpha-1}{\alpha} D_\alpha(p_{X|YZ}\|p_{X|Z})$ is increasing in α when $\alpha > 1$, this lemma follows from the monotone convergence theorem. □

Remark 3.1.1. In [IWK20], conditional maximal leakage is defined as a maximum over Z, while our conditional max-information is averaged over Z—which is always no larger than the conditional maximal leakage of [IWK20].

3.1.2 Comparison to Previous Definitions

All previous definitions of conditional α-information we are aware of are variations of the form (3.9) where α-divergence is minimized with respect to different probability measures $Q_{X|Z}$, $Q_{Y|Z}$, Q_Z or combinations. There are exactly $2^3 = 8$ possibilities:

$$\text{(o)}\quad I_\alpha^{000}(X;Y|Z) = \quad D_\alpha(P_{XYZ}\|P_{X|Z}P_{Y|Z}P_Z).$$

$$\text{(i)}\quad I_\alpha^{001}(X;Y|Z) = \min_{Q_Z} D_\alpha(P_{XYZ}\|P_{X|Z}P_{Y|Z}Q_Z).$$

$$\text{(ii)}\quad I_\alpha^{010}(X;Y|Z) = \min_{Q_{Y|Z}} D_\alpha(P_{XYZ}\|P_{X|Z}Q_{Y|Z}P_Z).$$

$$\text{(iii)}\quad I_\alpha^{011}(X;Y|Z) = \min_{Q_{YZ}} D_\alpha(P_{XYZ}\|P_{X|Z}Q_{YZ}).$$

$$\text{(iv)}\quad I_\alpha^{100}(X;Y|Z) = \min_{Q_{X|Z}} D_\alpha(P_{XYZ}\|Q_{X|Z}P_{Y|Z}P_Z).$$

$$\text{(v)}\quad I_\alpha^{101}(X;Y|Z) = \min_{Q_{XZ}} D_\alpha(P_{XYZ}\|Q_{XZ}P_{Y|Z}).$$

$$\text{(vi)}\quad I_\alpha^{110}(X;Y|Z) = \min_{Q_{X|Z}Q_{Y|Z}} D_\alpha(P_{XYZ}\|Q_{X|Z}Q_{Y|Z}P_Z).$$

$$\text{(vii)}\quad I_\alpha^{111}(X;Y|Z) = \min_{Q_{X|Z}Q_{YZ}} D_\alpha(P_{XYZ}\|Q_{X|Z}Q_{YZ}).$$

Definition (o) is mentioned in [TH18, Eq. (70)]. Definition (i) is the main proposal of Esposito et al. [EWG21]. Definition (ii) is discussed by Tomamichel and Hayashi [TH18, Eq. (74)] and is equivalent to definition (iv) by permuting the roles of X and Y:

$$I_\alpha^{100}(X;Y|Z) = I_\alpha^{010}(Y;X|Z). \tag{3.19}$$

Our definition (3.9) is definition (iii), and is equivalent to definition (v) by permuting the roles of X and Y:

$$I_\alpha^{101}(X;Y|Z) = I_\alpha^{011}(Y;X|Z). \tag{3.20}$$

Finally, definitions (vi) and (vii) seem new and related to a conditional version of the Lapidoth-Pfister mutual information [LP16]:

$$J_\alpha(X;Y) = \min_{Q_X Q_Y} D_\alpha(P_{XY} \parallel Q_X Q_Y). \tag{3.21}$$

Thus we need only to compare our definition to (o), (i), (ii), (vi) and (vii).

We now discuss various properties for these definitions, by decreasing order of importance: The fact that they admit or not a closed-form expression in terms of the involved probability densities; their consistency with respect to α-information $I_\alpha(X;Y|0) = I_\alpha(X;Y)$; the existence of a uniform expansion of the form $I_\alpha(U;Y|Z) = \log N - H_\alpha(U|YZ)$ when $U \sim \mathcal{U}(N)$ is independent of Z; and the fact that they satisfy data processing inequalities for conditional Markov chains.

Closed-Form and Consistency

Definition (o) is by itself a closed-form expression but is clearly *inconsistent* with respect to Sibson's α-information since $I_\alpha^{000}(X;Y|0) = D_\alpha(P_{XY}\|P_XP_Y)$ which by (1.47) is $\geq I_\alpha(X;Y)$ where the inequality is, in general, strict.

Definition (i) of Esposito et al. does admit a closed-form expression [EWG21, Thm. 2]. In fact, since

$$\begin{aligned}\langle p_{XYZ}\|p_{X|Z}p_{Y|Z}q_Z\rangle_\alpha^\alpha &= \iiint p_Z^\alpha p_{XY|Z}^\alpha (p_{X|Z}p_{Y|Z})^{1-\alpha} q_Y^{1-\alpha}\, d\mu \\ &= \big\langle p_Z \langle p_{XY|Z}\|p_{X|Z}p_{Y|Z}\rangle_\alpha \,\|\, q_Z\big\rangle_\alpha^\alpha,\end{aligned}$$

letting

$$q_Z^* = \frac{p_Z \langle p_{X|YZ}\|p_{X|Z}p_{Y|Z}\rangle_\alpha}{\mathbb{E}_Z \langle p_{X|YZ}\|p_{X|Z}p_{Y|Z}\rangle_\alpha} \tag{3.22}$$

and taking the logarithm gives the following variation of Sibson's identity (whose existence is mentioned but does not explicitly appear in [EWG21]):

Proposition 3.1.3.

$$D_\alpha(P_{XYZ}\|P_{X|Z}P_{Y|Z}Q_Z) = D_\alpha(Q_Z^*\|Q_Z) + I_\alpha^{001}(X;Y|Z), \tag{3.23}$$

with the following closed-form expression:

$$I_\alpha^{001}(X;Y|Z) = \tfrac{\alpha}{\alpha-1} \log \mathbb{E}_Z \langle p_{XY|Z}\|p_{X|Z}p_{Y|Z}\rangle_\alpha. \tag{3.24}$$

However, I_α^{001} is *inconsistent* (with respect to Sibson's α-information) for the same reason as in the case of I_α^{000}: From (3.24) we have

$$I_\alpha^{001}(X;Y|0) = D_\alpha(P_{XY}\|P_XP_Y) \geq I_\alpha(X;Y). \tag{3.25}$$

Definition (ii) of Tomamichel and Hayashi also admits a closed-form expression [TH18, Eq. (75)]. In fact by the (unconditional) Sibson identity (1.46) applied to all variables conditioned on $Z = z$ for any z, one easily sees that $D_\alpha(P_{XYZ}\|P_{X|Z}Q_{Y|Z}P_Z)$ achieves its minimum when for $q_{Y|Z} = q^*_{Y|Z} = p_{Y|Z}\langle p_{X|YZ}\|p_X\rangle_\alpha / \mathbb{E}_{Y|Z}\langle p_{X|YZ}\|p_{X|Z}\rangle_\alpha$ as given above in the proof of Proposition. (3.1.1), which gives

$$I_\alpha^{010}(X;Y|Z) = \tfrac{1}{\alpha-1}\log \mathbb{E}_Z(\mathbb{E}_{Y|Z}\langle p_{X|YZ}\|p_{X|Z}\rangle_\alpha)^\alpha. \tag{3.26}$$

From this it follows that $I_\alpha^{010}(X;Y|0) = I_\alpha(X;Y)$, proving that I_α^{010} is *consistent*.

Finally, definitions (vi) and (vii) are neither closed-form nor consistent; for when $Z \equiv 0$, the definitions reduce to the Lapidoth-Pfister mutual information:

$$J_\alpha(X;Y) = \min_{Q_XQ_Y} D_\alpha(P_{XY} \parallel Q_XQ_Y) \tag{3.27}$$

which already does not admit a closed-form expression, and for which $J_\alpha(X;Y) \leq I_\alpha(X;Y)$ where the inequality is, in general, strict [LP16]. In the following we focus on the other definitions which admit closed-form expressions.

Uniform Expansion Property

The uniform expansion property (UEP) is a crucial requirement in our subsequent derivations (Theorem 3.1.4). It is naturally satisfied for α-information (Property 1.3.7) and it is important that it is also satisfied for its conditional version.

Using the above closed-form expressions it is easy to check the UEP when $U \sim \mathcal{U}(N)$ is independent of Z, neither $I_\alpha^{000}(U;Y|Z)$, nor $I_\alpha^{001}(U;Y|Z)$, nor $I_\alpha^{010}(U;Y|Z)$ equals $\log N - H_\alpha(U|YZ)$. This is not surprising since in general, from the different minimizations of α-divergence,

$$\begin{aligned} I_\alpha(X;Y|Z) &= I_\alpha^{011}(X;Y|Z) \\ &\leq \min\{I_\alpha^{001}(X;Y|Z), I_\alpha^{010}(X;Y|Z)\} \\ &\leq I_\alpha^{000}(X;Y|Z) \end{aligned} \tag{3.28}$$

where inequalities are, in general, strict. Hence the only case where the UEP (which is crucial in our subsequent derivations) holds is for the definition (iii) proposed in this paper.

Data Processing Inequality

Finally, since definitions (o) and (i) are inconsistent with $I_\alpha^{000}(X;Y|0) = I_\alpha^{001}(X;Y|0) = D_\alpha(P_{XY}\|P_XP_Y)$, they do not even satisfy data processing inequalities for a unconditional Markov chain. Therefore, the only remaining candidate for DPI is definition (ii).

Property 3.1.4 (DPI for $I_\alpha^{010}(X;Y|Z)$)**.** If $W - X - Y - Z$ forms a conditional Markov chain given T, then $I_\alpha^{010}(X;Y|T) \geq I_\alpha^{010}(W;Z|T)$.

Proof. Let $P_{X,Y,T} \to \boxed{P_{X,Z,T|X,Y,T}} \to P_{X,Z,T} \to \boxed{P_{W,Z,T|X,Z,T}} \to P_{W,Z,T}$. By the conditional Markov condition, we have $P_{X,Z,T|X,Y,T} = P_{X,T|X,T}P_{Z|X,Y,T} = P_{X,T|X,T}P_{Z|Y,T}$ where $P_{X,T|X,T}$ is the identity operator; similarly $P_{W,Z,T|X,Z,T} = P_{W|X,Z,T}P_{Z,T|Z,T} = P_{W|X,T}P_{Z,T|Z,T}$. Thus if $Q_{Y|T} \to \boxed{P_{Z|Y,T}} \to Q_{Z|T}$, we find $P_{X|T}Q_{Y|T}P_T \to \boxed{P_{X,Z,T|X,Y,T}} \to P_{X|T}Q_{Z|T}P_T \to \boxed{P_{W,Z,T|X,Z,T}} \to P_{W|T}Q_{Z|T}P_T$. By the data processing inequality for α-divergence (Property 1.3.2), $D_\alpha(P_{X,Y,T}\|P_{X|T}Q_{Y|T}P_T) \geq D_\alpha(P_{W,Z,T}\|P_{W|T}Q_{Z|T}P_T) \geq I_\alpha(W;Z|T)$. Minimizing over $Q_{Y|T}$ gives the announced DPI. □

Table 3.1: Comparison of some properties for the various definitions.

Definition	Ref.	Closed-form	Consistency	UEP	DPI
o	[TH18]	yes	no	no	no
i	[EWG21]	yes	no	no	no
ii,iv	[TH18]	yes	yes	no	yes
iii,v	(this thesis)	**yes**	**yes**	**yes**	**yes**
vi,vii	—	no	no		

Table 1.1 summarizes the comparison between properties of (o)–(vii).

3.1.3 Fano Inequality for Conditional Alpha-Information

Now, the generalized Fano's inequality for conditional alpha-information can be derived as follows.

Theorem 3.1.4. *One has the following upper bound on the probability of success $\mathbb{P}_s$:*

$$I_\alpha(\boldsymbol{X},\boldsymbol{Y}|\boldsymbol{T}) = I_\alpha(K,\boldsymbol{Y}|\boldsymbol{T}) \geq d_\alpha(\mathbb{P}_s\|\frac{1}{N}) \tag{3.29}$$

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

$$d_\alpha(p\|q) = \tfrac{1}{\alpha-1}\log\big(p^\alpha q^{1-\alpha} + (1-p)^\alpha(1-q)^{1-\alpha}\big). \tag{3.30}$$

Proof. The chain $K - \boldsymbol{X} - \boldsymbol{Y}$ is Markov given $\boldsymbol{T}$ by assumption but since $\boldsymbol{X} = f(K,\boldsymbol{T})$, the chain $\boldsymbol{X} - K - \boldsymbol{Y}$ is also Markov given $\boldsymbol{T}$. Therefore, by the conditional DPI (Property 3.1.3), $I_\alpha(\boldsymbol{X},\boldsymbol{Y}|\boldsymbol{T}) = I_\alpha(K,\boldsymbol{Y}|\boldsymbol{T})$ (inequalities in both directions). Now since $K - \boldsymbol{Y} - \hat{K}$ is also Markov given $\boldsymbol{T}$, we have $I_\alpha(K;\boldsymbol{Y}|\boldsymbol{T}) \geq I_\alpha(K;\hat{K}|\boldsymbol{T})$. Since K is

equiprobable independent of $\boldsymbol{T}$, by the UEP (Property 3.1.2), $I_\alpha(K;\hat{K}|\boldsymbol{T}) = \log N - H_\alpha(K|\hat{K},\boldsymbol{T}) \geq \log N - H_\alpha(K|\hat{K}) = I_\alpha(K;\hat{K})$. Finally, using Lemma 1.3.1, $I_\alpha(K;\hat{K}) \geq d_\alpha(\mathbb{P}_s(K|\boldsymbol{Y},\boldsymbol{T})\|\mathbb{P}_s(K)) = d_\alpha(\mathbb{P}_s\|\frac{1}{N})$, which proves (3.29). □

Remark 3.1.2. From (3.30), $d_\alpha(p\|q)$ is increasing in p when $p \geq q$. Hence (3.29) gives an upper bound on $\mathbb{P}_s$ (which is obviously $\geq 1/N$ since $\mathbb{P}_s = 1/N$ corresponds to a blind guess when the attacker does not know Y).

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

Theorem 3.1.5 (Generalized Fano's Inequality at $\alpha = +\infty$)**.** *For a uniformly distributed secret K,*

$$\begin{aligned} I_\infty(\boldsymbol{X},\boldsymbol{Y}|\boldsymbol{T}) = I_\infty(K;\boldsymbol{Y}|\boldsymbol{T}) &= d_\infty(\mathbb{P}_s(K|\boldsymbol{Y},\boldsymbol{T})\|(\mathbb{P}_s(K))) \\ &= \log(N\mathbb{P}_s) \end{aligned}$$

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

Proof. Under the MAP rule, the probability of success writes

$$\begin{aligned} \mathbb{P}_s &= \mathbb{E}_{\boldsymbol{Y}\boldsymbol{T}}(\max_k\ p_{K|\boldsymbol{Y},\boldsymbol{T}}) \\ &= \mathbb{E}_{\boldsymbol{T}} \int_{\boldsymbol{y}} (\max_k\ p_{\boldsymbol{Y}|K,\boldsymbol{T}} p_{K|\boldsymbol{T}}) d\mu_{\boldsymbol{Y}}. \end{aligned} \tag{3.31}$$

Because the secret key K is uniformly distributed and independent from $\boldsymbol{T}$, Therefore, (3.31) becomes

$$\mathbb{P}_s = \frac{1}{N} \cdot \mathbb{E}_{\boldsymbol{T}} \int_{\boldsymbol{y}} (\max_k\ p_{\boldsymbol{Y}|K,\boldsymbol{T}}) d\mu_{\boldsymbol{Y}}. \tag{3.32}$$

Combining (3.13) and (3.32) we have $I_\infty(K;\boldsymbol{Y}|\boldsymbol{T}) = \log(N\mathbb{P}_s)$. Since $\mathbb{P}_s \geq 1/N$, one has $\mathbb{P}_s \cdot N \geq (1-\mathbb{P}_s)\cdot N/(N-1)$ and $d_\infty(\mathbb{P}_s(K|\boldsymbol{Y},\boldsymbol{T})\|(\mathbb{P}_s(K))) = \log(N\mathbb{P}_s)$. □

3.1.4 Numerical Simulations

We consider an implementation of the AES with a large number m of measurement traces. Here $N = 2^\ell = 256$ and the most commonly used leakage model is

$$Y_i = w_H(S(T_i \oplus K)) + \varepsilon_i \qquad (i = 1, 2, \ldots, m) \tag{3.33}$$

where w_H denotes the Hamming weight, S denotes a S-box permutation and ε_i are i.i.d $\sim \mathcal{N}(0, \sigma^2)$. Letting $\mathbf{X} = (X_i)_i$, $\mathbf{Y} = (Y_i)_i$, $\mathbf{T} = (T_i)_i$, both T_i and K belongs to $\mathbb{F}_{2^\ell}$. We next use Monte-Carlo simulation to compute $I_\alpha(\mathbf{X}, \mathbf{Y}|\mathbf{T}) = I_\alpha(K, \mathbf{Y}|\mathbf{T})$.

Recall that

$$\begin{aligned} I_\alpha(\mathbf{X},\mathbf{Y}|\mathbf{T}) = I_\alpha(K, \mathbf{Y}|\mathbf{T}) &= \tfrac{\alpha}{\alpha-1} \log \mathbb{E}_{\mathbf{YT}} \langle p_{K|\mathbf{YT}} \| p_{K|\mathbf{T}} \rangle_\alpha \\ &= \tfrac{\alpha}{\alpha-1} \log \Big(\int_{\boldsymbol{y}} \sum_{\mathbf{t}} p(\mathbf{y}, \mathbf{t}) \frac{\left(\sum_k p(k|\mathbf{t}) p^\alpha(\mathbf{y}|\mathbf{t}, k)\right)^{\frac{1}{\alpha}}}{p(\mathbf{y}|\mathbf{t})} \, \mathrm{d}\mu_{\mathbf{Y}} \Big). \end{aligned} \tag{3.34}$$

This value can be estimated by using Monte-Carlo simulation by the law of large numbers. Indeed, we have

$$\begin{aligned} \exp \tfrac{\alpha-1}{\alpha} I(\mathbf{X}, \mathbf{Y}|\mathbf{T}) &\approx \lim_{N_C \to \infty} \frac{1}{N_C} \sum_{j=1}^{N_C} \frac{\left(\sum_k p(k|\mathbf{t}^j) p^\alpha(\mathbf{y}^j|\mathbf{t}^j, k)\right)^{\frac{1}{\alpha}}}{p(\mathbf{y}^j|\mathbf{t}^j)} \\ &= \lim_{N_C \to \infty} \frac{1}{N_C} \sum_{j=1}^{N_C} \frac{\left(\sum_k p(k) p^\alpha(\mathbf{y}^j|\mathbf{t}^j, k)\right)^{\frac{1}{\alpha}}}{\sum_k p(k) p(\mathbf{y}^j|\mathbf{t}^j, k)}, \end{aligned} \tag{3.35}$$

where $\mathbf{t}^j \sim \mathcal{U}(\mathbb{F}_{2^\ell}^m)$ and $\mathbf{y}^j \sim \mathcal{N}(f(\mathbf{t}^j, k^j), \sigma^2 \mathbf{I}_m) \in \mathbb{R}^m$ by choosing $k^j \sim \mathcal{U}(\mathbb{F}_{2^\ell})$ and $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$.

By focusing on one draw $(\mathbf{t}, \mathbf{y})$ of Monte-Carlo simulation in (3.35), we have

$$\frac{\left(\sum_k p(k) p^\alpha(\mathbf{y}|\mathbf{t}, k)\right)^{\frac{1}{\alpha}}}{\sum_k p(k) p(\mathbf{y}|\mathbf{t}, k)} = p(k)^{\frac{1-\alpha}{\alpha}} \cdot \frac{\left(\sum_k p^\alpha(\mathbf{y}|\mathbf{t}, k)\right)^{\frac{1}{\alpha}}}{\sum_k p(\mathbf{y}|\mathbf{t}, k)} \tag{3.36}$$

$$= 2^{\ell \cdot \frac{\alpha-1}{\alpha}} \cdot \frac{\left(\sum_k \exp\left(-\frac{\alpha}{2\sigma^2} ||\mathbf{y} - f(\mathbf{t}, k)||_2\right)\right)^{\frac{1}{\alpha}}}{\sum_k \exp\left(-\frac{1}{2\sigma^2} ||\mathbf{y} - f(\mathbf{t}, k)||_2\right)}. \tag{3.37}$$

Considering independent Gaussian noise in each $\mathbf{y}^j$, we can simplify (3.35) and insert into (3.34), therefore,

$$I_\alpha(\mathbf{X}, \mathbf{Y}|\mathbf{T}) \approx \ell + \frac{\alpha}{\alpha - 1} \log \frac{1}{N_C} \sum_{j=1}^{N_C} \frac{\left(\sum_k p^\alpha(\mathbf{y}^j|\mathbf{t}^j, k)\right)^{\frac{1}{\alpha}}}{\sum_k p(\mathbf{y}^j|\mathbf{t}^j, k)} \tag{3.38}$$

$$= \ell + \frac{\alpha}{\alpha - 1} \log \frac{1}{N_C} \sum_{j=1}^{N_C} \frac{\left(\sum_k \exp\left(-\frac{\alpha}{2\sigma^2} ||\mathbf{y}^j - f(\mathbf{t}^j, k)||_2\right)\right)^{\frac{1}{\alpha}}}{\sum_k \exp\left(-\frac{1}{2\sigma^2} ||\mathbf{y}^j - f(\mathbf{t}^j, k)||_2\right)}, \tag{3.39}$$

given a larger enough N_C.

Figure 3.2 presents the numerical results of the *upper bounds* of success probability for diverse values of α, which compares them to the average performance of the optimal ML attack (with error bars).

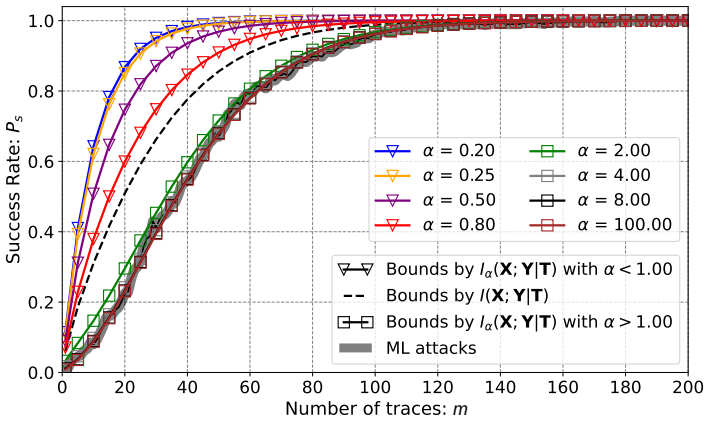


Figure 3.2: Comparison of upper bounds on success rate $\mathbb{P}_s$ given α-information $I_\alpha(\mathbf{X}, \mathbf{Y}|\mathbf{T})$ for different values of α.

With the increase in the value of α, the bound (3.29) becomes progressively tighter. And as we have already proven, the equality in (3.29) holds when $\alpha = \infty$. Compared to previous result [dCGRP19a] (represented by the black dashed line in the figure, when α=1), our generalized Fano's inequality has made significant improvements when $\alpha \geq 2$.

3.2 Side-Channel Leakage of Code-Based Masked Implementations

Let K, T, X represent the secret key, the public variable and sensitive variable respectively. The following model has been introduced in Section 1.2:

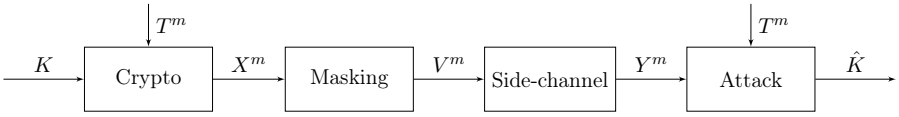


Figure 3.3: Side-channel seen as a communication channel (with masking).

In this section, the number of shares is denoted as n instead of $d+1$. This is because we will have plenty discussions about the masking code in this section, and using n as the notation aligns with the commonly used code parameter $[n, k]$.

To protect the key-dependent sensitive variable, X is masked into n shares: $V = (X_0, X_1, \ldots, X_{n-1})$. The side-channel leakage is denoted as $Y = (Y_0, Y_1, \ldots, Y_{n-1})$ where each Y_i deponding on X_i, $i = 0, 1, \ldots, n-1$. The attacker queries the cryptographic device m times and obtains corresponding traces Y^m. Finally, the attacker use Y^m and T^m to compute the guessed key $\hat{K}$.

For the code-based masking, as stated in Section. 1.2.1, randomness M is introduced in masking procedure and both X and M are multiplied by generator matrices of the masking codes. Let $\mathbb{F}_q$ be a finite field where q is a power of 2, and k, s be integers satisfying $k + s \leq n$. One has

$$V = X\mathcal{G}_{\mathcal{C}} + M\mathcal{G}_{\mathcal{D}}, \tag{3.40}$$

where $X \sim \mathcal{U}(\mathbb{F}_q^k)$, $M \sim \mathcal{U}(\mathbb{F}_q^s)$, Matrices $\mathcal{G}_{\mathcal{C}} \in \mathbb{F}_q^{k\times n}$ has rank k and $\mathcal{G}_{\mathcal{D}} \in \mathbb{F}_q^{s\times n}$ has rank s. The row space of $\mathcal{G}_{\mathcal{C}}$ and $\mathcal{G}_{\mathcal{D}}$ are denoted as $\mathcal{V}_{\mathcal{G}_{\mathcal{C}}} = \mathcal{C}$ and $\mathcal{V}_{\mathcal{G}_{\mathcal{D}}} = \mathcal{D}$, they are two linear codes with paremeters $[n, k]$ and $[n, s]$ respectively. We assume $\mathcal{C}$ and $\mathcal{D}$ are linear complementary codes:

$$\begin{aligned} \mathcal{V}_{\mathcal{G}_{\mathcal{C}}} \cap \mathcal{V}_{\mathcal{G}_{\mathcal{D}}} &= \mathcal{C} \cap \mathcal{D} = \{0_n\}, \\ \mathcal{V}_{\mathcal{G}_{\mathcal{C}}} \oplus \mathcal{V}_{\mathcal{G}_{\mathcal{D}}} &= \mathcal{C} \oplus \mathcal{D} = \mathbb{F}_q^n, \end{aligned} \tag{3.41}$$

where 0_n is the all-zero vector.

To simplify the derivations in the following sections, we represent $\mathbb{F}_q$ in $\mathbb{F}_2$ by the sub-field representation [MS77, § 7.7]. This allows us to focus only on *binary variables in* $\mathbb{F} = \mathbb{F}_2$ in later proof.

3.2.1 Background

In order to quantify the impact of probes, we adopt the following definition of the probing model in a κ-dimensional attack [ISW03].

Definition 3.2.1 (Probing Model)**.** Let $\kappa > 0$ be the *dimension* of the attack, and $\pi \in \mathbb{F}^n$ be a binary row vector of Hamming weight κ. The location of the nonzero elements in π represent the "location" of probes in a κ-dimensional attack. Let $\mathbf{A}^\pi$ denote the $m \times \kappa$ matrix obtained from the $m \times n$ matrix $\mathbf{A}$ by removing all columns $(A_{i,j})_j$ corresponding to zero elements $\pi_j = 0$ in π. The probing model is described as

$$V^\pi = X\mathcal{G}_\mathcal{C}^\pi + M\mathcal{G}_\mathcal{D}^\pi. \tag{3.42}$$

The concern is whether V^π (or some noisy version of it) leaks information about secret X in the presence of masking M.

Note that with our notation, $\pi^\pi \in \mathbb{F}^\kappa$ is the all-one vector and we have $\mathbf{A}\cdot\pi^T = \mathbf{A}^\pi\cdot(\pi^\pi)^T$. Let $w_H(\cdot)$ denote the Hamming weight of a vector. In particular $w_H(\pi) = \kappa$. The following notion of *generalized Hamming weight* is known to be a sound tool to characterize the leakage [Wei91], especially under the probing model in the noiseless scenario where the information leakage in code-based masking is modeled by a special case of wire-tap channel II (see Section. 1.2.2).

Definition 3.2.2 (Generalized Hamming Weight [Wei91])**.** For any linear code $\mathcal{C}$, the *support* $\chi(\mathcal{C})$ of $\mathcal{C}$ is the set of not-always-zero coordinates of $\mathcal{C}$. The r-th generalized Hamming weight of an $[n, k]$ linear code $\mathcal{C}$, where $1 \le r \le k$, is defined as the cardinality of the smallest support of a r-dimensional subcode of $\mathcal{C}$:

$$\delta_{r,\mathcal{C}} = \min_{\mathcal{C}'}\{|\chi(\mathcal{C}')| \ ; \ \mathcal{C}' \text{ is an } [n, r] \text{ subcode of } \mathcal{C}\}. \tag{3.43}$$

In particular $\delta_{1,\mathcal{C}}$ is the minimum Hamming weight of codewords in $\mathcal{C}$, i.e., the minimum distance of $\mathcal{C}$.

Definition 3.2.3 (Weight Enumerators)**.** Let $\{A_i\}_{i=0,\dots,n}$ be the Hamming weight distribution of the $[n, k]$ linear code $\mathcal{C}$. The weight enumerator of $\mathcal{C}$ is the polynomial

$$W_\mathcal{C}(x, y) = \sum_{i=0}^{n} A_i x^{n-i} y^i. \tag{3.44}$$

Let $\{B_i\}_{i=0,\dots,n}, \{A_i\}_{i=0,\dots,n}$ be the Hamming weight distribution of the masking code $\mathcal{D}$ and its dual code $\mathcal{D}^\perp$ respectively. Using MacWilliams theorem [MS77, Chap. 5, Theorem 1] for $x = p$ and $y = 1 - p$ one has

$$W_\mathcal{D}(p, 1-p) = \frac{1}{|\mathcal{D}^\perp|} W_{\mathcal{D}^\perp}(1, 2p-1). \tag{3.45}$$

3.2.2 Attacks Under Noiseless Measurements

In a noiseless attack, the attacker knows $Y = V^{\pi}$ without noise. The question is what quantity of information it can leak about the secret X. The information leakage is classically measured [VCS09, §5] by mutual information

$$I(X; V^{\pi}) = H(V^{\pi}) - H(V^{\pi}|X), \tag{3.46}$$

where $H(\cdot)$ denotes the Shannon entropy for discrete variable. Next, we will use (3.46) to estimate the value of $I(X; V^{\pi})$.

Lemma 3.2.1. *Let $\mathbf{A}$ be an $m \times n$ matrix of rank r over $\mathbb{F}$, and $\mathcal{V}_{\mathbf{A}} \subset \mathbb{F}^n$ its row space. If $X \sim \mathcal{U}(\mathbb{F}^m)$ is a row random vector uniformly distributed over $\mathbb{F}^m$, then $Y = X\mathbf{A} \sim \mathcal{U}(\mathcal{V}_{\mathbf{A}})$ is uniformly distributed over $\mathcal{V}_{\mathbf{A}}$.*

Proof. By the canonical decomposition of the linear application $\Phi : x \mapsto y = x\mathbf{A}$, $\operatorname{Im}\Phi = \mathcal{V}_{\mathbf{A}} \cong \mathbb{F}^m / \operatorname{Ker}\Phi$ where $\operatorname{Ker}\Phi$ has dimension $m - r$. In other words $\Phi^{-1}(y) = x + \operatorname{Ker}\Phi$ for any $y \in \mathcal{V}_{\mathbf{A}}$.

Now if $X \sim \mathcal{U}(\mathbb{F}^m)$ with $\mathbb{P}(X = x) = \frac{1}{2^m}$ and $Y = X\mathbf{A} = \Phi(X)$, then for any $y \in \mathcal{V}_{\mathbf{A}}$, we have

$$\begin{aligned} \mathbb{P}(Y = y) &= \mathbb{P}(X \in \Phi^{-1}(y)) = \mathbb{P}(X \in x + \operatorname{Ker}\Phi) \\ &= \frac{|\operatorname{Ker}\Phi|}{2^m} = \frac{2^{m-r}}{2^m} = \frac{1}{2^r} = \frac{1}{|\mathcal{V}_{\mathbf{A}}|}. \end{aligned} \tag{3.47}$$

□

Lemma 3.2.2. *One has*

$$H(V^{\pi}) = \kappa \text{ bits.} \tag{3.48}$$

Proof. From Lemma 3.2.1, $X\mathcal{G}_{\mathcal{C}} \sim \mathcal{U}(\mathcal{C})$, $M\mathcal{G}_{\mathcal{D}} \sim \mathcal{U}(\mathcal{D})$, hence $V = X\mathcal{G}_{\mathcal{C}} + M\mathcal{G}_{\mathcal{D}} \sim \mathcal{U}(\mathcal{C} \oplus \mathcal{D}) = \mathcal{U}(\mathbb{F}^n)$. It follows that $V^{\pi} \sim \mathcal{U}(\mathbb{F}^{\kappa})$, hence $H(V^{\pi}) = \log_2 |\mathbb{F}^{\kappa}| = \kappa$. □

Next consider the dual code $\mathcal{D}^{\perp}$. We have $\pi \in \mathcal{D}^{\perp}$ if and only if $\mathcal{G}_{\mathcal{D}} \cdot \pi^T = \mathcal{G}_{\mathcal{D}}^{\pi} \cdot (\pi^{\pi})^T = 0$. Thus to every codeword of $\pi \in \mathcal{D}^{\perp}$ of weight δ correspond to δ linearly dependent columns in $\mathcal{G}_{\mathcal{D}}^{\pi}$. Now if $\kappa < \delta_{\mathcal{D}^{\perp}}$, where $\delta_{\mathcal{D}^{\perp}}$ is the the dual distance of the code $\mathcal{D}$, every set of κ columns of $\mathcal{G}_{\mathcal{D}}$ are linearly independent so that $\mathcal{G}_{\mathcal{D}}^{\pi}$ always has full rank κ.

Theorem 3.2.3. *Let $\delta_{\mathcal{D}^{\perp}}$ be the dual distance of the code $\mathcal{D}$. If $\kappa < \delta_{\mathcal{D}^{\perp}}$ then*

$$I(X; V^{\pi}) = 0. \tag{3.49}$$

Proof. Since $\mathcal{G}_{\mathcal{D}}^{\pi}$ has full rank κ, by Lemma 3.2.1, $M\mathcal{G}_{\mathcal{D}}^{\pi} \sim \mathcal{U}(V_{\mathcal{G}_{\mathcal{D}}^{\pi}}) = \mathcal{U}(\mathbb{F}^{\kappa})$. The conditional distribution of V^{π} given $X = x$ is then $V^{\pi}|X = x \sim x\mathcal{G}_{\mathcal{C}}^{\pi} + \mathcal{U}(\mathbb{F}^{\kappa}) = \mathcal{U}(\mathbb{F}^{\kappa})$, which does not depend of x. Thus V^{π} is independent of X, that is, $I(X; V^{\pi}) = 0$. □

Hence V^{π} does not leak any information about the secret. In particular, we recover the following result from [CGC+21b]: If a polynomial P has numerical degree $< \delta_{\mathcal{D}^{\perp}}$, then $I(X; P(V^{\pi})) = 0$.

Theorem 3.2.4 (Noiseless Information Leakage)**.** *If an adversary chooses $\kappa = \delta_{\mathcal{D}^{\perp}}$ probes, then*

$$I(X; V^{\pi}) = \begin{cases} 1 \text{ bit}, & \text{if } \pi \in \mathcal{D}^{\perp}, \\ 0, & \text{otherwise.} \end{cases} \tag{3.50}$$

Proof. If $\pi \notin \mathcal{D}^{\perp}$ then the κ columns of $\mathcal{G}_{\mathcal{D}}^{\pi}$ are linearly independent and $\mathcal{G}_{\mathcal{D}}^{\pi}$ has full rank. Then as in the proof of Theorem 3.2.3, $I(X; V^{\pi}) = 0$.

If $\pi \in \mathcal{D}^{\perp}$ then the κ columns of $\mathcal{G}_{\mathcal{D}}^{\pi}$ are linearly dependent while every subset of less than κ columns of $\mathcal{G}_{\mathcal{D}}$ is linearly independent. Hence $\mathcal{G}_{\mathcal{D}}^{\pi}$ has rank $\kappa - 1$. By Lemma 3.2.1, $M\mathcal{G}_{\mathcal{D}}^{\pi} \sim \mathcal{U}(\mathcal{V}_{\mathcal{G}_{\mathcal{D}}^{\pi}})$ where $\mathcal{V}_{\mathcal{G}_{\mathcal{D}}^{\pi}}$ has dimension $\kappa - 1$, so that $H(V^{\pi}|X = x) = H(x\mathcal{G}_{\mathcal{C}}^{\pi} + M\mathcal{G}_{\mathcal{D}}^{\pi}) = \kappa - 1$ bits. Averaging over X gives $H(V^{\pi}|X) = \kappa - 1$ bits. From Lemma 3.2.2, $I(X; V^{\pi}) = H(V^{\pi}) - H(V^{\pi}|X) = \kappa - (\kappa - 1) = 1$ bit. □

Assuming the attacker chooses her probes' locations at random, let Π be a random vector chosen uniformly among all $\pi \in \mathbb{F}^n$ of weight κ. Then we have the following

Corollary 3.2.4.1. *If an adversary chooses κ positions of probe randomly and $\kappa = \delta_{\mathcal{D}^{\perp}}$, then on average*

$$I(X; V^{\Pi}) = \frac{A_{\kappa}}{\binom{n}{\kappa}} \text{ bits}, \tag{3.51}$$

where A_{κ} is the kissing number of $\mathcal{D}^{\perp}$.

Proof. From Theorem 3.2.4, $I(X; V^{\Pi} \mid \Pi = \pi) = 1$ or 0 according to whether $\pi \in \mathcal{D}^{\perp}$ (A_{κ} possibilities) or not ($\binom{n}{\kappa} - A_{\kappa}$ possibilities). Averaging over π gives

$$I(X; V^{\Pi}) = \frac{A_{\kappa}}{\binom{n}{\kappa}} \times 1 + \frac{\binom{n}{\kappa} - A_{\kappa}}{\binom{n}{\kappa}} \times 0 = \frac{A_{\kappa}}{\binom{n}{\kappa}} \text{ bits}. \tag{3.52}$$

□

Theorem 3.2.4 can be generalized as follows.

Theorem 3.2.5. *If an adversary can choose $\kappa \geq \delta_{\mathcal{D}^\perp} = \delta_{1,\mathcal{D}^\perp}$ probes, then the maximum amount of information she can extract is determined by:*

$$\max_{\pi} I(X; V^\pi) = \max\{r \ ; \ \delta_{r,\mathcal{D}^\perp} \leq w_H(\pi)\} \ \textit{bits}, \tag{3.53}$$

where $\delta_{r,\mathcal{D}^\perp}$ is the rth generalized Hamming weight of the code $\mathcal{D}^\perp$.

Proof. Probing $w_H(\pi)$ positions is equivalent with taping $w_H(\pi)$ coordinates of a codeword in the wiretap channel II. Therefore, it is straightforward from [Wei91] that the extractable information is determined by (3.53). □

3.2.3 Attacks Under Noisy Measurements

In classical side-channel analysis setups [HRG14a, BGHR14], the attacker exploits directly the noisy leakage, usually assumed to be equal to the leakage of V in the presence of some additive white Gaussian noise (AWGN) of variance σ^2. However, such setups require to make an ad-hoc assumption about the leakage model, i.e., a function that transduces a vector of field elements in $\mathbb{F}$ into a real number in $\mathbb{R}$. In order to be more general, we assume in this paper a narrower attack model, which digitizes the measured side-channel leakage for subsequent analysis. This corresponds to the situation of a hard detection or hard decision making—the side-channel is digitized prior to analysis.

Consider a AWGN channel with i.i.d noise $\sim \mathcal{N}(0, \sigma^2)$ in transmitting binary variables, followed by a binary detector. As is well known, the overall channel model becomes a memoryless binary symmetric channel (BSC) of probability $p = Q(\sqrt{\gamma})$ where

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-\frac{t^2}{2}} dt$$

is the Q-function and $\gamma = 1/\sigma^2$ is the actual signal-to-noise ratio (SNR). Therefore, in this section, we consider a discrete noise (a.k.a. binary error vector E) which follows the i.i.d. Bernoulli distribution and show how this discrete noise affects the amount of information an adversary can extract.

Let $E \in \mathbb{F}^n$ be the error vector with i.i.d components $E_i \sim \mathcal{B}(p)$ where $p = \mathbb{P}(E_i = 1)$. In short $E \sim \mathcal{B}(p)^{\otimes n}$. Let

$$V' = X\mathcal{G}_\mathcal{C} + M\mathcal{G}_\mathcal{D} + E$$

be the noisy leakage, and considering $\kappa = w_H(\pi)$ probes gives

$$V'^\pi = V^\pi + E^\pi = X\mathcal{G}_\mathcal{C}^\pi + M\mathcal{G}_\mathcal{D}^\pi + E^\pi. \tag{3.54}$$

The problem is to evaluate the mutual information $I(X; V'^\pi)$.

By Theorem 3.2.3, κ probes provide no information about the sensitive variable X when $\kappa < \delta_{\mathcal{D}^\perp}$. In this case $I(X; V'^\pi) = 0$. Therefore, we shall only consider the scenario for which $\kappa = \delta_{\mathcal{D}^\perp}$ with $\pi \in \mathcal{D}^\perp$. Then, from the analysis of the previous section, $\mathcal{G}_\mathcal{D}^\pi$ has rank $\kappa - 1$ and generates a $[\kappa, \kappa - 1]$ parity check code $\mathcal{D}^\pi = \mathcal{V}_{\mathcal{G}_\mathcal{D}^\pi}$, with the $[\kappa, 1]$ repetition code as the dual code $\mathcal{D}^{\pi\perp} = \{0, \pi^\pi\}$.

Theorem 3.2.6 (Noisy Information Leakage)**.** *In our hard decision probing model with* $\kappa = \delta_{\mathcal{D}^\perp}$ *with* $\pi \in \mathcal{D}^\perp$*, one has*

$$I(X; V'^{\pi}) = 1 - h(p^*) \tag{3.55}$$

where $h(p) = -p\log p - (1-p)\log(1-p)$ *denotes the binary entropy and* $p^* = W_{\mathcal{D}^\pi}(p, 1-p) = \sum_i B_i p^{\kappa-i}(1-p)^i$*, the weight enumerator polynomial of the code* $\mathcal{D}^\pi$ *generated by* $\mathcal{G}_{\mathcal{D}}^{\pi}$*.*

Notice that $0 \le p^* \le \sum_i \binom{\kappa}{i} p^{\kappa-i}(1-p)^i \le 1$ hence p^* is a probability.

Proof. Consider $I(X; V'^{\pi}) = H(V'^{\pi}) - H(V'^{\pi}|X)$. Because E is independent of V, $V'^{\pi} = V^{\pi} + E^{\pi}$ is, like V^{π}, uniformly distributed over $\mathcal{U}(\mathbb{F}^\kappa)$ so that $H(V'^{\pi}) = \kappa$.

The conditioned entropy $H(V'^{\pi}|X = x) = H(x\mathcal{G}_{\mathcal{C}}^{\pi} + M\mathcal{G}_{\mathcal{D}}^{\pi} + E^{\pi}) = H(M\mathcal{G}_{\mathcal{D}}^{\pi} + E^{\pi})$ is independent of the value of x because the probability distribution of $M\mathcal{G}_{\mathcal{D}}^{\pi} + E^{\pi}$ is only affected by the invertible shift operator which adds $x\mathcal{G}_{\mathcal{C}}^{\pi} \in \mathbb{F}^\kappa$. Hence averaging over X gives $H(V'^{\pi}|X) = H(M\mathcal{G}_{\mathcal{D}}^{\pi} + E^{\pi})$.

Now consider the κth extension of the memoryless BSC channel, which transforms each input vector $w \in \mathcal{V}_{\mathcal{G}_{\mathcal{D}}^{\pi}} = \mathcal{D}^{\pi}$ to some output $w' \in V_{\mathbb{F}^\kappa}$. Noting $p^* = \sum_i B_i p^{\kappa-i}(1-p)^i$, a direct inspection shows that there are two possible cases:

- $w' \in \mathcal{V}_{\mathcal{G}_{\mathcal{D}}^{\pi}}$: the probability of each w' is $\frac{p^*}{2^{\kappa-1}}$;
- $w' \notin \mathcal{V}_{\mathcal{G}_{\mathcal{D}}^{\pi}}$: the probability of each w' is $\frac{1-p^*}{2^{\kappa-1}}$.

Then we have

$$\begin{aligned} H(V'^{\pi}|X) &= H(M\mathcal{G}_{\mathcal{D}}^{\pi} + E^{\pi}) \\ &= \sum_{1}^{2^{\kappa-1}} \frac{p^*}{2^{\kappa-1}} \log \frac{2^{\kappa-1}}{p^*} + \sum_{1}^{2^{\kappa-1}} \frac{1-p^*}{2^{\kappa-1}} \log \frac{2^{\kappa-1}}{1-p^*} \\ &= d - 1 + h(p^*), \end{aligned}$$

hence $I(X; V'^{\pi}) = H(V'^{\pi}) - H(V'^{\pi}|X) = 1 - h(p^*)$. □

Theorem 3.2.6 shows that adding noise can only decrease the mutual information $I(X; V'^{\pi})$. In the sequel, we further detail the evaluation of mutual information under weak and strong noise, respectively.

3.2.3.1 Attacks Under Weak Noise

For weak noise we consider $\sigma \to 0$, $\gamma \to +\infty$, and, therefore [BS79],

$$p = Q(\sqrt{\gamma}) \sim \frac{\mathrm{e}^{-\gamma/2}}{\sqrt{2\pi\gamma}} \tag{3.56}$$

tends exponentially toward zero. As a result we have the following behavior.

Theorem 3.2.7 (Information Leakage Under Weak Noise)**.** *In our hard decision probing model with $\kappa = \delta_{\mathcal{D}^\perp}$ with $\mathrm{II} \in \mathcal{D}^\perp$, as $\sigma \to 0$ (hence $p \to 0$), one has the asymptotic equivalence*

$$1 - I(X; V'^{\mathrm{II}}) \sim \frac{\kappa \cdot \mathrm{e}^{-1/2\sigma^2}}{2\sqrt{2\pi\sigma^2}} \to 0. \tag{3.57}$$

Proof. Applying MacWilliams' identity (3.45) to the code $\mathcal{D}^{\mathrm{II}} = \mathcal{V}_{\mathcal{G}_{\mathcal{D}}^{\mathrm{II}}}$, whose dual code $\mathcal{D}^{\mathrm{II}\perp}$ is the $[\kappa, 1]$ repetition code, we obtain

$$p^* = \frac{1}{2}\sum_{i=0}^{\kappa} A_i (2p-1)^i = \frac{1 + (2p-1)^\kappa}{2}$$

Since $p \to 0$, according to whether κ is even or odd, $p^* \to 1$ or $p^* \to 0$. Hence $H_2(p^*) \to 0$ is equivalent to either $-(1-p^*)\log(1-p^*)$ or $-p^* \log p^*$. Therefore, $1 - I(X; Z'^{\mathrm{II}}) = H_2(p^*) \sim -p\kappa \log(p\kappa) \sim -p\kappa \log(p) \sim \frac{\gamma\kappa e^{-\gamma/2}}{2\sqrt{2\pi\gamma}}$ where $\gamma = \sigma^{-2}$, which yields the announced formula . □

Since $I(X; V'^{\mathrm{II}})$ will tend to 1 when the noise approaches zero, one recovers Theorem 3.2.4 in the noiseless case.

3.2.3.2 Attacks Under Strong Noise

One of the main benefits of masking is that, under sufficient strong noise, the number of measurements to recover the secret key used in a masked cryptographic implementation increases exponentially with the protection order (indicated by the dual distance in code-based masking). Herein we investigate the asymptotic features of information leakage quantified by $I(X; V'^{\mathrm{II}})$ under a strong noise, i.e., when $\sigma \to +\infty$, $\gamma \to 0$ so that $p \to \frac{1}{2}$.

Theorem 3.2.8 (Information leakage under strong noise)**.** *In our hard decision probing model with $\kappa = \delta_{\mathcal{D}^\perp}$ with $\mathrm{II} \in \mathcal{D}^\perp$, as $\sigma \to +\infty$, one has the following equivalence:*

$$I(X; V'^{\mathrm{II}}) \sim \frac{2^{\kappa-1}}{\pi^\kappa \cdot \ln 2} \cdot \sigma^{-2\kappa}. \tag{3.58}$$

Proof. By first-order Taylor expansion, $Q(x) = \frac{1}{2} - \frac{1}{\sqrt{2\pi}} \int_0^x \mathrm{e}^{\frac{-t^2}{2}} dt = \frac{1}{2} - \frac{x}{\sqrt{2\pi}} + o(x)$ and $p = Q(\sqrt{\gamma}) = \frac{1}{2} - \sqrt{\frac{\gamma}{2\pi}} + o(\sqrt{\gamma}) = \frac{1}{2} - \epsilon + o(\epsilon)$ where $\epsilon = \sqrt{\frac{\gamma}{2\pi}} = \frac{1}{\sqrt{2\pi\sigma^2}} \to 0$ when $\sigma \to +\infty$.

Applying MacWilliams' identity (3.45) to $\mathcal{D}^{\perp} = \mathcal{V}_{\mathcal{G}_{\mathcal{D}}^{\perp}}$, we obtain

$$p^* = \frac{1}{2} \sum_{i=0}^{\kappa} A_i (2p-1)^i = \frac{1}{2} + (-2)^{\kappa-1} \epsilon^{\kappa} + o(\epsilon^{\kappa}).$$

Now by Taylor's expansion at second order for the entropy $H_2(p^*)$ is $H_2(\frac{1}{2}) + H'(\frac{1}{2})(\frac{1}{2} - p^*) + \frac{1}{2} H''(\frac{1}{2}) \cdot (\frac{1}{2} - p^*)^2 = 1 - 2\log_2(\mathrm{e}) \cdot (\frac{1}{2} - p^*)^2$. Finally, we have

$$\begin{aligned} I(X; Z'^{\perp}) = 1 - H_2(p^*) &\sim 2\log_2(\mathrm{e}) \cdot \left(-2^{\kappa-1}(-\epsilon)^{\kappa}\right)^2 \\ &= 2^{2\kappa-1} \epsilon^{2\kappa} \cdot \log_2(\mathrm{e}) = \frac{2^{\kappa-1}}{\pi^{\kappa} \cdot \ln 2} \cdot \sigma^{-2\kappa} \end{aligned} \tag{3.59}$$

□

Theorem 3.2.8 shows that the mutual information between the sensitive variable X and the noisy measurements is exponentially decreasing in σ^2 with an exponent equal to the protection order κ (dual distance of $\mathcal{D}$).

CHAPTER 4

Attack Evaluation of Masked Implementations

The result of this chapter has been published in "Improved Alpha-Information Bounds for Higher-Order Masked Cryptographic Implementations" [LBC+23].

Contents

Abstract

In this chapter, the security of protected cryptographic implementations is evaluated for any masking order, using alpha-information measures. Universal upper bounds on the probability of success of any type of side-channel attack are derived. These also provide lower bounds on the minimum number of queries required to achieve a given success rate. An important issue, solved in this section, is to remove the loss factor due to the masking field size.

4.1 Alpha-Information Bounds for Higher-Order Masked Implementations

The security assessments in this section apply to implementations protected by arithmetic masking or boolean masking. The following model has been introduced in Section 1.2:

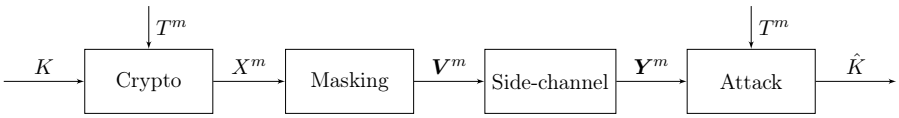


Figure 4.1: Side-channel seen as a communication channel (with masking).

As in the previous model, the secret K and the public variable T are independent of each other and uniformly distributed over a same finite field $\mathbb{F}_q = \mathbb{F}_{2^l}$. The field size is denoted as N. The sensitive variable $X \in \mathbb{F}_q$ is a deterministic function of K and T.

According to Definition.1.2.1 and Definition.1.2.2, both arithmetic masking and Boolean masking (for binary sequences with XOR operation) split the sensitive information X into $d+1$ shares $\boldsymbol{V} = (X_0, X_1, \ldots, X_d)$, satisfying

$$X = X_0 \oplus X_1 \oplus \cdots \oplus X_d, \tag{4.1}$$

where $\oplus$ is the additive operation in the underlying field. For example, notation $\oplus$ can be the bitwise XOR operation in Boolean masking.

During computation, side-channel information on $\boldsymbol{V} = (X_0, X_1, \ldots, X_d)$ is leaking and can be measured as a noisy "trace" by the attacker, denoted by $\boldsymbol{Y} = (Y_0, Y_1, \ldots, Y_d)$. We assume that $\boldsymbol{Y}$ is the output of a memoryless side-channel with input $\boldsymbol{V}$. Since masking shares are drawn uniformly and independently, both $\boldsymbol{V}$ and $\boldsymbol{Y}$ are i.i.d. sequences.

The attacker measures m traces $\boldsymbol{Y}^m = (\boldsymbol{Y}_1, \boldsymbol{Y}_2, \ldots, \boldsymbol{Y}_m)$ corresponding to text sequence $T^m = (T_1, T_2, \ldots, T_m)$, and exploits her knowledge of $\boldsymbol{Y}^m$ and T^m to estimate the secret key $\hat{K}$. Again, since the side channel is memoryless, $\boldsymbol{X}^m$ and $\boldsymbol{Y}^m$ are i.i.d. sequences. Let $\mathbb{P}_s = \mathbb{P}(K = \hat{K})$ be the probability of success of the attack upon observing T^m and $\boldsymbol{Y}^m$. In theory, maximum success is obtained by the MAP (maximum *a posteriori* probability) rule with success probability denoted by $\mathbb{P}_s = \mathbb{P}_s(K|\boldsymbol{Y}^m, T^m)$.

4.1.1 State-of-the-art

Duc et al. [DFS15] derived a lower bound on the minimum number m of queries required to achieve a given probability of success $\mathbb{P}_s$, which can be rewritten as:

$$m \geq \frac{\log(1-\frac{1}{N}) - \log(1-\mathbb{P}_s)}{-\log\left(1 - (\frac{N}{\sqrt{2\log e}})^{d+1} \prod_{i=0}^{d} \sqrt{I(X_i;Y_i)}\right)} \tag{4.2}$$

where $d+1$ is the number of shares, N is the field size, and $I(X_i;Y_i)$ is the mutual information between each share and its corresponding leakage. They also showed that this bound was quite loose in practice and conjectured that when the leakage of shares is

sufficiently noisy (and independent among shares), the lower bound on m should take the approximate form

$$m \gtrsim \frac{\beta(\mathbb{P}_s)}{\prod_{i=0}^{d} I(X_i;Y_i)} \tag{4.3}$$

where β is a "small constant depending on $\mathbb{P}_s$"[DFS19, p. 1279].

The bound (4.2) was improved recently in [MRS23]:

$$m \geq \frac{d_1(\mathbb{P}_s \| \frac{1}{N})}{\log\left(1 + \frac{N}{2}(\frac{2}{\log e})^{d+1} \prod_{i=0}^{d} I(X_i;Y_i)\right)} \tag{4.4}$$

where $d_1(\cdot\|\cdot)$ is the binary Kullback–Leibler divergence. A similar bound was derived independently in [IUH22]. Although this greatly improves (4.2) for small N, when the field size N is large, the N factor in the denominator loosens the bound by an substantial amount. Therefore, an important issue is to find out whether this factor N can be removed.

4.1.2 Lower Bounds on Sibson's Alpha-Information

4.1.2.1 Bounding Success by Sibson's Alpha-Information

In Fig. 4.1, the sensitive variable X^m is a function of K and T^m; $\hat{K}$ is a function of $(\boldsymbol{Y}^m, T^m)$. It is easily seen from the figure that the following Markov chains hold:

$$K \longleftrightarrow (\boldsymbol{Y}^m, T^m) \longleftrightarrow \hat{K}, \tag{4.5}$$

$$(K, T^m) \longleftrightarrow X^m \longleftrightarrow \boldsymbol{Y}^m. \tag{4.6}$$

The probability of success of the side-channel attack is $\mathbb{P}_s = \mathbb{P}_s(K|\boldsymbol{Y}^m, T^m)$. Using Lemma 1.3.1, one has $d_\alpha\big(\mathbb{P}_s \| \frac{1}{N}\big) \leq I_\alpha(K; \boldsymbol{Y}^m, T^m)$.

Based on the closed-form expression we have

Lemma 4.1.1. $I_\alpha(K; \boldsymbol{Y}^m, T^m) \leq I_\alpha(K, T^m; \boldsymbol{Y}^m)$. (4.7)

Proof. By definition, $\frac{\alpha-1}{\alpha} I_\alpha(K; \boldsymbol{Y}^m, T^m)$ can be writen as

$$\begin{aligned}\log \mathbb{E}_{\boldsymbol{Y}^m, T^m} \langle p_{K|\boldsymbol{Y}^m,T^m} \| p_K \rangle_\alpha &= \log \mathbb{E}_{T^m} \int_{\boldsymbol{Y}^m} p_{\boldsymbol{Y}^m|T^m} \Big(\sum_k p^\alpha_{K|\boldsymbol{Y}^m,T^m} p_K^{1-\alpha} \Big)^{\frac{1}{\alpha}} \\ &= \log \mathbb{E}_{T^m} \int_{\boldsymbol{Y}^m} \Big(\sum_k p^\alpha_{K,\boldsymbol{Y}^m|T^m} p_K^{1-\alpha} \Big)^{\frac{1}{\alpha}}.\end{aligned}$$

Then one has

$$\begin{aligned} I_\alpha(K;\boldsymbol{Y}^m,T^m) &\overset{(\star)}{=} \tfrac{\alpha}{\alpha-1}\log \mathbb{E}_{T^m}\int_{\boldsymbol{Y}^m}\Big(\sum_k p^\alpha_{\boldsymbol{Y}^m|K,T^m}p_{K|T^m}\Big)^{\frac{1}{\alpha}} \\ &\overset{(\star\star)}{\leq} \tfrac{\alpha}{\alpha-1}\log\int_{\boldsymbol{Y}^m}\Big(\mathbb{E}_{T^m}\sum_k p^\alpha_{\boldsymbol{Y}^m|K,T^m}p_{K|T^m}\Big)^{\frac{1}{\alpha}} \\ &= \tfrac{\alpha}{\alpha-1}\log\int_{\boldsymbol{Y}^m}\Big(\sum_{k,t^m} p^\alpha_{\boldsymbol{Y}^m|K,T^m}p_{K,T^m}\Big)^{\frac{1}{\alpha}} \\ &= \tfrac{\alpha}{\alpha-1}\log\int_{\boldsymbol{Y}^m} p_{\boldsymbol{Y}^m}\Big(\sum_{k,t^m} p^\alpha_{K,T^m|\boldsymbol{Y}^m}p^{1-\alpha}_{K,T^m}\Big)^{\frac{1}{\alpha}} = I_\alpha(K,T^m;\boldsymbol{Y}^m) \end{aligned}$$

where $(\star)$ holds since $p_K = p_{K|T^m}$ (K and T^m are independent) and $p^\alpha_{K,\boldsymbol{Y}^m|T^m}p^{-\alpha}_{K|T^m} = p^\alpha_{\boldsymbol{Y}^m|K,T^m}$; $(\star\star)$ is Jensen's inequality: when $\alpha > 1$, $x^{\frac{1}{\alpha}}$ is concave and $\frac{\alpha}{\alpha-1}$ is positive; when $0 < \alpha < 1$, $x^{\frac{1}{\alpha}}$ is convex and $\frac{\alpha}{\alpha-1}$ is negative. In both cases the inequality holds in the same direction. □

It follows that the generalized Fano inequality implies

$$d_\alpha\big(\mathbb{P}_s\|\frac{1}{N}\big) \leq I_\alpha(K,T^m;\boldsymbol{Y}^m). \tag{4.8}$$

Because $(K,T^m) \leftrightarrow X^m \leftrightarrow \boldsymbol{Y}^m$ forms a Markov chain, using the DPI of Sibson's α-information we have

$$I_\alpha(K,T^m;\boldsymbol{Y}^m) \leq I_\alpha(X^m;\boldsymbol{Y}^m). \tag{4.9}$$

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

$$I_\alpha(X^m;\boldsymbol{Y}^m) = mI_\alpha(X;\boldsymbol{Y}). \tag{4.10}$$

From (4.8), (4.9), and (4.10), we arrive at the main result of this section:

Theorem 4.1.2. *One has*

$$d_\alpha\big(\mathbb{P}_s\|\frac{1}{N}\big) \leq mI_\alpha(X;\boldsymbol{Y}) \tag{4.11}$$

Note that since $d_\alpha\big(p\|q\big)$ is increasing in p when $p \geq q$, Theorem 4.1.2 gives an upper bound on the probability of success $\mathbb{P}_s$.

4.1.2.2 Comparison with the Classical Bound

A natural question is to compare (4.11) with the classical bound for $\alpha = 1$, especially in terms of how it depends on N. Since d_α and I_α are non-decreasing in α, a precise answer is not obvious. One can argue as follows. Assume $\mathbb{P}_s$ is fixed in $(0,1)$. For $\alpha = 1$, one has at first order

$$d_1\big(\mathbb{P}_s\|\tfrac{1}{N}\big) = \log N - (1-\mathbb{P}_s)\log(N-1) - h(\mathbb{P}_s) \approx \mathbb{P}_s\log N \tag{4.12}$$

where $h(\mathbb{P}_s)$ is the binary entropy function. For $\alpha < 1$, $d_\alpha\big(\mathbb{P}_s\|\frac{1}{N}\big) \leq d\big(\mathbb{P}_s\|\frac{1}{N}\big)$ does not grow faster than $O(\log N)$. For $\alpha > 1$, one has at first order

$$d_\alpha\big(\mathbb{P}_s\|\tfrac{1}{N}\big) = \log N + \tfrac{1}{\alpha-1}\log\Big(\mathbb{P}_s^\alpha + \frac{(1-\mathbb{P}_s)^\alpha}{(N-1)^{\alpha-1}}\Big) \approx \log N \tag{4.13}$$

Thus the $O(\log N)$ term applies for any α, and the lower bound in (4.11) will not become looser than the classical bound as the field size N increases.

4.1.3 Upper Bound on Rényi Mutual Information

4.1.3.1 Euclidean Distance to the Uniform

In the field of cryptography, the *total variation distance* $\|P-U\|_1$ of a given N-ary distribution P to the uniform distribution $U \sim \mathcal{U}(N)$ is a common criterion to evaluate randomness. For $\alpha \neq 1$ we have the following

Definition 4.1.1 (α-Distance)**.** Let X be an N-ary random variable. The "α-distance" between P_X and a uniform distribution $U \sim \mathcal{U}(N)$ is defined as

$$\|P_X - U\|_\alpha = \Big(\sum_x \big|p_X(x) - \tfrac{1}{N}\big|^\alpha\Big)^{\frac{1}{\alpha}}. \tag{4.14}$$

In this section we focus on the Euclidean distance ($\alpha = 2$) because of the following

Lemma 4.1.3. *With the same notations, one has*

$$D_2(P_X\|U) = \log(1 + N\cdot\|P_X - U\|_2^2). \tag{4.15}$$

Proof. One has $\|P_X - U\|_2^2 = \sum_x (p_X(x) - \frac{1}{N})^2 = \sum_x p_X^2(x) - \frac{1}{N}$. Since $D_2(P_X\|U) = \log(N\cdot\sum_x p_X^2(x))$, the result follows. □

The following important Lemma is known as the XOR Lemma in the case of Boolean Masking[MRS23].

Lemma 4.1.4 (Group Lemma)**.** *Let X_1, X_2 be independent random variables over a finite Abelian group $\mathcal{X}$ of size N, and $U \sim \mathcal{U}(\mathcal{X})$. Let $X = X_1 \oplus X_2$, where $\oplus$ denotes the group operator in $\mathcal{X}$. One has*

$$\|P_X - U\|_2^2 \leq N\cdot\|P_{X_1} - U\|_2^2 \cdot \|P_{X_2} - U\|_2^2. \tag{4.16}$$

By finite induction, if X is split into $d+1$ independent shares: $X = X_0 \oplus X_1 \oplus \cdots \oplus X_d$, one has

$$\|P_X - U\|_2^2 \leq N^d \|P_{X_0} - U\|_2^2 \|P_{X_1} - U\|_2^2 \cdots \|P_{X_d} - U\|_2^2. \tag{4.17}$$

Proof. Since $X = X_1 \oplus X_2$, one has $P_X = P_{X_1} * P_{X_2}$ where $*$ denotes the convolution operator over the Abelian group. It is easy to check that $P_X - U = (P_{X_1} - U) * (P_{X_2} - U)$, and by the Cauchy-Schwarz inequality, $|(P_{X_1} - U) * (P_{X_2} - U)| \leq \|P_{X_1} - U\|_2 \|P_{X_2} - U\|_2$. Summing over the N values of X gives (4.16). □

Remark 4.1.1. Lemmas 4.1.3 and 4.1.4 do not seem to be easily generalized to other values of $\alpha \neq 2$. This is the main reason why we focus on $\alpha = 2$ in this paper.

4.1.3.2 Upper Bound of Rényi 2-Information for Each Share

Since Sibson's α-information does not exceed Rényi mutual information (inequality (1.49)), Theorem 4.1.2 implies

$$d_\alpha\big(\mathbb{P}_s \| \frac{1}{N}\big) \leq m I_\alpha^R(X; \boldsymbol{Y}). \tag{4.18}$$

We now upper bound $I_\alpha^R(X; \boldsymbol{Y})$ by noting that, by definition since X is uniformly distributed,

$$\begin{aligned} I_2^R(X; \boldsymbol{Y}) &= \log \mathbb{E}_{\boldsymbol{Y}} \exp D_2(P_{X|\boldsymbol{Y}} \| U) \\ &= \log(1 + N \cdot \mathbb{E}_{\boldsymbol{Y}} \|P_{X|\boldsymbol{Y}} - U\|_2^2). \end{aligned} \tag{4.19}$$

Since $\{X_i, Y_i\}_{i=0,\ldots,d}$ are mutually independent, (4.17) applies for $X|\boldsymbol{Y}$ and we have

$$I_2^R(X; \boldsymbol{Y}) \leq \log\big(1 + N \cdot \mathbb{E}_{\boldsymbol{Y}} N^d \prod_{i=0}^{d} \|P_{X_i|Y_i} - U\|_2^2\big) \tag{4.20}$$

$$= \log\big(1 + \prod_{i=0}^{d} N \cdot \mathbb{E}_{Y_i} \|P_{X_i|Y_i} - U\|_2^2\big) \tag{4.21}$$

$$= \log\big(1 + \prod_{i=0}^{d} (\exp I_2^R(X_i; Y_i) - 1)\big). \tag{4.22}$$

Putting all inequalities together yields the main result of this paper:

Theorem 4.1.5 (Main Result). *The number of traces m can be lower bounded by*

$$m \geq \frac{d_2(\mathbb{P}_s \| \frac{1}{N})}{\log\Big(1 + \prod_{i=0}^{d} (\exp I_2^R(X_i; Y_i) - 1)\Big)}. \tag{4.23}$$

Note that from Subsection 4.1.2.2 with $\alpha = 2$, the numerator does not lose tightness compared the case $\alpha = 1$ (compare (4.4)).

4.1.4 Numerical Results

In this subsection, we validate our results by simulation. The side-channel settings are as follows:

- the field of variables is the AES (Advanced Encryption Standard) field $\mathbb{F}_{2^\ell}$ with $\ell = 8$, thus $N = 256$;
- side-channel information is generated by taking the Hamming weight leakage model and additive white Gaussian noise (one of the most commonly adopted models [MOP07]);
- the Boolean masking is considered with orders $d \in 0, 1, 2$.

Shannon and Rényi mutual information (MI) is evaluated by Monte-Carlo simulation. In particular, we compare Rényi MI in (4.22) with the following

$$I(X;\boldsymbol{Y}) \leq \log\big(1 + \tfrac{N}{2}(\tfrac{2}{\log e})^{d+1} \prod_{i=0}^{d} I(X_i;Y_i)\big) \tag{4.24}$$

used in (4.4).

Fig. 4.2 confirms this on the performance bounds on the success rate as a function of m, for $d = 1$ and 2.

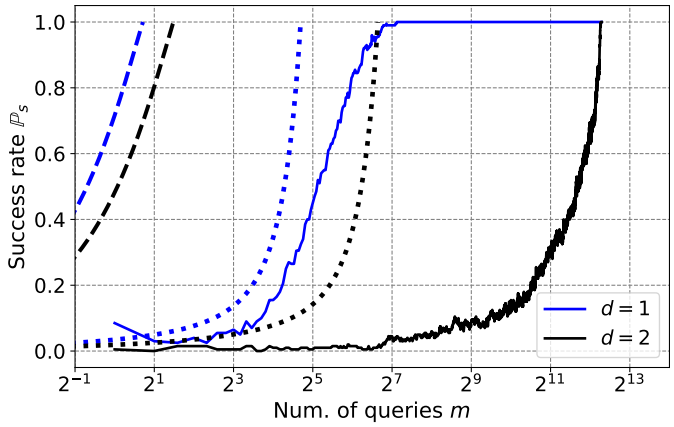


Figure 4.2: $\mathbb{P}_s$ vs m in attacks and the corresponding bounds for noise variance $\sigma^2 = 8$. The plain curves show the results of direct maximum likelihood (ML) attacks [HRG14b]; the *dotted* curves show the predictions by Theorem 4.1.5; the *dashed* curves are for the state-of-the-art bound (4.4).

Our new bounds are significantly more accurate than the state-of-the-art: For $\mathbb{P}_s = 80\%$ and $d = 1$, the ML attack gives about $m \geq 60$, our new bound gives $m \geq 25$, while (4.4) gives only $m \geq 1$. Much improvement can also be observed for $d = 2$.

This work uses traditional information-theoretic tools to evaluate the side-channel security of masked implementations, essentially providing bounds when $\alpha = 1$.

Interestingly, our result may also be related to [PGMP19], since equation (4.22) has the same form as [PGMP19, Theorem 3], but with different information-theoretic metrics. It would be worthwhile to investigate the relationships and compare the various information metrics used in security proofs.

4.2 Further Discussions

4.2.1 Possible Directions for Improvement

In Section 4.1, evaluating security of masked implementations mainly consists of three steps:

1. Firstly, using the Fano inequality to bound the success rate by the "global" mutual information between the sensitive variable and the side-channel leakage:

$$d_\alpha\big(\mathbb{P}_s\|\frac{1}{N}\big) \le I_\alpha(X^m;\boldsymbol{Y}^m). \tag{4.25}$$

   Here, "global" refers to the information between the leakage obtained from all m queries and their corresponding sensitive variables.

2. Secondly, using the information $I_\alpha(X;\boldsymbol{Y})$ obtained from a single query (times m) to bound the global mutual information $I_\alpha(X^m;\boldsymbol{Y}^m)$.

$$I_\alpha(X^m;\boldsymbol{Y}^m) \le m \cdot I_\alpha(X;\boldsymbol{Y}) \tag{4.26}$$

   Since the public variable used in each query is uniformly distributed and i.i.d. sequence, and the side-channel is memoryless, $(X^m;\boldsymbol{Y}^m)$ is also an i.i.d. sequence.

3. Lastly, we use the leakage information of each share to bound the information $I_\alpha(X;\boldsymbol{Y})$ when $\alpha = 2$:

$$I_2(X;\boldsymbol{Y}) \le \log\big(1+\prod_{i=0}^{d}(\exp I_2^R(X_i;Y_i)-1)\big). \tag{4.27}$$

   where $I_2^R(X_i;Y_i)$ is Rényi α-mutual information between each share and its leakage.

   The significance of this inequality lies in the fact that directly estimating the mutual information of high-dimensional vectors, i.e., $I_\alpha(X;\boldsymbol{Y})$ or $I_\alpha(\boldsymbol{V};\boldsymbol{Y})$, is too complicated. Thus we would like to bound $I_\alpha(X;\boldsymbol{Y})$ by the mutual information of each single share and its corresponding leakage, $I_\alpha(X_i;Y_i)$, $i = 0,1,\ldots,d$. This can significantly simplifies the estimation.

For certain reasons, we have only derived the bounds for $\alpha = 2$ in the third step. Ideally, we aim to derive:

$$d_\alpha\big(\mathbb{P}_s\|\frac{1}{N}\big) \le I_\alpha(X^m;\boldsymbol{Y}^m) \le mI_\alpha(X;\boldsymbol{Y}) \le mf\big(I_\alpha(X_0;\boldsymbol{Y}_0),\ldots,I_\alpha(X_d;\boldsymbol{Y}_d)\big) \tag{4.28}$$

for any α, where f is a function of $d+1$ variables.

Regarding this evaluation process, there are two points I'd like to clarify:

- Firstly, if we can replace the α-information in (4.28) with the corresponding conditional α-information (given the public variable T), we would obtain better bounds. In fact, the first inequality in (4.28) can be extended to conditional α-information, as shown in Section 3.1. However, somehow, it is challenging to obtain a result similar to the second inequality in (4.28) for the conditional α-information $I_\alpha(X^m;\boldsymbol{Y}^m|T^m)$.

- Secondly, it's challenging to determine which value of α yields the optimal solution for this bound. Intuitively, the first inequality in (4.28) might perform increasingly better as α grows, as indicated by the simulation results in Section 3.1. However, the second inequality might lose precision as α increases. The third inequality is even more elusive. Based on some other research results, the function f for different values of α appears to be distinct.

4.2.2 In the Case of Alpha = 1/2

Recall that α-entropy, α-divergence, and Sibson's α-information are defined as

$$H_\alpha(P_X) = \tfrac{\alpha}{1-\alpha} \log \|p_X\|_\alpha \tag{4.29}$$
$$D_\alpha(P_X\|Q_X) = \tfrac{1}{\alpha-1} \log \langle p_X \| q_X \rangle_\alpha^\alpha \tag{4.30}$$
$$I_\alpha(X;Y) = \tfrac{\alpha}{\alpha-1} \log \mathbb{E}_Y \langle p_{X|Y} \| p_X \rangle_\alpha. \tag{4.31}$$

where

$$\|p_X\|_\alpha = \Big(\int_{\mathcal{X}} p_X^\alpha(x)\ d\mu(x)\Big)^{1/\alpha} \tag{4.32}$$
$$\langle p_X \| q_X \rangle_\alpha = \Big(\int_{\mathcal{X}} p_X(x)^\alpha q_X(x)^{1-\alpha}\ d\mu(x)\Big)^{1/\alpha} \tag{4.33}$$

One has the following property:

Proposition 4.2.1. Let $X, U \sim \mathcal{U}(N)$ be uniform distributions, then

$$\exp\big(\tfrac{\alpha-1}{\alpha} I_\alpha(X;Y)\big) = \mathbb{E}_Y \exp\big(\tfrac{\alpha-1}{\alpha} D_\alpha(P_{X|Y}\|U)\big). \tag{4.34}$$

A natural question arises: Can the method used in Section 4.1 be applied to other values of α? Deriving similar bounds as in Section 4.1 for $\alpha = \frac{1}{2}$ seems to be promising, because $D_{\frac{1}{2}}(P_X\|Q_X)$ is also related to Euclidean distance:

$$\begin{aligned} D_{\frac{1}{2}}(P_X\|Q_X) &= -2\log\Big(\sum_x \sqrt{p_X(x)}\sqrt{q_X(x)}\Big) \\ &= -2\log\Big(1 - \frac{1}{2}\|\sqrt{p_X(x)} - \sqrt{q_X(x)}\|_2^2\Big) \end{aligned} \tag{4.35}$$

However, using the same method in Section 4.1, I only obtain a trivial bound for $\alpha = \frac{1}{2}$. The detailed derivation process is as follows.

When $\alpha = \frac{1}{2}$, Proposition 4.2.1 becomes

$$I_{\frac{1}{2}}(X;Y) = -\log\Big(\mathbb{E}_Y \exp\big(-D_{\frac{1}{2}}(P_{X|Y=y}\|U)\big)\Big). \tag{4.36}$$

We also have the following group lemma for $\alpha = \frac{1}{2}$:

Lemma 4.2.1 (Group Lemma)**.** *Let X_0, X_1, ...,X_d be independent random variables over a finite Abelian group $\mathcal{X}$ of size N, and $U \sim \mathcal{U}(\mathcal{X})$. Let $X = X_0 \oplus X_1 \oplus \cdots \oplus X_d$, where $\oplus$ denotes the group operator in $\mathcal{X}$. One has*

$$1 - \frac{1}{2}\|\sqrt{P_X} - \sqrt{U}\|_2^2 \geq \prod_{i=0}^{d}\Big(1 - \frac{1}{2}\|\sqrt{P_{X_i}} - \sqrt{U}\|_2^2\Big). \tag{4.37}$$

Proof. Proving this lemma for two variables ($X = X_0 \oplus X_1$) is enough:

$$\begin{aligned}
2 - \|\sqrt{P_X} - \sqrt{U}\|_2^2 &= 2 - \sum_x (\sqrt{p_X(x)} - \sqrt{\tfrac{1}{N}})^2 = \tfrac{2}{\sqrt{N}} \sum_x \sqrt{p_X(x)} && (4.38)\\
&= \frac{2}{\sqrt{N}} \sum_x \Big(\sum_{x_0} p_{X_0}(x_0) p_{X_1}(x \ominus x_0) \Big)^{\frac{1}{2}} && (4.39)\\
&\geq \frac{2}{\sqrt{N}} \sum_x \Big(\frac{1}{N} \cdot (\sum_{x_0} \sqrt{p_{X_0}(x_0) p_{X_1}(x \ominus x_0)})^2 \Big)^{\frac{1}{2}} && (4.40)\\
&= \frac{2}{N} \sum_{x_0} \sqrt{p_{X_0}(x_0)} \sum_x \sqrt{p_{X_1}(x \ominus x_0)} && (4.41)\\
&= \frac{1}{2} \prod_{i=0}^{1} \Big(\frac{2}{\sqrt{N}} \sum_{x_i} \sqrt{p_{X_i}(x_i)} \Big) && (4.42)\\
&= \frac{1}{2} \prod_{i=0}^{1} \Big(2 - \|\sqrt{P_{X_i}} - \sqrt{U}\|_2^2 \Big) && (4.43)
\end{aligned}$$

where (4.40) is using the Cauchy-Schwarz inequality $(\sum_{i=1}^{N} \sqrt{z_i})^2 \leq N \cdot \sum_{i=1}^{N} z_i$.

By induction, this lemma can be proved. □

Combine (4.35), (4.36) and (4.37) we can prove the following theorem:

Theorem 4.2.2. *With the same notations in Section 4.1, one has*

$$I_{\frac{1}{2}}(X; \boldsymbol{Y}) \leq \sum_{i=0}^{d} I_{\frac{1}{2}}(X_i; Y_i) \tag{4.44}$$

Proof.

$$\begin{aligned}
I_{\frac{1}{2}}(X; \boldsymbol{Y}) &= -\log\Big(\mathbb{E}_{\boldsymbol{Y}} \exp\big(2\log(1 - \tfrac{1}{2}\|\sqrt{P_{X|\boldsymbol{Y}=\boldsymbol{y}}} - \sqrt{U}\|_2^2)\big)\Big)\\
&\leq -\log\Big(\mathbb{E}_{\boldsymbol{Y}} \exp\big(2\log\big(\prod_{i=0}^{d}(1 - \tfrac{1}{2}\|\sqrt{P_{X_i|Y_i=y_i}} - \sqrt{U}\|_2^2)\big)\big)\Big)\\
&= -\log\Big(\mathbb{E}_{\boldsymbol{Y}} \big(\prod_{i=0}^{d}(1 - \tfrac{1}{2}\|\sqrt{P_{X_i|Y_i=y_i}} - \sqrt{U}\|_2^2)\big)^2\Big)
\end{aligned}$$

$$= -\log\Big(\mathbb{E}_{\boldsymbol{Y}}\big(\prod_{i=0}^{d}\exp(-\tfrac{1}{2}D_{\frac{1}{2}}(P_{X_i|Y_i=y_i}\|U))\big)^2\Big)$$
$$= -\log\Big(\prod_{i=0}^{d}\mathbb{E}_{Y_i}\exp\big(-D_{\frac{1}{2}}(P_{X_i|Y_i=y_i}\|U)\big)\Big) = \sum_{i=0}^{d} I_{\frac{1}{2}}(X_i;Y_i)$$

□

Unfortunately, this bound is useless because the upper bound of $I_{\frac{1}{2}}(X;\boldsymbol{Y})$ is even increasing as d increases. In fact, this is a trivial bound for any α, because we have Markov chain $X-(X_0,\ldots,X_d)-(Y_0,\ldots,Y_d)=\boldsymbol{Y}$, which implies

$$I_\alpha(X;\boldsymbol{Y}) \le I_\alpha(X_0,\ldots,X_d;Y_0,\ldots,Y_d) = \sum_{i=0}^{d} I_\alpha(X_i;Y_i). \tag{4.45}$$

Of course, it's always possible to find a better bound for $\alpha=\frac{1}{2}$ using other methods.

4.2.3 In the Case of Alpha =1

In [BCG+23], we used "Mrs. Gerber's lemma" to derive similar improved bounds (removing the field size loss). The following notations are used:

Definition 4.2.1 (Binary Entropy)**.** The Shannon entropy for a binary random variable is denoted as

$$H_b: \left|\begin{array}{rcl} [0,1] & \longrightarrow & [0,1] \\ p & \longmapsto & -p\log_2(p)-(1-p)\log_2(1-p)\end{array}\right.$$

Let $H_b^{-1}:[0,1]\mapsto\big[0,\frac{1}{2}\big]$ be the inverse of H_b restricted to $\big[0,\frac{1}{2}\big]$.

Definition 4.2.2 (Binary Convolution $\star$)**.** The convolution for a binary random variable is denoted as

$$\star: \left|\begin{array}{rcl} [0,1]^2 & \longrightarrow & [0,1] \\ x,y & \longmapsto & (1-x)y+x(1-y).\end{array}\right.$$

Definition 4.2.3 (Mrs. Gerber's functions)**.** For any positive integers ℓ,n, let $f_{\mathsf{MI},2^\ell}:[0,1]^{n+1}\to[0,1]$ be the function defined by

$$f_{\mathsf{MI},2^\ell}(\delta_0,\ldots,\delta_n) = 1-H_b\big(\overset{n}{\underset{i=0}{\star}} H_b^{-1}(1-\delta_i)\big).$$

Remark 4.2.1. The function $f_{MI,2^\ell}$ is decreasing with respect to each of its inputs, and is equal to 0 when every $\delta_i=0$.

The lower bound on the minimum number of queries required to achieve a given success rate is

Theorem 4.2.3 (Bounds on Number of Traces [BCG+23])**.**

$$m \geq \frac{d_1(\mathbb{P}_s \| \frac{1}{N})}{f_{\mathsf{MI},2^\ell}(I(X_0;Y_0),\ldots,I(X_d;Y_d))}. \tag{4.46}$$

4.2.4 In the Case of Alpha Tends to Infinity

The other paper [BLR+23] derives similar security bounds for $\alpha = \infty$, which might lead to a better bound because, as we pointed out earlier, the generalized Fano inequality for conditional Sibson's α-information becomes an equality when $\alpha = \infty$. More specifically, the upper bound in [BLR+23] is

Theorem 4.2.4 (Bounds on Number of Traces [BLR+23, Eq. (65)])**.**

$$m \geq \frac{d_\infty(\mathbb{P}_s \| \frac{1}{N})}{\log\Big(1 + (M-1)^d \prod_{i=0}^{d} (\exp I_\infty(X_i;Y_i) - 1)\Big)}. \tag{4.47}$$

Conclusions

The content of this thesis can be broadly classified into three aspects:

The first aspect centered around a widely-used countermeasure against side-channel attacks: generalized code-based masking. Prior research indicates that the security of code-based masking is related to the distance and kissing number of the dual of the masking code. Given a masking code with a specific length and dimension, the larger its dual distance, the more secure the corresponding masking becomes. If two masking codes have the same dual distance, then the one with a smaller kissing number of its dual code is even more secure. Based on this conclusion, **we first studied the possible values of the kissing number of a code with a determined length, dimension, and distance, and derived the upper and lower bounds of the kissing number using linear programming**. On the other hand, we considered: aside from the distance and kissing number, are there other parameters that impact the efficacy of masking? Experimental results indicate that the side-channel resistance of code-based masking can be better evaluated by considering the entire weight distribution of the dual code: **the smaller the lexicographical order based on prefixes of the weight distribution, the better the code-based masking.**

The second aspect involves the assessment of side-channel leakage. To more accurately estimate side-channel leakage, **we proposed a generalized definition of conditional mutual information** and compared it with other generalized conditional information definitions. **We then employed this new metric to estimate the side-channel leakage of unprotected cryptographic devices, deriving the corresponding Fano inequality**. This generalized definition introduces the parameter α: when $\alpha \to 1$, we can replicate the result in previous work; when $\alpha > 1$, this new metric gives better Fano inequality. This generalized definition introduces the parameter α: when $\alpha \to 1$, it replicates the results of previous work; when $\alpha > 1$, this new metric provides a more favorable Fano inequality. Experimental results, based on Hamming weight leakage and Gaussian noise, indicate that as alpha increases, the new Fano inequality can provide us with tighter bounds; notably, when α approaches infinity, the Fano inequality holds as an equality. **On the other hand, for cryptographic implementations protected by code-based masking, we utilized Fano's mutual information to evaluate their side-channel leakage**. We considered the probing model, demonstrating that an attacker needs to deploy at least as many probes as the dual distance of the mask to obtain useful information; when the side channel carries strong noise, the mutual information between the sensitive

variable and the noisy measurements decreases exponentially in σ^2 with an exponent equal to the protection order δ, where σ is the standard deviation of the noise and δ is the dual distance of the masking code.

The third aspect involves the attack evaluation of masked implementations for arithmetic masking and Boolean masking. Compared to the "leakage evaluation" in the second section, this "attack evaluation" section delves into the relationship between the amount of side-channel leakage information and the number of traces required to achieve a given success rate. Universal upper bounds on the probability of success for any type of side-channel attack are derived using alpha-information measures. In particular, when $\alpha = 2$, the generalized mutual information between the sensitive variable and the side-channel leakage can be upper-bounded by the leakage information quantity of each masking share. These also provide lower bounds on the minimum number of queries required to achieve a given success rate. An important issue, resolved in this part, is the removal of the loss factor due to the masking field size.

Other metrics have also been used in similar evaluations. Comparing the boundaries corresponding to different metrics, with the aim of identifying the optimal boundary, will be the objective of the next phase of research.

Appendix: Notes on Leftover Hash Lemma

This appendix is an attempt to improve the *leftover hash lemma* (LHL).

The LHL has been widely applied in various domains of cryptography and complexity theory. The proof of LHL employs the Pointcheval-Zimmer inequality. Since this inequality is not optimal, we tried to replace it with an optimal inequality that was recently proposed in [Rio22]. The improvement brought about by this change is negligible, but it may inspire the search for an α-information version of the leftover hash lemma.

5.1 Background, Notations and Definitions

Imagine a scenario: You have an n-bit secret key uniformly distributed over $\{0,1\}^n$, hence its Rényi entropy is n bits. However, t bits of the key have unfortunately been leaked to the adversary, and you don't know the specific positions of these t bits. Consequently, the entropy of this secret key decreases from n bits to $n-t$ bits. At this point, the key is like a "chicken rib", tasteless when consumed (it only has poor quality randomness), but a pity to discard (as it still contains $n-t$ bits of information entropy). So, how would you handle this partially leaked key?

A possible solution is to find a "randomness extractor" to extract the remaining randomness from this key. The *leftover hash lemma* shows (almost) universal hash functions are good randomness extractors: they can extract "almost uniform" random bits from any random variable X, and the length of extract bits is close to the min-entropy of X, i.e., $\boldsymbol{H}_\infty(X)$. Such extractors are called LHL-based extractors. Before delving into their details, we need to clarify the following concepts.

In this appendix we use H to represent Hash functions, and use $\boldsymbol{H}$ in bold to denote the entropy.

Total Variation Distance

In cryptography, the *total variation distance* (also known as *statistical distance*) is an essential metric that quantifies the disparity between two probability distributions by computing half the sum of their absolute probability differences.

Let P_X, Q_X be two probability distributions over a same sample space $\mathcal{X}$. The *total*

variation distance between P_X and Q_X is

$$\Delta(P_X, Q_X) = \frac{1}{2} \sum_{x \in \mathcal{X}} |p_X(x) - q_X(x)| \tag{5.48}$$

where the factor $\frac{1}{2}$ ensures $0 \leq \Delta(P_X, Q_X) \leq 1$.

The well-known Pinsker's inequality establishes the relationship between TV and Shannon entropy:

$$\Delta^2(P_X, U) \leq \frac{1}{2 \log \mathrm{e}} (\log |\mathcal{X}| - \boldsymbol{H}(P_X)) \tag{5.49}$$

where U represents a uniform distribution over the same sample space as P_X.

Universal Hash Functions

Let X,H be random variables over $\mathcal{X}$ and $\mathcal{H}$ respectively. The sample space $\mathcal{H}$ is a family of Hash functions and each $h \in \mathcal{H}$ maps $x \in \mathcal{X}$ to a bit sequence, $h : \mathcal{X} \to \{0,1\}^v$ where v is a positive integer.

Given $\varepsilon > 0$, if a uniform random variable H satisfy

$$\mathbb{P}\big(H(x_1) = H(x_2)\big) \leq \varepsilon \tag{5.50}$$

for any $x_1, x_2 \in \mathcal{X}$ with $x_1 \neq x_2$, we say $\mathcal{H}$ is ε-*almost universal*. Specifically, $\mathcal{H}$ is called *universal* if $\varepsilon = \frac{1}{2^v}$.

Extractors

A function $Ext : \mathcal{X} \times \{0,1\}^n \to \{0,1\}^m$ is a $(k, \varepsilon) - extractor$, if for all X distributed over $\mathcal{X}$ and $\boldsymbol{H}_\infty(X) \geq k$, we have

$$\Delta(Ext(X, U_n), U_m) \leq \varepsilon. \tag{5.51}$$

where U_n denotes the uniform distribution over $\{0,1\}^n$.

5.2 Leftover Hash Lemma

The leftover Hash lemma was first stated in [ILL89]. It has multiple versions. In this thesis we use the formulation of [Sho06, Thm 8.37].

Theorem 5.2.1 (Leftover Hash Lemma)**.** *Let X be a random variable over $\mathcal{X} = \{0,1\}^n$. $\mathcal{H}$ is a family of Hash functions, $|\mathcal{H}| = 2^d$, and these functions are denoted as $\{0,1\}^d$. Random variable H is uniformly distributed over $\mathcal{H}$ and every $h \in \mathcal{H}$ maps $\mathcal{X}$ to $\{0,1\}^m$.*

Let $Ext : (\mathcal{X}, \mathcal{H}) \to \{0,1\}^{m+d}$ be a function maps $(x, h) \in (\mathcal{X}, \mathcal{H})$ to

$$Ext(x, h) = (h(x), h). \tag{5.52}$$

If $\mathcal{H}$ is $\frac{1+\beta}{2^m}$-almost universal Hash functions, then the function Ext is a (k,ε)-extractor, where $\varepsilon = \frac{\sqrt{\beta+2^{m-k}}}{2}$.

Proof. By definition of total variation distance, one has

$$\Delta(Ext(X,H),U_{m+d}) = \frac{1}{2}\sum_{h\in H,x\in X}\Big|P_{(H(X),H)}(h(x),h) - \frac{1}{2^{m+d}}\Big|. \tag{5.53}$$

Use $V_i, i = 1,\ldots,2^{m+d}$ to denote the items in the summation of (5.53) without taking absolute value, i.e., $P_{(H(X),H)}(h(x),h) - \frac{1}{2^{m+d}}$. Use S_i to denote the corresponding sign vector of V_i, then (5.53) can be written as

$$\begin{aligned}&\Delta(Ext(X,H),U_{m+d})\\ &\quad= \frac{1}{2}\sum_{i=1}^{2^{m+d}} V_i\cdot S_i \le \frac{1}{2}\sqrt{(\sum_{i=1}^{2^{m+d}} V_i^2)(\sum_{i=1}^{2^{m+d}} S_i^2)} = \frac{1}{2}\sqrt{2^{m+d}\cdot(\sum_{i=1}^{2^{m+d}} V_i^2)}\end{aligned}$$

where

$$\begin{aligned}\sum_{i=1}^{2^{m+d}} V_i^2 &= \sum_{x\in\mathcal{X},h\in\mathcal{H}}\big(P_{(H(X),H)}(h(x),h) - \frac{1}{2^{m+d}}\big)^2\\ &= \sum_{x\in\mathcal{X},h\in\mathcal{H}} P^2_{(H(X),H)}(h(x),h) - \frac{2}{2^{m+d}} + \frac{1}{2^{m+d}}\\ &= \|P_{(H(X),H)}\|_2^2 - \frac{1}{2^{m+d}},\end{aligned}$$

where $\|\cdot\|_2$ is Euclidean norm, and $\boldsymbol{H}_2(P_X) = -\log\|P_X\|_2^2$.

Actually, the inequality we obtained above is the Pointcheval-Zimmer inequality [CFPZ09, Lemma 4] (which is also proved earlier in [Sho06, Thm 8.36]):

$$\begin{aligned}\Delta(Ext(X,H),U_{m+d}) &\le \frac{1}{2}\sqrt{N\cdot\exp\Big(-\boldsymbol{H}_2((H(X),H))\Big) - 1}\\ &= \frac{1}{2}\sqrt{N\cdot\|P_{(H(X),H)}\|_2^2 - 1}\end{aligned} \tag{5.54}$$

where $N = 2^{m+d}$. The next step is to bound $\|P_{(H(X),H)}\|_2^2$. One has

$$\begin{aligned}\|P_{(H(X),H)}\|_2^2 &= \sum_{x\in\mathcal{X},h\in\mathcal{H}} p^2_{(H(X),H)}\big(h(x),h\big) = \sum_h p_H^2(h)\sum_x p^2_{H(X)}\big(h(x)|H=h\big)\\ &= \mathbb{P}(H=H')\cdot\big(\mathbb{P}(X=X') + \mathbb{P}(X\neq X')\ \mathbb{P}(H(X)=H(X')|H=h,X\neq X')\big).\end{aligned}$$

Because $\mathbb{P}(H = H') = \frac{1}{2^d}$, $\mathbb{P}(X = X') \leq \frac{1}{2^k}$ when $\boldsymbol{H}_2(X) \geq \boldsymbol{H}_\infty(X) \geq k$, and $\mathbb{P}(H(x_1) = H(x_2)|H = h, x_1 \neq x_2) \leq \frac{1+\beta}{2^m}$ for any $h \in \mathcal{H}$ when $\mathcal{H}$ is $\frac{1+\beta}{2^m}$-almost universal, one has the following inequality

$$\begin{aligned}\Delta(Ext(X,H), U_{m+d}) &\leq \frac{1}{2}\sqrt{2^{m+d} \cdot \frac{1}{2^d}\,(\frac{1}{2^k} + \frac{1+\beta}{2^m}) - 1} \\ &= \frac{1}{2}\sqrt{\beta + 2^{m-k}}.\end{aligned}$$

Let $\varepsilon = \frac{1}{2}\sqrt{\beta + 2^{m-k}}$, then Ext is a (k, ε)-extractor. □

In particular, when $\beta = 0$, Theorem 5.2.1 becomes: If $\mathcal{H}$ is universal Hash functions, then the function Ext is a $(k, 2^{\frac{m-k}{2}-1})$-extractor.

Entropy Loss

The entropy loss of an extractor is one of the considerations for evaluating a randomness extractor. when extract m almost uniform bits from X with $H_\infty(X) = k$, the entropy loss is defined as $L = k - m$.

In Theorem 5.2.1, the input of Ext is random variable X with $H_\infty(X) = k$ and d bits turely random variable H, the output is almost uniform random bits of length $m + d$. So the entropy loss of this Ext is $L = k - m = \log_2 \frac{1}{4\varepsilon^2 - \beta}$. When $\beta = 0$, $L = 2\log_2(\frac{1}{\varepsilon}) - 2$.

5.3 Relation between Statistical Distance and Rényi Entropy

One of the most important steps in this proof is the Pointcheval-Zimmer inequality [CFPZ09, Lemma 4] (see (5.54)), which upper bounds the total variation distance by Rényi entropy of order 2. However, this inequality is not optimal. It can be improved as shown in [Rio22].

In [Rio22], the author derives bounds between statistical distance and Rényi entropy using majorization theory. This section will briefly introduce these bounds.

Majorization

Majorization is a widely-used concept in the theory of inequalities and in linear algebra, which provides a way to compare vectors in terms of the arrangement of their components.

Let X be a random variable over finite alphabet $\mathcal{X}$. The probability distribution of X is denoted as $P_X(x)$, and the probabilities $p_X(x)$ is rearranged as

$$p_{(1)} \geq p_{(2)} \geq \cdots \geq p_{(N)} \tag{5.55}$$

where $p_{(1)} = \max_x p_X(x)$ is the maximum probability, $p_{(i)}$ $(i = 2, 3, \ldots, N)$ is the i-th largest probability.

Given two probability distributions P_X and Q_X, we say that P_X is *majorized* by Q_X, denoted as $P_X \prec Q_X$, if their rearranged probabilities satisfy:

$$\sum_{i=1}^{k} p_{(i)} \leq \sum_{i=1}^{k} q_{(i)}, \qquad \text{for } k = 1, \ldots, n-1; \tag{5.56}$$

$$\sum_{i=1}^{n} p_{(i)} = \sum_{i=1}^{n} q_{(i)}. \tag{5.57}$$

S-Concavity

In [Rio22, Sec. 4], the author introduced the concept of "s-concavity": a quantity is called *s-concave* if it decreases with some certain transformations. As stated in [Rio22, Sec.4], Rényi entropy $\boldsymbol{H}_\alpha$ is s-concave because $p \prec q$ implies

$$\boldsymbol{H}_\alpha(p) \geq \boldsymbol{H}_\alpha(q). \tag{5.58}$$

Total Variation Distance v.s. Rényi Entropy

Use $\Delta = \Delta(P_X, U)$ to represent the total variation distance between P_X and uniform distribution $U \sim \mathcal{U}(N)$; let $K = |\{p \geq \frac{1}{N}\}|$, the following majorization was proved in [Rio22, Eq. (52)]:

$$(\underbrace{\tfrac{1}{N}+\tfrac{\Delta}{K}, \cdots, \tfrac{1}{N}+\tfrac{\Delta}{K}}_{K \text{ times}}, \underbrace{\tfrac{1}{N}-\tfrac{\Delta}{N-K}, \cdots, \tfrac{1}{N}-\tfrac{\Delta}{N-K}}_{N-K \text{ times}}) \prec P_X \prec (\Delta+\tfrac{1}{N}, \underbrace{\tfrac{1}{N}, \cdots, \tfrac{1}{N}}_{L-1 \text{ times}}, R-\tfrac{L}{N}, 0, \ldots, 0) \tag{5.59}$$

where $R = 1 - \Delta$ and $L = \lfloor NR \rfloor$.

Combine (5.58) and (5.59) one has the following lower and upper bounds for α-entropy of X:

$$\boldsymbol{H}_\alpha(X) \geq \tfrac{1}{1-\alpha} \log\big((\Delta + \tfrac{1}{N})^\alpha + \tfrac{L-1}{N^\alpha} + (1 - \Delta - \tfrac{L}{N})^\alpha\big), \tag{5.60}$$

$$\boldsymbol{H}_\alpha(X) \leq \tfrac{1}{1-\alpha} \log\big((\tfrac{1}{N} + \tfrac{\Delta}{K})^\alpha \cdot K + (\tfrac{1}{N} - \tfrac{\Delta}{N-K})^\alpha \cdot (N-K)\big). \tag{5.61}$$

Let $\alpha \to 1$ one has

$$\boldsymbol{H}(X) \geq \log N - (\Delta + \tfrac{1}{N}) \log(N\Delta + 1) - (R - \tfrac{L}{N}) \log(NR - L), \tag{5.62}$$

$$\boldsymbol{H}(X) \leq -(\Delta + \tfrac{K}{N}) \log(\tfrac{\Delta}{K} + \tfrac{1}{N}) - (\tfrac{N-K}{N} - \Delta) \log(\tfrac{1}{N} - \tfrac{\Delta}{N-K}). \tag{5.63}$$

When $\alpha > 1$, (5.61) becomes an upper bound on Δ:

$$\Delta \leq f_\alpha^{-1}\Big(\exp\big((1-\alpha)\boldsymbol{H}_\alpha(X)\big)\Big) \tag{5.64}$$

where f_α^{-1} is the inverse of the function:

$$f_\alpha(\Delta) = (\tfrac{1}{N} + \tfrac{\Delta}{K})^\alpha \cdot K + (\tfrac{1}{N} - \tfrac{\Delta}{N-K})^\alpha \cdot (N-K). \tag{5.65}$$

This inverse exists because $f_\alpha(\Delta)$ is an increasing function and continuous on $[0,1]$ (as $f'_\alpha(\Delta) \geq 0$ when $\Delta \in [0,1]$).

In particular, $f_2(\Delta) = \frac{1}{N} + \frac{N\Delta^2}{K(N-K)}$ and

$$f_2^{-1}(y) = \sqrt{(y - \tfrac{1}{N})\tfrac{K(N-K)}{N}}. \tag{5.66}$$

Combine (5.64) and (5.66) one has a new inequality between total variation distance and 2-entropy $\boldsymbol{H}_2(X)$:

$$\Delta \leq \tfrac{\sqrt{K(N-K)}}{N}\sqrt{N \cdot \exp\Big(-\boldsymbol{H}_2(X)\Big) - 1}. \tag{5.67}$$

It is better than the Pointcheval-Zimmer inequality because $\frac{\sqrt{K(N-K)}}{N}$ is always no greater than $\frac{1}{2}$.

5.4 Applying New Bounds to LHL

By replacing (5.54) in the proof of LHL with (5.64), one obtains the following theorem:

Theorem 5.4.1 (Leftover Hash Lemma)**.** *Let X be a random variable over $\mathcal{X} = \{0,1\}^n$. $\mathcal{H}$ is a family of Hash functions, $|\mathcal{H}| = 2^d$, and these functions are denoted as $\{0,1\}^d$. Random variable H is uniformly distributed over $\mathcal{H}$ and every $h \in \mathcal{H}$ maps $\mathcal{X}$ to $\{0,1\}^m$.*

Let $Ext : (\mathcal{X}, \mathcal{H}) \to \{0,1\}^{m+d}$ be a function maps $(x,h) \in (\mathcal{X}, \mathcal{H})$ to

$$Ext(x,h) = (h(x), h). \tag{5.68}$$

Let $N = 2^{m+d}$ and $K = |p\big((h(x),h)\big) \geq \frac{1}{M}|$. If $\mathcal{H}$ is $\frac{1+\beta}{2^m}$-almost universal Hash functions, then the function Ext is a (k,ε')-extractor, where $\varepsilon' = \frac{\sqrt{K(N-K)}}{N} \cdot \sqrt{\beta + 2^{m-k}}$.

Remark 5.4.1. In this theorem, the entropy loss of Ext is

$$L = k - m = \log_2 \frac{1}{\frac{N^2}{K(N-K)}\varepsilon^2 - \beta}. \tag{5.69}$$

When $\beta = 0$, $L = 2\log_2(\frac{1}{\varepsilon}) - 2\log_2 \frac{N}{\sqrt{K(N-K)}}$. The entropy loss in this theorem is always less than what is presented in Theorem 5.2.1 when $K \neq N/2$.

This theorem tells us that if there is a family of universal hash functions with more extreme value of K (means K is far away from $\frac{N}{2}$), then its corresponding LHL-based extractor has less entropy loss. However, such an improvement is negligible: the new LHL, compared to the original, both achieve an entropy loss of $2\log_2(\frac{1}{\varepsilon}) - O(1)$ when $\beta = 0$. And it has been proven in [RT00] that $2\log_2(\frac{1}{\varepsilon})$ is the smallest possible entropy loss for any extractor when one is concerned with general distinguishers.

5.5 Perspective

In the proof of Theorem 5.4.1 we only use the optimal bound (5.61) when $\alpha = 2$. Since the optimal bound (5.61) is actually valid for all $\alpha > 0$, it might be possible to obtain an α-version of LHL using this bound. This may require proposing the α-version of universality of Hash functions. This is a future research direction.

Bibliography

[ABL01] A Askikhmin, Alexander Barg, and Simon Litsyn. Estimates of the distance distribution of codes and designs. *IEEE Transactions on Information Theory*, 47(3):1050–1061, 2001.

[ABV01] Alexei Ashikhmin, Alexander Barg, and Serge Vladut. Linear codes with exponentially many light vectors. *Journal of combinatorial theory. Series A*, 96(2):396–399, 2001.

[Ari75] Suguru Arimoto. Information measures and capacity of order α for discrete memoryless channels. In Antoine Joux, editor, *Topics in Information Theory, Proc. 2nd Colloq. Math. Societatis János Bolyai*, volume 16, pages 41–52, 1975.

[BCC+14] Julien Bringer, Claude Carlet, Hervé Chabanne, Sylvain Guilley, and Houssem Maghrebi. Orthogonal direct sum masking - a smartcard friendly computation paradigm in a code, with builtin protection against side-channel and fault attacks. In *Information Security Theory and Practice. Securing the Internet of Things - 8th IFIP WG 11.2 International Workshop, WISTP 2014, Heraklion, Crete, Greece, June 30 - July 2, 2014. Proceedings*, pages 40–56, 2014.

[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 Elif Bilge Kavun and Michael Pehl, editors, *Constructive Side-Channel Analysis and Secure Design - 14th International Workshop, COSADE 2023, Munich, Germany, April 3-4, 2023, Proceedings*, volume 13979 of *Lecture Notes in Computer Science*, pages 86–104. Springer, 2023.

[BCP97] Wieb Bosma, John Cannon, and Catherine Playoust. The Magma algebra system. I. The user language. *J. Symbolic Comput.*, 24(3-4):235–265, 1997. Computational algebra and number theory (London, 1993).

[BFG+17] Josep Balasch, Sebastian Faust, Benedikt Gierlichs, Clara Paglialonga, and François-Xavier Standaert. Consolidating inner product masking. In *Advances in Cryptology - ASIACRYPT 2017 - 23rd International Conference on*

the Theory and Applications of Cryptology and Information Security, Hong Kong, China, December 3-7, 2017, Proceedings, Part I, pages 724–754, 2017.

[BGHR14] Nicolas Bruneau, Sylvain Guilley, Annelie Heuser, and Olivier Rioul. Masks will fall off – higher-order optimal distinguishers. In *Advances in Cryptology – ASIACRYPT 2014 - 20th International Conference on the Theory and Application of Cryptology and Information Security, Kaoshiung, Taiwan, R.O.C., December 7-11, 2014, Proceedings, Part II*, pages 344–365, 2014.

[BH01] Koichi Betsumiya and Masaaki Harada. Binary optimal odd formally self-dual codes. *Designs, Codes and Cryptography*, 23:11–22, 2001.

[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.

[BS79] Per Ola Börjesson and Carl-Erik W. Sundberg. Simple approximations of the error function $Q(x)$ for communications applications. *IEEE Trans. Commun.*, 27(3):639–643, 1979.

[Cac97] Christian Cachin. *Entropy measures and unconditional security in cryptography*. PhD thesis, ETH Zurich, 1997.

[CCD00] Christophe Clavier, Jean-Sébastien Coron, and Nora Dabbous. Differential power analysis in the presence of hardware countermeasures. In *Cryptographic Hardware and Embedded Systems—CHES 2000: Second International Workshop Worcester, MA, USA, August 17–18, 2000 Proceedings 2*, pages 252–263. Springer, 2000.

[CFPZ09] Céline Chevalier, Pierre-Alain Fouque, David Pointcheval, and Sébastien Zimmer. Optimal randomness extraction from a diffie-hellman element. In *Annual International Conference on the Theory and Applications of Cryptographic Techniques*, pages 572–589. Springer, 2009.

[CG18] Claude Carlet and Sylvain Guilley. Statistical properties of side-channel and fault injection attacks using coding theory. *Cryptography and Communications*, 10(5):909–933, 2018.

[CGC+21a] Wei Cheng, Sylvain Guilley, Claude Carlet, Jean-Luc Danger, and Sihem Mesnager. Information leakages in code-based masking: a unified quantification approach. *IACR Trans. Cryptogr. Hardw. Embed. Syst.*, 2021(3):465–495, 2021.

[CGC+21b] Wei Cheng, Sylvain Guilley, Claude Carlet, Sihem Mesnager, and Jean-Luc Danger. Optimizing inner product masking scheme by a coding theory approach. *IEEE Trans. Inf. Forensics Secur.*, 16:220–235, 2021.

[CJRR99] Suresh Chari, Charanjit S. Jutla, Josyula R. Rao, and Pankaj Rohatgi. Towards sound approaches to counteract power-analysis attacks. In *CRYPTO*, pages 398–412, 1999.

[CK78] Imre Csiszár and Janos Korner. Broadcast channels with confidential messages. *IEEE transactions on information theory*, 24(3):339–348, 1978.

[CLGR22] Wei Cheng, Yi Liu, Sylvain Guilley, and Olivier Rioul. Toward finding best linear codes for side-channel protections (extended version). *Journal of Cryptographic Engineering*, pages 1–15, 2022.

[Cop94] Don Coppersmith. The data encryption standard (des) and its strength against attacks. *IBM journal of research and development*, 38(3):243–250, 1994.

[CRL+22] Wei Cheng, Olivier Rioul, Yi Liu, Julien Béguinot, and Sylvain Guilley. Side-channel information leakage of code-based masked implementations. In *17th Canadian Workshop on Information Theory, CWIT 2022, Ottawa, ON, Canada, June 5-8, 2022*, pages 51–56. IEEE, 2022.

[Csi95] Imre Csiszár. Generalized cutoff rates and Rényi's information measures. *IEEE Trans. Inf. Theory*, 41(1):26–34, 1995.

[dCGRP19a] 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.

[dCGRP19b] Éloi de Chérisey, Sylvain Guilley, Olivier Rioul, and Pablo Piantanida. An information-theoretic model for side-channel attacks in embedded hardware. In *2019 IEEE International Symposium on Information Theory (ISIT 2019)*, 7 2019.

[Del72] Philippe Delsarte. Bounds for unrestricted codes, by linear programming. *https://api.semanticscholar.org/CorpusID:23131591*, 1972.

[Del75] Philippe Delsarte. The association schemes of coding theory. In *Combinatorics*, pages 143–161. Springer, 1975.

[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.

[DH22] Whitfield Diffie and Martin E Hellman. New directions in cryptography. In *Democratizing Cryptography: The Work of Whitfield Diffie and Martin Hellman*, pages 365–390. ACM, 2022.

[DKL+00] Jean-Francois Dhem, Francois Koeune, Philippe-Alexandre Leroux, Patrick Mestré, Jean-Jacques Quisquater, and Jean-Louis Willems. A practical implementation of the timing attack. In *Smart Card Research and Applications: Third International Conference, CARDIS'98, Louvain-la-Neuve, Belgium, September 14-16, 1998. Proceedings 3*, pages 167–182. Springer, 2000.

[DSVC14] François Durvaux, François-Xavier Standaert, and Nicolas Veyrat-Charvillon. How to certify the leakage of a chip? In *Advances in Cryptology–EUROCRYPT 2014: 33rd Annual International Conference on the Theory and Applications of Cryptographic Techniques, Copenhagen, Denmark, May 11-15, 2014. Proceedings 33*, pages 459–476. Springer, 2014.

[EGI21] Amedeo Roberto Esposito, Michael Gastpar, and Ibrahim Issa. Generalization error bounds via Rényi-, f-divergences and maximal leakage. *IEEE Transactions on Information Theory*, 67, 2021.

[EWG21] Amedeo Roberto Esposito, Diyuan Wu, and Michael Gastpar. On conditional Sibson's α-mutual information. *CoRR*, abs/2102.00720, 2021.

[FB14] Serge Fehr and Stefan Berens. On the conditional Rényi entropy. *IEEE Trans. Inf. Theory*, 60(11):6801–6810, 2014.

[For65] G. David Forney. *Concatenated codes.* PhD thesis, M.I.T. Dept. of Electrical Engineering, 12 1965.

[GM11] Louis Goubin and Ange Martinelli. Protecting aes with shamir's secret sharing scheme. In *Cryptographic Hardware and Embedded Systems–CHES 2011: 13th International Workshop, Nara, Japan, September 28–October 1, 2011. Proceedings 13*, pages 79–94. Springer, 2011.

[GMO01] Karine Gandolfi, Christophe Mourtel, and Francis Olivier. Electromagnetic analysis: Concrete results. In *Cryptographic Hardware and Embedded Systems—CHES 2001: Third International Workshop Paris, France, May 14–16, 2001 Proceedings 3*, pages 251–261. Springer, 2001.

[GMOP15] Andreas Gornik, Amir Moradi, Jürgen Oehm, and Christof Paar. A hardware-based countermeasure to reduce side-channel leakage: Design, implementation, and evaluation. *IEEE Trans. Comput. Aided Des. Integr. Circuits Syst.*, 34(8):1308–1319, 2015.

[GPY09] Leila Golshani, Einollah Pasha, and Gholamhossein Yari. Some properties of Rényi entropy and Rényi entropy rate. *Information Sciences*, 179(14):2426–2433, 2009. Including Special Section – Linguistic Decision Making.

[Gra07] Markus Grassl. Bounds on the minimum distance of linear codes and quantum codes, 2007.

[Hay11] Masahito Hayashi. Exponential decreasing rate of leaked information in universal random privacy amplification. *IEEE Transactions on Information Theory*, 57(6):3989–4001, 2011.

[HRG14a] Annelie Heuser, Olivier Rioul, and Sylvain Guilley. Good is not good enough - deriving optimal distinguishers from communication theory. In *Cryptographic Hardware and Embedded Systems - CHES 2014 - 16th International Workshop, Busan, South Korea, September 23-26, 2014. Proceedings*, pages 55–74, 2014.

[HRG14b] Annelie Heuser, Olivier Rioul, and Sylvain Guilley. Good is not good enough - Deriving optimal distinguishers from communication theory. In Lejla Batina and Matthew Robshaw, editors, *Cryptographic Hardware and Embedded Systems - CHES 2014 - 16th International Workshop, Busan, South Korea, September 23-26, 2014. Proceedings*, volume 8731 of *Lecture Notes in Computer Science*, pages 55–74. Springer, 2014.

[ILL89] Russell Impagliazzo, Leonid A Levin, and Michael Luby. Pseudo-random generation from one-way functions. In *Proceedings of the twenty-first annual ACM symposium on Theory of computing*, pages 12–24, 1989.

[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, 8 2003. Santa Barbara, California, USA.

[IUH22] Akira Ito, Rei Ueno, and Naofumi Homma. On the success rate of side-channel attacks on masked implementations: information-theoretical bounds and their practical usage. In Heng Yin, Angelos Stavrou, Cas Cremers, and Elaine Shi, editors, *Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security, CCS 2022, Los Angeles, CA, USA, November 7-11, 2022*, pages 1521–1535. ACM, 2022.

[IWK20] Ibrahim Issa, Aaron B. Wagner, and Sudeep Kamath. An operational approach to information leakage. *IEEE Transactions on Information Theory*, 66(3):1625–1657, 2020.

[JA04] Petr Jizba and Toshihico Arimitsu. The world according to rényi: Thermodynamics of multifractal systems. *Annals of Physics*, 312(1):17–59, 2004.

[KJJ99] Paul Kocher, Joshua Jaffe, and Benjamin Jun. Differential power analysis. In *Advances in Cryptology—CRYPTO'99: 19th Annual International Cryptology Conference Santa Barbara, California, USA, August 15–19, 1999 Proceedings 19*, pages 388–397. Springer, 1999.

[Kob87] Neal Koblitz. Elliptic curve cryptosystems. *Mathematics of computation*, 48(177):203–209, 1987.

[Koc96] Paul C Kocher. Timing attacks on implementations of diffie-hellman, rsa, dss, and other systems. In *Advances in Cryptology—CRYPTO'96: 16th Annual International Cryptology Conference Santa Barbara, California, USA August 18–22, 1996 Proceedings 16*, pages 104–113. Springer, 1996.

[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.

[LBM15] Liran Lerman, Gianluca Bontempi, and Olivier Markowitch. A machine learning approach against a masked aes: Reaching the limit of side-channel attacks with a learning model. *Journal of Cryptographic Engineering*, 5:123–139, 2015.

[LCGR21] Yi Liu, Wei Cheng, Sylvain Guilley, and Olivier Rioul. On conditional alpha-information and its application to side-channel analysis. In *IEEE Information Theory Workshop, ITW 2021, Kanazawa, Japan, October 17-21, 2021*, pages 1–6. IEEE, 2021.

[Lem75] Abraham Lempel. Matrix factorization over GF(2) and trace-orthogonal bases of GF(2^n). *SIAM J. Comput.*, 4(2):175–186, 1975.

[LLC] Wolfram Alpha LLC. Wolfram|Alpha. (access June 10, 2016).

[LN97] Rudolf Lidl and Harald Niederreiter. *Finite field*. Encyclopedia of mathematics and its applications. Cambridge University Press, 1997. ISBN 10: 0521392314, ISBN 13: 9780521392310.

[LP16] Amos Lapidoth and Christoph Pfister. Two measures of dependence. In *2016 IEEE International Conference on the Science of Electrical Engineering (ICSEE)*, pages 1–5, 2016.

[Mes00] Thomas S Messerges. Securing the aes finalists against power analysis attacks. In *International Workshop on Fast Software Encryption*, pages 150–164. Springer, 2000.

[MOP07] Stefan Mangard, Elisabeth Oswald, and Thomas Popp. *Power analysis attacks — Revealing the secrets of smartcards*. Springer, 2007.

[MRS23] Loïc Masure, Olivier Rioul, and François-Xavier Standaert. A nearly tight proof of duc et al.'s conjectured security bound for masked implementations. In *Smart Card Research and Advanced Applications: 21st International Conference, CARDIS 2022, Birmingham, UK, November 7–9, 2022, Revised Selected Papers*, pages 69–81. Springer, 2023.

[MS77] Florence Jessie MacWilliams and Neil James Alexander Sloane. *The theory of error-correcting codes*, volume 16. Elsevier, 1977.

[OW84] Lawrence H Ozarow and Aaron D Wyner. Wire-tap channel II. In *Workshop on the Theory and Application of of Cryptographic Techniques*, pages 33–50. Springer, 1984.

[PGMP19] Thomas Prest, Dahmun Goudarzi, Ange Martinelli, and Alain Passelégue. 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.

[PGS+17] Romain Poussier, Qian Guo, François-Xavier Standaert, Claude Carlet, and Sylvain Guilley. Connecting and improving direct sum masking and inner product masking. In *Smart Card Research and Advanced Applications - 16th International Conference, CARDIS 2017, Lugano, Switzerland, November 13-15, 2017, Revised Selected Papers*, pages 123–141, 2017.

[PR13] Emmanuel Prouff and Matthieu Rivain. Masking against side-channel attacks: A formal security proof. In *Annual International Conference on the Theory and Applications of Cryptographic Techniques*, pages 142–159. Springer, 2013.

[PV10] Yury Polyanskiy and Sergio Verdú. Arimoto channel coding converse and Rényi divergence. In *2010 48th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 1327–1333, 2010.

[RD01] Vincent Rijmen and Joan Daemen. Advanced encryption standard. *Proceedings of federal information processing standards publications, national institute of standards and technology*, 19:22, 2001.

[Rén61] Alfréd Rényi. On measures of entropy and information. In Jerzy Neyman, editor, *Berkeley Symposium on Mathematical Statistics and Probability*, volume 4.1, pages 547–561. Springer, 1961.

[Rio21] Olivier Rioul. A primer on alpha-information theory with application to leakage in secrecy systems. In *5th conference on Geometric Science of Information (GSI'21), Paris, France, 21-23 July 2021*, Lecture Notes in Computer Science, 2021.

[Rio22] Olivier Rioul. What is randomness? the interplay between alpha entropies, total variation and guessing. In *Physical Sciences Forum*, volume 5, page 30. MDPI, 2022.

[RSA78] Ronald L Rivest, Adi Shamir, and Leonard Adleman. A method for obtaining digital signatures and public-key cryptosystems. *Communications of the ACM*, 21(2):120–126, 1978.

[RT00] Jaikumar Radhakrishnan and Amnon Ta-Shma. Bounds for dispersers, extractors, and depth-two superconcentrators. *SIAM J. Discret. Math.*, 13(1):2–24, 2000.

[Sha53] Claude E. Shannon. The lattice theory of information. *Transactions of the IRE Professional Group on Information Theory*, 1(1):105–107, 2 1953.

[Sho06] Victor Shoup. *A computational introduction to number theory and algebra.* Cambridge University Press, 2006.

[Sib69] Robin Sibson. Information radius. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 14(2):149–160, 1969.

[Sin64] Richard C. Singleton. Maximum distance q -nary codes. *IEEE Trans. Information Theory*, 10(2):116–118, 1964.

[SL80] Gadiel Seroussi and Abraham Lempel. Factorization of symmetric matrices and trace-orthogonal bases in finite fields. *SIAM J. Comput.*, 9(4):758–767, 1980.

[SLC+21] Patrick Solé, Yi Liu, Wei Cheng, Sylvain Guilley, and Olivier Rioul. Linear programming bounds on the kissing number of q-ary codes. In *IEEE Information Theory Workshop, ITW 2021, Kanazawa, Japan, October 17-21, 2021*, pages 1–5. IEEE, 2021.

[SMY09] François-Xavier Standaert, Tal Malkin, and Moti Yung. A unified framework for the analysis of side-channel key recovery attacks. In Antoine Joux, editor, *Advances in Cryptology - EUROCRYPT 2009, 28th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Cologne, Germany, April 26-30, 2009. Proceedings*, volume 5479 of *Lecture Notes in Computer Science*, pages 443–461. Springer, 2009.

[SSYA19] Ilia Shumailov, Laurent Simon, Jeff Yan, and Ross Anderson. Hearing your touch: A new acoustic side channel on smartphones. *CoRR*, abs/1903.11137, 2019.

[TH18] Marco Tomamichel and Masahito Hayashi. Operational interpretation of Rényi information measures via composite hypothesis testing against product and markov distributions. *IEEE Trans. Inf. Theory*, 64(2):1064–1082, 2018.

[VCS09] Nicolas Veyrat-Charvillon and François-Xavier Standaert. Mutual information analysis: how, when and why? In *Cryptographic Hardware and Embedded Systems-CHES 2009: 11th International Workshop Lausanne, Switzerland, September 6-9, 2009 Proceedings*, pages 429–443. Springer, 2009.

[vEH14] Tim van Erven and Peter Harremoës. Rényi divergence and Kullback-Leibler divergence. *IEEE Trans. Inf. Theory*, 60(7), 2014.

[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.

[Wei91] Victor K.-W. Wei. Generalized Hamming weights for linear codes. *IEEE Trans. Inf. Theory*, 37(5):1412–1418, 1991.

[WMCS20] Weijia Wang, Pierrick Méaux, Gaëtan Cassiers, and François-Xavier Standaert. Efficient and private computations with code-based masking. *IACR Trans. Cryptogr. Hardw. Embed. Syst.*, 2020(2):128–171, 2020.

[Wyn75] Aaron D Wyner. The wire-tap channel. *Bell system technical journal*, 54(8):1355–1387, 1975.

[XBZ12] Zhi Xu, Kun Bai, and Sencun Zhu. Taplogger: Inferring user inputs on smartphone touchscreens using on-board motion sensors. In *Proceedings of the fifth ACM conference on Security and Privacy in Wireless and Mobile Networks*, pages 113–124, 2012.

ECOLE
DOCTORALE

Titre : Évaluation de la Sécurité Contre les Attaques par Canaux Auxiliaires : Perspectives a partir d'une Vision Théorique de l'Information

Mots clés : Canaux Auxiliaires, Théorie de l'Information, Masquage, Évaluation de la sécurité

Résumé : L'utilisation répandue des dispositifs cryptographiques met en évidence le besoin de leur fonctionnement sécurisé sur des plateformes physiques. Des fuites d'informations involontaires, telles que la durée d'exécution, la puissance, et les émissions électromagnétiques, peuvent permettre aux attaquants de déduire les clés secrètes utilisés via des attaques par canaux cachés (SCAs). L'importance des SCAs a intensifié la recherche sur la sécurité des dispositifs cryptographiques, avec l'émergence de mesures théoriques de l'information comme outils d'évaluation efficaces. Dans ce contexte, les objectifs centraux de cette thèse sont de **quantifier les fuites par canaux cachés**, **évaluer la sécurité des dispositifs cryptographiques face aux SCAs** (à la fois non protégés et masqués), et de **trouver une méthode pour élaborer des codes de masquage plus efficaces**. Pour la construction du code de masquage, nous trouvons des bornes par programmation linéaire sur le nombre de contact des codes q-aires. Nous montrons également que le code est d'autant plus performant que le polynome enumérateur des poids du code dual est minimal pour l'ordre lexicographique.

Concernant l'évaluation des fuites par canaux cachés, nous introduisons une nouvelle métrique d'information, appelée alpha-information conditionnelle de Sibson. Elle peut exprimer par une formule explicite propice aux évalution numériques et vérifie plusieurs propriétés utiles. En utilisant cette mesure, nous examinons les fuites par canaux cachés des dispositifs non protégés. De plus, nous utilisons l'information mutuelle de Fano pour évaluer les fuites par canaux cachés des implémentations masquées basées sur un code sous un modèle de sondage.

Enfin, pour l'évaluation de la sécurité des implémentations masquées, nous utilisons l'alpha-information pour évaluer les implémentations de masquage arithmétique et booléen. Nous définissons des limites inférieurs universelles sur le nombre de requêtes nécessaires pour atteindre un taux de succès donné.

Title : Security Assessment Against Side-Channel Attacks: Insights from an Information-Theoretic Perspective

Keywords : Side Channel, Information Theory, Masking, Security Assessment

Abstract : In today's world, the widespread use of cryptographic devices highlights the need for their secure operation. Unintended leakages, like time, power, and electromagnetic emissions, can allow attackers to deduce secret keys via side-channel attacks (SCAs). Evaluating the security of cryptographic devices against SCAs is important for both the industrial and academic sectors, and information-theoretic metrics turn out to be effective tools. "Masking" stands out as a key countermeasure, with ongoing discussions on its optimization and the security of its implementations.

In light of this context, the central aims of this thesis are to **quantify side-channel leakage**, **appraise the security of cryptographic devices against SCAs** (both unprotected and masked), and to **explore methodologies for formulating more potent masking codes**.

For masking code construction, we establish linear programming bounds for the kissing number of q-ary linear codes, guided by recent findings on optimized code-based masking performance related to the dual code's kissing number. In addition, we demonstrate the connection between code-based masking efficacy and the whole weight enumeration of the dual of the masking code. The lexicographical order based on weight distribution prefixes is proposed for selecting ideal masking codes.

Regarding side-channel leakage evaluation, we introduce a novel information metric, called conditional Sibson's alpha-information. This metric has an explicit expression and possesses several beneficial properties. Utilizing this metric, we delve into the side-channel leakage of unprotected devices. Additionally, we use Fano's mutual information to evaluate the side-channel leakage of code-based masked implementations under probing model.

Lastly, when considering the security assessment of masked implementations, we utilize the alpha-information measure to appraise the security of both arithmetic and Boolean masking implementations. We derive universal bounds on the probability of success of any type of side-channel attack. These also provide lower bounds on the minimum number of queries required to achieve a given success rate.

Institut Polytechnique de Paris
91120 Palaiseau, France