What can information guess ? : Towards information leakage quantification in side-channel analysis

Wei Cheng

INSTITUT POLYTECHNIQUE DE PARIS

NNT : 2021IPPAT044

Thèse de doctorat

What Can Information Guess?

Towards Information Leakage Quantification in Side-Channel Analysis

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

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

Thèse présentée et soutenue à Palaiseau, le December 10, 2021, par

WEI CHENG（成玮）

Composition du Jury :

David NACCACHE Professeur, École Normale Supérieure de Paris, FRANCE	Président
François-Xavier STANDAERT Professeur, Université Catholique de Louvain La Neuve, BELGIQUE	Rapporteur
Michael C. GASTPAR Professeur, École Polytechnique Fédérale de Lausanne, SUISSE	Rapporteur
Sihem MESNAGER Maitre de Conférence (HDR), Université Paris 8, FRANCE	Examinatrice
Joseph BOUTROS Professeur, Texas A&M University at Qatar, QATAR	Examinateur
Annelie HEUSER CR CNRS, IRISA, Rennes, FRANCE	Examinatrice
Olivier RIOUL Professeur, Télécom Paris, FRANCE	Directeur de thèse
Sylvain GUILLEY Professeur, Secure-IC, FRANCE	Co-directeur de thèse
Yongbin ZHOU Professeur, Nanjing University of Science and Technology, CHINA	Invité
Jean-Luc DANGER Professeur, Télécom Paris, FRANCE	Co-directeur de thèse

Abstract

Cryptographic algorithms are nowadays prevalent in establishing secure connectivity in our digital society. Such computations handle sensitive information like encryption keys, which are usually very exposed during manipulation, resulting in a huge threat to the security of the sensitive information concealed in cryptographic components. Eventually, the compromission of keys results in the compromission of the whole connected system. In the field of embedded systems security, side-channel analysis is one of the most powerful techniques against cryptographic implementations. It consists in the extracting the sensitive data by exploiting certain physically observable leakages during the execution of the cryptographic components. As such attacks pose a real threat, various protections have been studied and developed, wherein the random masking is one of the most well-established. Basically, masking provides provable security against side-channel analysis by splitting (at a systematic level) any sensitive variable into several random shares. A fundamental parameter for a masking scheme is its security order t. Thereby, any adversaries aiming at key-recovering must employ strictly more than t shares to launch a higher-order attack, whereas the data complexity increases exponentially in t (under noisy conditions), which in turn is providing security against side-channel analysis.

The main subject of this thesis is the measurable side-channel security of cryptographic implementations, particularly in the presence of random masking. Overall, this thesis consists of two topics. One is the leakage quantification of the most general form of masking equipped with the linear codes, so-called code-based masking; the other one is exploration of applying more generic information measures in a context of side-channel analysis. Two topics are inherently connected to each other in assessing and enhancing the practical security of cryptographic implementations.

Regarding the former, we propose a unified coding-theoretic framework for measuring the information leakage in code-based masking. Specifically, our framework builds formal connections between coding properties (including the dual distance and the kissing number of codes) and leakage metrics in side-channel analysis (including signal-to-noise ratio and mutual information). As has been reported in literature, different linear codes have distinct impact on the side-channel

resistance of a code-based masking. Those formal connections enable us to push forward the quantitative evaluation on how the linear codes can affect the concrete security of all code-based masking schemes, including non-redundant cases (e.g., inner product masking, direct sum masking, etc) and redundant cases (e.g., polynomial masking based on Shamir's secret sharing, etc). Moreover, relying on our framework, we consolidate code-based masking by providing the optimal linear codes in the sense of maximizing the side-channel resistance of the corresponding masking scheme. Our framework is finally verified by attack-based evaluation, where the attacks utilize maximum-likelihood based distinguishers and are therefore optimal.

Regarding the latter, we present a full spectrum of application of alpha-information, a generalization of (Shannon) mutual information, for assessing side-channel security. In side-channel analysis, mutual information is frequently adopted in an information-theoretic evaluation of side-channel security level. By a communication channel model, mutual information provides an upper bound on the success rate of any attacks given a fixed set of side-channel measurements, or it gives a lower bound on the number of measurements to achieve a specific success rate. However, those bounds are loose, even much looser in highly noisy scenarios. In this thesis, we propose to utilize a more general information-theoretic measure, namely alpha-information (α-information) of order α. The new measure also gives the upper bound on success rate and the lower bound on the number of measurements. More importantly, with proper choices of α, α-information provides very tight bounds, in particular, when α approaches to positive infinity, the bounds will be exact. As a matter of fact, maximum-likelihood based distinguishers will converge to the bounds. Therefore, we demonstrate how the two world, information-theoretic measures (bounds) and maximum-likelihood based side-channel attacks, are seamlessly connected in side-channel analysis.

In summary, our study in this thesis pushes forward the evaluation and consolidation of side-channel security of cryptographic implementations. From a protection perspective, our quantitative outputs allow to empower practical masked implementations with the highest achievable side-channel resistance when equipped with the optimal linear codes. Therefore, we provides a best-practice guideline for the application of code-based masking. From an evaluation perspective, the application of alpha-information enables practical evaluators and designers to have a more accurate (or even exact) estimation of concrete side-channel security level of their cryptographic chips.

Résumé

Les algorithmes cryptographiques jouent un rôle prédominant pour établir une connectivité sécurisée dans notre société numérique actuelle. Ces calculs traitent des informations sensibles telles que des clés de chiffrement, qui sont généralement très exposées lors de la manipulation, ce qui représente une menace énorme pour la sécurité des informations sensibles dans les composants cryptographiques et l'ensemble des systèmes connectés. Dans le domaine de la sécurité des systèmes embarqués, l'analyse des canaux auxiliaires est l'une des techniques les plus puissantes contre les implémentations cryptographiques. Elle consiste à extraire les données sensibles en exploitant certaines fuites physiquement observables lors de l'exécution des composants cryptographiques. Comme ces attaques représentent une menace réelle, diverses protections ont été étudiées et développées. Parmi celles-ci, le masquage aléatoire est l'une des mieux établies. Fondamentalement, le masquage offre une sécurité prouvable contre l'analyse des canaux auxiliaires en divisant toute variable sensible en plusieurs parts aléatoires. Un paramètre fondamental pour un schéma de masquage est son ordre de sécurité $t \geq 0$. Ainsi, tout adversaire visant à récupérer des clés doit utiliser strictement plus de t parts pour lancer une attaque d'ordre supérieur, alors que la complexité des données augmente de façon exponentielle en t (sous des conditions liées à la puissance du bruit de mesure), ce qui à son tour fournit une sécurité contre les attaques par canaux auxiliaires.

Le sujet principal de cette thèse concerne la sécurité mesurable des canaux auxiliaires des implémentations cryptographiques, en particulier en présence de masquage aléatoire. Globalement, cette thèse se compose de deux sujets. L'un est la quantification des fuites de la forme la plus générale de masquage équipé des codes linéaires, dit masquage à base de code ; l'autre est l'exploration de l'application de mesures d'information plus génériques dans un contexte d'analyse de canaux auxiliaires. Ces deux sujets sont intrinsèquement liés l'un à l'autre dans l'évaluation et l'amélioration de la sécurité pratique des implémentations cryptographiques.

Pour ce qui concerne le premier sujet, nous proposons un cadre théorique de codage unifié pour mesurer la fuite d'informations dans le masquage basé sur les codes. Plus précisément, notre cadre établit des connexions formelles entre les propriétés de codage (y compris la distance

duale et le nombre de points de contact permettant un sondage) et les métriques de fuite dans l'analyse des canaux auxiliaires (y compris le rapport signal sur bruit et l'information mutuelle). Comme cela a été rapporté dans la littérature, différents codes linéaires ont un impact distinct sur la résistance aux attaques par canal auxiliaires d'un masquage basé sur un code. Ces connexions formelles nous permettent de faire avancer l'évaluation quantitative sur la façon dont les codes linéaires peuvent affecter la sécurité concrète de tous les schémas de masquage basés sur les codes, y compris les cas non-redondants (par exemple, masquage basé sur le produit scalaire, masquage par somme directe, etc.) et les cas redondants (par exemple, masquage polynômial basé sur le partage secret de Shamir, etc.). De plus, en nous appuyant sur notre cadre théorique, nous consolidons le masquage basé sur le code en fournissant les codes linéaires optimaux dans le sens qu'ils maximisent la résistance des canaux auxiliaires du schéma de masquage correspondant. Notre formalisation est finalement vérifiée par une évaluation basée sur les attaques, où les attaques utilisent des distingueurs basés sur le maximum de vraisemblance et sont donc optimales.

Concernant le deuxième sujet, nous présentons un spectre complet d'applications d'une variante de l'information mutuelle de Shannon, appelée "alpha-information". Il s'agit d'une généralisation de l'information mutuelle permettant d'évaluer la sécurité d'une implémentation face aux attaques par canaux auxiliaires. Dans l'analyse des canaux auxiliaires, l'information mutuelle est fréquemment adoptée dans une évaluation du point de vue de la théorie de l'information afin d'établir le niveau de sécurité des attaques par canaux auxiliaires. Par un modèle de canal de communication, l'information mutuelle fournit une limite supérieure sur le taux de succès de toute attaque étant donné un ensemble fixe de mesures de canaux auxiliaires. Alternativement, elle donne une limite inférieure sur le nombre de mesures pour atteindre un taux de succès spécifique. Cependant, ces limites sont souples, encore plus souples dans des scénarios à fort bruit de mesure. Dans cette thèse, nous proposons d'utiliser une mesure plus générale du point de vue de la théorie de l'information, à savoir l'information alpha (α-information) d'ordre α. La nouvelle mesure donne également la limite supérieure du taux de succès et la limite inférieure du nombre de mesures. Ce qui est remarquable, c'est qu'avec des choix appropriés de α, l'information α fournit des bornes très proches de la réalité ; en particulier, lorsque α tend vers l'infini (positif), les limites seront exactes. En fait, les distingueurs basés sur le maximum de vraisemblance convergeront vers les limites. Par conséquent, nous démontrons comment les deux mondes, à savoir les mesures du point de vue de la théorie de l'information (limites) et

les attaques par canaux auxiliaires basées sur le maximum de vraisemblance, sont parfaitement connectés dans l'analyse par canaux auxiliaires.

En résumé, notre étude dans cette thèse fait avancer l'évaluation et la consolidation de la sécurité des canaux auxiliaires des implémentations cryptographiques. Du point de vue de la protection, nos sorties quantitatives permettent de mettre en œuvre des implémentations masquées concrètes, implémentant une résistance vis-à-vis des attaques par canal auxiliaire la plus élevée possible lorsqu'elles sont équipées des codes linéaires optimaux. Par conséquent, nous fournissons un guide des meilleures pratiques pour l'application du masquage basé sur le code. Du point de vue de l'évaluation, l'application de l'alpha-information permet aux évaluateurs et concepteurs (développeurs) d'avoir une estimation plus précise (voire exacte) du niveau de sécurité concret des canaux auxiliaires émanant de leurs puces cryptographiques.

This dissertation is dedicated to my respectful parents,
my beloved soulmate Tong Zhao,
my brother and my whole family for everything.

I would never have made it this far without your selfless love,
constant support and inspiring encouragement.

This dissertation is especially dedicated to the memory of my grandfather,
those indelible and warm memories will stay with me all my life.

Acknowledgments

It has been a long journey to be here. I am really happy that I have enjoyed so much at research work and in life, which gives me courage and faith in being a mature researcher and a better person. I could not make it this far without help, support and encouragement from many people and I would like to express my sincere gratitude to them.

First and foremost, I'm deeply grateful to my Ph.D supervisors: Olivier Rioul, Sylvain Guilley and Jean-Luc Danger for granting me this great opportunity and the honor to work with them. Thank you for supporting me, guiding me and encouraging me, even when faced with some uncontrollable difficulties. They always bring me to a new place full of insights and inspiring ideas. I really appreciate that each of them taught me different aspects of research in distinct approaches from more theoretical to very real-world instances. They have set excellent examples for being a scientific researcher and more globally a gentle person. Moreover, special thanks to A. Prof. Laurent Sauvage for helping me at the beginning of my Ph.D journey.

I would like to express my gratitude to juries for their willingness of being committee members, their time and efforts: Prof. David Naccache for kindly presiding it; Prof. François-Xavier Standaert and Prof. Michael C. Gastpar for reviewing the manuscript and providing valuable feedback, which has improved this thesis; Prof. Sihem Mesnager, Prof. Joseph Boutros, A. Prof. Annelie Heuser and Prof. Yongbin Zhou for their special interests, discussions, knowledge and recommendations.

My gratitude also goes to many incredible collaborators and several subjects would not have been possible without them. I am very fortunate to have worked and continue to work with many excellent professors, colleagues and labmates. Special thanks to Prof. Claude Carlet, Prof. Sihem Mesnager, Prof. Patrick Solé, Prof. Yongbin Zhou and Prof. Naghmeh Karimi for insightful discussions and exchanges. I would like to specifically thank my friends, Dr. Oualid Trabelsi, Dr. Alexander Schaub, Dr. Sébastien Carré, Dr. Khaled Karray, Dr. Youssef Souissi, Dr. Sofiane Takarabt, Mme. Yi Liu, M. Julien Béguinot, and M. Trevor Kroeger for sharing new ideas and interesting discussions in various aspects.

A special acknowledgment goes to Prof. Yongbin Zhou for introducing and leading me into the world of cryptography and side-channel analysis, and for motivating me to explore more in

this field. I would like to thank all my former colleagues for those days we worked together and had fun together, especially M. Chao Zheng, Dr. Yiwen Gao, M. Jingdian Ming, Mme. Huizhong Li, Mme. Qian Zhang, M. Guang Yang and all the others. I have enjoyed and learned a lot in working and collaborating with all of you.

My special thanks to the Chinese Scholarship Council (CSC) and Télécom Paris. CSC offered me a scholarship which covered my living expenses for three years. Moreover, Télécom Paris offered me complementary support and a three-month extension for finalizing the thesis. I would like to thank the doctoral school, especially Prof. Bruno Thedrez, Mme. Florence Besnard and Prof. Alain Sibille for their kind help and support in the last three years.

This thesis will end but the list of acknowledgments will never come to an end. At last, I would like to thank my girlfriend Tong Zhao, my parents, my brother and my family. It is a very decree by destiny to get to know you and we have crossed oceans and mountains to meet each other, finally we share everything together. This journey of exploration and discovery is full of challenges and exceptions and I am so fortunate with your companion and support. I would like to thank my parents for their unconditional love and continuous support since my very first glance in this world. Family is more important than anything and I always appreciate what they have done for me in those difficult times. My gratitude finally goes to my big loving family, the love, support and freedom from them have motivated me to be a better person and will be with me for my whole life.

Wei Cheng

Massy, December 2021

Contents

List of Figures

List of Tables

List of Abbreviations

AES	Advanced Encryption Standard.
ASIC	Application Specific Integrated Circuit.
AWGN	Additive White Gaussian Noise.
BKLC	Best Known Linear Code.
BM	Boolean Masking.
CI	Correlation Immunity, usually with a order d.
CPA	Correlation Power Analysis.
DES	Data Encryption Standard.
DoM	Difference of Means.
DPA	Differential Power Analysis.
DPI	Data Processing Inequality.
DSM	Direct Sum Masking.
ECC	Elliptic-Curve Cryptography.
FIA	Fault Injection Attack/ Analysis.

FPGA	Field-Programmable Gate Array.
GCM	Generalized Code-based Masking.
GE	Guessing Entropy.
HOOD	Higher-Order Optimal Distinguisher.
IC	Integrated Circuit.
IPM	Inner Product Masking.
LEMS	Low Entropy Masking Scheme.
LS	Leakage Squeezing.
MDS matrix	Maximum Distance Separable matrix.
MI	Mutual Information.
MIA	Mutual Information Analysis.
ML	Maximum-Likelihood, e.g., ML-based distinguishers.
RS code	Reed-Solomon code.
RSA	Rivest-Shamir-Adleman algorithm.
RSM	Rotating S-Box Masking, a special instance of LEMS.
SCA	Side Channel Attack/ Analysis.
SNI	Strong Non-Inference, resp. NI for Non-Inference.
SNR	Signal-to-Noise Ratio.
SR	Success Rate.
SSS scheme	Shamir's Secret Sharing scheme, e.g., with parameters (n, t).
UEP	Uniform Expansion Property.

Part I

Introduction and Contributions

CHAPTER 1

Introduction

Contents

1.1 Cryptography & Cybersecurity

Nowadays, abundant electronic devices are proliferating in our daily life, such as SIM cards, cell-phones, bank cards, etc. For instance, from Eurosmart's survey [1], there are about 9.54 billion shipped units of secure elements. Particularly, the Telecom market has closed 2020 with around 5,1 billion units shipped, including 309 million units shipped for eSIM and a 4,8 billion units for SIM, which experienced a significant increase. Those secure elements are widely deployed in

[1] Eurosmart, https://www.eurosmart.com/2019-shipments-and-2020-outlook/

telecom, financial services, device manufacturers, etc. However, such secure elements usually handle some sensitive information, which are very exposed during computations when loading, manipulating, and storing them, resulting in massive scales of vulnerabilities and attacks in practice. As a consequence, improving their security has become the highest priority task.

In this respect, modern cryptography is the cornerstone to build the chain of trust and security. It plays a fundamental and pivotal role in establishing secure connectivity in this emerging digital era. In other words, cryptography makes secure communications between different parties possible and evolves along with computation and communication technologies. Basically, cryptography provides five primary functionalities including confidentiality, integrity, authentication, non-repudiation and key exchange. Those functionalities are well-established on the basis of various concepts of mathematics such as information-theoretic security, computational-complexity theory, number theory, coding theory, probability theory and so on.

Relying on mathematical tools, it is feasible to devise and construct theoretically secure cryptographic algorithms or protocols. In the field of symmetric key cryptography, the Data Encryption Standard (DES) [116] and its successor Advanced Encryption Standard (AES) [117] is one of the most important algorithms that was published two decades ago by National Institute of Standards and Technology (NIST). On the contrary, in the field of public key cryptography, RSA [141] and ECC [92, 109] are two well-known instances that are based on the intractability of the corresponding mathematical problems.

1.2 The Root of Security & the Chain of Security

As a basic rule and common consensus in cryptography, Kerckhoffs's principle, dated back to 19th century, states that a cryptosystem should be secure, even if everything about the system is accessible to adversaries except the key [90, 91]. It is followed and reformulated by Claude E. Shannon in 1949 known as Shannon's maxim: "one ought to design (crypto) systems under the assumption that the enemy will immediately gain full familiarity with them" [147]. The keys in a cryptosystem form the basis for the root of trust that is critical to the system. Theoretically, above constructions (e.g., AES, RSA, ECC, etc) are computationally secure in this regard under the black-box assumption, wherein an adversary can only access to inputs and outputs of a cryptosystem.

However, in practical applications, the keys are not static but manipulated dynamically in the digital world. Indeed, each stage of manipulations (computations) shall expose those keys,

which leads to the demand of chain of security to guarantee security in reality. As a matter of fact, any digital devices will leak physically observable information [108] of internal states during executions. Although mathematical proofs of security for cryptographic algorithms are fundamental and indispensable, they usually cannot guarantee the practical security of the corresponding cryptographic implementations. In reality, those cryptographic algorithms must be run in some physical devices. Therefore, those physical observations usually violate the black-box setting assumption that an adversary can only access to the inputs and outputs of a cryptographic algorithm. As knowing certain observable information makes it advantageous to adversary, the black-box model is lifted to gray-box setting by considering any (abstract) form of observable leakages existing in practice. Accordingly, the attacks exploiting those physically observable leakages are called *physical attacks*.

1.3 Side-Channel Analysis

Side-channel analysis (SCA) is among the most powerful physical attacks against cryptographic implementations. Since the seminal works [94, 95], a very large amount of SCAs have been proposed by exploiting various observable physical leakages in practice. Those physical leakages include but not limited to the running time [59, 94], the power consumption [46, 95], the electromagnetic emanations [71, 132], the acoustic emission [20, 73], the photonic emission [27, 67, 96], etc., and more exploitable leakages emerge as technology improves (e.g., Nanotechnology) and in-depth understanding of behaviors of elementary circuits, like micro-architectural data leakages [72, 93, 98, 104]. Essentially, any measurable secret-dependent information or behaviors of the underlying cryptographic devices can be exploited to launch a successful side-channel attack.

In principle, side-channel analysis consists of extracting the sensitive information from noisy measurements. It is commonly classified into two classes depending on the ability of the adversary and corresponding setting.

- **Non-profiling attacks.** An adversary attempts to extract the sensitive information by correlating side-channel measurements and hypothetical leakages. Several well-known attacks are simple power analysis (SPA) [94], differential power analysis (DPA) [95], correlation power analysis (CPA) [16], mutual information analysis (MIA) [74, 163], etc.

- **Profiling attacks.** They are two-phrase attacks. an adversary is assumed to possess an identical device to build some exact profiles on the leakage behaviors and then apply these

profiles during the attack phrase. Some well-known instances are template attack [31], stochastic attack [143], etc. In particular, the template attack is known as the most powerful side-channel attack knowing the leakage model.

Additionally, machine learning (including deep learning) techniques have been adapted into side-channel analysis in both non-profiling [121, 133, 155] and profiling settings [10, 19, 106, 168, 171]. In essence, side-channel classifies different key hypotheses relying on observations, in which learning-based techniques shall amplify those attacks dramatically. However, those learning-based attacks tolerate a loss of interpretability on results, even in some restricted scenarios.

1.4 Protections and Code-based Masking

In order to protect cryptographic chips (implementations) against SCA, many countermeasures have been proposed, wherein three main routines are masking, shuffling and hiding. Specifically, masking schemes [30, 49, 102, 139] randomize the dependency between sensitive data and leakages by dividing each sensitive variable into several random shares to thwart SCA, while Shuffling schemes [50, 83, 140] randomize the order of operations during the executions. Quite differently, by circuit-level alteration, hiding-based countermeasures [46, 102, 134] attempt to make the leakages uniformly independent to the data processed, while it is difficult to have any guarantee [85]. Among them, masking schemes are a class of the most attractive and frequently used techniques against SCA, since they provide formally provable security and could be implemented on algorithmic-level without any hardware alteration.

1.4.1 Masking Schemes

Featured with the favorable provable security, masking has triggered a fruitful line of works, ranging from theoretical constructions of secure components (usually called gadget) to practical resilience evaluations by side-channel attacks. Typically, the key parameter of a masking scheme is the security order t under the probing model [86], which indicates the least order $(t+1)$ of a successful attack must have. In a t-th order secure masking, each sensitive variable into at least $t+1$ shares. The rationale is that, the attack complexity increases exponentially with the number of shares [30, 128] given a sufficient amount of noise, while the implementation cost increases only quadratically (or cubically in higher-order glitch-free implementations [79]).

Various masking schemes have been proposed since 1999, as shown in Fig. 1.1. Typically instances include Boolean masking [30], Inner Product masking (IPM) [2, 3], Leakage Squeezing (LS) [23, 24] and Direct Sum masking (DSM) [17, 123]. Note that those proposals marked in blue are the first proposals of the corresponding schemes. An exception exists for the original IPM [4], since there exist some first-order information leakages that are fixed in the improved one [2]. To the best of our knowledge, the generalized code-based masking (GCM) [35, 164] is the most generic scheme in this respect [1]. In particular, polynomial masking [77, 131] is also a special case of GCM, which is built upon Shamir's secret sharing (SSS) scheme [145].

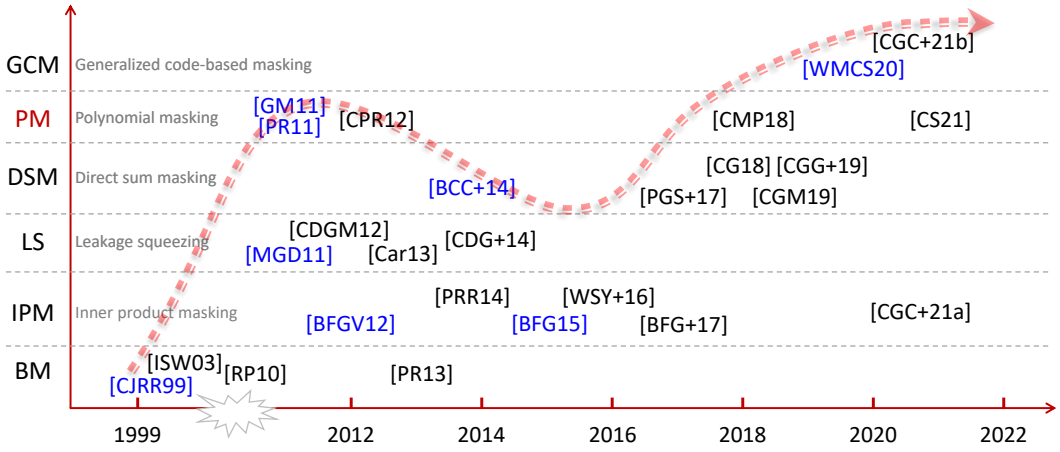


Figure 1.1: Various proposals of masking schemes with corresponding constructions, security assessment, and some variants.

Naturally, two questions arise: first, *how to measure information leakage in different schemes?* and second, *for each scheme, how to choose optimal codes (or parameters)?*

1.4.2 Generalizing to Code-based Masking

Code-based masking follows the generalization trend and unifies many schemes by concentrating on the encodings in sharing. In code-based masking, two linear codes are involved, namely $\mathcal{C}$ and $\mathcal{D}$. The only requirement is that there is no nonzero codeword in their intersections [35, 164]. As a result, the resilience of a code-based masking against side-channel analysis depends highly on the two linear codes, in which the coding-theoretic properties shall be connected to algebraic complexity from a view of (pseudo)-Boolean function.

[1] In the sequel, we call the code-based masking in the most general scenario for simplicity.

The first representative scheme is IPM, in which the encoding is similar to the simplest Boolean masking except that each share is equipped with a linear function (multiplied by a public constant). It consumes n parameters in an n-share setting and enjoys the simple structure that can be implemented quite efficiently [3]. As a special instance of non-redundant code-based masking, two linear codes in IPM are complementary, resulting in a super simplification when evaluating its side-channel resistance. Indeed, we demonstrate that the side-channel security of IPM only depends on properties of the code $\mathcal{D}$ [37]. More generally, only the code $\mathcal{D}$ matters in any non-redundant code-based masking like DSM.

Another typical example is the polynomial masking that is based on the SSS scheme. It also employs n public parameters in an n-share setting, but forms an entirely different encoding. Essentially, the encoding in SSS-based masking can be reformulated and connected to the Reed-Solomon (RS) codes [29, 101]. Considering an (n, t)-SSS based sharing as depicted in Fig. 1.2, it forms n shares while provides a t-th order privacy (side-channel resistance) rather than nt parameters in a random setting. From a coding-theoretic perspective, the RS code is optimal in a given finite field which achieves the Singleton bound [149]. However, as shown in [29], distinct public points play a role in the resilience and the efficiency of the protection. Therefore, the questions above still remain.

In the above two representative schemes, we can refine the second question: *how to choose n parameters to maximize the side-channel protection?* More straightforwardly, how to choose public points in the case of SSS-based masking.

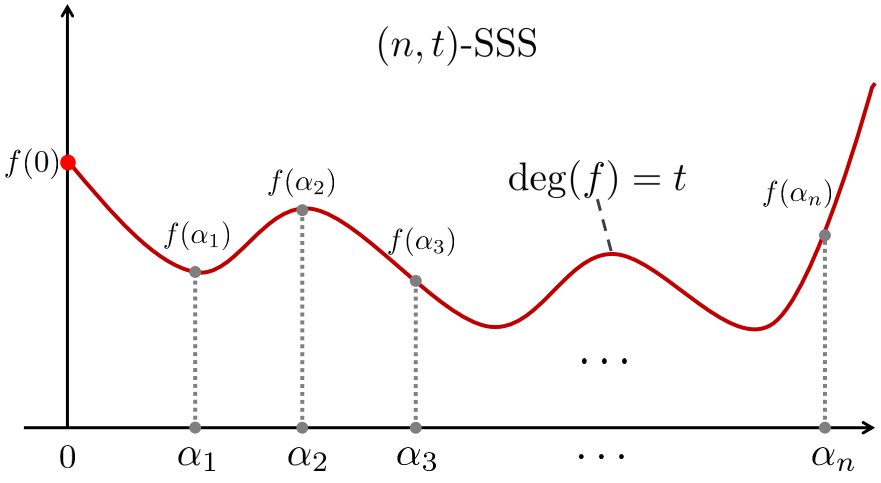


Figure 1.2: Illustration of an instance of redundant masking. In an (n, t)-SSS based polynomial masking, the sensitive variable $X = f(0)$ is encoded into n shares with a security order t.

1.5 Towards Measurable Side-Channel Security

Side-channel attacks pose a considerable threat to cryptographic devices that are physically or remotely accessible by an attacker. Naturally, the side-channel leakage is at the core of evaluating the practical security of a given cryptographic implementation. That is, *how much information can an adversary collect* and/or *how much of it can be exploited in practice?* Regarding the former, we refer to *information leakage quantification*, in which we aim at measuring side-channel leakage in a quantification way. Those leakages might be independent of specific attacks relying on different statistical tools in the corresponding side-channel distinguishers. On the contrary, the latter is much more relevant to the probability of success in extracting sensitive variables (like secret keys) in real scenarios.

1.5.1 Information Leakage Quantification

Quantifying the information leakage is essential in assessing the concrete side-channel security of a cryptographic chip. Typically, the performance of a side-channel distinguisher is highly determined by the amount of information leakage that is usually measured by leakage metrics like signal-to-noise ratio, correlation coefficients, mutual information, etc.

According to different leakage models and the abstraction level of cryptographic implementations, the strategies for quantifying side-channel leakages are roughly classified into five categories as follows.

- Firstly, the conformance-based leakage detection aims at answering the following question at a high abstraction level: *does the device under test leak side-channel information?* [13, 47, 112]. Those statistical tools include Welch's t-test [13, 47], χ^2-test [112, 136], etc. A similar approach is detailed in ISO/IEC 17825 [87].

- Secondly, the proof-based evaluation intends to prove the side-channel resistance of a masked design under abstract models like the probing model [86] and related variants [60, 62, 126, 128]. Typically, under independence assumption and large noise condition, several leakage models are equivalent with certain forms of constants [126] in providing formal security guarantees of the masked implementation. However, physical defaults like couplings, glitches, etc., usually contradict assumptions behind the probing model [5, 103]. As a consequence, it is recommended to launch more quantitative evaluations in assessing practical side-channel security.

- Thirdly, the moment-based evaluation attempts to find the least order of moments of side-channel measurement that depend on the sensitives. Representative metrics including signal-to-noise ratio (SNR) [151] under proper definitions and the normalized inter-class variance (NICV) [14], etc. Particularly, NICV is connected to SNR in the sense that both of them evaluate the key-dependent variance of leakage. With proper definition, SNR can also be used to measure the leakage in presence of higher-order masking schemes.

- Fourthly, the information-theoretic evaluation aims at measuring side-channel leakages by utilizing information-theoretic measures [151, 166]. In essential, it usually provides information-theoretic bounds on the probability of success for any side-channel distinguishers given a set side-channel measurements [41, 57]. The frequently used measures include Shannon mutual information (MI), Kullback-Leibler divergence, conditional entropy, etc. Additionally, some more general measures like Rényi entropy and Rényi divergence shall be considered for a purpose of more accurate evaluation.

- Finally, the fifth category of attack-based evaluation is at the core of side-channel security evaluation, which aims at assessing the probability of success of a specific side-channel distinguisher. We shall detail more in the following subsection.

Summing up, the conformance-based leakage detection only provides qualitative assertion on whether the masked circuits leak or not, while other evaluations give quantitative assessment of concrete side-channel security. In the following, we present the last evaluation strategy, which quantifies information leakage by exploitation.

1.5.2 Information Leakage Exploitation by Attacks

As is argued frequently, the leakage detected information-theoretic evaluation (e.g., using mutual information) might not lead to a successful attack in practice. Eventually, the exploitability of the side-channel leakage determines the success rate of certain attacks.

In this respect, the last category, namely the attack-based evaluation is at the core of side-channel security evaluation, which aims at assessing the probability of success of a specific side-channel distinguisher. Relying on large variety of side-channel distinguishers like correlation power analysis [16], template attacks [31], stochastic attacks [143], higher-order optimal distinguisher [18], etc, the attack-based evaluation provides more accurate assessment of leakage, which captures device-specific features of side-channel leakage. In particular, some metrics

can help including the signal-to-noise ratio (SNR) for leakage detection and success rate as an ultimate attacking metric.

Although attack-based evaluation allows us a more accurate characterization on the concrete side-channel security, it is much more relying on the expertise of evaluators and measurement environment (acquisition equipment, set-ups, etc). Moreover, compared to empirical attacks, the bounds on the success rate given by information-theoretic tools are usually very loose, even much looser when the noise level is high or the leakage model is not accurate (with model mismatches). To the best of our knowledge, it is still an open problem: *how to narrow down or even bridge the gap between theoretical bounds and success rate in practical?*

Table 1.1: Summary of evaluation strategies in assessing side-channel resilience of cryptographic devices (implementations).

Rationale	**Quantitative?**	**Target**	**Tools / Metrics**
Conformance-based	✗	Impl.	t-test, χ^2-test, etc.
Proof-based	✓	Abstract	Probing model, etc.
Moment-based	✓	Impl.	**SNR,** NICV, etc.
Information-theoretic	✓	Abstract & Impl.	**Entropy**, **MI**, α-information, etc.
Attack-based	✓	Impl.	**Success rate,** Guessing entropy, etc.

- *Impl.* is short for implementations;
- *Abstract* denotes abstract or theoretical constructions.

All above five strategies are summarized in Tab. 1.1. To a large extent, all five evaluation approaches are complementary to each other in practical application, varying with different evaluation requirements and necessary expertise on launching evaluations.

1.6 Measuring Leakage in a General Context

Shannon information theory (e.g., entropy, conditional entropy, mutual information, etc) is frequently adopted in side-channel analysis for measuring the leakage from an information-theoretic perspective. More generally, the problem of information leakage quantification shall be extended into a more general one: given two dependent random variables X and Y, how to measure the information that Y brings on X. Intuitively, it is to measure the difference between the amount of uncertainty on X alone and the remaining uncertainty when Y is known (the

equivocation). There is consensus, since the founding work of Shannon [146, 147], that this information must be given by the mutual information $I(X;Y) = H(X) - H(X|Y)$ where $H(X)$ is the entropy and $H(X|Y)$ is the equivocation.

Mutual information has been successfully applied to solve many telecommunication problems. However, recent theories in computer science and information theory show that this paradigm is not always satisfying in practice [150]. Indeed, the "operational" definition of the quantities $I(X,Y)$, $H(X)$, $H(X|Y)$ is linked to the size of the corresponding typical sets (via the AEP, the asymptotic equipartition property), thanks to the law of large numbers. It thus supposes to constitute sequences i.i.d. infinitely long of X and Y to be operational. We would prefer more practical definitions for much shorter sequences, typically for discrete variables, where the knowledge of the information implies knowing at least partially X. For instance, in side-channel analysis, the key-recovery attack is essentially to recover a discrete sub-key by utilizing unintentional side-channel leakages (like computation time, electromagnetic emanation, power consumption for embedded implementations).

In this regard, another promising approach is using Rényi entropy and divergence [135], which are known as α-information theory with a flexible order α (such that $\alpha > 0$ and $\alpha \neq 1$). It is indeed a generalization of Shannon information theory. Particularly, Sibson's α-information [148] is more appropriate for some applications [88, 89] than other proposals, which generalizes mutual information (without mutuality). More recently, several conditional versions of α-information have been proposed [64, 99, 157] for generalizing conditional mutual information. Especially, [99] shows great potentials when applied into side-channel analysis by providing much tighter bounds on the probability of success, although only few candidates of order α are provided.

However, the open problem still remains: *how to derive a more accurate or even exact bound on the success rate of empirical attacks?* In this thesis, we shall explore more possibilities in applications of α-information theory in side-channel analysis and answer this open problem in a formal and exact way.

CHAPTER 2

Contributions

Contents

During this thesis, the main subject targets the measurable security of cryptographic chips. More specifically, the first topic is how to unify and quantify the information leakage of cryptographic implementations in the presence of masking protections. In this regard, we present a coding-theoretic framework to concretely measure the information leakage in code-based masking. Secondly, how the general information-theoretic measures (e.g., Rényi entropy, α-divergence, α-information, etc) can be exploited to evaluate and understand the concrete security level of cryptographic devices. In this respect, we present information-theoretic bounds on the maximum success rate of key-recovery attacks and the minimum number of side-channel traces to achieve a specific success rate.

2.1 Empowering Inner Product Masking by Optimal Codes

The first contribution of this thesis lies in optimizing Inner Product Masking (IPM) by providing the optimal codes for it [36, 37, 40]. IPM is proposed to strengthen the frequently used Boolean

masking by improving the algebraic complexity of encoding (or sharing). We propose a coding-theoretic approach to quantitatively assess the side-channel security of the IPM. Specifically, starting from the expression of IPM in a coded form, we use two defining parameters of the code (namely the *dual distance* and the *kissing number*) to characterize its side-channel resistance. We then connect it to two leakage metrics, namely signal-to-noise ratio (SNR) and mutual information (MI) from an information-theoretic aspect. Next, we show how to systematically choose optimal codes (in the sense of maximizing the resilience) to optimize IPM. We present a simple but effective algorithm for choosing optimal codes for IPM, which should be of special interest for designers when selecting optimal parameters for IPM.

2.2 Leakage Quantification of the Code-based Masking

In this thesis, we follow a generalization approach by targeting the most general code-based masking called generalized code-based masking (GCM) [35], which includes Boolean masking, IPM, Leakage Squeezing (LS), Direct Sum masking (DSM), Shamir's Secret Sharing (SSS)-based masking, etc. We follow the above coding-theoretic approach and propose a unified leakage quantification framework for GCM by connecting the side-channel resistance of GCM with two coding properties of the corresponding linear codes used in GCM. The two coding properties are the *dual distance* and the *adjusted kissing number*. We demonstrate that the two properties are analytically linked to commonly used leakage metrics, namely signal-to-noise ratio and mutual information, in the case of GCM.

As straightforward applications, we show that our extended framework is consistent with the above in IPM. Particularly, the adjusted kissing number converges to the kissing number when the masking is non-redundant, e.g., in cases of IPM and DSM. Secondly, we illustrate how the redundancy in SSS-based masking affects its side-channel resistance [35]. We highlight that the public interpolation points (see in Fig. 1.2) significantly impact the side-channel resistance of SSS-based masking. We then provide an information-theoretic evaluation on public points and show the optimal public points for SSS-based masking.

2.3 Bounding Success Rate in Recovering Secret Key

The third part of this thesis completes side-channel resistance of the code-based masking by providing attack-based evaluations and present information-theoretic bounds when attacking masked cryptographic implementations. In this respect, success rate (SR) is one of the ultimate

metrics in side-channel analysis. Firstly, we employ the higher-order optimal distinguisher (HOOD) against instances of code-based masking, namely IPM and SSS-based masking corresponding to redundant and non-redundant cases, respectively. The experimental results of HOOD exactly confirm our previous information-theoretic evaluations. We emphasize that the redundancy in code-based masking can only decrease its resilience against practical attacks. We also provide some optimal and worst cases of both IPM and SSS-based masking, especially the worst cases of $(3, 1)$-SSS based masking is less resilient than the first-order Boolean masking in spite of the same security order.

Secondly, we derive information-theoretic bounds on the success rate following a communication channel model [41, 138]. When evaluating the practical side-channel security of chips, it is extremely useful to have an upper bound on success rate of any attack given a (fixed) number of side-channel measurements. Or conversely, it is equivalent to derive a lower bound on the number of queries for a given success rate of any attacks. In this thesis, we derive several bounds in both directions by using information-theoretic tools, particularly for cryptographic implementations protected by masking schemes (including the code-based masking). In particular, those bounds are bidirectional by either providing upper bounds on success rate of any attacks or lower bounds on the number of traces to achieve a certain success rate.

2.4 Generic Information-Theoretic Measures in SCA

In the final part of this thesis, we investigate more general information-theoretic measures in the context of side-channel analysis [43, 99], particularly in comparison with Shannon entropy and mutual information [35, 138]. Those measures include Rényi entropies, guessing entropy and α-information, etc. In the problem of guessing a cryptographic key, we illustrate a full spectrum of upper bounds on the probability of success by using conditional α-information between the secret key and information leakage. Especially, we show that the success rate is tightly upper bounded by α-information of a larger enough order (e.g., when $\alpha \geq 100.00$, as shown in Fig. 2.1).

More importantly, we demonstrate that the success rate of the maximum-likelihood (ML) based attack converges to the exact upper bound by the conditional α-information of order $\alpha \to \infty$ (also called maximal information). The ML-based attacks are optimal, for instance, consider HOOD when the leakage model is known. Therefore, our derivatives imply that this bound is also achievable. To the best of our knowledge, we shall for the first time seamlessly connect information-theoretic measures and real attacks in side-channel analysis. Taking the

Hamming weight leakage with additive white Gaussian noises, numerical results confirm our findings and show meaningful indications in practice. As a perspective, it would be extremely interesting to extend our evaluation into protected scenarios, e.g., in the presence of masking.

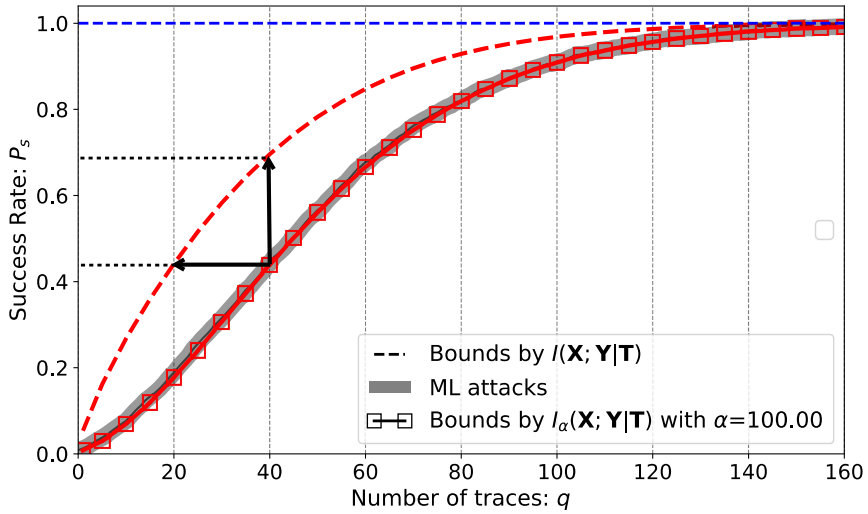


Figure 2.1: Illustration of information-theoretic bounds on success rate [2]by Shannon mutual information and α-information of order $\alpha = 100.00$ in an unprotected AES cases, with Gaussian noise of variance $\sigma^2 = 10.00$.

2.5 Outline of the Thesis

This thesis mainly consists of four parts and accompanied with the introduction and conclusion parts. The overall structure of this thesis is diagrammed as in Fig. 2.2.

In Part II, we focus on optimizing inner product masking by a coding-theoretic approach. Some basics on the linear codes, Pseudo-Boolean functions and information theory are firstly revisited in Chap. 3 and they will be used throughout this thesis. Secondly, we dive into IPM in Chap. 4 by forming it into a coding-theoretic fashion. Then the information leakage is measured by both signal-to-noise ratio and mutual information, along with numerical results. Moreover, we present the rationale of selecting optimal codes on the basis of our leakage quantification.

Next, in Part III we present the generalization of several masking schemes into the most general scenario, and also extend our coding-theoretic approach into this general scenario. Specifically, in Chap. 5, we extend the leakage quantification approach to redundant scenarios,

[2]Note that the success rate is evaluated $10,000$ times to be more accurate and the curve is much smoother.

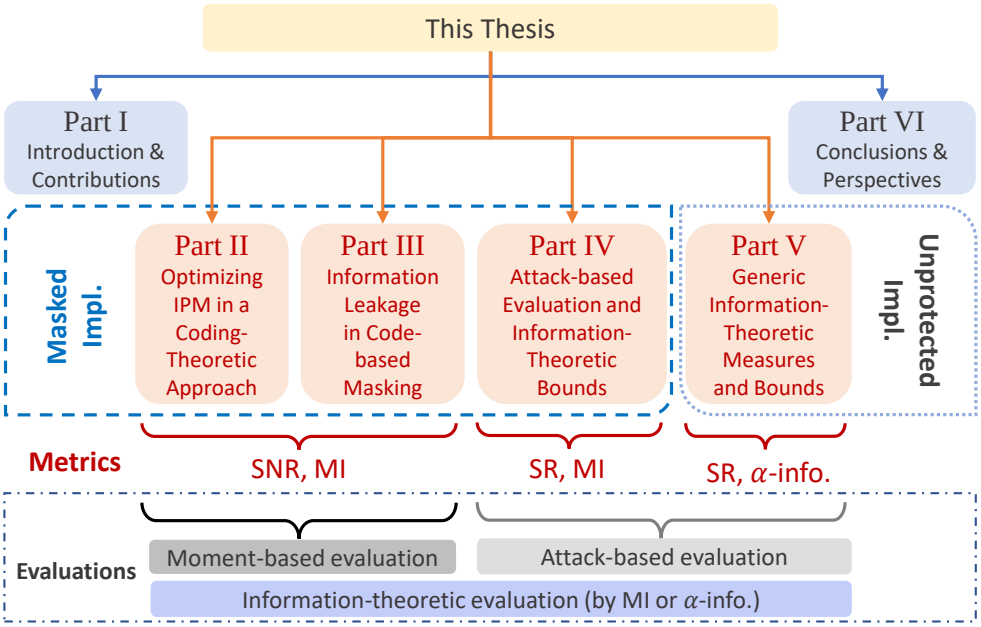


Figure 2.2: The overall structure of this thesis.

for instance in the case of SSS-based masking. As an application, in Chap. 6, we present an example that exactly verifies the effectiveness of our new framework and provides optimal linear codes in several cases.

Then, in Part IV we first present an attack-based evaluation on code-based masking in Chap. 7. In order to amplify the attack, we employ the optimal one that is based on the maximum-likelihood rule. We highlight that those numerical results are well-coincided with our theoretical derivatives. Second, we show how to derive several information-theoretic bounds on the success rate of any attacks in Chap. 8.

Last in Part V or in Chap. 9, we illustrate how those more general information-theoretic tools can be applied in side-channel analysis. We first explore the application of Rényi information theory (including Rényi entropy, Rényi divergence and related extensions) into the problem of guessing the secret key from its Hamming weight leakage. Second, we propose to use conditional α-information to assess the concrete side-channel security.

The conclusions of this thesis are in Part VI along with possible investigations in the future. Additional proofs and materials are included in appendices, Part VII.

Part II

Optimizing Inner Product Masking in a Coding-Theoretic Approach

CHAPTER 3

Preliminaries: Basics on Linear Codes, Vector Space and Pseudo-Boolean Functions

In this chapter, we revisit basics of the linear codes, the vector spaces and the pseudo-Boolean functions that are used through this thesis.

Contents

3.1 Linear Codes

We recall several known definitions and properties of linear codes, which hold respectively when the base field is $\mathbb{K} = \mathbb{F}_2$ or $\mathbb{K} = \mathbb{F}_{2^\ell}$. Let $n,\ k,\ d \in \mathbf{N}^*$ be positive integers such that $k \leq n$. The linear code is defined as follows.

Definition 3.1 (Linear code [101])**.** A linear code $\mathcal{C}$ is a set of vectors, also called codewords, which form a vector space. The parameters of a linear code $\mathcal{C}$ is a triple (n, k, d), where n is the code length, k denotes its dimension, and d is its minimum distance. The parameters are denoted as $[n, k, d]_q$ to refer to the finite field $\mathbb{F}_q$ the code is defined on.

The minimum distance of a code $\mathcal{C}$ is defined as $d_{\mathcal{C}} = \min_{c,c' \in \mathcal{C}} d_H(c, c')$ where d_H denotes the Hamming distance. In particular, $d_{\mathcal{C}}$ equals to the minimum weight of its nonzero codewords.

Given a linear code $\mathcal{C}$ with parameters $[n, k, d_{\mathcal{C}}]$, its weight enumerator is defined as follows.

Definition 3.2 (Weight Enumerator [101, §5.2])**.** The weight enumerator of a linear code specifies the number of codewords $\mathcal{C}$ of each possible Hamming weight in $\mathcal{C}$. Specifically, we have

$$W_{\mathcal{C}}(\mathsf{X}, \mathsf{Y}) = \sum_{i=0}^{n} B_i \mathsf{X}^{n-i} \mathsf{Y}^i \tag{3.1}$$

where $B_i = |\{c \in \mathcal{C} | w_H(c) = i\}|$ and $w_H(\cdot)$ denotes the Hamming weight function. In particular, $B_{d_{\mathcal{C}}}$ is called the kissing number of $\mathcal{C}$.

Lemma 3.1. *Basic properties of $B_i \in \mathbb{N}$:*

- $B_0 = 1$, $B_1 = \cdots = B_{d_{\mathcal{C}}-1} = 0$,
- $B_{d_{\mathcal{C}}} > 0$, *meaning the kissing number is nonzero,*
- $B_n = 1$ *if and only if the code $\mathcal{C}$ has a codeword with all ones* (*e.g.*, $[1, \ldots, 1]$).

Note that two linear codes are said to be equivalent if one can be obtained from the other by a series of operations of the following two types: 1) an arbitrary permutation of the coordinate positions and, 2) in any coordinate position, multiplication by any nonzero scalar. Straightforwardly, equivalent linear codes have the same weight enumerator.

Definition 3.3 (Dual Code [101, §1.8])**.** The dual code of $\mathcal{C}$ is the linear code $\mathcal{C}^{\perp} = \{u \in \mathbb{K}^n \,|\, \forall c \in \mathcal{C},\, c \cdot u = 0\}$, where $c \cdot u$ is the standard inner product.

Definition 3.4 (Dual Distance [101])**.** The dual distance $d_{\mathcal{C}}^{\perp}$ of a linear code $\mathcal{C}$ is the minimum Hamming weight $w_H(u)$ of nonzero $u \in \mathbb{K}^n$, such that $\sum_{c \in \mathcal{C}} (-1)^{c \cdot u} \neq 0$.

Let $\mathcal{E}$ a vector space of $\mathbb{K}^n$. The indicator of $\mathcal{E}$ is the application

$$x \in \mathbb{K}^n \mapsto \mathbb{1}_{\mathcal{E}}(x) = \begin{cases} 1 & \text{if } x \in \mathcal{E}, \\ 0 & \text{otherwise.} \end{cases}$$

Then we introduce a well-known property of the linear code as follows.

Lemma 3.2. *For arbitrary linear code $\mathcal{C}$ and $u \in \mathbb{K}^n$, we have $\sum_{c \in \mathcal{C}} (-1)^{c \cdot u} = |\mathcal{C}| \mathbb{1}_{\mathcal{C}^{\perp}}(u)$.*

Proof. We give this well-known proof for the self-contained content. For $u \in \mathcal{C}^{\perp}$, it is straightforward to see that $\sum_{c \in \mathcal{C}} (-1)^{c \cdot u} = |\mathcal{C}|$.

Suppose that $u \notin \mathcal{C}^{\perp}$, thus $\exists v \in \mathcal{C}$ such that $u \cdot v = 1$. We denote $\mathcal{C} = \mathcal{C}' \cup (\mathcal{C}' + v)$ and $\dim(\mathcal{C}') = \dim(\mathcal{C}) - 1$. Then $\sum_{c \in \mathcal{C}} (-1)^{c \cdot u} = \sum_{c' \in \mathcal{C}'} (-1)^{c' \cdot u} + \sum_{c' \in \mathcal{C}'} (-1)^{(c'+v) \cdot u} = \sum_{c' \in \mathcal{C}'} (-1)^{c' \cdot u} - \sum_{c' \in \mathcal{C}'} (-1)^{c' \cdot u} = 0$. □

Corollary 3.1. *For a linear code $\mathcal{C}$, we have $d_{\mathcal{C}}^{\perp} = d_{\mathcal{C}^{\perp}}$.*

Therefore, the dual distance of a linear code $\mathcal{C}$ is the same as the minimum distance of the dual code $\mathcal{C}^{\perp}$.

In this thesis, we consider two classes of linear codes: $[n, k, d_w]_{2^\ell}$ and $[n\ell, k\ell, d_b]_2$ with minimal distances d_w and d_b at word- and bit-level, respectively. More precisely, the latter is the expanded code of the former from $\mathbb{F}_{2^\ell}$ into $\mathbb{F}_2$. Indeed, if z is a codeword of the former code, then the corresponding codeword $[z]_2$ of the latter code is obtained by replacing each term in z by its coordinates with respect to some fixed basis $(e_1, \ldots, e_\ell)$ of $\mathbb{F}_{2^\ell}$ over $\mathbb{F}_2$. Let $(b_1, \ldots, b_k)$ be a basis of the former code, then a basis of the latter code is $([e_i b_j]_2)_{i=1,\ldots,\ell;\, j=1,\ldots,k}$.

According to [113, Theorem 5.1.18], there exists a self-dual basis of $\mathbb{F}_{q^\ell}$ over $\mathbb{F}_q$ if and only if either q is even or both q and ℓ are odd. For the sake of simplicity, we herein fix $q = 2$. We call above expansion the sub-field representation defined as follows.

Definition 3.5 (Sub-field representation [101, §7.7])**.** Let $x \in \mathbb{F}_{2^\ell}$, the sub-field representation of x is $[x]_2 \in \mathbb{F}_2^\ell$.

Definition 3.6 (Code Expansion [101, §7.7])**.** By using sub-field representation, the elements in $\mathbb{F}_{2^\ell}$ are decomposed over $\mathbb{F}_2$. Consider a generating matrix of a linear code of size $k \times n$ in $\mathbb{F}_{2^\ell}$. It becomes a generating matrix of size $k\ell \times n\ell$ in $\mathbb{F}_2$. Any linear codes of parameters $[n, k]_{2^\ell}$ contain $(2^\ell)^k = 2^{k\ell}$ codewords, hence is turned into a $[n\ell, k\ell]_2$ linear code in $\mathbb{F}_2$. The latter code is called the expansion code of the former.

Correspondingly, two kinds of security order t_w and t_b are at word- and bit-level, respectively. Summing up, the two definitions build a direct link between word- and bit-level representation of a linear code and the corresponding conversion. This allows to connect the word (or register)-level probing and the bit-level probing security models, depending on the granularity of the attacker spying tool.

3.2 Complementary Vector Space

In this section, we introduce relevant properties of complementary vector space that will be needed to derive our results. The set of n-bit vectors is denoted by $\mathbb{F}_2^n$, which is an n-dimensional vector space over the finite field $\mathbb{K} = \mathbb{F}_2$.

An $[n, k, d]_q$ linear code $\mathcal{C}$ over $\mathbb{K}$ is a k-dimensional subspace of $\mathbb{K}^n$, therefore, we use the same notations as for the linear codes.

Definition 3.7 (Complementary Vector Space)**.** Two subspaces $\mathcal{C}$ and $\mathcal{D}$ are complementary in direct sum (denoted by $\mathcal{C} \oplus \mathcal{D} = \mathbb{K}^n$) if $\mathcal{C} + \mathcal{D} = \mathbb{K}^n$, and $\mathcal{C} \cap \mathcal{D} = \{0\}$, that is: $\forall z \in \mathbb{K}^n, \exists!(c, d) \in \mathcal{C} \times \mathcal{D},$ such that $z = c + d$.

Accordingly, we shall define the complementary codes as follows.

Definition 3.8 (Complementary Linear Codes [169])**.** Two linear codes $\mathcal{C}$ and $\mathcal{D}$ are complementary if $\mathcal{C} + \mathcal{D} = \mathbb{K}^n$, and $\mathcal{C} \cap \mathcal{D} = \{0\}$.

Lemma 3.3. *Let $\mathcal{C}$ and $\mathcal{D}$ be two vector spaces in $\mathbb{K}^n$ built from independent bases, meaning that $\mathcal{C} \cap \mathcal{D} = \{0\}$. Then $\mathcal{C}^\perp \cap \mathcal{D}^\perp = (\mathcal{C} \oplus \mathcal{D})^\perp$.*

Proof. First of all, we notice that $(\mathcal{C} \oplus \mathcal{D})^\perp \subseteq \mathcal{C}^\perp$. Indeed, a vector orthogonal to all vectors of $\mathcal{C} \oplus \mathcal{D}$ is in particular orthogonal to all vectors of $\mathcal{C} + 0 = \mathcal{C}$. In a symmetric way, we have that $(\mathcal{C} \oplus \mathcal{D})^\perp \subseteq \mathcal{D}^\perp$. Therefore, $(\mathcal{C} \oplus \mathcal{D})^\perp \subseteq \mathcal{C}^\perp \cap \mathcal{D}^\perp$.

Let us now prove the converse inclusion. Let $x \in \mathcal{C}^\perp \cap \mathcal{D}^\perp$. For any vector y in $\mathcal{C} \oplus \mathcal{D}$, there exists a unique pair $(c, d) \in \mathcal{C} \times \mathcal{D}$ (owing to the complementarity of vector spaces $\mathcal{C}$ and $\mathcal{D}$), such that $y = c + d$. Now, $x \cdot y = x \cdot (c + d) = x \cdot c + x \cdot d = 0 + 0 = 0$. Indeed, $x \cdot c = 0$ because $x \in \mathcal{C}^\perp$ and $x \cdot d = 0$ because $x \in \mathcal{D}^\perp$. Therefore, we also have $\mathcal{C}^\perp \cap \mathcal{D}^\perp \subseteq (\mathcal{C} \oplus \mathcal{D})^\perp$. □

Lemma 3.4. *Let $\mathcal{C}$ and $\mathcal{D}$ two complementary vector spaces, namely: $\mathcal{C} \cap \mathcal{D} = \{0\}$, and $\mathcal{C} \oplus \mathcal{D} = \mathbb{K}^n$. Then we have: $\mathcal{C}^\perp \cap \mathcal{D}^\perp = \{0\}$.*

Proof. By application of Lemma 3.3, we have that $\mathcal{C}^\perp \cap \mathcal{D}^\perp = (\mathcal{C} \oplus \mathcal{D})^\perp = (\mathbb{K}^n)^\perp$. Now, as $\mathbb{K}^n$ is the universe code, we have $(\mathbb{K}^n)^\perp = \{0\}$. □

Actually, for the general case when $\mathcal{C}$ and $\mathcal{D}$ are not complementary, we can complement $\mathcal{C} \oplus \mathcal{D}$ with a vector space $\mathcal{E}$, such that:

- $\mathcal{C} \cap \mathcal{D} = \{0\}$, $\mathcal{C} \cap \mathcal{E} = \{0\}$, $\mathcal{D} \cap \mathcal{E} = \{0\}$,
- $\mathcal{C} \oplus \mathcal{D} \oplus \mathcal{E} = \mathbb{K}^n$.

Then, a similar result as Lemma 3.4 holds:

Lemma 3.5. $\mathcal{C}^\perp \cap \mathcal{D}^\perp \cap \mathcal{E}^\perp = \{0\}$.

Proof. Similar with the proof of Lemma 3.4, first treat $\mathcal{C} \oplus \mathcal{D}$ together and then straightforwardly apply Lemma 3.3, which gives the results. □

In this thesis, we consider two cases in the code-based masking:

- In the generalized code-based masking as a general case [35, 164]: $\mathcal{C} \cap \mathcal{D} = \{0\}$, and $\mathcal{C} \oplus \mathcal{D} \subseteq \mathbb{K}^n$, where two linear codes $\mathcal{C}$ and $\mathcal{D}$ are not necessarily complement to each other. In particular, the redundant case when $n > t+1$ corresponds to the strict condition: $\mathcal{C} \oplus \mathcal{D} \subsetneq \mathbb{K}^n$ and then $\{0\} \subsetneq \mathcal{C}^\perp \cap \mathcal{D}^\perp$.
- In inner product masking or direct sum masking as special cases: $\mathcal{C} \cap \mathcal{D} = \{0\}$, and $\mathcal{C} \oplus \mathcal{D} = \mathbb{K}^n$, meaning that $\mathcal{C}$ and $\mathcal{D}$ are complementary. This is the case of [37], where we have $\mathcal{C}^\perp \cap \mathcal{D}^\perp = \{0\}$ as shown in Lemma 3.4.

3.3 Pseudo-Boolean Functions

Leakage functions turn a bitvector into a real value, which the attacker measures. Those functions are pseudo-Boolean functions $P : \mathbb{K}^{n\ell} \mapsto \mathbb{R}$, where $\mathbb{K} = \mathbb{F}_2$.

It is well-known that a pseudo-Boolean function P can be *uniquely* expressed in a *monomial basis* [25] called *Numerical Normal Form* (NNF) [115]:

$$P(Z) = \sum_{I \in \{0,1\}^{n\ell}} \beta_I Z^I, \tag{3.2}$$

where $Z^I = \prod_{i \in \{1,\ldots,n\ell\} \text{ s.t. } I_i=1} Z_i$, and $\beta_I \in \mathbb{R}$. For instance, $Z^{(000\cdots0)_2} = 1$, $Z^{(100\cdots0)_2} = Z_1$ and $Z^{(110\cdots0)_2} = Z_1 Z_2$.

In fact, P is a nice abstraction of practical attacks. For example, in differential power analysis [95] against the most/least significant bit of the sensitive variable, then P equals $Z^{(100\cdots0)_2}$ or $Z^{(000\cdots1)_2}$. Moreover, in correlation power analysis [16] when the Hamming weight model is adopted, P equals $w_H(Z) = Z^{(100\cdots0)_2} + Z^{(010\cdots0)_2} + \cdots + Z^{(000\cdots1)_2}$.

Thanks to the existence and the uniqueness of NNF, we can define the numerical degree of P as follows.

Definition 3.9 (Numerical Degree [25])**.** The numerical degree of a pseudo-Boolean function P denoted by $\deg(P)$ equals: $\deg(P) := d = \max\{w_H(I) | \beta_I \neq 0\}$.

Definition 3.10 (Fourier Transform [22, §2.2])**.** The Fourier transform of a pseudo-Boolean function $P : \mathbb{K}^{n\ell} \mapsto \mathbb{R}$ is denoted by $\widehat{P} : \mathbb{K}^{n\ell} \mapsto \mathbb{R}$, and is defined as: $\widehat{P}(z) = \sum_{y \in \mathbb{K}^{n\ell}} P(y)(-1)^{y \cdot z}$.

Recall from [22, 25] that, $\widehat{P}(z) = (-1)^{w_H(z)} \sum_{I \subseteq \{1,\ldots,n\ell\}; supp(z) \subseteq I} 2^{n\ell - |I|} \beta_I$ where $\beta_I = 2^{-n\ell}(-2)^{|I|} \sum_{z \in \mathbb{F}_2^{n\ell}; I \subseteq supp(z)} \widehat{P}(z)$.

Definition 3.11 (Convolution [22, §2.2])**.** The convolution of two pseudo-Boolean functions f and g is defined as: $(f \otimes g)(z) = \sum_{y \in \mathbb{K}^{n\ell}} f(y) g(y + z)$.

We recall below two well-known properties of *Fourier transform* as well as a property on the convolution. We omit the proofs for the sake of brevity and refer to [22] for details.

Proposition 3.1 (Involution Property [22, §2.2])**.** $\widehat{\widehat{P}}(z) = |\mathbb{K}^{n\ell}| P(z) = 2^{n\ell} P(z)$, $\forall z \in \mathbb{K}^{n\ell}$.

Proposition 3.2 (Inverse Fourier Transform [22, §2.2])**.** $P(z) = 2^{-n\ell} \sum_{y \in \mathbb{K}^{n\ell}} \widehat{P}(y)(-1)^{y \cdot z}$, $\forall z \in \mathbb{K}^{n\ell}$.

Proposition 3.3 (Fourier Transform and Convolution [22, §2.2, Prop. 8])**.** $\widehat{f \otimes g}(z) = \widehat{f}(z) \cdot \widehat{g}(z)$, $\forall z \in \mathbb{K}^{n\ell}$.

3.4 Shannon Information-Theoretic Measures

We also define some information theoretic tools. The entropy of a random vector $\mathbf{X}$ of length q is defined by:

$$H(\mathbf{X}) = -\sum_{\mathbf{x}\in\mathcal{X}^q} \Pr(\mathbf{x}) \log_2 \Pr(\mathbf{x}).$$

The conditional entropy of a random vector $\mathbf{X}$ knowing vector $\mathbf{Y}$ is defined by:

$$\begin{aligned} H(\mathbf{X} \mid \mathbf{Y}) &= -\sum_{\mathbf{y}\in\mathcal{Y}^q} \Pr(\mathbf{y}) H(\mathbf{X} \mid \mathbf{Y} = \mathbf{y}) \\ &= -\sum_{\mathbf{y}\in\mathcal{Y}^q} \Pr(\mathbf{y}) \sum_{\mathbf{x}\in\mathcal{X}^q} \Pr(\mathbf{x} \mid \mathbf{y}) \log_2 \Pr(\mathbf{x} \mid \mathbf{y}). \end{aligned}$$

The Mutual Information between two random vectors $\mathbf{X}$ and $\mathbf{Y}$ is defined as $I(\mathbf{X};\mathbf{Y}) = H(\mathbf{X}) - H(\mathbf{X} \mid \mathbf{Y})$. The conditional Mutual Information $I(\mathbf{X};\mathbf{Y} \mid \mathbf{T})$ where $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{T}$ are random vectors is defined as $I(\mathbf{X};\mathbf{Y} \mid \mathbf{T}) = H(\mathbf{X} \mid \mathbf{T}) - H(\mathbf{X} \mid \mathbf{Y},\mathbf{T})$. Last, the Kullback-Leibler divergence between two distributions p and q over the same set $\mathcal{X}$ is defined as:

$$\mathrm{D}(p\|q) = \sum_{x\in\mathcal{X}} \mathbb{P}(x) \log_2 \frac{p(x)}{q(x)}.$$

CHAPTER 4

Measuring the Leakages in IPM and Optimal Codes

The results presented in this chapter have been published in collaboration with Sylvain Guilley, Claude Carlet, Jean-Luc Danger, and Sihem Mesnager in the *IEEE Transactions on Information Forensics and Security* (T-IFS) [37] and the journal of *Cryptography and Communications Discrete Structures, Boolean Functions and Sequences* (CCDS) [38].

Contents

4.1 Introduction of Inner Product Masking

Masking is one of the most investigated countermeasures against side-channel analysis, allowing all cryptographic operations to be performed on the masked data. Essentially, masking is a sound way to improve the side-channel security of cryptographic implementations, since given high enough noise, the attack complexity increases exponentially with the number of shares [128], while the implementation cost increases only quadratically (or only cubically in higher-order glitches free implementations [79]). For instance, the Boolean masking scheme is the simplest one which enables high performance when implemented on real circuits. The first provably secure higher-order masking scheme has been introduced by Ishai et al. [86] for the protection of single bits in $\mathbb{F}_2$. Then, this scheme has been extended to the protection of words (e.g. bytes in $\mathbb{F}_{2^8}$) with higher-order security by Rivain et al. [139]. Interestingly, it has been noticed later that this masking scheme can be further improved by mixing bits in each share (of $\ell = 8$ bits). In brief, the main idea is to elevate the bit-level algebraic complexity of the masking scheme. Thus in this respect, *Inner Product Masking* (IPM) scheme has been proposed as an alternative, in which inner product is adopted as a mixing operation.

The IPM scheme has been first introduced by Balasch et al. at ASIACRYPT'12 [4] as an alternative to masking schemes like Boolean or multiplicative masking and has been further improved by Balasch et al. at EUROCRYPT'15 [2] and at ASIACRYPT'17 [3]. In IPM, the random masks are not used plain, but a mixing between the bits is carried out by the multiplication with a public vector $\alpha = (1, \alpha_2, \ldots, \alpha_n)$ and then involved into the cryptographic computation ($Z = X + \alpha_2 Y_2 + \cdots + \alpha_n Y_n$, where X is the sensitive data and Y_i are the $n-1$ masks). Interestingly, by different settings of vector L and mask materials, Balasch et al. [4] pointed out that IPM is the generalization of four typical kinds of masking schemes, namely the Boolean one, the multiplicative one [76], the affine one [70] and the polynomial one [77, 131].

4.2 The-State-of-the-Arts

The concrete security order of a masking scheme depends not only on the number of shares but also on the encodings involving the sensitive variables and mask materials into cryptographic operations. With the same number of shares, Balasch et al. [4] observed that the IPM leaks consistently less than Boolean masking, and further demonstrated this observation in [2, 3]. In fact, this observable feature originates from the encoding of IPM, in which the random masks are multiplied by the coordinates of the public parameter $\alpha \in \mathbb{F}_{2^\ell}^n$. Therefore several bits in each share are mixed together, which increases the algebraic complexity of the encoding. By contrast, in Boolean masking the masks are directly involved by bit-wise XOR operation. This is the primary advantage of IPM. Furthermore, another interesting effect in [3, Fig. 3] is that the different choices of the L vector in IPM significantly affect its concrete bit-level security. For instance, with $n = 2$ shares made up of $\ell = 8$ bits (byte-oriented), the security order in *bounded moment model* [7] can be $t_{bound} = 3$, while the security order in (word-level) *probing model* is only $t_w = 1$.

In fact, this parameter effect in IPM has been studied firstly by Wang et al. [165], named as "Security Order Amplification". Wang et al. propose the parameter O_{min}, the lowest key-dependent statistical moment, as a metric to measure the amplified security order. This metric O_{min} is directly related to the bit-level security order t_b in *bit-level probing model* proposed by Poussier et al. [123] since $O_{min} = t_b + 1$. More importantly, Poussier et al. firstly introduce the coding form of IPM as: $Z = X\mathbf{G} + Y\mathbf{H}$ where X, Y, Z are the sensitive variable, random mask(s) and masked variable, $\mathbf{G}$ and $\mathbf{H}$ are the generator matrices of two codes $\mathcal{C}$ and $\mathcal{D}$, respectively. Then they prove that the bit-level security of IPM is related to one of the defining parameters of the code $\mathcal{D}$ (namely its dual distance $d_{\mathcal{D}}^{\perp}$). This result gives an explanation of the security order amplification discussed in [165].

The other line of research on the encoding and parameter effect of masking schemes is about the *Leakage Squeezing* (LS) which stems from Carlet et al. [24]. Particularly, Carlet et al. show that IPM is an instance of LS. They statistically studied the security order of LS scheme by linking the *correlation immunity* [22] of the indicator of the code (that equals the dual distance $d_{\mathcal{D}}^{\perp}$ minus 1), the mutual information (MI) and the success rate (SR) of side-channel attacks together. More precisely, in logarithmic form, mutual information $\log(MI)$ is a linear function of the logarithmic noise variance $\log(\sigma^2)$, and the slope (security order) of this linear function equals the dual distance of $\mathcal{D}$. To summarize, the bit-level security order t_b of IPM is $d_{\mathcal{D}}^{\perp} - 1$,

where $d^{\perp}_{\mathcal{D}}$ is the dual distance of the code $\mathcal{D}$ in the coding form. Related works are summarized in Tab. 4.1 (note that *SNR* is short for attack signal-to-noise ratio [102, § 4.3.2, page 73]).

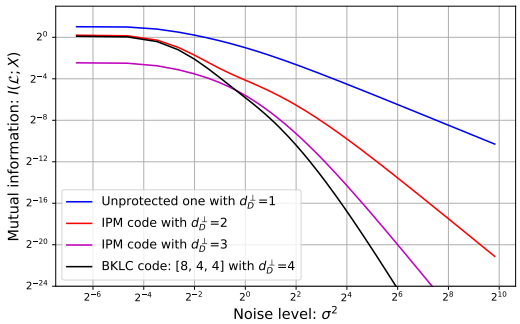
Table 4.1: Summary of side-channel security analysis on IPM.

	Security Orders	Code Parameters	Metrics	Comments
Balasch et al. [2]	t_w	–	*MI*	*MI* varies for different L vectors
Wang et al. [165]	t_b	$d^{\perp}_{\mathcal{D}}$	*MI*	O_{min} $(= d^{\perp}_{\mathcal{D}})$ was used (the lowest key-dependent statistical moment)
Poussier et al. [123]	$t_w,\ t_b$	$d^{\perp}_{\mathcal{D}}$	*MI*	
Balasch et al. [3]	t_w, t_b	–	*MI*	t_{bound} $(= t_b + 1)$ is in the bounded moment model
Carlet et al. [24]	$t_w,\ t_b$	$d^{\perp}_{\mathcal{D}}$	*MI*, *SR*	*SR* of the *optimal attack* [18]
This work	$t_w,\ t_b$	$d^{\perp}_{\mathcal{D}}$, $\boldsymbol{B_{d^{\perp}_{\mathcal{D}}}}$	***SNR***, *MI*, *SR*	A unified framework to analyze all IPM codes by closed-form expression

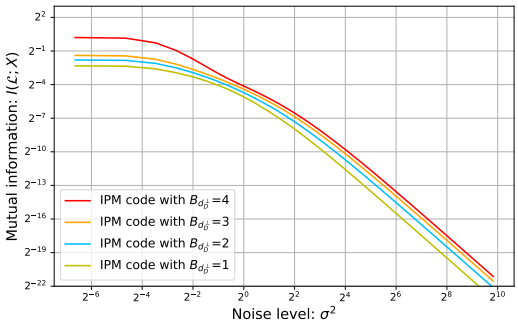
* Here $t_w,\ t_b$ are word- and bit-level security orders, where $t_w = n-1$. Bit-level security order t_b equals to $d^{\perp}_{\mathcal{D}}-1$ as in [24, 123] and in this paper.

Actually, the security order of IPM depends on the code $\mathcal{D}$ involved in the scheme, which can be easily demonstrated by information-theoretic metric. As shown in Fig. 4.1(a), the security order (the slope) of IPM depends on the dual distance of the chosen code $\mathcal{D}$, namely $d^{\perp}_{\mathcal{D}}$. Specifically, the slope in the log-log plot representation of *MI* as a function of noise variance σ^2 is $-d^{\perp}_{\mathcal{D}}$. However, it can be observed that for different choices of the code $\mathcal{D}$ with the same dual distance, the *MI*s are distinctly different as shown in Fig. 4.1(b). The smaller the number of nonzero codewords of minimal weight ($B_{d^{\perp}_{\mathcal{D}}}$), the smaller the *MI* consistently over the full range of noise variance σ^2. Similar situations happen with success rates of *optimal attacks* [18], indicating that only parameter of $\mathcal{D}$ equal to the dual distance $d^{\perp}_{\mathcal{D}}$ is not enough to characterize the side-channel resistance of IPM. Therefore, a natural question is: *What is/are other defining parameter(s) of* $\mathcal{D}$ *that influence the concrete side-channel security level of IPM?* Since the different choices of the code $\mathcal{D}$ have critical impacts on the concrete security order of IPM, then another question that comes with it is: *how to choose optimal codes in the sense of side-channel resistance for IPM?*

[2]Note that the only criteria is the highest minimum Hamming distance [158].



(a) Codes with different $d_{\mathcal{D}}^{\perp}$.



(b) Codes with different $B_{d_{\mathcal{D}}^{\perp}}$ but with the same $d_{\mathcal{D}}^{\perp}=2$.

Figure 4.1: Systematic investigation of linear codes of IPM over $\mathbb{F}_{2^4}$ grouped by $d_{\mathcal{D}}^{\perp}$ and $B_{d_{\mathcal{D}}^{\perp}}$, and one *BKLC* code (Best Known Linear Code[2]).

4.3 IPM in a Coding-Theoretic Form

Let an information word $X \in \mathbb{K} = \mathbb{F}_{2^\ell}$ and $n-1$ masks $Y_i \in \mathbb{K}$, the Boolean masking scheme [123] protects X as:

$$Z = \left(X + \sum_{i=2}^{n} Y_i, Y_2, Y_3, \ldots, Y_n \right) = X\mathbf{G} + Y\mathbf{H}, \tag{4.1}$$

where $\mathbf{G}$, $\mathbf{H}$ are generator matrices of two linear codes $\mathcal{C}$ and $\mathcal{D}$ as follows, respectively. Moreover, $\mathcal{C}$ and $\mathcal{D}$ are supplementary codes such that $\mathcal{C} \oplus \mathcal{D} = \mathbb{K}^n$.

$$\begin{aligned} \mathbf{G} &= \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \end{pmatrix} \in \mathbb{K}^{1\times n}, \\ \mathbf{H} &= \begin{pmatrix} 1 & 1 & 0 & \cdots & 0 \\ 1 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & \cdots & 1 \end{pmatrix} \in \mathbb{K}^{(n-1)\times n}. \end{aligned} \tag{4.2}$$

IPM is an encoding to improve the algebraic complexity by mixing bits in each share together. In IPM, linear functions are applied to mask materials y_i to construct only the first share. We define a family of bijective linear functions $f_i : \mathbb{K} \mapsto \mathbb{K}$ defined by $f_i(y_i) = \alpha_i y_i$ where $\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{K}^n$, $\alpha_1 = 1$ and $\alpha_i \in \mathbb{K}\backslash\{0\}$ for $i \in \{2, 3, \ldots, n\}$. Then the IPM scheme [2] with n shares is expressed as:

$$Z = (X + \sum_{i=2}^{n} f_i(Y_i), Y_2, Y_3, \ldots, Y_n) = X\mathbf{G} + Y\mathbf{H}. \tag{4.3}$$

Remark 4.1 (Word-level security order). In the first share Z_1 of Z, X is masked only by mask Y_1, where Y_1 is a uniformly distributed mask equal to $Y_1 \stackrel{def}{=} \sum_{i=2}^{n} f_i(Y_i)$. But still, the masking scheme is more than second-order secure since the attacker cannot directly measure a leakage

arising from Y_1. Instead, to get information from Y_1, the attacker should measure the leakage from shares $(Z_2, \ldots, Z_n) = (Y_2, \ldots, Y_n)$, hence a n-order attack.

Two generator matrices $\mathbf{G}$ and $\mathbf{H}$ of the above linear codes are as follows. Note that this encoding[3] was first introduced in [123] and we borrow it here.

$$\mathbf{G} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \end{pmatrix} \in \mathbb{K}^{1\times n},$$
$$\mathbf{H} = \begin{pmatrix} \alpha_2 & 1 & 0 & \cdots & 0 \\ \alpha_3 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \alpha_n & 0 & 0 & \cdots & 1 \end{pmatrix} \in \mathbb{K}^{(n-1)\times n}. \tag{4.4}$$

IPM is a generalization of Boolean masking (by choosing $\alpha_i = 1$, $1 \leq i \leq n$). Both schemes ensure the property that X cannot be deduced from $d < n$ shares provided Y_i are uniformly distributed (see Prop. 4.1 for a detailed formulation).

4.4 Quantifying Leakages of IPM via SNR

In this section, we focus on quantitatively assessing the leakages of IPM by SNR. Let $P : \mathbb{K}^{n\ell} \mapsto \mathbb{R}$ where $\mathbb{K} = \mathbb{F}_2$ with numerical degree $d^\circ P$ be the leakages collected (and manipulated) by the attacker. In practice, $d^\circ P$ reflects the strength of the attacker, because it is the number of masked bits which shall be combined together to unveil a dependency on the key. Two typical situations are:

- The devices leak bits individually, as in the *probing model* [86]. Therefore, the degree d of the leakage function is the number of probed bits.
- The devices leak bits as words in parallel through a leakage function ϕ. The attackers subsequently apply their strategy (a composition function) ψ on top of ϕ. For instance, ϕ is the Hamming weight and ψ consists of raising the result at some power d, resulting in $P = \psi \circ \phi = w_H(\cdot)^d$.

4.4.1 Leakage Model & Attack Strategy

In practice, the security of a cryptographic implementation not only depends on its leakages during execution but also highly relates to the capability of an adversary to exploit these leakages. For instance, for a t-th order secure masking scheme, an adversary can launch a successful d-th order attack against it when d is greater than t.

[3] Note that Equ. 5 in [123] contains a mistake, namely $\mathbf{G}$ should be $(I_\ell, 0, \ldots, 0)$, and not $(1, \ldots, 1, 0, \ldots, 0)$.

As the first step, we clarify the leakage model of a device and the attack strategy of an adversary in practical scenarios. An illustration is provided in Fig. 4.2. Taking noisy leakage

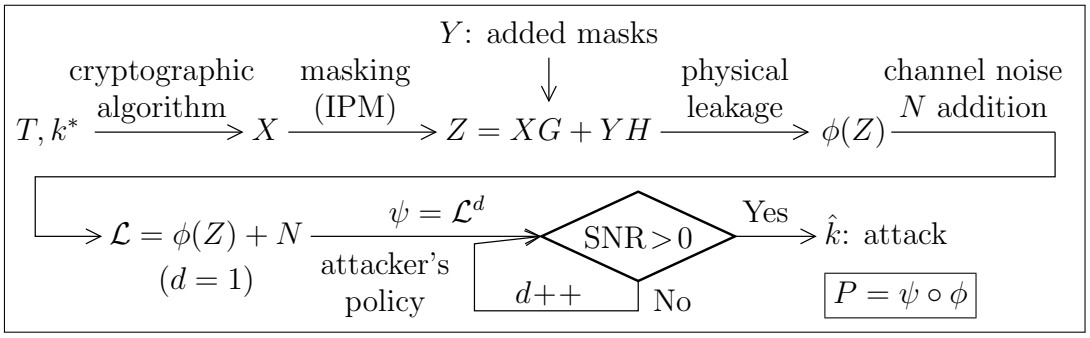


Figure 4.2: Overview of the attacker's strategy in the higher-order (moments) side-channel attacks to extract the secret key k^*, using side-channel leakages and the plain/cipher-text T.

model with additive Gaussian noise into consideration, we specify the leakages of a real device in two cases:

- In a serial implementation, all shares are manipulated at different times (clock cycles). We denote the leakages from the device as $\mathcal{L}_i = \phi(z_i) + N_i$, where $z_i \in \mathbb{K} = \mathbb{F}_{2^\ell}$ is the ith share and $N_i \sim \mathcal{N}(0, \sigma^2)$ for $i \in \{1, \ldots, n\}$ are associated noises. From the attacker's point of view, the best strategy is to combine these leakages together to launch higher-order attacks. As is known, the centered product combination is the most efficient combination function [129] [4]. Thus,

$$\mathcal{L}_{ser} = \prod_{i=1}^{d} \mathcal{L}_i = \prod_{i=1}^{d} (\phi(z_i) + N_i) = \underbrace{\prod_{i=1}^{d} \phi(z_i)}_{P(z):\, z \in \mathbb{K}^n} + \zeta + \prod_{i=1}^{d} N_i,$$

  where the adversary combines leakages of d shares over all n shares. ζ denotes intermediate terms with numerical degree $d^\circ\zeta < d$ that does not depend on the sensitive variables thus have no positive impact on attacks. Assume that N_i for $i \in \{1, 2, \ldots, n\}$ are i.i.d, then $\mathbb{V}\left[\prod_{i=1}^{d} N_i\right] = \sigma^{2d}$.

- In a fully parallel implementation, all shares are manipulated at the same time (the same clock cycle). Thus we have $\mathcal{L} = \phi(z) + N = \sum_{i=1}^{n} \phi(z_i) + N$ by assuming the device leaks in linear leakage model, where $z \in \mathbb{K}^n$ and $z_i \in \mathbb{K} = \mathbb{F}_{2^\ell}$. In this case, the best strategy is to use

[4] It is worth noting that Pearson correlation coefficient is invariant under affine transformation, although authors used the centered product in [129] to launch the correlation power analysis (CPA).

the least d-th order of statistical moments to launch the attack. Therefore,

$$\mathcal{L}_{par} = \mathcal{L}^d = \left(\sum_{i=1}^{n} \phi(z_i) + N\right)^d = \underbrace{\prod_{i=1}^{d} \phi(z_i)}_{P(z):\, z\in\mathbb{K}^n} + \zeta' + N^d,$$

where ζ' denotes intermediate terms with numerical degree $d^\circ\zeta' < d$. For Gaussian noise $N \sim \mathcal{N}(0, \sigma^2)$, by higher-order moments for Gaussian variables [120, § 5.4], we have:

$$\begin{aligned}\mathbb{V}\left[N^d\right] &= \mathbb{E}\left[N^{2d}\right] - \mathbb{E}\left[N^d\right]^2 \\ &= \begin{cases} \sigma^{2d}(2d-1)!! & \text{if } d \text{ is odd,} \\ \sigma^{2d}\big((2d-1)!! - (d-1)!!\big) & \text{if } d \text{ is even.} \end{cases}\end{aligned}$$

Hence, the variance of noise by raising to power d is proportional to σ^{2d}, namely:

$$\mathbb{V}\left[N^d\right] \propto \sigma^{2d}. \tag{4.5}$$

In summary, we formalize the leakage function (with the attacker's strategy) by a pseudo-Boolean function $P : \mathbb{K}^n \mapsto \mathbb{R}$ such that $P(z) = \prod_{i=1}^{d} \phi(z_i)$, which can be decomposed into $P(Z) = \sum_{I\in\mathbb{F}_2^{n\ell}} \alpha_I Z^I$ as in Eqn. 5.4. In both cases, we have $\mathbb{V}\left[N^d\right] \propto \sigma^{2d}$. Thanks to this model, we are able to explain the link between leakages at word-level and at bit-level. We also give an explanation on the physical defaults like physical couplings in a quantitative way. For instance, in AES implemented on a 32-bit embedded device (e.g. ARM Cortex 4), leakages of four bytes of intermediates may interfere with each other because of couplings, thus could leak the sensitive data from the joint distribution of leakages. This kind of joint distributions corresponds to the assignment of different values for α_I of $P(Z)$ as in Eqn. 5.4.

Definition of SNR. The SNR [102] is a critical security metric in the field of side-channel analysis, which is the ratio between the signal variance and the noise variance.

Let $\mathcal{L} = P(Z)+N$ denote the leakage which is irrespective to serial or parallel implementations. N denotes the independent noise with variance $\mathbb{V}[N] = \sigma^2_{total} \propto \sigma^{2d}$ as shown in Eqn. 4.5. We have $\mathbb{V}[\mathbb{E}[P(Z) + N|X]] = \mathbb{V}[\mathbb{E}[P(Z)|X]]$, then the SNR of leakages is defined as:

$$SNR = \frac{\mathbb{V}[\mathbb{E}[\mathcal{L}|X]]}{\mathbb{V}[N]} = \frac{\mathbb{V}[\mathbb{E}[P(Z)|X]]}{\sigma^2_{total}}. \tag{4.6}$$

In side-channel analysis, if SNR is null, attacks are merely impossible. Otherwise, attacks are possible and are all the more powerful as the SNR is larger.

4.4.2 Quantifying Leakages of IPM by SNR

Recall the coding form of IPM, whereby the sensitive variable X is encoded into Z by:

$$Z = X\mathbf{G} + Y\mathbf{H} \in \mathbb{K}^n = \mathbb{F}_{2^\ell}^n.$$

Let us consider this equation in $\mathbb{F}_2$ basefield, and thus let $\mathcal{X} = \mathbb{F}_2^\ell$, $\mathcal{Y} = \mathbb{F}_2^{(n-1)\ell}$ and $\mathcal{Z} = \mathbb{F}_2^{n\ell}$. We clarify the computations as follows (where $\mathcal{D}$ is expanded as per Def. 3.6):

- $\mathbb{E}\left[P(Z)|X = x\right]$ for a given $x \in \mathcal{X}$ is: $\mathbb{E}\left[P(x\mathbf{G} + Y\mathbf{H})\right] = \sum_{y\in\mathcal{Y}} \mathbb{P}(Y = y)P(x\mathbf{G} + y\mathbf{H}) = \frac{1}{|\mathcal{Y}|}\sum_{y\in\mathcal{Y}} P(x\mathbf{G} + y\mathbf{H}) = \frac{1}{|\mathcal{D}|}\sum_{d\in\mathcal{D}} P(x\mathbf{G} + d)$,
- For any variable X, we have that $\mathbb{V}\left[\mathbb{E}\left[P(Z)|X\right]\right] = \mathbb{E}\left[\mathbb{E}\left[P(Z)|X\right]^2\right] - \mathbb{E}\left[\mathbb{E}\left[P(Z)|X\right]\right]^2$.

Hence, we have the following two lemmas to compute terms $\mathbb{E}\left[\mathbb{E}\left[P(Z)|X\right]^2\right]$ and $\mathbb{E}\left[\mathbb{E}\left[P(Z)|X\right]\right]$ for IPM.

Lemma 4.1. $\mathbb{E}\left[\mathbb{E}\left[P(Z)|X\right]\right] = \frac{1}{2^{n\ell}}\widehat{P}(0)$.

Lemma 4.2. $\mathbb{E}\left[\mathbb{E}\left[P(Z)|X\right]^2\right] = \frac{1}{2^{2n\ell}}\sum_{x\in\mathcal{D}^\perp}\left(\widehat{P}(x)\right)^2$.

The proofs of Lemma 4.1 and 4.2 are in Appendix A.1. Therefore for the SNR of IPM scheme we have the following theorem.

Theorem 4.1. *Let a device be protected by the IPM scheme as $Z = X\mathbf{G} + Y\mathbf{H}$. Assume the leakages of the device can be represented in the form: $\mathcal{L} = P(Z) + N$ and an adversary may launch a d-th order attack by using higher-order moments (e.g., in parallel scenarios) or multivariate combinations (e.g., in serial scenarios). Hence the SNR of the exploitable leakages is:*

$$SNR = \frac{2^{-2n\ell}}{\sigma_{total}^2}\sum_{x\in\mathcal{D}^\perp\setminus\{0\}}\left(\widehat{P}(x)\right)^2,$$

where $\sigma_{total}^2 \propto \sigma^{2d}$.

Proof. On the basis of Lemma 4.1 & 4.2, we have that

$$\begin{aligned} SNR &= \frac{\mathbb{V}\left[\mathbb{E}\left[\mathcal{L}|X\right]\right]}{\mathbb{V}\left[N\right]} \\ &= \frac{\mathbb{E}\left[\mathbb{E}\left[P(Z)|X\right]^2\right] - \mathbb{E}\left[\mathbb{E}\left[P(Z)|X\right]\right]^2}{Var(N)} \\ &= \frac{2^{-2n\ell}}{\sigma_{total}^2}\left(\sum_{x\in\mathcal{D}^\perp}\widehat{P}^2(x) - \widehat{P}^2(0)\right) \\ &= \frac{2^{-2n\ell}}{\sigma_{total}^2}\sum_{x\in\mathcal{D}^\perp\setminus\{0\}}\widehat{P}^2(x). \end{aligned} \tag{4.7}$$

□

Remarkably, this quantity does not depend on the properties of the code $\mathcal{C}$, except for the fact it is supplementary to $\mathcal{D}$ in $\mathbb{K}^n$ (recall Lemma 4.1). The only factor for the SNR of IPM that makes all the differences is the choice of $\mathcal{D}$. The special interest of Theorem 4.1 is that it allows quantifying the leakages of any IPM and its variants (e.g. [165]). In a nutshell, Theorem 4.1 works under any form of P. Indeed, as shown in Fig. 4.2, the function P is composed of the leakage function ϕ from a device and the attack strategy ψ from an adversary, where ϕ and ψ can be any functions. In particular, a real device may produce nonlinear leakages rather than the simple Hamming weight one, where Theorem 4.1 can be applied straightforwardly.

4.4.3 Link between SNR and Security Order t

In fact, it is easy to build the connection between SNR and the side-channel security order of an implementation by checking whether SNR equals 0. From Theorem 4.1, we deduce the security order t of IPM from SNR as follows.

Theorem 4.2. *If $d^\circ P < d_{\mathcal{D}}^\perp$, the attack exploiting leakage function P fails (i.e., $SNR = 0$), thus the security order of IPM scheme in the bounded moment model is $t = d_{\mathcal{D}}^\perp - 1$.*

Proof. We know from [17, Lemma 1] (in fact, this is a direct consequence of results of [25]) that, given a pseudo-Boolean function P, one has $\widehat{P}(z) = 0$ for all $z \in \mathbb{K}^n$ such that $w_H(z) > d^\circ P$. Let $z \in \mathcal{D}^\perp \backslash \{0\}$; then $w_H(z) \geq d_{\mathcal{D}}^\perp$. Assuming that the numerical degree of P is strictly less than $d_{\mathcal{D}}^\perp$, we then have $w_H(z) \geq d_{\mathcal{D}}^\perp > d^\circ P$, which means that $\widehat{P}(z)$ equals 0, resulting in the fact that $SNR = \frac{1}{2^{2n\ell}\sigma_{total}^2} \sum_{x \in \mathcal{D}^\perp \backslash \{0\}} \widehat{P}(x)^2 = 0$. Hence, the security order t in *bounded moment model* equals $d_{\mathcal{D}}^\perp - 1$. □

Let us assume that the attacker builds its attack by tweaking P. For example, if the device leaks the sensitive variable Z through a noisy leakage function ϕ, the attacker can choose to use $P = \phi$ or $P = \phi^2$, ..., or $P = \phi^d$ (see illustration in Fig. 4.2), or actually any composition $P = \psi \circ \phi$. Therefore the security order is the minimum value of $d^\circ P$ such that $SNR \neq 0$. Although the Theorem 4.2 is essentially the same as [123, Proposition 1], we obtain this theorem in a different way. More importantly, by combining with Theorem 4.1, the quantitative leakages can be assessed straightforwardly. In practice, we can directly compare, for a given leakage model, two countermeasures: if $SNR_1 < SNR_2$, then the first countermeasure is more secure than the second one.

With Theorem 4.2, we directly link the dual distance $d_{\mathcal{D}}^\perp$ of codes $\mathcal{D}$ in IPM to the security order in bounded moment model. Furthermore, the quantitative expression in Theorem 4.1 allows designers to assess easily the security order of an IPM scheme by using properties of

the code $\mathcal{D}$. Since IPM is the generalization of Boolean masking, Theorem 4.1 & 4.2 are also applicable to the Boolean masking and other variants [165].

4.4.4 Connecting SNR with Code Parameters

As common leakage models, Hamming weight and affine models have been validated in practice [123] for side-channel analysis. We set $\phi(z) = w_H(z)$, then use $P(z) = w_H(z)^d$ as a leakage model. Clearly, the numerical degree of P is $d^\circ P = d$. Moreover, one can write:

$$\begin{aligned} P(z) = w_H(z)^d &= \sum_{J_1+\cdots+J_{n\ell}=d} \binom{d}{J_1,\ldots,J_{n\ell}} \prod_{i=1}^{n\ell} z_i^{J_i} \\ &= \sum_{\substack{J\in\mathbf{N}^{n\ell},\ \text{s.t.}\ \sum_{i=1}^{n\ell} J_i = d \\ w_H(J)<d}} \binom{d}{J} z^J + d! \sum_{\substack{I\in\{0,1\}^{n\ell} \\ w_H(I)=d}} z^I, \end{aligned} \tag{4.8}$$

where $\mathbf{N} = \{0, 1, \ldots\}$ is the set of integers. The multinomial coefficient $\binom{d}{J_1,\ldots,J_{n\ell}}$ is defined as $\frac{d!}{J_1!\cdots J_{n\ell}!}$ (recall that $J = (J_1, \ldots, J_{n\ell}) \in \mathbf{N}^{n\ell}$ with $\sum_{i=1}^{n\ell} J_i = d$). This coefficient equals to $d!$ as long as for all i ($1 \le i \le n\ell$), $J_i = 0$ or 1. Now, the terms in $P(z)$ are categorized into two cases:

- z^J where $J \in \mathbf{N}^{n\ell}$, $w_H(J) < d$, which consists in products of $< d$ bits of z, as $z^J = \prod_{i\in\{1,\ldots,n\ell\}\ s.t.\ J_i>0} z_i$,
- z^I where $I \in \{0,1\}^{n\ell}$, $w_H(I) = d$ which consists in products of d bits of z, as $z^I = \prod_{i\in\{1,\ldots,n\ell\}\ s.t.\ I_i=1} z_i$.

Indeed, let $i \in \{1, \ldots, n\ell\}$, then $z_i^{J_i} = 1$ if $J_i = 0$, and $z_i^{J_i} = z_i$ if $J_i > 0$. The first terms z^J have numerical degree $d^\circ(z^J) < d$, hence can be discarded in the analysis (they contribute nothing to the SNR). Remaining terms of numerical degree d are: $\sum_{I\in\{0,1\}^{n\ell},\, w_H(I)=d} z^I$. Hence we have following theorem for quantifying the leakages of IPM.

Theorem 4.3. *Let a device leak in Hamming weight model, which is protected with IPM at bit-level security order $t = d_{\mathcal{D}}^{\perp} - 1$. A higher-order attack is possible only if the attacker uses a leakage function P with $d^\circ P = d > t$. Moreover, the SNR can be quantified by:*

$$SNR = \begin{cases} 0 & \text{if } d^\circ P \le t \\ \frac{B_{d_{\mathcal{D}}^{\perp}}}{\sigma^2_{total}} \left(\frac{d_{\mathcal{D}}^{\perp}!}{2^{d_{\mathcal{D}}^{\perp}}}\right)^2 & \text{if } d^\circ P = t+1 = d_{\mathcal{D}}^{\perp}\ . \end{cases} \tag{4.9}$$

Proof. Let $\varphi_I(z) = z^I$ where $I \in \{0,1\}^{n\ell}$. Thus

$$z^I = \prod_{i\in I} z_i = \prod_{i\in I} \frac{(1-(-1)^{z_i})}{2} = \frac{1}{2^d} \prod_{i\in I} (1-(-1)^{z_i}). \tag{4.10}$$

By Theorem 4.2, all monomials with numerical degree $d^\circ < d$ have $SNR = 0$, hence we only focus on monomials with $d^\circ = d$. We have $\varphi_I(z) = \phi_I(z) + \frac{(-1)^d}{2^d}(-1)^{\sum_{i\in I} z_i}$ where $\phi_I(z)$ is linear combination of monomials with numerical degree $d^\circ < d$ in $\varphi_I(z)$. The *Fourier transform* of $\varphi_I(z)$ is

$$\begin{aligned}\widehat{\varphi}_I(y) &= \widehat{\phi}_I(y) + \frac{(-1)^d}{2^d}\sum_z (-1)^{z\cdot I}(-1)^{z\cdot y} \\ &= \widehat{\phi}_I(y) + \frac{(-1)^d}{2^d}\sum_z (-1)^{z\cdot (I+y)} \\ &= \widehat{\phi}_I(y) + (-1)^d 2^{n\ell-d}\mathbf{1}_{\{I\}}(y).\end{aligned} \tag{4.11}$$

We have $\widehat{\phi}_I(y) = 0$ for y with $w_H(y) \geq d_{\mathcal{D}}^\perp = t+1 > d$ (recall the proof of Theorem 4.2).

Thus by combining Eqn. 4.11 with Eqn. 4.7, we have the following equation for $\mathbb{V}\left[\mathbb{E}\left[P(Z)|X\right]\right]$:

$$\begin{aligned}\mathbb{V}\left[\mathbb{E}\left[P(Z)|X\right]\right] &= \sum_{y\in\mathcal{D}^\perp\setminus\{0\}} \frac{\widehat{P}^2(y)}{2^{2n\ell}} \\ &= 2^{-2n\ell}\sum_{y\in\mathcal{D}^\perp\setminus\{0\}}\left[\sum_{I|w_H(I)=d}(-1)^d 2^{n\ell-d}\binom{d}{I}\mathbf{1}_{\{I\}}(y)\right]^2 \\ &= 2^{-2d}\sum_{y\in\mathcal{D}^\perp,\, w_H(y)=d}\left[\sum_{I|w_H(I)=d}\binom{d}{I}\mathbf{1}_{\{I\}}(y)\right]^2 \\ &= 2^{-2d}\sum_{y\in\mathcal{D}^\perp,\, w_H(y)=d}(d!)^2 \\ &= B_d\left(\frac{d!}{2^d}\right)^2.\end{aligned} \tag{4.12}$$

Finally, using Theorem 4.2, it appears that the only possible solution of d is $d = d_{\mathcal{D}}^\perp$ such that $SNR \neq 0$, thus $\mathbb{V}\left[\mathbb{E}\left[P(Z)|X\right]\right] = B_{d_{\mathcal{D}}^\perp}\left(\frac{d_{\mathcal{D}}^\perp!}{2^{d_{\mathcal{D}}^\perp}}\right)^2$, then

$$SNR = \frac{\mathbb{V}\left[\mathbb{E}\left[P(Z)|X\right]\right]}{\mathbb{V}\left[N\right]} = \frac{B_{d_{\mathcal{D}}^\perp}}{\sigma^2_{total}}\left(\frac{d_{\mathcal{D}}^\perp!}{2^{d_{\mathcal{D}}^\perp}}\right)^2. \tag{4.13}$$

□

In a nutshell, Theorem 4.3 provides a quantitative way for assessing the side-channel security level of an implementation under Hamming weight leakages. More importantly, the SNR is linked to two parameters of the code used in IPM, which brings great convenience on simplifying the assessment. In practice, the designer can easily select a better or even optimal code for IPM which amplifies the side-channel resistance of the implementation protected by IPM.

In fact, the quantitative result in Theorem 4.3 can be extended to all linear (affine) leakages [97] which can be expressed as:

$$P : z \in \mathbb{F}_2^{n\ell} \mapsto P(z) = \beta_0 + \langle \beta, z\rangle \in \mathbb{R}, \tag{4.14}$$

where $\beta_0 \in \mathbb{R}$ is an unimportant additive constant that can be considered null (which is dropped in the sequel), $\beta = (\beta_1, \cdots, \beta_{n\ell}) \in \mathbb{R}^{n\ell}$ (note that β_i is normalized by $\beta'_i = \frac{\sqrt{n\ell}}{\|\beta\|_2}\beta_i$, where $\|\beta\|_2 = \sqrt{\sum_{i=1}^{n\ell} \beta_i^2}$ is the L$_2$-norm of β) are coordinate-wise weights, and $\langle x, y\rangle = \sum_{i=1}^{n\ell} x_i \cdot y_i \in \mathbb{R}$ is the canonical scalar product. This leakage model is also known as UWSB model ("unevenly weighted sum of the bits") as in [172]. The validity of this leakage model can be tested easily by stochastic profiling [143] on practical samples. Therefore, we denote the leakage function as:

$$\begin{aligned} P(z) &= \Big(\sum_{i=1}^{n\ell} \beta_i z_i\Big)^d = \sum_{\substack{J_1+J_2+\cdots \\ +J_{n\ell}=d}} \binom{d}{J_1, \ldots, J_{n\ell}} \prod_{i=1}^{n\ell} (\beta_i z_i)^{J_i} \\ &= \sum_{\substack{J\in\mathbf{N}^{n\ell},\, w_H(I)<d \\ J_1+\cdots+J_{n\ell}=d}} \binom{d}{J} (\beta z)^J + d! \sum_{\substack{I\in\{0,1\}^{n\ell} \\ w_H(I)=d}} (\beta z)^I. \end{aligned} \tag{4.15}$$

Thus, we deduce the following corollary for SNR under UWSB leakage model as follows:

Corollary 4.1. *Let a device leak in UWSB model, which is protected with IPM at bit-level security $t = d_{\mathcal{D}}^{\perp} - 1$. A higher-order attack is possible only if the attacker uses a leakage function P with numerical degree $d^\circ P = d > t$. Moreover, the SNR is:*

$$SNR = \lambda \cdot \frac{1}{\sigma_{total}^2} \left(\frac{d!}{2^d}\right)^2 = \begin{cases} 0 & \text{if } d^\circ P \le t \\ \lambda \cdot \frac{1}{\sigma_{total}^2} \left(\frac{d_{\mathcal{D}}^{\perp}!}{2^{d_{\mathcal{D}}^{\perp}}}\right)^2 & \text{if } d^\circ P = t+1 = d_{\mathcal{D}}^{\perp} \end{cases}, \tag{4.16}$$

where $\lambda = \sum_{\substack{y\in\mathcal{D}^{\perp} \\ w_H(y)=d}} \Big(\prod_{\substack{1\le i\le n\ell \\ s.t.\ y_i=1}} \beta_i\Big)^2$, *and* $\lambda = 0$ *if* $d < d_{\mathcal{D}}^{\perp}$.

For instance, as the Hamming weight model is a special case of UWSB model with $\beta_i = 1$ for $i \in \{1, 2, \ldots, n\ell\}$, we obtain $\lambda = B_{d_{\mathcal{D}}^{\perp}}$, which is exactly the Theorem 4.3.

In summary, by Corollary 4.1, the SNR of IPM scheme under affine leakage model depends only on the two parameters of $\mathcal{D}$ and the leakage model β. In practice, β is fixed for a given device and mainly depends on the device itself that an adversary has no control on. Hence the special interest is that designers can choose optimal codes $\mathcal{D}$ for IPM with maximized side-channel resistance by simply selecting optimal $d_{\mathcal{D}}^{\perp}$ and $B_{d_{\mathcal{D}}^{\perp}}$.

Numerical Comparison with codes by SNR in Theorem 4.3. First, we show the SNR of different codes for IPM in Tab. 4.2. We omit the full table of all codes with different $B_{d_{\mathcal{D}}^{\perp}}$, but only showing codes with the maximal and the minimal values of $B_{d_{\mathcal{D}}^{\perp}}$. The last column of Tab. 4.2 shows the possible candidates of α_2 in IPM, by which the generator matrix of corresponding code $\mathcal{D}$ is $\mathbf{H} = (\alpha_2,\ 1)$.

Table 4.2: Demonstration of categorizing codes in IPM by SNR and *MI*.

$d_{\mathcal{D}}^{\perp}$	$B_{d_{\mathcal{D}}^{\perp}}$	σ^2·SNR when $P(Z) = w_H(Z)^d$				$MI \cdot 10^3$	α_2
		$d=1$	$d=2$	$d=3$	$d=4$	($\sigma=3$)	(Non-equivalent codes)
2	max B_2=8	0	2.0000	>0	>0	4.4025	$\{\alpha^0\}$ [1]
	min B_2=1	0	0.2500	>0	>0	3.8410	$\{\alpha^7\}$
3	max B_3=7	0	0	3.9375	>0	0.5470	$\{\alpha^{24}, \alpha^{25}\}$
	min B_3=1	0	0	0.5625	>0	0.2374	$\{\alpha^{18}, \alpha^{39}, \alpha^{43}, \ldots\}$ [2]
4	max B_4=15	0	0	0	33.750	0.0494	$\{\alpha^{59}, \alpha^{60}\}$
	min B_4=3	0	0	0	6.7500	0.0392	$\{\alpha^{8}, \alpha^{126}, \alpha^{127}\}$

In Tab. 4.2, the green part shows where SNR equals 0 given different $d_{\mathcal{D}}^{\perp}$. Clearly, given a $d_{\mathcal{D}}^{\perp}$ for IPM codes, the SNR of the corresponding IPM decreases along with $B_{d_{\mathcal{D}}^{\perp}}$. Moreover, the *MI* decreases when $d_{\mathcal{D}}^{\perp}$ increases and/or $B_{d_{\mathcal{D}}^{\perp}}$ decreases. As a result, the code with the maximized $d_{\mathcal{D}}^{\perp}$ and minimized $B_{d_{\mathcal{D}}^{\perp}}$ performs best against side-channel analysis, which validates our approach on selecting optimal codes for IPM.

To summarize, there are three optimal codes (up to equivalence) for 2-share IPM on $\mathbb{F}_{2^8}$. Those codes have dual distance 4 and only 3 codewords of nonzero minimum weight equal to 3. Those codes are the optimal for 2-share IPM operating on bytes, and were previously not specifically distinguished amongst binary codes of parameters $[16, 8]_2$. Their generating matrix are provided in Appendix B.1.

4.5 Measuring Leakages by Mutual Information

We investigate the security order of IPM at both word- and bit-level, and show the essential reason of the "Security Order Amplification" which has been observed and described in [3, 123, 165]. We here go further by using an information-theoretic metric, the standard notion of mutual information (MI), to quantify the leakages of IPM.

4.5.1 Security Orders at Word-level t_w and Bit-level t_b

The first important property of IPM is its higher security order at bit-level than at word-level, namely $t_b \geq t_w$. Here we start from a very well-known property of the generator matrix.

[1] This code corresponds to the Boolean masking where $\alpha_2 = \alpha^0 = 1$ and $\mathbf{H} = (1,\ 1)$.
[2] There are 36 codes (including equivalent codes) with $d_{\mathcal{D}}^{\perp} = 3$ and $\min B_3 = 1$ [39].

Proposition 4.1. *The maximal number of linearly independent columns of the generator matrix* $\mathbf{H}$ *of the code* $\mathcal{D}$ *is* $d_{\mathcal{D}}^{\perp} - 1$.

It is a well-known theorem in error-correcting codes [101, Theorem 10]. Hence, if the attacker probes up to $d < d_{\mathcal{D}}^{\perp}$ (inclusive) wires, the sensitive variable X is encoded as a codeword in $\mathbb{F}_{2^\ell}^n$ and is perfectly masked. Therefore no information on X can be recovered.

Lemma 4.3. *The IPM is secure at the maximized order* t_w *in the terms of probing model if and only if the code generated by the* $1 \times n$ *matrix* $\mathbf{H}^{\perp} = (\alpha_1 = 1, \alpha_2, \alpha_3, \ldots, \alpha_n)$ *is a code with parameters* $[n, 1, d_w]_{2^\ell}$*, where* $d_w = t_w + 1$.

Proof. Note that $\mathbf{H}^{\perp} = (1, \alpha_2, \alpha_3, \ldots, \alpha_n)$ is the generator matrix of the dual code of $\mathcal{D}$ generated by matrix $\mathbf{H}$ in Eqn. 4.4. The masking scheme is secure at order t_w under *probing model* means that any tuple of Z's coordinates of size $\leq t_w$ leaks no information on X. Now, $Z = f(X) + Y\mathbf{H}$, i.e., similar to additive masking, which is secure at order t_w meaning that any t_w tuple of $Y\mathbf{H}$ is uniformly distributed ("Vernam code"). By definition, this means that $d_{\mathcal{D}}^{\perp} > t_w$.

Since for IPM, $d_{\mathcal{D}}^{\perp} = d_{\mathcal{D}^{\perp}}$ where the later is the minimum distance of the dual code $\mathcal{D}^{\perp}$. By the definition of the dual distance, we have $d_{\mathcal{D}}^{\perp} = t_w + 1$. □

Obviously, we have $d_{\mathcal{D}}^{\perp} = n$ if and only if $\alpha_i \neq 0$ for $i \in \{1, 2, \ldots, n\}$. Therefore, the security order of IPM scheme is $t_w = d_{\mathcal{D}}^{\perp} - 1 = (n - 1)$ over $\mathbb{K} = \mathbb{F}_{2^\ell}$. This has been formally proved in [2, 3] and pointed out in [123]. We put it here in Lemma 4.3 for completeness of this thesis and we can directly obtain the word-level security order t_w by Theorem 4.2. In brief, IPM is optimal in term of word-level security and has the same security order as Boolean masking (where $\alpha_i = 1$).

4.5.2 Bit-Level Security Order t_b

By code expansion as Def. 3.6, we can expand the code $\mathcal{D}$ from $\mathbb{K} = \mathbb{F}_{2^\ell}$ to $\mathbb{K} = \mathbb{F}_2$, which turns a code $[n, 1, d_w]_{2^\ell}$ to $[n\ell, \ell, d_b]_2$. At first, we show the connection between the two security orders t_w and t_b as follows.

Lemma 4.4. *In IPM, the word-level security order is not greater than bit-level security order, namely* $t_w \leq t_b$.

In fact, this is essentially the "Security Order Amplification" as explained in [123]. Here we give another proof as follows.

Proof. With code expansion, the generator matrix $\mathbf{H}^{\perp} = (1, \alpha_2, \cdots, \alpha_n)$ is expanded to $[\mathbf{H}^{\perp}]_2 = (I_\ell, [\alpha_2], \cdots, [\alpha_n])$, as per Def. 3.6. Since $\alpha_i \neq 0$, we have at least one 1 in each row of $[\alpha_i]$.

Otherwise, if one row of $[\alpha_i]$ is all zeros, we have:

$$[\alpha_i] = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ - & - & - & - \\ - & - & - & - \\ - & - & - & - \end{pmatrix} \in \mathbb{F}_2^{\ell \times \ell}$$

then $[\alpha_i]$ is not invertible which indicates that L_i does not exist. □

Remark 4.2. It is worth mentioning that since the Boolean masking is a special case of IPM with $\alpha_i = 1$ for $i \in \{1, \ldots, n\}$, we always have $t_w = n - 1$, $t_b = n - 1$. While for IPM, t_b can be greater than $n - 1$ if there exists at least one $i \in \{1, \ldots, n\}$ such that $\alpha_i \notin \{0, 1\}$, where the algebraic complexity of IPM is greater than Boolean masking.

To optimize IPM scheme, we aim at choosing t_b as large as possible compared to t_w, thereby increasing the security order as much as possible. For instance, with $n = 2$ shares of $\ell = 4$ bits, we have $\mathbb{F}_{16} = \{0, 1, \alpha, \ldots, \alpha^{14}\}$, where $\mathbb{F}_{16} := \mathbb{F}_2[\alpha]/\langle \alpha^4 + \alpha + 1 \rangle$. There are 15 candidates for $\alpha_2 \in \mathbb{F}_{16} \backslash \{0\}$. All codes have the same word-level security since $t_w = 1$ ($d_w = 2$). While for bit level security, we have $t_b = 1$ ($d_b = 2$) for 7 candidates and $t_b = 2$ ($d_b = 3$) for 8 candidates (refer to Tab. 4.5 for all codes), respectively. Therefore, in this case, the optimal t_b for IPM is 2.

4.5.3 Linking Mutual Information with Code Parameters

The other primary means to evaluate the security of a cryptographic implementation is to utilize the information-theoretic analysis. In this sense, mutual information is a well-known metric in the field of side-channel analysis [151]. Therefore we use it to assess the leakages of IPM as follows.

Theorem 4.4. *For a device leaking under the Hamming weight model that is protected by IPM scheme with $Z = X\mathbf{G} + Y\mathbf{H}$, the mutual information $\mathsf{I}(\mathcal{L}; X)$ between the leakage $\mathcal{L} = P(Z) + N$ and the sensitive variable X is approximately equal to the first nonzero term: $\mathsf{I}(\mathcal{L}; X) \approx \frac{d_{\mathcal{D}}^{\perp}! B_{d_{\mathcal{D}}^{\perp}}}{2 \ln 2 \cdot 2^{2 d_{\mathcal{D}}^{\perp}}} \cdot \frac{1}{\sigma^{2 d_{\mathcal{D}}^{\perp}}}$ when the leakage function P of a higher-order attack has numerical degree $d^{\circ} P = d_{\mathcal{D}}^{\perp}$. Specifically,*

$$\mathsf{I}(\mathcal{L}; X) = \begin{cases} 0, & \text{if } d^{\circ} P < d_{\mathcal{D}}^{\perp} \\ \frac{d_{\mathcal{D}}^{\perp}! B_{d_{\mathcal{D}}^{\perp}}}{2 \ln 2 \cdot 2^{2 d_{\mathcal{D}}^{\perp}}} \times \frac{1}{\sigma^{2 d_{\mathcal{D}}^{\perp}}} + \mathcal{O}\left(\frac{1}{\sigma^{2(d_{\mathcal{D}}^{\perp}+1)}}\right), & \text{if } d^{\circ} P = d_{\mathcal{D}}^{\perp}, \text{ when } \sigma \to +\infty \end{cases} \tag{4.17}$$

where σ is the standard deviation of noise.

Proof. It is obvious that there is no leakage when $d^{\circ} P = d < d_{\mathcal{D}}^{\perp}$. We assess the leakages in an information-theoretic sense as the mutual information between $P(Z)$ and X, defined by $\mathsf{I}(P(Z); X) = \mathsf{H}(P(Z)) - \mathsf{H}(P(Z)|X)$, where:

- the entropy is $\mathsf{H}(P(Z)) = -\sum_z \mathbb{P}(P(z)) \log_2 \mathbb{P}(P(z))$,

- the conditional entropy $\mathsf{H}(P(Z)|X)$ is:

$$\mathsf{H}(P(Z)|X) = -\sum_{x \in \mathbb{F}_2^\ell} \mathbb{P}_X(x) \sum_z \mathbb{P}(P(z)|x) \log_2 \mathbb{P}(P(z)|x).$$

In the presence of noise N, the mutual information between the noisy leakage $\mathcal{L} = P(Z) + N$ and the sensitive variable X can be developed using a Taylor's expansion[5] [23]:

$$\begin{aligned} \mathsf{I}(\mathcal{L}; X) &= \sum_{d=0}^{+\infty} \frac{1}{2\, d! \ln 2} \sum_{x \in \mathbb{F}_2^\ell} \mathbb{P}_X(x) \frac{\left(k_d(P(Z)|x) - k_d(P(Z))\right)^2}{\left(Var(P(Z)) + \sigma^2\right)^d} \\ &= \frac{1}{\ln 2} \sum_{d=0}^{+\infty} \frac{1}{2\, d!} \frac{\mathbb{V}\left[k_d(P(Z)|X)\right]}{\left(\mathbb{V}\left[P(Z)\right] + \sigma^2\right)^d} \ , \end{aligned} \tag{4.18}$$

where k_d is the d-th order cumulant [21].

As for a d-CI (*Correlation Immune*) function [22] that is not $(d+1)$-CI, all moments of order $i \leq d$ are centered, so are the cumulants. Hence the first nonzero cumulant $k_{d_{\mathcal{D}}^\perp}(X)$ is equal to $\mu_{d_{\mathcal{D}}^\perp}(X)$. It results that in Eqn. 4.18, the term $\mathbb{V}\left[k_d(P(Z)|X)\right]$ is null for all $d < d_{\mathcal{D}}^\perp$, and it is equal to $\mathbb{V}\left[\mu_{d_{\mathcal{D}}^\perp}(P(Z)|X)\right] = \mathbb{V}\left[\mathbb{E}\left[P(Z)^{d_{\mathcal{D}}^\perp}|X\right]\right]$ for $d = d_{\mathcal{D}}^\perp$. Thus, assuming the device is leaking in Hamming weight model, the mutual information can be developed at the first order in $1/\sigma^{2d_{\mathcal{D}}^\perp}$ by Eqn. 4.18:

$$\mathsf{I}(\mathcal{L}; X) = \frac{d_{\mathcal{D}}^\perp! B_{d_{\mathcal{D}}^\perp}}{2 \ln 2 \cdot 2^{2d_{\mathcal{D}}^\perp}} \times \frac{1}{\sigma^{2d_{\mathcal{D}}^\perp}} + \mathcal{O}\left(\frac{1}{\sigma^{2(d_{\mathcal{D}}^\perp+1)}}\right), \tag{4.19}$$

when $\sigma \to +\infty$. This proves Theorem 4.4. □

Particularly from Theorem 4.3 and 4.4, it is noteworthy that reducing $B_{d_{\mathcal{D}}^\perp}$ allows both to reduce the SNR and the MI, which demonstrates our intuition for the impact of $B_{d_{\mathcal{D}}^\perp}$ on the concrete security level of IPM. In summary, two parameters that determine the leakages of IPM are depicted in Fig. 4.3:

- the slope in the log-log representation of the MI versus the noise standard deviation is all the more steep as $d_{\mathcal{D}}^\perp$ is higher,
- the vertical offset is adjusted by $B_{d_{\mathcal{D}}^\perp}$: the smaller $B_{d_{\mathcal{D}}^\perp}$ is, the smaller the MI.

When the noise variance σ^2 tends to infinity, $\mathsf{I}(\mathcal{L}; X)$ is converging to the dominating term in the expansion given in Eqn. 4.19. Hence, there is an affine law in the log-log representation, in which the slope equals to the negative order of the first nonzero moment of random variable $\mathcal{L}|X$, namely the least order of key-dependent moments.

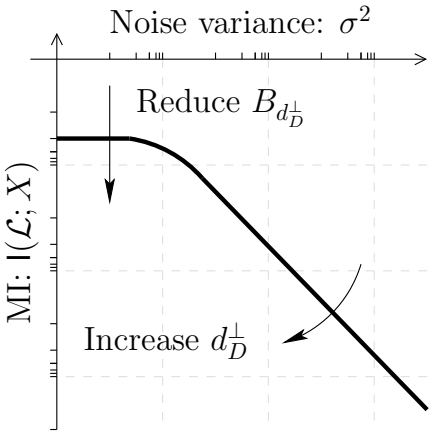


Figure 4.3: Two concomitant objectives to reduce the mutual information.

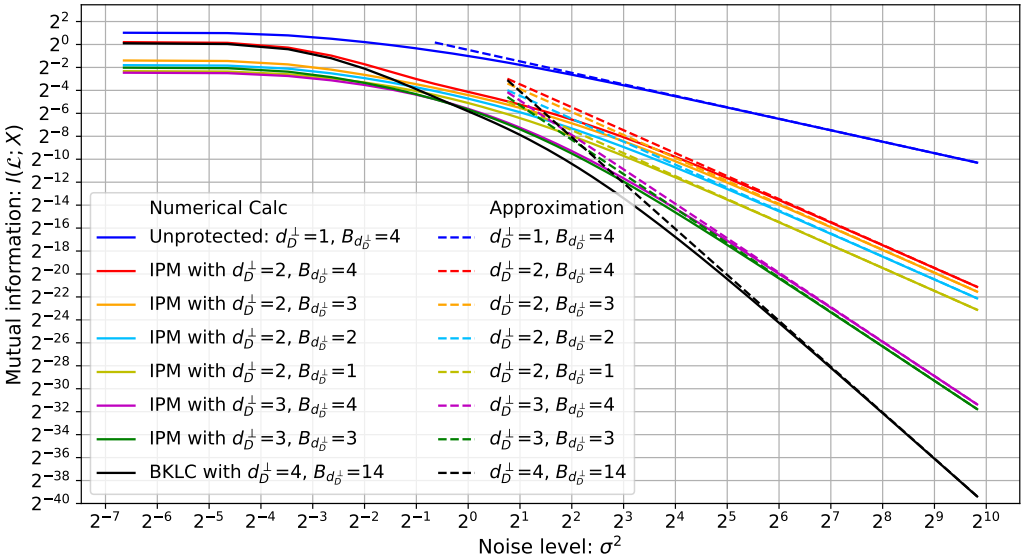


Figure 4.4: Numerical calculation and approximation of $\mathsf{I}(\mathcal{L}; X)$ between leakages and the sensitive variable $X \in \mathbb{F}_{2^4}$ in IPM. The *BKLC* code [8, 4, 4] cannot be used in IPM. We put it here to show the code with $d_{\mathcal{D}}^{\perp} = 4$.

Numerical Evaluation of MI. By information-theoretic analysis, we connected the mutual information with two defining parameters of $\mathcal{D}$, namely $d_{\mathcal{D}}^{\perp}$ and $B_{d_{\mathcal{D}}^{\perp}}$. In order to further demonstrate Theorem 4.4, we numerically compute the MI for $n = 2$ shares and $\ell = 4$ bits. The value of $\mathsf{I}(\mathcal{L}; X)$ is shown in Fig. 4.4, where $\mathcal{L}$ takes the *"Hamming weight + Gaussian noise"* as leakages. Several illustrations of leakage distributions are depicted in Appendix C.1 for IPM with $\alpha_1 \in \{1, \alpha, \alpha_5\}$.

[5]The normalization by $\ln 2$ allows the mutual information expressed in bits instead of nats.

Obviously, $d_{\mathcal{D}}^{\perp}$ and $B_{d_{\mathcal{D}}^{\perp}}$ clearly indicate the concrete security level of IPM as measured by MI. Furthermore, our estimation of MI by Eqn. 4.19 is in accordance with the numerical calculations. From Fig. 4.4, the codes for practical applications can be chosen according to the noise level of real devices (situations). For instance, if the noise level is $\sigma^2 > 2$, the $d_{\mathcal{D}}^{\perp}$ is more dominant on choosing optimal codes, while if $\sigma^2 < 2^{-1}$, $B_{d_{\mathcal{D}}^{\perp}}$ is more important; for noise level σ^2 in $[2^{-1}, 2]$, more efforts are needed in choosing a good code.

4.6 A Unified Leakage Assessment Framework for IPM

We introduce a unified framework, consisting in the two parameters $d_{\mathcal{D}}^{\perp}$ and $B_{d_{\mathcal{D}}^{\perp}}$ of code $\mathcal{D}$, to quantify the linear (e.g., Hamming weight) and affine leakages of IPM. By Theorem 4.3 and Corollary 4.1, we propose the unified framework for assessing the leakages of IPM as follows:

Framework 1 (Unified Leakage Assessment Framework for IPM)**.** *The leakages of IPM with a linear code $\mathcal{D}$ can be quantified by the assessment framework consisting of two defining parameters of $\mathcal{D}$, namely its dual distance $d_{\mathcal{D}}^{\perp}$ and the coefficient $B_{d_{\mathcal{D}}^{\perp}}$ in its weight enumerator (recall Theorem 4.3).*

In summary, when the leakage model is Hamming weight or affine model, the side-channel resistance of IPM scheme is straightforwardly related to two defining parameters of the selected code $\mathcal{D}$, namely $d_{\mathcal{D}}^{\perp}$ and $B_{d_{\mathcal{D}}^{\perp}}$, which are core ingredients of our unified framework. From the attacker's perspective, the only way to compromise a countermeasure is to perform attacks with order no less than $d_{\mathcal{D}}^{\perp}$. From the other side of the coin, designers can use this framework in practice, namely to enhance the side-channel security of IPM by choosing appropriate $d_{\mathcal{D}}^{\perp}$ and $B_{d_{\mathcal{D}}^{\perp}}$. Hereafter, we show how to use this framework to select the optimal codes for IPM.

4.6.1 Selecting Optimal Codes for IPM

Recall that the generator matrix of dual code $\mathcal{D}^{\perp}$ is $\mathbf{H}^{\perp} = (\alpha_1 = 1, \alpha_2, \alpha_3, \ldots, \alpha_n)$. From above, two ingredients of our unified framework are $d_{\mathcal{D}}^{\perp}$ and $B_{d_{\mathcal{D}}^{\perp}}$ of the code $\mathcal{D}$. Since our framework straightforwardly indicates the concrete security order of IPM, we propose an algorithm to choose optimal code for IPM as Alg. 1.

Summing up, our framework is generic and applicable to IPM under Hamming weight and affine leakages. From the perspective of designers, it would be advantageous to choose the optimal codes with proper $d_{\mathcal{D}}^{\perp}$ and $B_{d_{\mathcal{D}}^{\perp}}$ by using Alg. 1 instead of finding them via long and tedious design then evaluation cycles. Some optimal codes are shown in Tab. 4.3.

Algorithm 1: Conceptional Selection of Optimal Code for IPM

Input : All codes of $\mathcal{D}^{\perp}$ generated by $\mathbf{H}^{\perp}$

Output : Code(s) with Optimized $d_{\mathcal{D}}^{\perp}$ and $B_{d_{\mathcal{D}}^{\perp}}$

$\mathscr{D}$ $\leftarrow$ all codes $\mathcal{D}$ with $\mathcal{D}^{\perp}$: $[n, 1, n]_{2^\ell}$ over $\mathbb{F}_{2^\ell}$; // Optimize d_w in word-level, $d_w = n$ if $\alpha_i \neq 0$

$\mathscr{D}_2 \leftarrow \{[\mathcal{D}]_2 |\ \mathcal{D} \in \mathscr{D}\}$ over $\mathbb{F}_{2^\ell}$ with $[\mathcal{D}^{\perp}]_2$: $[n\ell, \ell, d_b]_2$;

$d_b \leftarrow \max\{d_{\mathcal{D}}^{\perp} |\ \mathcal{D} \in \mathscr{D}_2\}$; // Optimize in bit-level (d_b)

$\mathscr{D}' = \{\mathcal{D} |\ d_{\mathcal{D}}^{\perp} = d_b,\ \mathcal{D} \in \mathscr{D}\}$; // Only keep codes with maximized $d_{\mathcal{D}}^{\perp}$

$B_{min} \leftarrow \min\{B_{d_{\mathcal{D}}^{\perp}} |\ \mathcal{D} \in \mathscr{D}'\}$;

$\mathscr{D}'' = \{\mathcal{D} |\ B_{d_{\mathcal{D}}^{\perp}} = B_{min},\ \mathcal{D} \in \mathscr{D}'\}$; // Only keep codes with minimized $B_{d_{\mathcal{D}}^{\perp}}$

return $\mathscr{D}''$;

For $n = 2$ and $n = 3$ shares for 4 & 8-bit variables, the best IPM codes and *BKLC* codes are tabulated in Tab. 4.3.

Table 4.3: The optimal codes for IPM in several scenarios with BKLCs and Boolean one in comparison (refer to [39] for list of all codes).

	$\mathbb{F}_{2^\ell}$	IPM Codes	*BKLC* Codes	$d_{\text{ipm}}^{\perp} - d_{\text{bool}}^{\perp}$	$d_{\text{ipm}}^{\perp} - d_{\text{bklc}}^{\perp}$	SR_{bool}=0.8	SR_{ipm}=0.8	SR_{bklc}=0.8	Comments
$n=2$	$\ell = 4$	$\mathbf{H}^{\perp}$=(1, α^5): $d_{\mathcal{D}}^{\perp} = 3$, $B_{d_{\mathcal{D}}^{\perp}} = 3$	[8, 4, 4]: (unique) $d_{\mathcal{D}}^{\perp} = 4$, $B_{d_{\mathcal{D}}^{\perp}} = 14$	**1**	-1	800 (1D) 340 (2D)	4,000 (1D) 1,320 (2D)	8,400 (1D) 2,500 (2D)	[24, 165]
	$\ell = 8$	$\mathbf{H}^{\perp}$=(1, α^8): $d_{\mathcal{D}}^{\perp} = 4$, $B_{d_{\mathcal{D}}^{\perp}} = 3$	[16, 8, 5]: (unique) $d_{\mathcal{D}}^{\perp} = 5$, $B_{d_{\mathcal{D}}^{\perp}} = 24$	**2**	-1	1,900 (1D) 870 (2D)	>80,000 (1D) >20,000 (2D)	>100,000 (1D) >40,000 (2D)	[123]. We introduce one nonlinear code (16,256,6)
$n=3$	$\ell = 4$	$\mathbf{H}^{\perp}$=(1, α^5, α^{10}): $d_{\mathcal{D}}^{\perp} = 6$, $B_{d_{\mathcal{D}}^{\perp}} = 12$	[12, 4, 6]: $d_{\mathcal{D}}^{\perp} = 6$, $B_{d_{\mathcal{D}}^{\perp}} = 12$	**3**	0	4,600 (1D) 310 (3D)	>45,000 (1D) 3,050 (3D)	>45,000 (1D) 3,050 (3D)	**New**, the best IPM code is equivalent to *BKLC* code
	$\ell = 8$	$\mathbf{H}^{\perp}$=(1, α^{18}, α^{183}): $d_{\mathcal{D}}^{\perp} = 8$, $B_{d_{\mathcal{D}}^{\perp}} = 7$	[24, 8, 8]: $d_{\mathcal{D}}^{\perp} = 8$, $B_{d_{\mathcal{D}}^{\perp}} = 130$	**5**	0	–	–	–	[123], the best IPM codes is better than *BKLC* one [39]

4.6.2 The Completeness of Our Unified Framework

In this chapter, we quantify the side-channel security of IPM using two complementary metrics, namely the SNR and the MI, since they depict different aspects of the side-channel leakage. Specifically,

- the SNR measures the amount of leakage at a given moment (mean, variance, etc.) in the bounded leakage model;

- the MI measures the total leakage distribution, namely depending on all orders of moments of the leakage.

Both metrics are correlated with the attack metric *SR*, which is the pragmatic evaluation of the exploitability of the leakage. This means that the smaller the SNR or the MI, the smaller the *SR* for a given number of traces used to attack. Moreover, the two complementary metrics are utilized to thoroughly validate our unified framework.

Our framework shows that security can be assessed only in terms of dual distance $d_{\mathcal{D}}^{\perp}$ and parameter $B_{d_{\mathcal{D}}^{\perp}}$ of code $\mathcal{D}$:

- regarding SNR, whatever the value of σ — refer to Theorem 4.3 for the Hamming weight leakages and Corollary 4.1 for the affine leakages; moreover, for general leakages, e.g., nonlinear leakages, refer to Theorem 4.1;

- regarding MI, when σ is large and the leakage model is Hamming weight — refer to Theorem 4.4.

Furthermore, equivalent codes feature the same SNR and MI when the leakage model is Hamming weight, since permuting coordinates does not change the Hamming weight. So, we have that the MI of two equivalent codes is the same whatever the value of σ when the leakage model is Hamming weight. But the converse does not hold, as shown in Tab. 4.4.

It is interesting to notice that for $n = 2$ and $\ell = 4$, all codes $(1, \alpha_2)_{2^\ell}$ represented in $\mathbb{F}_2$ with the same weight enumerator are equivalent[6] as shown in Tab. 4.5. Note that the codes in the same rows are equivalent, so they have the same MI.

Table 4.4: Example of Non-equivalent IPM codes with $n = 3$, $\ell = 4$ that have the same weight enumerator but different MI (noiseless).

α_2	α_3	Weight Enumerators	$\mathsf{I}(Z;X)$
α^1 α^2	α^5 α^5	[(0, 1), (4, 2), (5, 3), (6, 2), (7, 4), (8, 3), (9, 1)]	0.016494
α^1 α^8	α^7 α^9	[(0, 1), (4, 2), (5, 3), (6, 2), (7, 4), (8, 3), (9, 1)]	0.016377

[6] From viewpoint of coding theory as described in Sec. 3.1.

Table 4.5: The weight enumerators of IPM codes with $m = 2$, $\ell = 4$ and MI in a noiseless case.

α_2	$d_{\mathcal{D}}^{\perp}$	$B_{d_{\mathcal{D}}^{\perp}}$	Weight Enumerators	$\mathsf{I}(Z;X)$
α^0 (Boolean masking)	2	4	[(0, 1), **(2, 4)**, (4, 6), (6, 4), (8, 1)]	1.151963
$\alpha^1,\ \alpha^{14}$	2	3	[(0, 1), **(2, 3)**, (3, 2), (4, 3), (5, 4), (6, 1), (7, 2)]	0.380288
$\alpha^2,\ \alpha^{13}$	2	2	[(0, 1), **(2, 2)**, (3, 3), (4, 3), (5, 4), (6, 2), (7, 1)]	0.287149
$\alpha^3,\ \alpha^{12}$	2	1	[(0, 1), **(2, 1)**, (3, 4), (4, 3), (5, 4), (6, 3)]	0.199569
$\alpha^4, \alpha^6, \alpha^7, \alpha^8, \alpha^9, \alpha^{11}$	3	4	[(0, 1), **(3, 4)**, (4, 5), (5, 4), (6, 2)]	0.181675
$\alpha^5,\ \alpha^{10}$	3	3	[(0, 1), **(3, 3)**, (4, 7), (5, 4), (7, 1)]	0.246318

4.7 Categorizing Linear Codes of 2-Share IPM over $\mathbb{F}_{2^8}$

The key takeaway from the mathematical analysis of the previous section is that the two metrics SNR and MI concur, in that show that $d_{\mathcal{D}}^{\perp}$ and $B_{d_{\mathcal{D}}^{\perp}}$ are the two relevant parameters to consider when seeking for an optimal code $\mathcal{D}$. Namely, theorems 4.3 and 4.4 agree in that SNR and MI decrease when $d_{\mathcal{D}}^{\perp}$ increases and when $B_{d_{\mathcal{D}}^{\perp}}$ decreases. Therefore, we deduce an algorithm to sort codes $\mathcal{D}$ of given length n with respect to their suitability in terms of IPM resistance. It is sketched in Alg. 1 (borrowed from [37]).

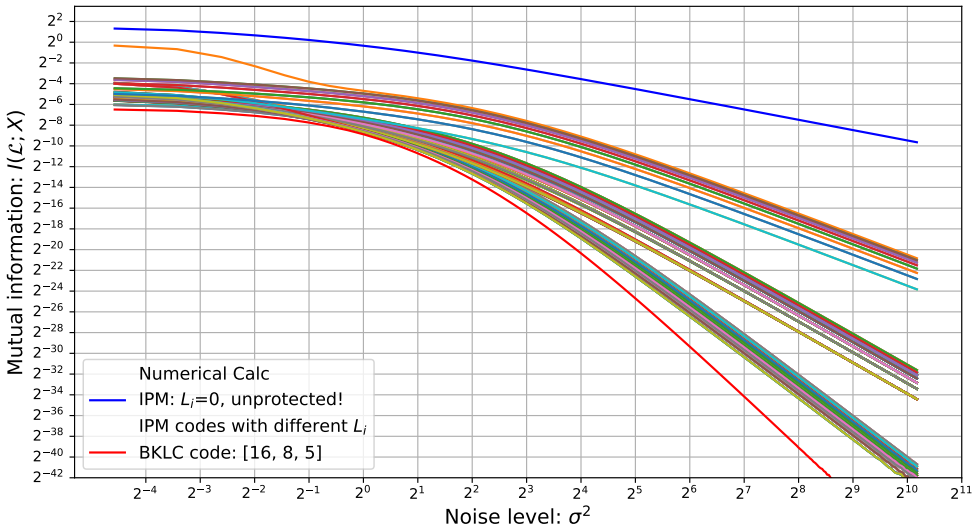


Figure 4.5: Numerical simulation of mutual information $\mathsf{I}(\mathcal{L};X)$ between leakages and the sensitive variable $X \in \mathbb{F}_{2^8}$ of all linear codes in IPM, and a *BKLC* code of parameters $[16, 8, 5]$. The blue curve is the one with $L_2 = 0$ corresponding to unprotected case.

Clearly, the dual distance $d_{\mathcal{D}}^{\perp}$ and the number of minimum weight nonzero codewords $B_{d_{\mathcal{D}}^{\perp}}$

are good indicators of the side-channel resistance of IPM. From Fig. 4.5, the linear codes in IPM can be classified into three categories with different $d_{\mathcal{D}}^{\perp}$, namely $d_{\mathcal{D}}^{\perp} = 2,\ 3$ and 4. Note that we call the linear codes used in IPM as IPM codes, where all of them are at word-level over the field $\mathbb{F}_{2^\ell}$, while all codes in DSM [17] are at bit-level over $\mathbb{F}_2$. In addition, we present the unprotected one and the *BKLC* code of parameters $[16, 8, 5]$. It is worth noting that the code with parameters $[16, 8, 5]$ is unique [11] (up to equivalence). In Fig. 4.5, when dual distance $d_{\mathcal{D}}^{\perp}$ gets larger, the slopes of mutual information curves get steeper. Therefore, the *BKLC* code has the best side-channel resistance and it is better than all IPM codes (but this code cannot be used to carry out secure and provable computations, as is the case of IPM codes). Among IPM codes, these with $d_{\mathcal{D}}^{\perp} = 4$ are better than others ($d_{\mathcal{D}}^{\perp} = 3$ or $d_{\mathcal{D}}^{\perp} = 2$).

Remark 4.3. In Fig. 4.5, all curves of codes in IPM are between the blue and the red ones over all range of noise variances σ^2. Moreover, the Boolean one (the first orange curve under the blue one) is the highest among all codes in IPM, which indicates clearly that Boolean masking is the worst case of IPM in the sense of side-channel resistance.

However, we observe that there exist distinct differences in each of classes categorized by $d_{\mathcal{D}}^{\perp}$, which are affected by $B_{d_{\mathcal{D}}^{\perp}}$. Hereafter, we investigate each group of IPM codes with the same dual distances by further studying the other code property $B_{d_{\mathcal{D}}^{\perp}}$.

4.7.1 IPM Codes with $d_{\mathcal{D}}^{\perp} = 2$

As the first investigation, we move into the IPM codes with $d_{\mathcal{D}}^{\perp} = 2$ shown in Fig. 4.6. There are fifteen codes that can be classified into eight classes, each of which have two equivalent codes with same $B_{d_{\mathcal{D}}^{\perp}}$. Recall that the IPM codes are determined by α_2, thus we search α_2 as $\alpha_2 = \alpha^i$ and we get $i \in \{0, 1, 2, 3, 4, 5, 6, 7, 248, 249, 250, 251, 252, 253, 254\}$. Correspondingly, the $B_{d_{\mathcal{D}}^{\perp}}$ are in the set $\{8, 7, 6, 5, 4, 3, 2, 1, 1, 2, 3, 4, 5, 6, 7\}$. The best two codes are equivalent and have $B_{d_{\mathcal{D}}^{\perp}} = 1$. It is worthy noting that Boolean masking is special case of IPM with $\alpha_2 = \alpha^0$, which is the highest curves in Fig. 4.6 (the worst case of IPM). Hence in the sense of side-channel resistance, IPM is more advantageous than the Boolean masking.

In summary, the (sub-)optimal codes in the class of $d_{\mathcal{D}}^{\perp} = 2$ are these with the minimized $B_{d_{\mathcal{D}}^{\perp}} = 1$ (meaning $\alpha_2 \in \{\alpha^7, \alpha^{248}\}$). This is in consistent with Theorem 4.4.

4.7.2 IPM Codes with $d_{\mathcal{D}}^{\perp} = 3$

Secondly, we investigate all linear codes with $d_{\mathcal{D}}^{\perp} = 3$. There are 146 codes (68 non-equivalent codes) in IPM.

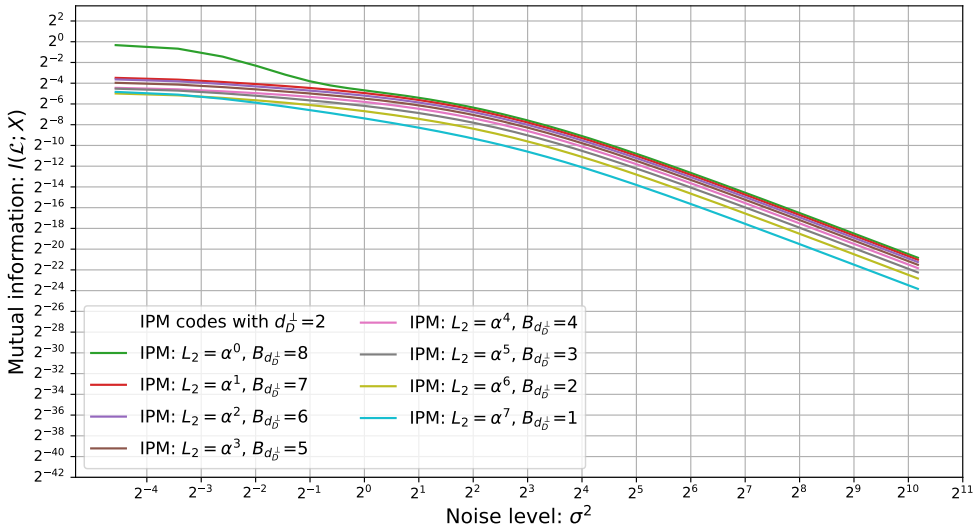


Figure 4.6: Numerical simulation of mutual information $\mathsf{I}(\mathcal{L}; X)$ between leakages and the sensitive variable $X \in \mathbb{F}_{2^8}$ in IPM where all codes have $d_{\mathcal{D}}^{\perp} = 2$ but different $B_{d_{\mathcal{D}}^{\perp}}$.

Although there are 68 non-equivalent codes, the $B_{d_{\mathcal{D}}^{\perp}}$ only takes seven values in set $\{1, 2, 3, 4, 5, 6, 7\}$. Again, we say that the codes with $B_{d_{\mathcal{D}}^{\perp}} = 1$ are (sub-)optimal in the sense of side-channel resistance. In order to show this clearly, we only choose seven codes with different $B_{d_{\mathcal{D}}^{\perp}}$ as in Fig. 4.7. For all codes with $d_{\mathcal{D}}^{\perp} = 3$, there are 36 optimal candidates of α_2 with $B_{d_{\mathcal{D}}^{\perp}} = 1$.

However, there are two IPM codes (mutual equivalent) with $\alpha_2 \in \{\alpha^{95}, \alpha^{160}\}$ which have different side-channel resistance under low noise situations ($\sigma^2 < 2^{-1}$), while with higher noise level they are in accordance with other codes. Their weight enumerator is shown as Eqn. 4.20:

$$\begin{aligned} W(\mathsf{X}, \mathsf{Y}) = \mathsf{X}^{16} + 3\mathsf{X}^{13}\mathsf{Y}^{3} + 4\mathsf{X}^{12}\mathsf{Y}^{4} + 16\mathsf{X}^{11}\mathsf{Y}^{5} + 36\mathsf{X}^{10}\mathsf{Y}^{6} + 43\mathsf{X}^{9}\mathsf{Y}^{7} + 45\mathsf{X}^{8}\mathsf{Y}^{8} \\ + 48\mathsf{X}^{7}\mathsf{Y}^{9} + 36\mathsf{X}^{6}\mathsf{Y}^{10} + 17\mathsf{X}^{5}\mathsf{Y}^{11} + 6\mathsf{X}^{4}\mathsf{Y}^{12} + \mathsf{X}\mathsf{Y}^{15}. \end{aligned} \tag{4.20}$$

The takeaway point for all codes with $d_{\mathcal{D}}^{\perp} = 3$ is that, they are preferable when the noise level is very low (e.g., $\sigma^2 < 2^{-3}$). Nevertheless, the optimal dual distance for 2-share IPM over $\mathbb{F}_{2^8}$ is equal to 4 as shown in next subsection.

4.7.3 IPM Codes with $d_{\mathcal{D}}^{\perp} = 4$

As the last part, we investigate the rest of IPM codes with $d_{\mathcal{D}}^{\perp} = 4$ where there are 94 codes (40 non-equivalent codes).

Interestingly, there are 12 candidates of $B_{d_{\mathcal{D}}^{\perp}}$ in set $\{3, 4, 5, 6, 7, 8, 9, 10, 11, 13, 14, 15\}$. In order to show the differences between IPM codes with different $B_{d_{\mathcal{D}}^{\perp}}$, we choose one code for

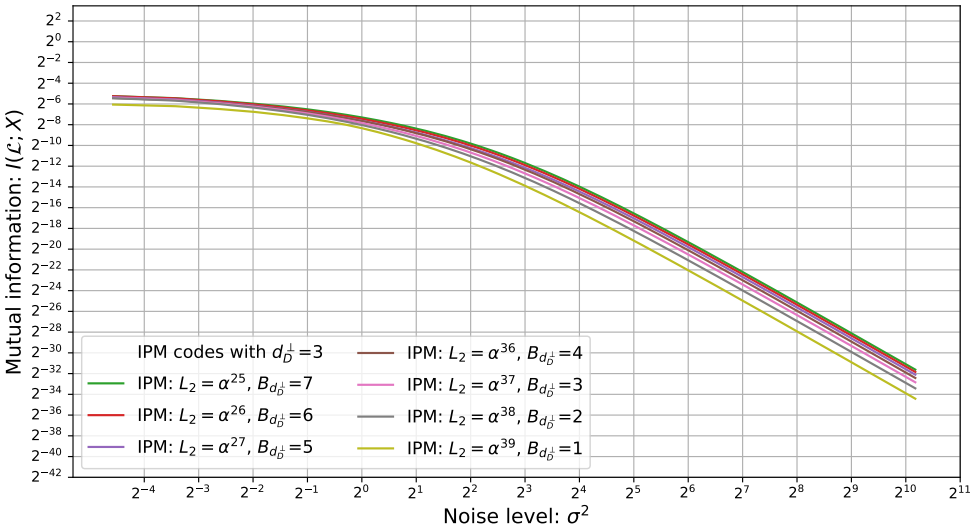


Figure 4.7: Numerical simulation of mutual information $\mathsf{I}(\mathcal{L};X)$ of IPM codes with $d_{\mathcal{D}}^{\perp}=3$ but different $B_{d_{\mathcal{D}}^{\perp}}$.

each of distinct $B_{d_{\mathcal{D}}^{\perp}}$ as shown in Fig. 4.8.

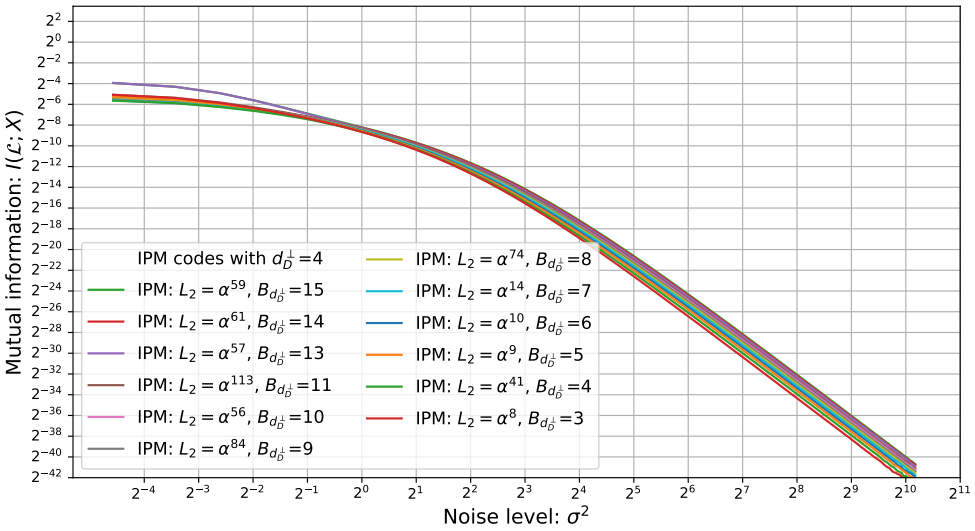


Figure 4.8: Numerical simulation of mutual information $\mathsf{I}(\mathcal{L};X)$ of IPM codes with $d_{\mathcal{D}}^{\perp}=4$ but different $B_{d_{\mathcal{D}}^{\perp}}$.

Clearly, the optimal codes for IPM is the code with minimal $B_{d_{\mathcal{D}}^{\perp}}=3$. These codes are optimal with $\sigma^2>1$ and they correspond to $\alpha_2\in\{\alpha^8,\alpha^{247},\alpha^{126},\alpha^{129},\alpha^{127},\alpha^{128}\}$. In summary, a takeaway conclusion is that the optimal codes are with $\alpha_2\in\{\alpha^8,\alpha^{126},\alpha^{128}\}$ (only three

non-equivalent codes), in which the $d_{\mathcal{D}}^{\perp}$ equals to 4 is maximized while $B_{d_{\mathcal{D}}^{\perp}}$ equals to 3 is minimized among all possible codes in the 2-share IPM.

We underline that all above codes with corresponding α_2 in IPM, dual distance $d_{\mathcal{D}}^{\perp}$, the parameter $B_{d_{\mathcal{D}}^{\perp}}$ and weight enumerators are available on `Github` [39] (all codes for 2-share and 3-share IPM over both $\mathbb{F}_{2^4}$ and $\mathbb{F}_{2^8}$ are included).

4.7.4 Estimation of MI by Theorem 4.4

In this section, we use the MI to show the impact of $B_{d_{\mathcal{D}}^{\perp}}$ in each of above three classes. In addition, we add the unprotected one, the Boolean one ($\alpha_2 = \alpha^0 = 1$) and the *BKLC* one for comparison and shown in Fig. 4.9.

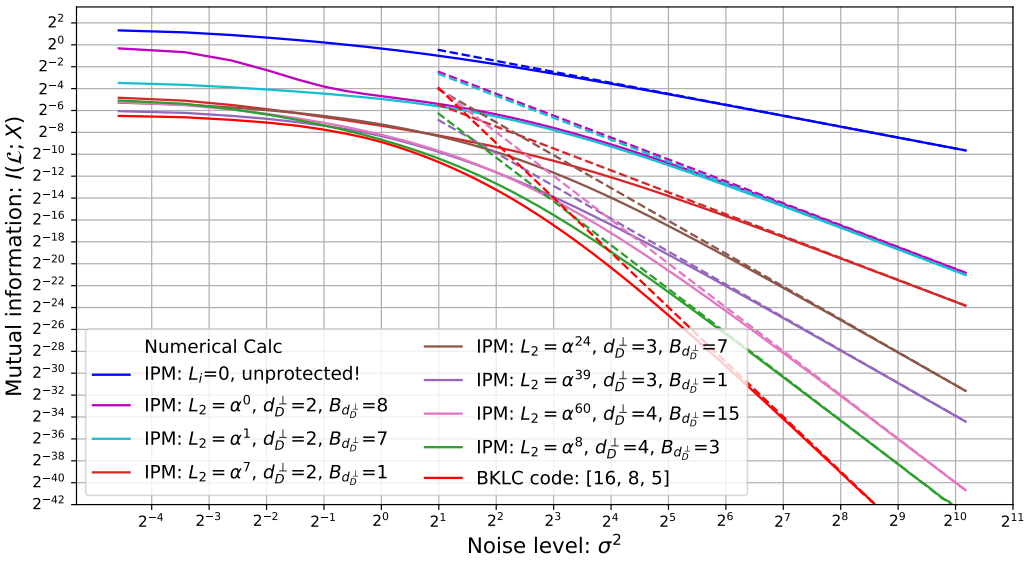


Figure 4.9: Comparing seven codes of IPM and one *BKLC* code of parameters $[16, 8, 5]$ where all codes have different $d_{\mathcal{D}}^{\perp}$ and $B_{d_{\mathcal{D}}^{\perp}}$. The solid curves are from numerical simulation, while the dotted lines are estimated by using Eqn. 4.17.

From Fig. 4.9, we can choose directly the optimal codes for IPM (the green curve). Furthermore, the *BKLC* code of parameters $[16, 8, 5]$ is better than all IPM codes and it is the best one among all linear codes (but it cannot be used in IPM since there is no IPM code of parameter $(1,\ \alpha_2)$ corresponding to the generator matrix of this *BKLC* code). This again confirms the advantages of DSM beyond IPM in 2-share setting. However, the best IPM codes in 3-share setting could be as good as the codes used in DSM.

In Fig. 4.9, the estimated MI (dotted curves) are close to numerical calculation (solid curves) when the noise level is high. Moreover, the side-channel resistance of IPM is highly depend on the linear codes used in it, which can be quantified easily by using two properties of the codes. In a nutshell, even with 2-share setting, the side-channel resistance of IPM can be significantly different. Hence, a dedicated choice of optimal codes for IPM is more than preferable, where Alg. 1 provides a good solution.

4.8 Further Applications to More General Masking Schemes

We show in this work the optimal codes for 2-share IPM over $\mathbb{F}_{2^8}$. In fact, our approach also allows analyzing all codes used in DSM, which is the generalization of IPM. Consequently, these results would be interesting for designers in practice, since the selection of the best parameters of DSM is simple but very effective.

More generally speaking, IPM is a special case of Leakage Squeezing (LS) [23] and Direct Sum Masking (DSM) [17, 24]. The connections between these masking schemes are shown in Fig. 4.10. Moreover, the efficient and secure algorithms for performing the elementary operations like addition and multiplication on shared data are proposed in [164] for DSM. In particular, in the case of IPM, more efficient computations have been proposed in [3], in which the multiplication part can be simplified. From a performance perspective, the overhead of IPM is about 40% and 60% more than the Boolean masking when deployed on AES-128, for 2-share and 3-share implementations, respectively [3, Tab. 2].

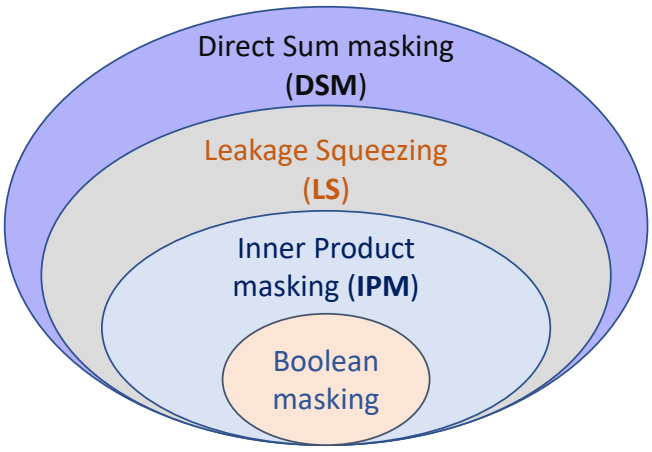


Figure 4.10: Connections between IPM, LS and DSM from a generalization perspective.

Furthermore, the optimality of the selected codes not only holds under Hamming weight model, but also holds under the "unevenly weighted sum of the bits" (UWSB) model, in which each bit in the sensitive variables leaks differently. Indeed, in both cases the leakage function P have the degree equal to 1. Eventually, Theorem 4.3 is more general and can be used to assess the codes under the nonlinear or higher order moments leakages where $d^{\circ}P > 1$, e.g., $P(Z) = w_H(Z)^d$ has $d^{\circ}P = d$.

The Impact of Different Irreducible Polynomials. Although all representations of $\mathbb{F}_{2^8}$ are isomorphic and therefore equivalent, they do not preserve their properties after sub-field extension. In particular, there are 30 irreducible polynomials over $\mathbb{F}_{2^8}$ and two typical cases of them are:

- $g_1(\alpha) = \alpha^8 + \alpha^4 + \alpha^3 + \alpha + 1$: which is the standard irreducible polynomial in AES[7],
- $g_2(\alpha) = \alpha^8 + \alpha^4 + \alpha^3 + \alpha^2 + 1$: which is the default irreducible polynomial in `Magma` and `Matlab`, also the one used in this chapter.

In IPM, the irreducible polynomial plays an important role in expanding codes from $\mathbb{F}_{2^8}$ to $\mathbb{F}_2$, which determines the linear codes over $\mathbb{F}_2$. As a result, the optimal choices of the linear code for IPM may vary for different irreducible polynomials. Taking above two polynomials as an example, the best achievable values of $B_{d_{\mathcal{D}}^{\perp}}$ with $d_{\mathcal{D}}^{\perp} = 4$ for 2-share IPM are different, where the best $B_{d_{\mathcal{D}}^{\perp}}$ for $g_1(\alpha)$ and $g_2(\alpha)$ are 4 and 3, respectively. Moreover, an information-theoretic comparison on side-channel resistance of the corresponding IPM is shown in Fig. 4.11, which shows a slight advantage of using $g_2(\alpha)$.

However, the different irreducible polynomials have marginal impact on the best linear codes with respect to $d_{\mathcal{D}}^{\perp}$ and $B_{d_{\mathcal{D}}^{\perp}}$. For instance, the possible values for $d_{\mathcal{D}}^{\perp}$ are the same for both $g_1(\alpha)$ and $g_2(\alpha)$, and the difference on the best values of $B_{d_{\mathcal{D}}^{\perp}}$ is only one when other $g_1(\alpha)$ or $g_2(\alpha)$ is deployed.

4.9 Conclusions

In this part, we followed a quantitative approach to characterize the side-channel resistance of IPM scheme. In particular, we proposed a unified framework and linked it to two theoretical metrics (*SNR* and mutual information), and also an attack metric (success rate). The framework

[7] As an example, this irreducible polynomial is used to construct the optimal codes in [33].

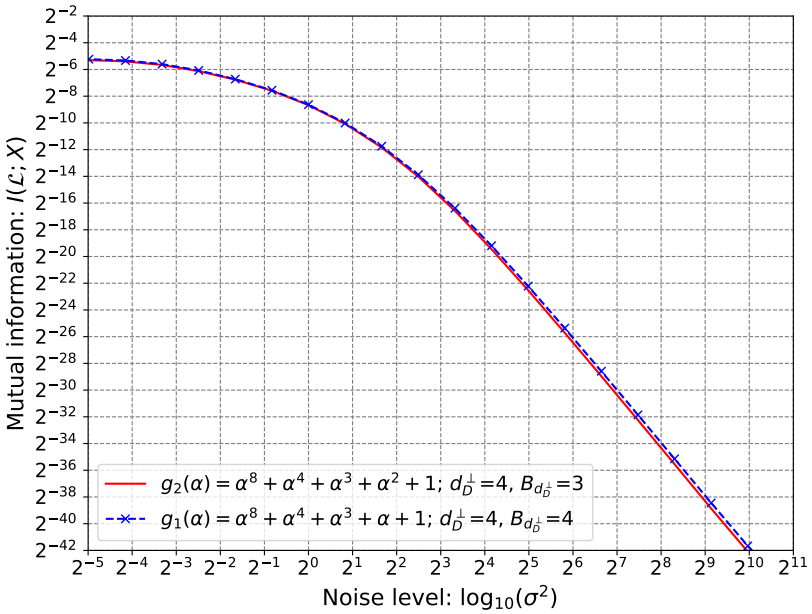


Figure 4.11: Comparison of the impact of two irreducible polynomials (of the finite field) on the best linear codes for IPM.

is based on two parameters of the code $\mathcal{D}$, namely the dual distance $d_{\mathcal{D}}^{\perp}$ and the coefficient $B_{d_{\mathcal{D}}^{\perp}}$ in its weight enumeration polynomial. We showed that the concrete security level of IPM can be fully depicted by our framework. By our framework, we provided a quantitative explanation for "Security Order Amplification", which has been observed in previous works including CARDIS'16, CARDIS'17 and ASIACRYPT'17. At last, we proposed an effective method to select the optimal codes for IPM and validated by experiments.

Although we validated our framework by simulated leakages with realistic noise parameters, it is still not clearly verified on real devices. As a perspective, we will consider the practical validations of our findings. Moreover, we show in Tab. 4.3 and 4.5 *optimal* codes obtained by an *exhaustive* study, which is very time-consuming. Such method to find the optimal codes becomes computationally impossible when the number of shares n gets larger (e.g., $n > 5$). Hence, a systematic (e.g., algebraic) construction of better codes than mere random codes is much preferable and could be leveraged. However, it is still an open problem to construct optimal or suboptimal codes for IPM or LS with a larger number of shares.

Part III

Information Leakage in Code-based Masking: Formalization, Quantification and Applications

CHAPTER 5

Quantifying Leakage in Code-based Masking

This chapter presents the work [35] published at *IACR Trans. Cryptogr. Hardw. Embed. Syst.* (TCHES) 2021, issue 3. Part of results are also been demonstrated in [42].

Contents

5.1 Introduction

Masking is one of the most well-studied countermeasures to protect cryptographic implementations against side-channel attacks due to the favorable provable security it provides. The core idea underlying any masking scheme is to split the sensitive (key-dependent) variables into several shares and perform independent computations on masked variables only. Indeed, the rationale is that, given a sufficient amount of noise, the attack complexity increases exponentially with the number of shares [30, 128], while the implementation cost increases only quadratically (or only cubically in higher-order glitch-free implementations [79]).

Two key ingredients of a masking scheme are the encoding for randomizing the sensitive variables, and the masked operations for manipulating the random shares. Regarding the latter, the secure masked operations can be constructed effectively [86, 139] for both bit- and word-oriented variables. Furthermore, thanks to the well-established concept of (Strong) Non-Inference (NI and SNI) introduced by Barthe et al. [6], the basic gadgets carrying out the elementary operations (e.g., addition, multiplication, etc.) can be composed to construct the whole implementation without losing the claimed security properties. Regarding the former, the encoding is a more fundamental ingredient in masking that provides the achievable upper bounds of side-channel security order with tunable public parameters. Indeed, firstly, the side-channel security order of the full implementation cannot exceed the security order of the corresponding encoding, and secondly, when implemented ideally, the security order of an implementation can be guaranteed by its encoding. However, evaluating the concrete side-channel resistance of the encoding in general cases remains an open problem since many different encodings in various masking schemes behave differently when fed with diverse parameters. Therefore, a unified quantification approach would formalize and compare the security of different encodings and find optimal parameters for a specific masking scheme.

5.1.1 Unifying Masking Schemes by Generalization

Generalization is a promising approach to unify different masking schemes. In this trend, the code-based masking generalizes many existing schemes, including Boolean masking, Inner Product masking (IPM)[1] [2, 3], Leakage Squeezing (LS) [23, 24] and Direct Sum masking (DSM) [17, 123]. To the best of our knowledge, the generalized code-based masking (GCM) [164]

[1]We consider the improved IPM [2] rather than the original one [4], since firstly, there exist some first-order information leakages in the latter [130], and secondly the performance of the latter is much lower than the former, which makes it impractical.

is the most generic scheme in this respect. In particular, polynomial masking [77, 131] is also a special case of GCM, which is built upon Shamir's secret sharing (SSS) scheme [145].

Let $X \in \mathbb{F}_{2^\ell}^k$ and $Y \in \mathbb{F}_{2^\ell}^t$ be respectively the sensitive variable and t random masks. Then the encoded variable in GCM writes:

$$Z = X\mathbf{G} + Y\mathbf{H} \in \mathbb{F}_{2^\ell}^n,$$

given that $k + t \leq n$, where $\mathbf{G}$ and $\mathbf{H}$ are generator matrices of two codes $\mathcal{C}$ and $\mathcal{D}$, respectively. For the sake of simplicity, we take $k = 1$, but essentially, the GCM can use packed secret sharing techniques [79, 164] to improve the performance by parallelism. However, the side-channel security evaluation of encoding is similar to any k, since each of the k sensitive variables is encoded similarly. The overview of connections between these masking schemes is shown in Fig. 5.1, where the four intersecting areas are:

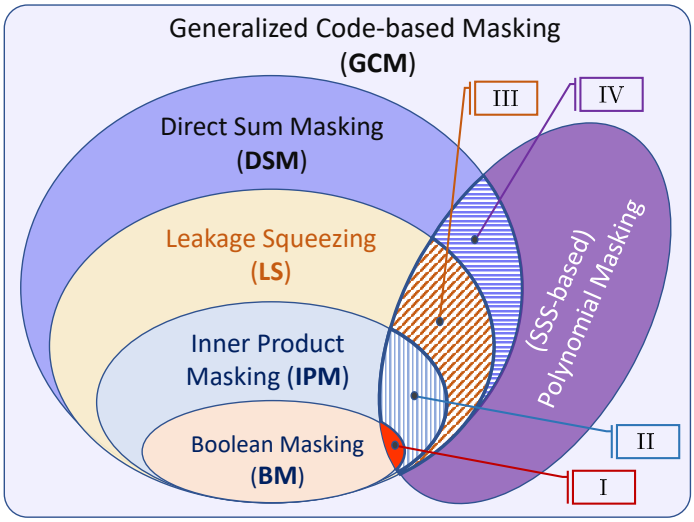


Figure 5.1: Overview of code-based masking schemes. In particular, all intersections I, II, III, and IV mean that $n = t + 1$ in SSS-based masking, where the two codes $\mathcal{C}$ and $\mathcal{D}$ are complementary.

- Intersection I: as pointed out in [51], Boolean masking can be considered as a special case of polynomial masking for small enough parameters ($n \leq 6$ or equivalently $t \leq 5$).

- Intersection II: in [2], the authors claimed that the polynomial masking is a special case of IPM. However, this generalization does not indicate the exact connections between SSS-scheme and RS codes. Indeed, if we take the polynomial evaluations in encoding into consideration, the generalization from SSS-based masking to IPM is valid only when $n = 2$ and $t = 1$.

- Intersections III and IV: in SSS-based masking, if $n = t + 1$, the codes $\mathcal{C}$ and $\mathcal{D}$ are complementary, therefore they can be viewed as DSM (or LS) scheme. Otherwise, if $n > t + 1$, the corresponding masking schemes are out of DSM's scope. On the other side, the linear codes for DSM may not be converted into SSS-based schemes since the codes in SSS are endowed with a specific algebraic structure.

The most significant benefit of utilizing code-based masking is the higher security order than the simple Boolean masking given the same number of shares. Taking 2-share IPM over $\mathbb{F}_{2^8}$ [3, 37] as an instance, when appropriate public parameters are chosen, the side-channel security order can be maximized to 3 under the bit-probing model [123], which is higher than 1 in Boolean masking. Moreover, the security orders are enlarged to 7 vs. 2 (IPM vs. Boolean one) in 3-share scenarios [37, Tab. 2].

Currently, the side-channel security order of GCM has been connected to the dual distance of $\mathcal{D}$ [24, 123], which is denoted as $d_{\mathcal{D}}^{\perp}$. As a special case, the security order t in IPM and DSM is equal to $d_{\mathcal{D}}^{\perp} - 1$ since the two codes $\mathcal{C}$ and $\mathcal{D}$ are complementary. However, as pointed out in [37], the dual distance of $\mathcal{D}$ is not sufficient to characterize the concrete side-channel resistance of IPM, hence a new framework with a new parameter (more precisely $B_{d_{\mathcal{D}}^{\perp}}$, which counts the number of codewords of Hamming weight equal to $d_{\mathcal{D}}^{\perp}$ in $\mathcal{D}^{\perp}$) is proposed to model IPM's concrete security level more accurately. Nevertheless, this framework is not applicable to GCM since $\mathcal{C}$ and $\mathcal{D}$ may not be complementary anymore.

5.1.2 Public Points in SSS and Polynomial Masking

To construct a t-th order secure polynomial masking, a polynomial of degree t is firstly selected: $f_X(\mathsf{X}) = X + \sum_{i=1}^{t} u_i \mathsf{X}^i$, where the secret X is then associated as the constant term in $f_X(\mathsf{X})$. Secondly, $f_X(\mathsf{X})$ is evaluated in n distinct points α_i for $1 \leq i \leq n$, which are called "public points" in the scheme. As a result, the secret X is encoded by using the private parameters u_i (which are random masks viewed in the context of masking).

As observed in [29], the public points in SSS play a significant role in the side-channel resistance of SSS-based masking schemes. In fact, this problem of public points is inherent in the SSS scheme and can be dated back to Massey [105] who claimed that SSS scheme "*can be attacked with the well-developed tools of algebraic coding theory*". The SSS-based masking provides a practical example whereby changing the public points in polynomial masking, the concrete security level can be significantly different.

However, to the best of our knowledge, there are neither qualitative principles for selecting good or even optimal public points in SSS-based masking nor a quantitative approach to evaluate the role of public points played in the side-channel resistance of SSS-based masking. In this chapter, we propose solutions to the two problems by utilizing a coding-theoretic quantitative approach.

5.1.3 Independence Assumption behind Masking Schemes

The independence assumption is an indispensable condition behind the security proofs when extending from the probing model to the bounded moment model or noisy leakage models [8, 60]. For instance, if this independence condition is violated due to physical defaults (e.g., couplings through the ground or parasitic capacitances, glitches, etc.), the side-channel security order will decrease accordingly [61]. However, this independence condition is essentially related to inter-share leakages from different shares in masking and treats each share as a whole.

Moreover, the independence issue also happens in intra-share cases where the leakages of different bits in the same share leak jointly. This kind of leakage is often called non-linear leakages and comes, e.g., from registers or memory units of real devices. In fact, both intra-share and inter-share independence issues can happen simultaneously. Taking AES implemented on ARM Cortex-M4 as an example, where the registers are 32-bit, and each share is in $\mathbb{F}_{2^8}$, four shares can be manipulated at the same time. Consequently, the register will leak jointly, including intra-share and inter-share leakages. To the best of our knowledge, the intra-share independence issue has not yet been studied thoroughly in the sense of security order reduction. We will show that essentially, the intra-share independence is the condition for higher security orders under the bounded moment model [8].

5.2 Our Contributions

In view of the above state-of-the-art, our contributions are threefold as follows.

A Unified Leakage Quantification Approach for GCM. We derive a closed-form expression for SNR to quantify the information leakages in GCM for any leakage functions. In particular, we present a simplified expression for the Hamming weight leakage model. In fact, this new result generalizes the framework proposed in [37] for IPM. Furthermore, we use mutual information (MI) to quantify the information leakages of GCM in an information-theoretic sense. Both SNR and MI are connected to two properties (namely the dual distance and the number

of conditioned codewords) of the linear codes used in GCM. Relying on a theoretical analysis of SNR and MI, we propose a unified approach to quantify information leakage in GCM. Then we show how to select optimal codes for GCM by optimizing the two properties. The experimental results confirm that the MI can be minimized by utilizing optimal codes, which indicates the improved concrete security level of the corresponding masking scheme.

Optimal Public Points for SSS-based Polynomial Masking. As an application of our unified approach, we characterize the side-channel resistance of polynomial masking from a coding-theoretic point of view. The first outcome is a more accurate characterization of information leakage and the second outcome is a straightforward method to choose optimal linear codes (parameters) for SSS-based masking. For the first time, we quantify the impact of combining different public points in SSS-based masking in the context of side-channel analysis and show that more shares leak more information (given a specific t). In particular, our coding-theoretic approach can exactly depict the observations made in [29]. Using MI, we present the quantitative results of information leakages in SSS-based masking, which again validate our unified approach. For the first time, we exhibit several optimal tuples of public points (the linear codes in a coding-theoretic perspective) for SSS-based masking in the sense of side-channel resistance.

Revisited Independence Condition in Masking Schemes. Independence condition requires that the information leakages from different variables are statistically independent. In the context of masking, it exists in two cases: inter-share and intra-share. Specifically, the former means that the leakages of different shares are independent, which is well-studied in literature [8]. The latter deals with the leakages from one share, in which different bits in this share may leak independently or not. To capture both of them, we introduce the leakage function P, where its numerical degree indicates both cases' independence conditions. For instance, the commonly assumed Hamming weight leakage model has a numerical degree equal to 1, a perfect independent case. Moreover, we show how the degree of P affects the side-channel security order of a masking scheme.

We underline that all mathematical derivations presented in this chapter have been verified formally with `Magma` computational algebra system [158]. The open sources of this work are available on `Github` [34].

Differences between Chap. 4 or [37] with GCM. In this work, we study GCM by using a similar coding-theoretic approach as in [37]. However, two key differences make this work significantly different from [37].

Firstly, GCM generalizes IPM by allowing $\mathcal{C}$, and $\mathcal{D}$ to be non-complementary, which allows deriving security metrics in a more general manner. In [37], the authors prove that the side-channel security of IPM only depends on the code $\mathcal{D}$. While in this work, for the first time, we show that the side-channel security depends on both $\mathcal{C}$ and $\mathcal{D}$. In particular, the quantitative findings enable us to put forward optimal GCM encodings which are new upon [51]: given the same parameters n and t (the number of shares and security order), we decrease the information leakage in GCM to the lesser possible extent.

Secondly, GCM allows for protections in much more general contexts. Namely, GCM can be used to withstand glitches [131] and to detect errors against fault injection attacks on top of preventing side-channel attacks. Therefore, our work has broader implications for the protection of realistic platforms. In a nutshell, GCM opens a new path to derive unified countermeasures against both fault injection and side-channel attacks.

5.3 Encodings in Code-based Masking

5.3.1 Technical Overview

Let n, k be positive integers and $\mathbb{K} = \mathbb{F}_{2^\ell}$ be a finite field. Let $\mathcal{C}$ be an $[n,k]_q$ linear code parameter with generator matrix $\mathbf{G}$ defined over $\mathbb{F}_q$ (here we use $q = 2^\ell$). Let the irreducible polynomial be $g(\alpha) = \alpha^8 + \alpha^4 + \alpha^3 + \alpha^2 + 1$ to generate the field $\mathbb{K} = \mathbb{F}_{2^8}$. Recall that for an (n, t)-SSS scheme, the secret X is split into n shares, and the sharing is t-privacy, where any $t+1$ shares can be used to recover the secret but not for less than t shares. Note that the (n, t)-SSS scheme is also connected to the Reed-Solomon (RS) code with parameters $[n, t+1]$.

Let $X \in \mathbb{K}^k$, $Y \in \mathbb{K}^t$ and $Z \in \mathbb{K}^n$ be the sensitive variable, the random masks, and the shared variable; we use Eqn. 5.1 as the uniform representation of encoding in GCM which is used throughout the chapter:

$$Z = X\mathbf{G} + Y\mathbf{H} \in \mathbb{F}_{2^\ell}^n, \tag{5.1}$$

where $k+t \leq n$, $\mathbf{G}$ and $\mathbf{H}$ are two generator matrices of the two codes $\mathcal{C}$ and $\mathcal{D}$ with $\mathcal{C} \cap \mathcal{D} = \{0\}$.

In this work, we focus on GCM, which is the most general case of code-based maskings[1]. By using the uniform representation as Eqn. 5.1, we revisit the encodings of code-based masking

[1] As a special case of IPM, a Boolean masking can be obtained by taking $\alpha_i = 1$ for $1 \leq i \leq t$ in Tab. 5.1.

schemes as in Tab. 5.1.

Table 5.1: Encodings in IPM, LS, DSM, SSS-based masking and GCM, revisited.

	IPM [2, 3]	LS [1] [23]	DSM [17, 123]	SSS-based masking [77, 131]	**GCM** [164]
Conditions on $\mathcal{C}$ and $\mathcal{D}$	$\mathcal{C} \cap \mathcal{D} = \{0\}$, $\mathcal{C} + \mathcal{D} = \mathbb{K}^n$	$\mathcal{C} \cap \mathcal{D} = \{0\}$, $\mathcal{C} + \mathcal{D} = \mathbb{K}^n$	$\mathcal{C} \cap \mathcal{D} = \{0\}$, $\mathcal{C} + \mathcal{D} = \mathbb{K}^n$	$\mathcal{C} \cap \mathcal{D} = \{0\}$	$\mathcal{C} \cap \mathcal{D} = \{0\}$
$\mathbf{G} \in \mathbb{K}^{k \times n}$	$\begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \end{pmatrix}$	$\mathbf{G} \in \mathbb{K}^{k \times n}$	$\mathbf{G} \in \mathbb{K}^{k \times n}$	$\begin{pmatrix} 1 & 1 & \cdots & 1 \end{pmatrix}$	$\mathbf{G} \in \mathbb{K}^{k \times n}$
$\mathbf{H} \in \mathbb{K}^{t \times n}$	$\begin{pmatrix} \alpha_1 & 1 & 0 & \cdots & 0 \\ \alpha_2 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \alpha_t & 0 & 0 & \cdots & 1 \end{pmatrix}$	$\mathbf{H} \in \mathbb{K}^{t \times n}$	$\mathbf{H} \in \mathbb{K}^{t \times n}$	$\begin{pmatrix} \alpha_1^1 & \alpha_2^1 & \cdots & \alpha_n^1 \\ \alpha_1^2 & \alpha_2^2 & \cdots & \alpha_n^2 \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_1^t & \alpha_2^t & \cdots & \alpha_n^t \end{pmatrix}$	$\mathbf{H} \in \mathbb{K}^{t \times n}$
Security parameters: n, k, t	$k = 1$, $n = t + 1$	$n = k + t$. **G**, **H** can be any matrices	$n = k + t$. **G**, **H** can be any matrices	$n \geq k + t$ and $f_X(\mathsf{X})$. In glitch-free case, $n \geq 2t + 1$ [131]	$n \geq k + t$. **G**, **H** can be any matrices

5.3.2 Connecting SSS Scheme to the RS code

We recall the (n, t)-SSS scheme by mainly referring to [28, 29]. Let $X \in \mathbb{K}$ again be the secret and can be split into n shares such that no tuple of shares with cardinality lower than t depends on X. The SSS scheme consists in selecting a random polynomial $f_X(\mathsf{X}) \doteq X + \sum_{i=1}^{t} u_i \mathsf{X}^i$ of degree t where u_i with $1 \leq i \leq t$ are t random coefficients (masks) in $\mathbb{K}$. The secret X is the constant term: $X = f_X(0)$. Then a (n, t)-sharing $(Z_1, Z_2, \ldots, Z_n)$ of X is defined by evaluating the polynomial $f_X(\mathsf{X})$ in n distinct public non-zero points $\alpha_1, \alpha_2, \ldots, \alpha_n$ in $\mathbb{K}$ such that $Z_i = f_X(\alpha_i)$. The recovery of X from its sharing consists in two steps: $f_X(\mathsf{X})$ is first recovered by using the Lagrange interpolation and second, $f_X(\mathsf{X})$ is evaluated in 0. Since in an (n, t)-SSS, any tuple of shares with cardinality greater than t can be used to recover X, we denote by U the selected shares ($|U| \geq t + 1$), which is called the interpolation set.

Actually, these two steps can be combined into one [29]:

$$X = \sum_{Z_i \in U} Z_i \cdot \gamma_i \ , \tag{5.2}$$

where the public constants γ_i are computed from α_i by: $\gamma_i = \prod_{j=1 \, s.t. \, j \neq i, \, Z_j \in U}^{n} \frac{\alpha_j}{\alpha_j - \alpha_i}$.

[1] LS consists of the application of an arbitrary bijection on the shares. Although it has only been studied on vectors of bits (on $\mathbb{F}_2$), it can be trivially extended to vectors on $\mathbb{F}_{2^\ell}$. When the bijections are linear, LS is thus equivalent to DSM.

Remark 5.1. Note that in $\mathbb{K} = \mathbb{F}_{2^\ell}$, the subtraction is the same operation as the addition.

Remark 5.2. In an (n, t)-SSS scheme, any combination of more than t shares, meaning $|U| \geq t+1$, can be used to recover the polynomial $f_X(\mathsf{X})$ and the secret X. Hence, in each combination (e.g., each set U), γ_i should be computed correspondingly.

Next, we recall the Reed-Solomon codes.

Definition 5.1 (Reed-Solomon Code [29])**.** The Reed-Solomon code $\mathrm{RS}(\mathcal{S}, t+1) \subset \mathbb{K}^n$ of dimension $t+1$ over a finite field $\mathbb{K}$ and with evaluation subset $\mathcal{S} = \{\alpha_0, \alpha_1, \ldots, \alpha_n\}$ of $\mathbb{K}$ is the subspace:

$$RS(\mathcal{S}, t+1) = \{(f(\alpha_0), f(\alpha_1), \ldots, f(\alpha_n)); f(\mathsf{X}) \in \mathbb{K}[\mathsf{X}] \text{ and } \deg(f) \leq t\} \ .$$

Given the degree of $f(\mathsf{X})$ is t, then $t+1$ evaluations of it can be used to recover $f(\mathsf{X})$ itself and the codewords. In terms of RS codes, the sharing of X with SSS scheme is an encoding with a RS code RS $(\{\alpha_1, \ldots, \alpha_n\}, t+1)$:

$$Z = (Z_1, Z_2, \ldots, Z_n) = (X,\, Y) \begin{pmatrix} \mathbf{G} \\ \mathbf{H} \end{pmatrix} = X\mathbf{G} + Y\mathbf{H}, \tag{5.3}$$

where $\begin{pmatrix} \mathbf{G} \\ \mathbf{H} \end{pmatrix}$ is the generator matrix $(\alpha_i^j)_{i\in[1;\,n],\, j\in[0;\,t]}$. More precisely, $\mathbf{G}$ is an 1-by-n matrix equal to $(1, 1, \ldots, 1)$ and $\mathbf{H}$ is a Vandermonde matrix. By denoting $\mathbf{G}_i$ and $\mathbf{H}_i$ the i-th column of $\mathbf{G}$ and $\mathbf{H}$ respectively, we have: $Z_i = f_X(\alpha_i) = X + \sum_{j=1}^{t} Y_j \alpha_i^j = X\mathbf{G}_i + (Y_1, \ldots, Y_t)\ \mathbf{H}_i$.

Accordingly, the reconstruction of X from $Z = (Z_1, Z_2, \ldots, Z_n)$ is done by taking Z_i to obtain an interpolation set U such that $|U| \geq t+1$. We also call this scheme the redundant sharing when $n > t+1$ since at least $t+1$ shares can recover X. We will show in Sec. 6.2 that more redundancies in sharing of SSS-based masking leak more information on X.

5.4 Quantifying Information Leakages in GCM

In this section, we use SNR as a leakage metric to evaluate the information leakages in GCM. In particular, SNR quantifies the key-dependent leakage at certain degrees. SNR is thus attractive in that if SNR at a given degree d is null, then one can conclude that the scheme is secure at order d.

5.4.1 Uniform Representation of Leakage Function

As the first step, we formalize the information leakages from a device. In this respect, we rely on the clarification on serial and parallel implementations proposed in [8].

Before formalization, we give an example to provide some intuition for the uniform leakage function P. Let $Z = (Z_1, Z_2, \ldots, Z_n)$ denote the encoded intermediate with n shares and X be the secret. By ignoring the noise, we assume the leakage of each share is $\mathcal{L}_i = Z_i$ under the identity leakage model and $\mathcal{L} = \sum_i \mathcal{L}_i$ is the total leakage. To launch a successful attack, an adversary needs to find the (smallest) key-depend statics, namely raising d such that $\mathbb{E}\left[\mathcal{L}^d|X\right] \neq \mathbb{E}\left[\mathcal{L}^d\right]$, but $\mathbb{E}\left[\mathcal{L}^i|X\right] = \mathbb{E}\left[\mathcal{L}^i\right]$ for all $i < d$. Equivalently, an adversary needs the smallest d such that $\mathbb{V}\left[\mathbb{E}\left[\mathcal{L}^d|X\right]\right] \neq 0$, which measures the informative part in $\mathcal{L}$.

Formally, let $P = \varphi_P \circ \phi_P$ denote the leakage function, where ϕ_P is the leakage model for each share, and φ_P is the combination function that assembles the leakages from selected shares. In this thesis, we call ϕ_P and φ_P the intra-share and inter-share leakage model, respectively. For instance, in serial implementations, the leakage of each share is: $\mathcal{L}_i = \phi_P(Z_i) + N_i$, then the exploitable leakages can be combined by φ_P. For instance, taking the Hamming weight model and centered product as leakage model and combination function, respectively, then $\mathcal{L}_i = \phi_P(Z_i) + N_i = w_H(Z_i) + N_i$ and $\mathcal{L} = \prod_{c=1}^{d}(\mathcal{L}_c - \mathbb{E}\left[\mathcal{L}_c\right]) = P(Z) + N_{total}$ where the latter combines leakages of d shares by the normalized product. Consequently, the highest order of key-dependent leakages is captured by P with numerical degree d.

Therefore, we use the following representation of P as a pseudo-Boolean function:

$$P(Z) = \sum_{I \in \{0,1\}^{n\ell}} \beta_I Z^I, \tag{5.4}$$

where $Z^I = \prod_{i \in \{1,\ldots,n\ell\} \text{ s.t. } I_i = 1} Z_i$, and $\beta_I \in \mathbb{R}$ and $\deg(P) = \max\{w_H(I) \,|\, \beta_I \neq 0\}$.

Two Probing Models. For the purpose of a finer-grain analysis, we clarify the two kinds of probing model (see also [54, §2.2]) and corresponding security orders as follows:

- **Bit-probing model**: each probe only gets one bit at a time where each bit leaks independently or jointly. Correspondingly, ϕ_P is defined at bit-level and φ_P at certain degrees are used to combine the bit-level leakages. The security order in the bit-probing model is denoted by t_b.
- **Word-probing model**: each probe gets an ℓ-bit word at a time, where an ℓ-bit variable leaks as a whole. As a result, the degree of ϕ_P implies how many numbers of bits leaked jointly, in which the intra-share independence condition plays a role in security order reduction, as shown above. Similarly, the security order is then denoted by t_w.

When connected to coding-theoretic properties, the security orders t_b and t_w are related to the dual distance of the code $\mathcal{D}$ used in GCM over $\mathbb{F}_2$ and $\mathbb{F}_{2^\ell}$, respectively [37, 123]. More

precisely[1], we have $t_w = d_{\mathcal{D}}^{\perp} - 1$ and $t_b = d_{\mathcal{D}_2}^{\perp} - 1$ where $\mathcal{D}_2$ is the sub-field representation of $\mathcal{D}$. In the sequel, we call t the security order for the sake of simplicity, t_b and t_w should be unambiguous from the context (e.g., variables in $\mathbb{F}_2$ or $\mathbb{F}_{2^\ell}$).

5.4.2 SNR-based Information Leakage Quantification

Let $P(Z)$ be a leakage function as in Eqn 5.4 and let N denote the independent noise with zero mean and variance $\mathbb{V}[N] = \sigma_{total}^2 \propto \sigma^{2d}$ ($\propto$ means proportional to σ^{2d}) [37]. Then, the leakage is:

$$\mathcal{L} = P(Z) + N.$$

We have $\mathbb{V}[\mathbb{E}[P(Z) + N|X]] = \mathbb{V}[\mathbb{E}[P(Z)|X]]$, where $Z = X\mathbf{G} + Y\mathbf{H} \in \mathbb{K}^n = \mathbb{F}_{2^\ell}^n$ is the encoding in GCM (Equ. 5.1). The SNR of leakages is defined as:

$$\text{SNR} = \frac{\mathbb{V}[\mathbb{E}[\mathcal{L}|X]]}{\mathbb{V}[N]} = \frac{\mathbb{V}[\mathbb{E}[P(Z)|X]]}{\sigma_{total}^2}. \tag{5.5}$$

Therefore, we propose the following theorem to quantify the leakages in the GCM scheme by SNR.

Theorem 5.1. *Let a device be protected by the GCM scheme as $Z = X\mathbf{G} + Y\mathbf{H}$. Assume the leakages of the device can be represented in the form: $\mathcal{L} = P(Z) + N$. Then the SNR of the exploitable leakages is:*

$$SNR = \frac{\mathbb{V}[\mathbb{E}[P(Z)|X]]}{\sigma_{total}^2} = \frac{1}{2^{2n\ell} \cdot \sigma_{total}^2} \left(\sum_{x,\, y \in \mathcal{D}^{\perp} \setminus \mathcal{C}^{\perp};\ x+y \in \mathcal{C}^{\perp}} \widehat{P}(x)\widehat{P}(y) \right), \tag{5.6}$$

where $\sigma_{total}^2 \propto \sigma^{2d}$ is the total noise and $\widehat{P}(\cdot)$ is the Fourier transform *of $P(\cdot)$*

The demonstration of Theorem 5.1 involves computing $\mathbb{V}[\mathbb{E}[P(Z)|X]]$, which can be derived by the following Lemma 5.1. In order to have the paper read fluently, its proof is relegated in Appendix A.1.1 which also proves Theorem 5.1.

Lemma 5.1. *Let a pseudo-Boolean function $P(Z)$ denote the leakage function, and taking the same notations as above, we have*

$$\mathbb{V}[\mathbb{E}[P(Z)|X]] = \frac{1}{2^{2n\ell}} \sum_{x,\, y \in \mathcal{D}^{\perp} \setminus \mathcal{C}^{\perp};\, x+y \in \mathcal{C}^{\perp}} \widehat{P}(x)\widehat{P}(y). \tag{5.7}$$

[1] In [164], a special case is presented with $t > d_{\mathcal{D}}^{\perp} - 1$. However, we always have $t = d_{\mathcal{D}}^{\perp} - 1$ if the optimal codes are used in GCM. Especially, the equality holds for all RS codes in SSS-based masking.

Remark 5.3. Note that Lemma 5.1 encompasses the core result in [37]. Indeed, as a special case, if $n = t + 1$ in SSS-based masking, the two codes $\mathcal{C}$ and $\mathcal{D}$ are complementary, as well as $\mathcal{C}^\perp$ and $\mathcal{D}^\perp$. Since by Lemma 3.4, we have $\mathcal{C}^\perp \cap \mathcal{D}^\perp = \{0\}$ and the only possible solution in Eqn. 5.7 is $x = y \neq 0$. Therefore, $\mathbb{V}[\mathbb{E}[P(Z)|X]]$ can be simplified into:

$$\mathbb{V}[\mathbb{E}[P(Z)|X]] = \frac{1}{2^{2n\ell}} \sum_{x \in \mathcal{D}^\perp \setminus \{0\}} \widehat{P}(x)^2, \tag{5.8}$$

which is exactly the same result as in [37].

As a nutshell, the information leakages from GCM can be quantified by Theorem 5.1 under the generic leakage model characterized by P, which evaluates the SNR of the leakages. As a direct result, we have the following proposition, which connects the code property $d_{\mathcal{D}}^\perp$ and the security order in GCM.

Proposition 5.1. *The GCM is secure at the order* $t = d_{\mathcal{D}}^\perp - 1$ *under the bounded moment model and the probing model if* $\deg(P) < d_{\mathcal{D}}^\perp$.

Proof. Given a pseudo-Boolean function P, one has $\widehat{P}(z) = 0$ for all $z \in \mathbb{K}^n$ such that $w_H(z) > \deg(P)$ [25]. As a result, SNR will be zero since $\deg(P) < d_{\mathcal{D}}^\perp$ and all codewords of $\mathcal{D}^\perp \setminus \mathcal{C}^\perp$ as in Eqn. 5.6 have Hamming weight no less than $d_{\mathcal{D}}^\perp$. □

Consequently, the attacks on GCM fail if $\deg(P) < d_{\mathcal{D}}^\perp$. Conversely, for an attack to succeed, one must have $\deg(P) \geq d_{\mathcal{D}}^\perp$. This is, however, only a necessary condition, but not a sufficient one. Indeed, it is possible that attacks in the setting $\deg(P) \geq d_{\mathcal{D}}^\perp$ fail. This is illustrated in the next remark.

Remark 5.4. The security order can be even higher than $d_{\mathcal{D}}^\perp - 1$ when there is no x, $y \in \mathcal{D}^\perp \setminus \mathcal{C}^\perp$ such that $x + y \in \mathcal{C}^\perp$ which have weight $d_{\mathcal{D}}^\perp$. Indeed, in Eqn. 5.6, the sum will be empty if the degree of P is equal to $\deg(P) = d_{\mathcal{D}}^\perp$. Thus the SNR is equal to zero, and the security order increases accordingly. A specific example can be found in [164, Example 1] (shown in Appendix 6.1), in which $d_{\mathcal{D}}^\perp$ equals 2 and the security order equals 2 as well.

5.5 Quantifying Hamming Weight Leakages

One realistic leakage model is the so-called "Hamming weight" leakage: each bit is leaking in a similar amount, though independently from others. It has been demonstrated to be practical in many works, such as [16]. In this case, the attacker can measure a quantity $P(Z) = w_H(X\mathbf{G} + Y\mathbf{H})$. However, $\mathbb{E}[P(Z)|X] = \mathbb{E}[P(Z)]$ if the masking is perfect. But there exists a $d > 1$ such that for some x, $\mathbb{E}\left[P(Z)^d|X = x\right] \neq \mathbb{E}\left[P(Z)^d\right]$.

5.5.1 Simplifications

We use $P(z) = w_H(z)^d$ as the informative part in a leakage model, which captures the higher-order leakages where the numerical degree $\deg(P)$ equals d. Moreover, we have:

$$\begin{aligned} P(z) &= w_H(z)^d \\ &= \sum_{J_1+\cdots+J_{n\ell}=d} \binom{d}{J_1,\ldots,J_{n\ell}} \prod_{i=1}^{n\ell} z_i^{J_i} \\ &= \sum_{\substack{J\in\mathbf{N}^{n\ell},\ \text{s.t. } w_H(J)<d; \\ \sum_{i=1}^{n\ell} J_i = d}} \binom{d}{J} z^J + d! \sum_{\substack{I\in\{0,1\}^{n\ell}; \\ w_H(I)=d}} z^I \end{aligned} \tag{5.9}$$

where $\mathbf{N} = \{0, 1, \ldots\}$ is the set of integers. The multinomial coefficient $\binom{d}{J_1,\ldots,J_{n\ell}}$ is defined as $\frac{d!}{J_1!\cdots J_{n\ell}!}$ (recall that $J = (J_1, \ldots, J_{n\ell}) \in \mathbf{N}^{n\ell}$ with $\sum_{i=1}^{n\ell} J_i = d$). This coefficient equals $d!$ as long as for all i ($1 \le i \le n\ell$), $J_i = 0$ or 1. Now, the terms in $P(z)$ are categorized into two cases:

- z^J where $J \in \mathbf{N}^{n\ell}$, $w_H(J) < d$, which consists in products of $< d$ bits of z, as $z^J = \prod_{i\in\{1,\ldots,n\ell\}\ s.t.\ J_i>0} z_i$,
- z^I where $I \in \{0,1\}^{n\ell}$, $w_H(I) = d$ which consists in products of d bits of z, as $z^I = \prod_{i\in\{1,\ldots,n\ell\}\ s.t.\ I_i=1} z_i$.

Indeed, let $i \in \{1, \ldots, n\ell\}$, then $z_i^{J_i} = 1$ if $J_i = 0$, and $z_i^{J_i} = z_i$ if $J_i > 0$. The first terms z^J have numerical degree $\deg(z^J) < d$, hence can be discarded in the analysis (they contribute nothing to the SNR). Remaining terms of numerical degree d are: $\sum_{I\in\{0,1\}^{n\ell},\, w_H(I)=d} z^I$.

Relying on decomposition in Eqn. 5.9, we can simplify lemma 5.1 as follows.

Lemma 5.2. *Let a pseudo-Boolean function $P(Z) = w_H(Z)^d$ denote the leakage function, and taking the same notations as above, we have*

$$\mathbb{V}\left[\mathbb{E}\left[P(Z)|X\right]\right] = B'_d \left(\frac{d!}{2^d}\right)^2. \tag{5.10}$$

where B'_d denotes the adjusted coefficient in weight enumerator which is defined in Def. 5.2.

Before diving into the proof of Lemma 5.2, we define the parameter $B'_{d\frac{\perp}{\mathcal{D}}}$ which count the number of codewords under certain conditions in $\mathcal{C}^\perp$ and $\mathcal{D}^\perp$.

Definition 5.2 (Adjusted coefficient in weight enumerator)**.** Let $\mathcal{C}$ and $\mathcal{D}$ denote two linear codes. The adjusted coefficient B'_d is defined as:

$$B'_d = \left|\{(x, y) \in (\mathcal{D}^\perp \backslash \mathcal{C}^\perp)^2 \,|\, x + y \in \mathcal{C}^\perp,\ w_H(x) = w_H(y) = d\}\right|. \tag{5.11}$$

To be more precise, we use subscript 2 (if necessary) to indicate the subfield representation of a linear code. For instance, $\mathcal{D}_2$ denotes the subfield representation of $\mathcal{D}$ over $\mathbb{F}_2$. Therefore, we have the following lemma for B'_d.

Lemma 5.3. *Recall that $B_{d^\perp_{\mathcal{D}_2}}$ is the coefficient in weight enumerator of $\mathcal{D}_2^\perp$ defined in Def. 3.2, then we have the following inequality in SSS-based masking:*

$$B'_{d^\perp_{\mathcal{D}_2}} \geq B_{d^\perp_{\mathcal{D}_2}} \ .$$

Proof. $B'_{d^\perp_{\mathcal{D}_2}}$ is the number of pairs of codewords (x, y) in $\mathcal{D}^\perp \backslash \mathcal{C}^\perp$ which satisfy the two conditions: their sum is in $\mathcal{C}^\perp$ and their weights are equal to $d^\perp_{\mathcal{D}_2}$. Clearly, this number is greater or equal to the same number of pairs where in addition, x and y are chosen to be identical. In the latter case, the number of codewords is equal to:

$$\left|\{x \in \mathcal{D}^\perp \backslash \mathcal{C}^\perp | w_H(x) = d^\perp_{\mathcal{D}_2}\}\right| , \tag{5.12}$$

because $x + y = 0$ does always belong to $\mathcal{C}^\perp$ and that x and y have the same Hamming weight since they are equal. Now, Eqn. 5.12 is the minimum nonzero coefficient in the weight enumerator of $\mathcal{D}^\perp \backslash \mathcal{C}^\perp$, which is equal to B_d in SSS-based masking. □

Hereafter, we demonstrate Lemma 5.2 by utilizing Eqn. 5.9 to simplify Lemma 5.1.

Proof of Lemma 5.2. Let $\varphi_I(z) = z^I$ where $I \in \{0,1\}^{n\ell}$. Thus

$$z^I = \prod_{i \in I} z_i = \prod_{i \in I} \frac{(1 - (-1)^{z_i})}{2} = \frac{1}{2^d} \prod_{i \in I} (1 - (-1)^{z_i}). \tag{5.13}$$

Since all monomials with numerical degree smaller than d have SNR $= 0$, we only focus on monomials with numerical degree equal to d. Taking $\varphi_I(z) = \phi_I(z) + \frac{(-1)^d}{2^d}(-1)^{\sum_{i \in I} z_i}$ where $\phi_I(z)$ is linear combination of monomials with numerical degree smaller than d in $\varphi_I(z)$, then the *Fourier transform* of $\varphi_I(z)$ is:

$$\begin{aligned} \widehat{\varphi}_I(y) &= \widehat{\phi}_I(y) + \frac{(-1)^d}{2^d} \sum_z (-1)^{z \cdot I} (-1)^{z \cdot y} = \widehat{\phi}_I(y) + \frac{(-1)^d}{2^d} \sum_z (-1)^{z \cdot (I+y)} \\ &= \widehat{\phi}_I(y) + \frac{(-1)^d}{2^{d-n\ell}} \mathbb{1}_{\{I\}}(y). \end{aligned} \tag{5.14}$$

We have $\widehat{\phi}_I(y) = 0$ for y with $w_H(y) \geq d^\perp_{\mathcal{D}} = t + 1 > d$, since given a pseudo-Boolean function P, one has $\widehat{P}(z) = 0$ for all $z \in \mathbb{K}^n$ with $w_H(z) > \deg(P)$ [17, Lemma 1]. As a result, by combining

Eqn. 5.14 with Eqn. A.3, we have the following equation:

$$\begin{aligned}
\mathbb{V}[\mathbb{E}[P(Z)|X]] &= \frac{1}{2^{2n\ell}} \sum_{\substack{x,\, y\in\mathcal{D}^\perp\setminus\mathcal{C}^\perp \\ x+y\in\mathcal{C}^\perp}} \widehat{P}(x)\widehat{P}(y) \\
&= \frac{1}{2^{2n\ell}} \sum_{\substack{x,\, y\in\mathcal{D}^\perp\setminus\mathcal{C}^\perp \\ x+y\in\mathcal{C}^\perp}} \left(\sum_{I|w_H(I)=d} \frac{(-1)^d}{2^{d-n\ell}} \binom{d}{I} \mathbb{1}_{\{I\}}(x)\right) \left(\sum_{I|w_H(I)=d} \frac{(-1)^d}{2^{d-n\ell}} \binom{d}{I} \mathbb{1}_{\{I\}}(y)\right) \\
&= 2^{-2d} \sum_{\substack{x,\, y\in\mathcal{D}^\perp\setminus\mathcal{C}^\perp \\ x+y\in\mathcal{C}^\perp}} \left(\sum_{I|w_H(I)=d} \binom{d}{I} \mathbb{1}_{\{I\}}(x)\right) \left(\sum_{I|w_H(I)=d} \binom{d}{I} \mathbb{1}_{\{I\}}(y)\right) \\
&= \left(\frac{d!}{2^d}\right)^2 \sum_{x,\, y\in\mathcal{D}^\perp\setminus\mathcal{C}^\perp;\ x+y\in\mathcal{C}^\perp} 1 \\
&= B'_d \left(\frac{d!}{2^d}\right)^2,
\end{aligned} \tag{5.15}$$

where B'_d is the adjusted coefficient in weight enumerator defined in Def. 5.2. □

5.5.2 Connecting SNR with Code Properties

Taking Lemma 5.2 as an input to Theorem 5.1, we have the following theorem for Hamming weight leakages in GCM.

Theorem 5.2. *Let a device be protected by the GCM scheme as $Z = X\mathbf{G} + Y\mathbf{H}$. Assume the device is leaking in Hamming weight model in the form: $\mathcal{L} = P(Z) + N$. Then the SNR of the exploitable leakages is:*

$$SNR = \frac{\mathbb{V}\left[\mathbb{E}\left[P(Z)|X\right]\right]}{\sigma^2_{total}} = \begin{cases} 0\,, & \text{if } \deg(P) < d^\perp_{\mathcal{D}} \\ \frac{B'_{d^\perp_{\mathcal{D}}}}{\sigma^2_{total}} \left(\frac{d^\perp_{\mathcal{D}}!}{2^{d^\perp_{\mathcal{D}}}}\right)^2, & \text{if } \deg(P) = d^\perp_{\mathcal{D}} \end{cases} \tag{5.16}$$

where σ^2_{total} is the total noise such that $\sigma^2_{total} \propto \sigma^{2d}$ with $\deg(P) = d$.

Proof. Obviously, substituting the expression of $\mathbb{V}\left[\mathbb{E}\left[P(Z)|X\right]\right]$ in Theorem 5.1 by Lemma 5.2 gives the proof. □

The takeaway point is, the Hamming weight leakages, in which $\deg(P) = 1$, are quantified by Theorem 5.2, in which the two parameters that have an impact on SNR are the dual distance $d^\perp_{\mathcal{D}}$ and the coefficient $B'_{d^\perp_{\mathcal{D}}}$. Therefore, the two parameters also affect the concrete security level of GCM. As a straightforward application of Theorem 5.2, the side-channel resistance of GCM can be optimized by increasing $d^\perp_{\mathcal{D}}$ and/or decreasing $B'_{d^\perp_{\mathcal{D}}}$.

5.5.3 MI-based Information-Theoretic Leakage Quantification

We extend the leakage quantification approach by using another metric, namely MI, in an information-theoretic sense. Let the secret X be encoded as in Eqn. 5.1, and let the leakages be $\mathcal{L} = P(Z) + N$, then the MI between $\mathcal{L}$ and X is defined as $\mathsf{I}(\mathcal{L}; X) = \mathsf{H}(\mathcal{L}) - \mathsf{H}(\mathcal{L}|X)$ where:

- the total entropy is: $\mathsf{H}(\mathcal{L}) = -\int_l \mathbb{P}l \log_2 \mathbb{P}l \mathrm{d}l$,
- the conditional entropy $\mathsf{H}(\mathcal{L}|X)$ is: $\mathsf{H}(\mathcal{L}|X) = -\sum_{x \in \mathbb{F}_2^\ell} \mathbb{P}x \int_l \mathbb{P}l|x \log_2 \mathbb{P}l|x \mathrm{d}l$.

In multivariate cases, two entropies are computed on $\mathcal{L} = (\mathcal{L}_1, \mathcal{L}_2, \ldots, \mathcal{L}_d)$ for a d-variate MI by a d-D integral on continuous variables. While in monovariate cases, two entropies are computed by 1-D integrals. Moreover, $\mathsf{I}(\mathcal{L}; X)$ can be expanded using a Taylor's expansion[1] [23]:

$$\mathsf{I}(\mathcal{L}; X) = \sum_{d=0}^{+\infty} \frac{1}{2\, d! \ln 2} \sum_{x \in \mathbb{F}_2^\ell} Pr(x) \frac{\left(k_d(P(Z)|x) - k_d(P(Z))\right)^2}{\left(\mathbb{V}\left[P(Z)\right] + \sigma^2\right)^d} \tag{5.17}$$

where k_d is the d-th order cumulant [21].

Assuming the device is leaking in the Hamming weight model, we have the following theorem for quantifying the information leakages in GCM.

Theorem 5.3. *Let a device be protected by the GCM scheme as $Z = X\mathbf{G} + Y\mathbf{H}$. Assume the leakages of the device can be represented in the form: $\mathcal{L} = P(Z) + N$. Then the MI between $\mathcal{L}$ and X is estimated as:*

$$\mathsf{I}(\mathcal{L}; X) = \begin{cases} 0\,, & \text{if } \deg(P) < d_{\mathcal{D}}^{\perp} \\ \frac{d_{\mathcal{D}}^{\perp}! B'_{d_{\mathcal{D}}^{\perp}}}{2 \ln 2 \cdot 2^{2 d_{\mathcal{D}}^{\perp}}} \times \frac{1}{\sigma^{2 d_{\mathcal{D}}^{\perp}}} + \mathcal{O}\left(\frac{1}{\sigma^{2(d_{\mathcal{D}}^{\perp}+1)}}\right), & \text{if } \deg(P) = d_{\mathcal{D}}^{\perp}, \text{ when } \sigma \to +\infty \end{cases} \tag{5.18}$$

where σ is the standard deviation of noise in the leakage of each share.

Proof. Since for a d-CI (*Correlation Immune*) function [22, Def. 1], all moments of order $i \leq d$ are centered, so are the cumulants. Therefore, the first nonzero cumulant $k_d(X)$ is $k_{d_{\mathcal{D}}^{\perp}}(X)$ and it equals $\mu_{d_{\mathcal{D}}^{\perp}}(X)$. As a consequence, the term $\mathbb{E}\left[\left(k_d(P(Z)|X) - k_d(P(Z))\right)^2\right]$ in Eqn. 5.17 is null for all $d < d_{\mathcal{D}}^{\perp}$ and it is equal to $\mathbb{E}\left[\left(\mu_d(P(Z)|X) - \mu_d(P(Z))\right)^2\right] = \mathbb{V}\left[\mu_{d_{\mathcal{D}}^{\perp}}(P(Z)|X)\right] = \mathbb{V}\left[\mathbb{E}\left[P(Z)^{d_{\mathcal{D}}^{\perp}}|X\right]\right]$ for $d = d_{\mathcal{D}}^{\perp}$.

Assume that the device leaks in Hamming weight model, then $P(Z)^{d_{\mathcal{D}}^{\perp}}$ has a degree equal to $d_{\mathcal{D}}^{\perp}$. Hence the MI is equal to:

$$\mathsf{I}(\mathcal{L}; X) = \frac{1}{2 \ln 2 \cdot d_{\mathcal{D}}^{\perp}!} \frac{\mathbb{V}\left[\mathbb{E}\left[P(Z)^{d_{\mathcal{D}}^{\perp}}|X\right]\right]}{\left(\mathbb{V}\left[P(Z)\right] + \sigma^2\right)^{d_{\mathcal{D}}^{\perp}}} + \mathcal{O}\left(\frac{1}{\left(\mathbb{V}\left[P(Z)\right] + \sigma^2\right)^{d_{\mathcal{D}}^{\perp}+1}}\right), \tag{5.19}$$

[1] The normalization by $\ln 2$ allows the mutual information to be expressed in unit of bits.

when $\sigma \to +\infty$. Finally, Eqn. 5.19 can be further developed at the first order in $1/\sigma^{2d_{\mathcal{D}}^{\perp}}$ as follows after involving Eqn. 5.15:

$$\mathsf{I}(\mathcal{L}; X) = \frac{d_{\mathcal{D}}^{\perp}! B'_{d_{\mathcal{D}}^{\perp}}}{2 \ln 2 \cdot 2^{2d_{\mathcal{D}}^{\perp}}} \times \frac{1}{\sigma^{2d_{\mathcal{D}}^{\perp}}} + \mathcal{O}\left(\frac{1}{\sigma^{2(d_{\mathcal{D}}^{\perp}+1)}}\right) ,$$

when $\sigma \to +\infty$, which proves Theorem 5.3. □

A comparison of MIs by estimation and numerical calculation is shown in Fig. 5.2. More precisely, the estimated MIs are converging to numerical one when $\log_{10} \sigma^2 \approx 1.5$, which verifies Theorem 5.3 numerically.

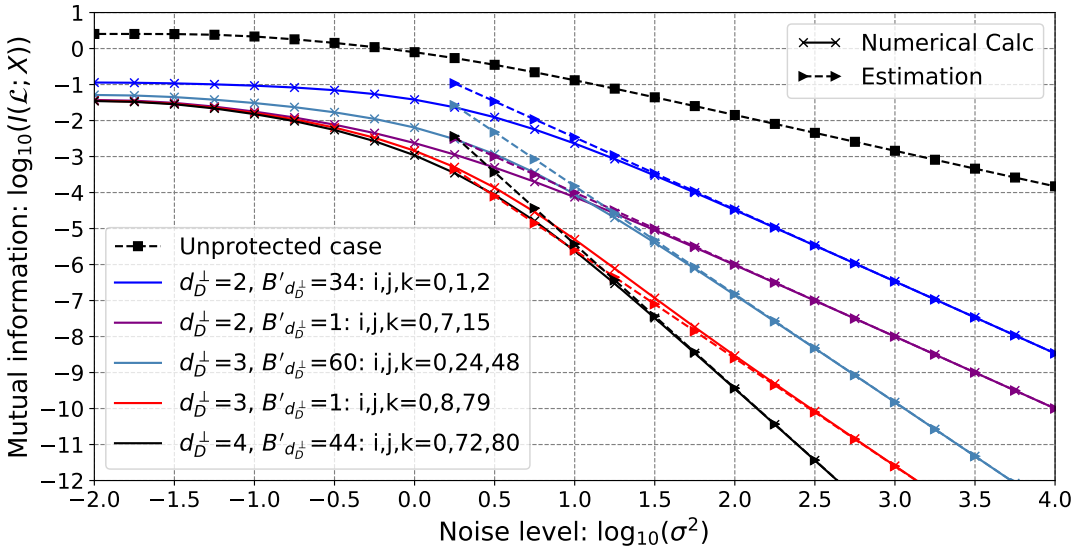


Figure 5.2: Numerical calculation and approximation of $\mathsf{I}(\mathcal{L}; X)$ between leakage $\mathcal{L}$ and the sensitive variable X in $(3, 1)$-SSS based masking. The three public points are $\alpha_1 = \alpha^i$, $\alpha_2 = \alpha^j$, $\alpha_3 = \alpha^k$.

Summing up, the information leakages of GCM under the Hamming weight model can be estimated by the two parameters $d_{\mathcal{D}}^{\perp}$ and $B'_{d_{\mathcal{D}}^{\perp}}$ in an information-theoretic sense. In the general case of leakage function P, the MI can be estimated similarly by applying different forms of P into Eqn. 5.19 to derive connections to coding properties correspondingly.

5.6 Optimal Codes for GCM

Thanks to Theorem 5.1, 5.2 and 5.3, we can compare the information leakages of GCM in a quantitative manner. More importantly, relying on the analytic characterization of information

leakages, the three theorems enable us to choose optimal linear codes for GCM. Specifically, the codes with maximized $d_{\mathcal{D}}^{\perp}$ and minimized $B'_{d_{\mathcal{D}}^{\perp}}$ are the best candidates for GCM. Considering the SSS-based masking as a special case, the optimal public points can be determined straightforwardly by applying the two theorems.

To thoroughly validate the optimal codes, we consider multivariate leakages. In particular, it is shown in [153] that comparing to sum, absolute difference, and normalized product, the joint distribution is the most efficient way to combine the multivariate leakages in side-channel analysis. In this work, we consider both sum and joint distribution to exploit the multivariate leakages. A comparison of the two combination functions in an information-theoretic sense is presented in Appendix B.3.

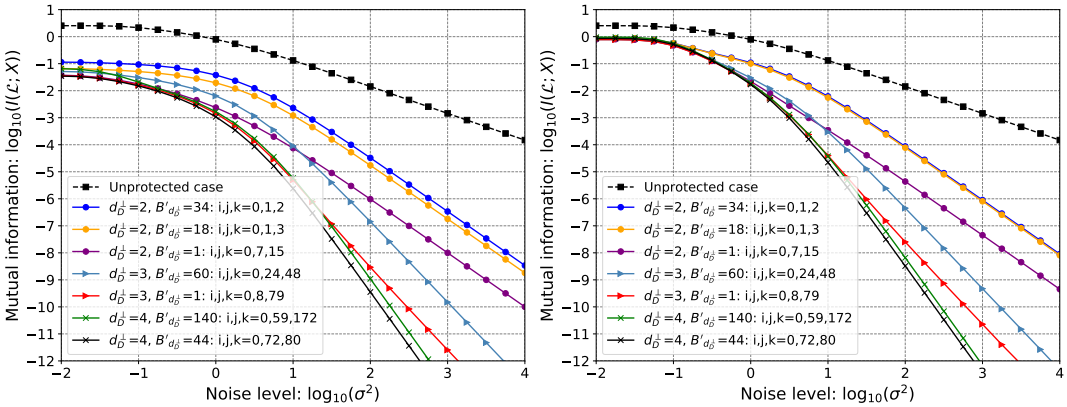
We take $(3, 1)$-SSS based masking as an example of GCM and specify it as follows. Let X be encoded into $Z = X\mathbf{G} + Y\mathbf{H}$ with $n = 3$ shares, the two generator matrices are:

$$\begin{aligned} \mathbf{G} &= \begin{pmatrix} 1 & 1 & 1 \end{pmatrix}, \\ \mathbf{H} = \begin{pmatrix} \alpha_1 & \alpha_2 & \alpha_3 \end{pmatrix} &= \begin{pmatrix} 1 & \alpha^j & \alpha^k \end{pmatrix}. \end{aligned} \tag{5.20}$$

Considering the common "*Hamming weight + Gaussian noise*" model, the side-channel leakages are simulated as follows. Let $\mathcal{L} = (\mathcal{L}_1, \mathcal{L}_2, \mathcal{L}_3)$ be 3-D leakages where $\mathcal{L}_i = \phi_P(Z_i) + N_i = w_H(Z_i) + N_i$ for $1 \le i \le 3$ and $N_i \sim \mathcal{N}(0, \sigma^2)$ is the Gaussian noise. To combine the 3-D leakages, other sum or joint distribution are applied wherein $\varphi_P(\mathcal{L}) = \sum_{i=0}^{3} \mathcal{L}_i$ is called 1-D leakages or $\varphi_P(\mathcal{L}) = (\mathcal{L}_1, \mathcal{L}_2, \mathcal{L}_3)$ is called 3-D leakages, respectively.

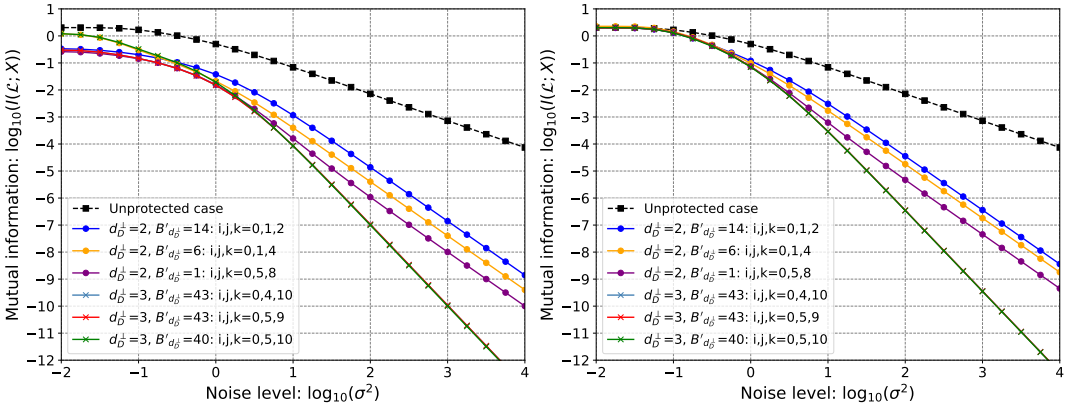
The results are shown in Fig. 5.3(a) and 5.3(b) are 1-D MI and 3-D MI, respectively (more results over $\mathbb{F}_{2^4}$ are in Fig. 5.4). The first observation is that the 3-D MI utilizing joint distribution exploits more key-dependent information existed in leakages, therefore the attack is more efficient when using the joint distribution of leakages [18]. Secondly, the numerical results in Fig. 5.3 are in accordance with the Theorem 5.2 and 5.3, where the two parameters $d_{\mathcal{D}}^{\perp}$ and $B'_{d_{\mathcal{D}}^{\perp}}$ in codes play a significant role in determining the side-channel resistance of GCM.

Thirdly, the strategy to choose the optimal codes for GCM is to maximize the dual distance $d_{\mathcal{D}}^{\perp}$ and/or to minimize the conditioned number of codewords $B'_{d_{\mathcal{D}}^{\perp}}$. Moreover, the concrete side-channel security level of GCM will be improved by optimizing either of the two parameters. Interestingly, when the noise levels are at certain intervals, the codes with smaller $d_{\mathcal{D}}^{\perp}$ (also with smaller $B'_{d_{\mathcal{D}}^{\perp}}$) may be better than that with larger $d_{\mathcal{D}}^{\perp}$. For instance, for the curves in purple (the fourth one) and in sky-blue (the fifth one) of Fig. 5.3, the corresponding $d_{\mathcal{D}}^{\perp}$ are 2 and 3,



(a) 1-D MI on 8-bit case. (b) 3-D MI on 8-bit case.

Figure 5.3: An information-theoretic evaluation of the leakages $\mathcal{L}$ and the sensitive variable X in $(3,1)$-SSS based masking. We choose seven codes with different values of $d_{\mathcal{D}}^{\perp}$ and/or $B'_{d_{\mathcal{D}}^{\perp}}$. The three public points are $\alpha_1 = \alpha^i$, $\alpha_2 = \alpha^j$, $\alpha_3 = \alpha^k$.



(a) 1-D MI on 4-bit case. (b) 3-D MI on 4-bit case.

Figure 5.4: An information-theoretic evaluation of the leakages $\mathcal{L}$ and the sensitive variable $X \in \mathbb{F}_{2^4}$. Six codes are chosen with different $d_{\mathcal{D}_2}^{\perp}$ and/or $B'_{d_{\mathcal{D}_2}^{\perp}}$.

respectively. When $\sigma^2 < 10$, the purple curve shows a better side-channel resistance than the sky-blue one.

5.7 Conclusions and Perspectives

This chapter presented a unified approach to quantifying the information leakages of code-based masking in the most general case, namely GCM, which already encompasses many state-of-the-art masking schemes. Firstly, by a uniform representation of encodings in GCM, we proposed a

quantitative approach to evaluate the concrete security level of GCM. The signal-to-noise ratio and mutual information are used as two complementary metrics to quantify the lowest degree of key-dependent leakages. By this unified approach, we were able to quantify the impact of different codes in GCM and optimize it by choosing optimal codes for it. Next, we evaluated the impact of public points in Shamir's Secret Sharing in the context of masking. Thanks to the unified analytic approach, we showed the impact of public points in side-channel security orders of the corresponding masking. More importantly, we provided a roadmap to optimal linear codes for designers to optimize the SSS-based masking (also GCM) soundly. Lastly, we revisited the independence condition behind the masking scheme and showed that the intra-share dependence could ruin higher-order security under the bounded moment model. In particular, we showed how the higher-order intra-share leakages affect the side-channel security orders precisely.

CHAPTER 6

Redundancy in Code-based Masking

This chapter presents the work [35] published at *IACR Trans. Cryptogr. Hardw. Embed. Syst.* (TCHES) 2021, issue 3. Part of results are also been demonstrated in [42].

Contents

6.1 A Starter Example

As shown in Remark 5.4, there are some cases of GCM in which the side-channel security order can be greater than the dual distance of $\mathcal{D}$ minus one. In particular, Wang et al. [164] presented

an example where the generator matrices of $\mathcal{C}$ and $\mathcal{D}$ as follows, respectively,

$$\begin{aligned}\mathbf{G} &= \begin{pmatrix}1&1&1&1&0&0&0&0\\0&0&0&0&1&1&0&0\end{pmatrix} \in \mathbb{F}_2^{2\times 8},\\ \mathbf{H} &= \begin{pmatrix}1&1&0&0&0&0&0&0\\0&0&1&1&0&0&0&1\\0&0&0&1&1&1&0&0\\0&0&0&0&0&1&1&1\end{pmatrix} \in \mathbb{F}_2^{4\times 8}.\end{aligned} \tag{6.1}$$

We can compute the generator matrices of the dual codes $\mathcal{C}^\perp$ and $\mathcal{D}^\perp$ as follows, respectively,

$$\begin{aligned}\mathbf{G}^\perp &= \begin{pmatrix}1&0&0&1&0&0&0&0\\0&1&0&1&0&0&0&0\\0&0&1&1&0&0&0&0\\0&0&0&0&1&1&0&0\\0&0&0&0&0&0&1&0\\0&0&0&0&0&0&0&1\end{pmatrix} \in \mathbb{F}_2^{6\times 8},\\ \mathbf{H}^\perp &= \begin{pmatrix}1&1&0&0&0&0&0&0\\0&0&1&0&0&0&1&1\\0&0&0&1&0&1&0&1\\0&0&0&0&1&1&1&0\end{pmatrix} \in \mathbb{F}_2^{4\times 8},\end{aligned} \tag{6.2}$$

where $\mathcal{C}^\perp$ is a code with parameters $[8,6,1]$ and $\mathcal{D}^\perp$ is of parameters $[8,4,2]$. We have $d_{\mathcal{D}}^\perp = d_{\mathcal{D}^\perp} = 2$ and $B_2 = 1$ for $\mathcal{D}^\perp$. Therefore, there is only one codeword $u = [1,1,0,0,0,0,0,0] \in \mathcal{D}^\perp$ such that $w_H(u) = 2$. Since u is also in $\mathcal{C}^\perp$, which indicates that B_2' equals 0. As a consequence, applying Theorem 5.2 gives that SNR equals 0 for $\deg(P) = d_{\mathcal{D}}^\perp = 2$ under Hamming weight leakages (e.g., $P(Z) = w_H(Z)$) and then the security order is at least equal to $d_{\mathcal{D}}^\perp$, rather than $d_{\mathcal{D}}^\perp - 1$. More generally, taking Theorem 5.1 gives the same conclusion for any leakage function P with $\deg(P) = 2$.

In particular, we checked that the first nonzero $B'_{d_{\mathcal{D}}^\perp}$ for nonzero codewords is $B_3' = 3$. Therefore the security order is exactly 2 in above example.

6.2 Enhancing the SSS-based Polynomial Masking

In the context of masking, the random masks in SSS-based masking are u_i for $1 \le i \le t$ where $\alpha_1, \alpha_2, \ldots, \alpha_n$ are n public points. Two main observations made in [29] are:

- the choices of public points α_i can have an impact on side-channel resistance of the corresponding masking scheme, therefore, combining different $t+1$ tuples of Z_i, the efficiencies of corresponding template attacks are different,
- combining more than $t+1$ tuples of Z_i may improve the attack efficiency in the sense of the number of traces needed to recover the secret key.

Recall that the generator matrices in SSS-based masking (e.g., the RS code) from Tab. 5.1, $\mathbf{G}$ and $\mathbf{H}$ are the same as the generator matrices in DSM when $n = t + 1$. In the context of masking, we only care about $\mathbf{G}$ and $\mathbf{H}$, since the former is used to encode the secret X and the latter is for encoding the random masks (e.g., $u_1, \ldots, u_t$ in the case of SSS-based masking).

Note that $\mathbf{H}$ is a Vandermonde matrix, resulting in that the code $\mathcal{D}$ is a maximum distance separable (MDS) code, it is optimal at word-level. However, with different parameters α_i for $1 \leq i \leq n$, the codes have different impacts on side-channel resistance when they are adopted in masking schemes.

6.2.1 Further Clarifications

We further clarify the properties of the code $\mathcal{D}$ and its dual as follows. Let $\mathcal{D}$ be an RS code of parameters $[n, t, n - t + 1]$ which is generated by $\mathbf{H}$ in Eqn. 5.3. Then its dual code $\mathcal{D}^{\perp}$ is also an RS code of parameters $[n, n - t, t + 1]$ [101]. Recall the connections between the RS code and SSS scheme, $\mathcal{D}$ can be used to construct an (n, t)-SSS scheme.

Given that $n \geq t + 1$, we assume that $t + 1 \leq n' \leq n$, the code $\mathcal{D}'$ is constructed by selecting n' columns from the generator matrix $\mathbf{H}$ of $\mathcal{D}$ (or equivalently, remove $n - n'$ columns in $\mathbf{H}$). Subsequently, the code $\mathcal{D}'$ has parameters $[n', t, n' - t + 1]$. It is also an RS code and its dual code $\mathcal{D}'^{\perp}$ has parameters $[n', n' - t, t + 1]$. Therefore, the dual distance of $\mathcal{D}'$ is equal to $\mathcal{D}$, namely $d^{\perp}_{\mathcal{D}'} = d^{\perp}_{\mathcal{D}} = t + 1$. In summary, removing some coordinates ($n' \geq t + 1$) in RS code does not decrease its dual distance (at word-level).

Remark 6.1. Note that for two arbitrary linear codes $\mathcal{D}$ and $\mathcal{D}'$ where the latter is generated from the former as above (by selecting some coordinates), we have the following lemma for their dual distances.

Lemma 6.1. $d^{\perp}_{\mathcal{D}} \leq d^{\perp}_{\mathcal{D}'}$.

Proof. Assume $u \in \mathcal{D}'^{\perp}$, by appending $n - n'$ zeros to u, then the new codeword $(u,\ 0_{n-n'})$ is also a codeword of $\mathcal{D}^{\perp}$. Therefore we have $d^{\perp}_{\mathcal{D}} \leq d^{\perp}_{\mathcal{D}'}$ [26]. □

Interestingly, Lemma 6.1 implies that given a fixed t, adding more shares in an (n, t)-SSS based masking cannot increase the security order of the corresponding masking scheme and can be more likely to lower the security order, especially under the bit-probing model.

6.2.2 Representing Linear Codes in Subfield $\mathbb{F}_2$

We take $\mathbb{F}_2$ as the subfield, then any codes over $\mathbb{F}_{2^\ell}$ can be expanded into subfields by code expansion Def. 3.6. We further investigate the properties of codes $\mathcal{D}$ and $\mathcal{D}'$.

Let $\mathcal{D}_2$ and $\mathcal{D}'_2$ denote the expanded codes of $\mathcal{D}$ and $\mathcal{D}'$ over $\mathbb{F}_2$, respectively. Since they are not MDS codes at the bit level, there is no straightforward method to compare the dual distances of $\mathcal{D}_2$ and $\mathcal{D}'_2$. However, by Lemma 6.1, it is obvious to have $d^{\perp}_{\mathcal{D}_2} \leq d^{\perp}_{\mathcal{D}'_2}$. This connection helps in SSS-based masking since, by increasing n, the dual distance at word-level keeps the same, but the dual distance at bit-level cannot be larger than in the case with $n' = t + 1$. Moreover, from the adversary's viewpoint, combining more than $t + 1$ shares may be more efficient when attacking a specific SSS-based implementation.

From the quantitative results in Sec. 5.4, two parameters that have an impact on the side-channel resistance of GCM is the dual distance $d^{\perp}_{\mathcal{D}_2}$ and the coefficient $B'_{d^{\perp}_{\mathcal{D}_2}}$. Hereafter, we use the information-theoretic metric to show how the more redundant shares affect the concrete security level in SSS-based masking.

6.2.3 More Redundancy in Sharing Leaks More

We present an information-theoretic evaluation on $(3, 1)$-SSS based polynomial masking. Taking $n = 3$ and $t = 1$, then the three public points $(\alpha_1, \alpha_2, \alpha_3)$ can be derived by setting $\alpha_1 = \alpha^i$, $\alpha_2 = \alpha^j$ and $\alpha_j = \alpha^k$, where i, j, k must be distinct integers. Due to the equivalence of the linear codes (Sec. 3.1), we can choose $i = 0$, $1 \leq j < k \leq 254$ and obtain 32131 candidates rather than $\binom{255}{3} = 2731135$ in total. Recall that the generator matrices $\mathbf{G}$ and $\mathbf{H}$ are as in Eqn. 5.20. Therefore, taking a random mask u_1, the X is encoded into:

$$Z = (Z_1, Z_2, Z_3) = X\mathbf{G} + u_1\mathbf{H} = (X + u_1\alpha_1, X + u_1\alpha_2, X + u_1\alpha_3) \ . \tag{6.3}$$

For all possible values of $\alpha_1, \alpha_2, \alpha_3 \in \mathbb{F}_{2^8}$, we study the dual distance $d^{\perp}_{\mathcal{D}}$ and the coefficient $B_{d^{\perp}_{\mathcal{D}}}$ at both word-level and bit-level. As expected, all codes have the same weight enumerator at word-level (they are all MDS codes and optimal at word-level). However, there are three possible values for $d^{\perp}_{\mathcal{D}}$ at bit-level, namely $d^{\perp}_{\mathcal{D}_2} \in \{2, 3, 4\}$. Hence, for each possible $d^{\perp}_{\mathcal{D}_2}$, we further study the possible values for the other parameter $B_{d^{\perp}_{\mathcal{D}_2}}$. In particular, for each case of $d^{\perp}_{\mathcal{D}_2}$, we show two or three codes with maximal and minimal values of $B_{d^{\perp}_{\mathcal{D}_2}}$. The specific properties of the codes are listed in Tab. 6.1 [1] and the MI between the leakages $\mathcal{L}$ and X are depicted in Fig. 5.3. The complete details of all linear codes for the $(3, 1)$-SSS based masking are available in [34].

[1] The data in Tab. 6.1 is formally verified by `Magma` [158]. Moreover, the scripts for calculating B'_d are also available on `Github` [34].

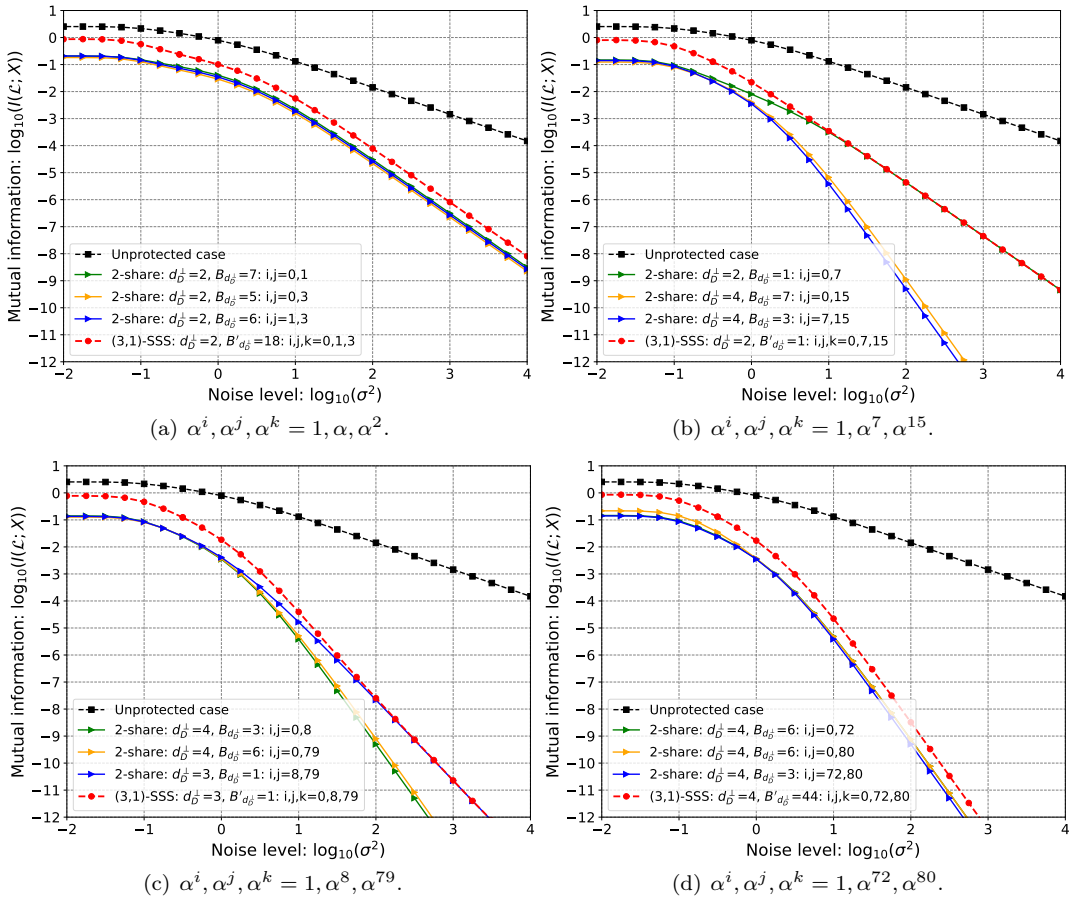
Table 6.1: Exhibiting different codes in $(3,1)$-SSS scheme generated by Eqn. 6.3. Note that we take $\alpha_1 = \alpha^i = 1$, $\alpha_2 = \alpha^j$ and $\alpha_3 = \alpha^k$.

	$j=1$ $k=2$	$j=1$ $k=3$	$j=7$ $k=15$	$j=24$ $k=48$	$j=8$ $k=79$	$j=59$ $k=172$	$\boldsymbol{j=72}$ $\boldsymbol{k=80}$
Minimum distance $d_{\mathcal{D}}$	3	3	3	3	3	3	3
Dual distance (word) $d^{\perp}_{\mathcal{D}}$	2	2	2	2	2	2	2
Dual distance (bit) $d^{\perp}_{\mathcal{D}_2}$	2	2	2	3	3	**4**	**4**
Coefficient (bit) $B_{d^{\perp}_{\mathcal{D}_2}}$	20	18	**1**	22	**1**	76	**36**
Coefficient (bit) $B'_{d^{\perp}_{\mathcal{D}_2}}$	34	18	**1**	60	**1**	140	**44**

As shown in Tab. 6.1, for the first time, we exhibit an approach to find the optimal codes for SSS-based masking and present optimal codes for $(3,1)$-SSS based masking. Specifically, the code with $\alpha_1 = 1$, $\alpha_2 = \alpha^{72}$ and $\alpha_3 = \alpha^{80}$ (in the last column of Tab. 6.1) is one of the best candidates for $(3,1)$-SSS based masking. In addition, the generator matrices of all three optimal (nonequivalent) codes are shown in Appendix B.2. It is worth noting that the codes obtained by permuting the order of α_i for $1 \leq i \leq 3$ are equivalent, resulting in only three optimal codes for $(3,1)$-SSS based masking over $\mathbb{F}_{2^8}$.

Using the same settings of $(3,1)$-SSS based masking as in Sec. 5.6, the results of MI on the information leakages of 3-share and corresponding 2-share combinations are shown in Fig. 6.1. In each of four cases, the main takeaway point is that given a specific t in (n,t)-SSS based masking, all the more shares leak more key-dependent information. Specifically, we first highlight that the smallest security order determines the side-channel security of SSS-based masking among all $\binom{n}{t+1}$ combinations. In the context of coding theory, the dual distance of n-share SSS-based masking is determined by the minimum value of dual distances in truncated codes $\mathcal{D}'$. Two instances are in Fig. 6.1(b) and 6.1(c) where the minimum of dual distances are 2 and 3, respectively.

Secondly, when the codes in SSS and its truncated variants have the same dual distance, the parameter $B'_{d^{\perp}_{\mathcal{D}}}$ plays a role in side-channel resistance. More precisely, smaller $B'_{d^{\perp}_{\mathcal{D}}}$ brings improved concrete security for GCM. Two instances are shown in Fig. 6.1(a) and 6.1(d) where the dual distances of $\mathcal{D}$ are 2 and 4, respectively. Interestingly, a recent work [51] provides empirical comparisons on some instances of $(2,1)$-SSS and $(3,1)$-SSS based masking, which confirms our information-theoretic evaluation.



(a) $\alpha^i, \alpha^j, \alpha^k = 1, \alpha, \alpha^2$. (b) $\alpha^i, \alpha^j, \alpha^k = 1, \alpha^7, \alpha^{15}$.

(c) $\alpha^i, \alpha^j, \alpha^k = 1, \alpha^8, \alpha^{79}$. (d) $\alpha^i, \alpha^j, \alpha^k = 1, \alpha^{72}, \alpha^{80}$.

Figure 6.1: More shares leak more information, two study-cases on $(3,1)$-SSS based masking, where the three public points are: $\alpha_1 = \alpha^i$, $\alpha_2 = \alpha^j$, $\alpha_3 = \alpha^k$.

In summary, the information-theoretic evaluations in Fig. 6.1 confirms that more redundancy in sharing of GCM would leak more information. Besides, one way to find optimal codes for GCM is to build up from (sub-)optimal choices of the codes with less shares.

6.2.4 Different Codes for $(3,1)$-SSS and $(5,2)$-SSS based Masking

We present further results for both $(3,1)$-SSS and $(5,2)$-SSS based masking schemes which are supplementary to Tab. 6.1.

Note that in Tab. 6.3 we fix both α_1 and α_2 since there are too many candidates for enumeration (more accurately, $\binom{255}{5} = 8,637,487,551$ candidates in total). In addition, the reason for taking $\alpha_2 = \alpha^8$ is that $(1 \quad \alpha^8) \in \mathbb{F}_{2^8}^2$ is one of the optimal code for $(2,1)$-SSS based masking.

Table 6.2: Exhibiting different codes in $(3, 1)$-SSS scheme over $\mathbb{F}_{2^4}$ generated by Eqn. 5.20. Note that we take $\alpha_1 = \alpha^i = 1$, $\alpha_2 = \alpha^j$ and $\alpha_3 = \alpha^k$.

	$j = 1$ $k = 2$	$j = 1$ $k = 3$	$j = 3$ $k = 7$	$j = 4$ $k = 8$	$j = 5$ $k = 10$
Minimum distance $d_{\mathcal{D}}$	3	3	3	3	3
Dual distance (word) $d^{\perp}_{\mathcal{D}}$	2	2	2	2	2
Dual distance (bit) $d^{\perp}_{\mathcal{D}_2}$	2	2	2	**3**	**3**
Coefficient (bit) $B_{d^{\perp}_{\mathcal{D}_2}}$	8	6	**1**	17	**16**
Coefficient (bit) $B'_{d^{\perp}_{\mathcal{D}_2}}$	14	6	**1**	45	**40**

Table 6.3: Exhibiting different codes in $(5, 2)$-SSS scheme over $\mathbb{F}_{2^8}$. Note that we fix $\alpha_1 = \alpha^i = 1$, $\alpha_2 = \alpha^8$ and enumerate all possible $\alpha_3 = \alpha^k$, $\alpha_4 = \alpha^l$ and $\alpha_5 = \alpha^r$.

	$k = 116$ $l = 169$ $r = 214$	$k = 1$ $l = 3$ $r = 184$	$k = 139$ $l = 172$ $r = 225$	$k = 1$ $l = 3$ $r = 12$	$k = 18$ $l = 52$ $r = 219$	$k = 1$ $l = 5$ $r = 51$	$k = 14$ $l = 111$ $r = 219$	$k = 90$ $l = 92$ $r = 192$
Minimum distance $d_{\mathcal{D}}$	4	4	4	4	4	4	4	4
Dual distance (word) $d^{\perp}_{\mathcal{D}}$	3	3	3	3	3	3	3	3
Dual distance (bit) $d^{\perp}_{\mathcal{D}_2}$	3	3	4	4	5	5	**6**	**6**
Coefficient (bit) $B_{d^{\perp}_{\mathcal{D}_2}}$	19	**1**	29	**1**	43	**1**	115	**30**
Coefficient (bit) $B'_{d^{\perp}_{\mathcal{D}_2}}$	35	**1**	39	**1**	55	**1**	215	**32**

Table 6.4: Exhibiting different codes in $(5, 2)$-SSS scheme over $\mathbb{F}_{2^4}$. Note that we take $\alpha_1 = \alpha^i = 1$, $\alpha_2 = \alpha^j$, $\alpha_3 = \alpha^k$, $\alpha_4 = \alpha^l$ and $\alpha_5 = \alpha^r$.

	$j = 1,\ k = 4$ $l = 6,\ r = 12$	$j = 1,\ k = 4$ $l = 6,\ r = 11$	$j = 1,\ k = 2$ $l = 3,\ r = 11$	$j = 3,\ k = 6$ $l = 9,\ r = 12$	$j = 1,\ k = 3$ $l = 5,\ r = 8$
Minimum distance $d_{\mathcal{D}}$	4	4	4	4	4
Dual distance (word) $d^{\perp}_{\mathcal{D}}$	3	3	3	3	3
Dual distance (bit) $d^{\perp}_{\mathcal{D}_2}$	3	3	3	**4**	**4**
Coefficient (bit) $B_{d^{\perp}_{\mathcal{D}_2}}$	12	11	**1**	25	**17**
Coefficient (bit) $B'_{d^{\perp}_{\mathcal{D}_2}}$	20	19	**1**	225	**39**

6.3 Revisiting the Independence Condition

Failing to ensure the independence of the shares can ruin a masking scheme by revealing a lower order of key-dependent leakages than the designed security order. For instance, the unintentional

physical coupling [8] in the hardware device can combine leakages from different shares, hence degrade the concrete security level of a masked implementation. In this section, we investigate the intra-share independence issue and show the theoretical condition of higher-order security of code-based masking, especially in GCM as it is the most general case.

Another reason why the independence condition might be broken is the existence of glitches. Let us reason on a canonical example, namely that of the exclusive-or (XOR) gate. Let Z_1 and Z_2 be two single-bit shares, which enter an XOR gate. Recall that the leakage function is $P = \varphi_P \circ \phi_P$ as introduced in Sec. 5.4. Taking $\phi_P = 1$, then the leakage function is the pseudo-Boolean function φ_P, which lives in $\mathbb{F}_2 \times \mathbb{F}_2 \to \mathbb{R}$. It is equal to:

$$\varphi_P(Z_1, Z_2) = Z_1 \times Z_2 + (1 - Z_1) \times (1 - Z_2) = 2Z_1 \times Z_2 - Z_1 - Z_2 + 1. \tag{6.4}$$

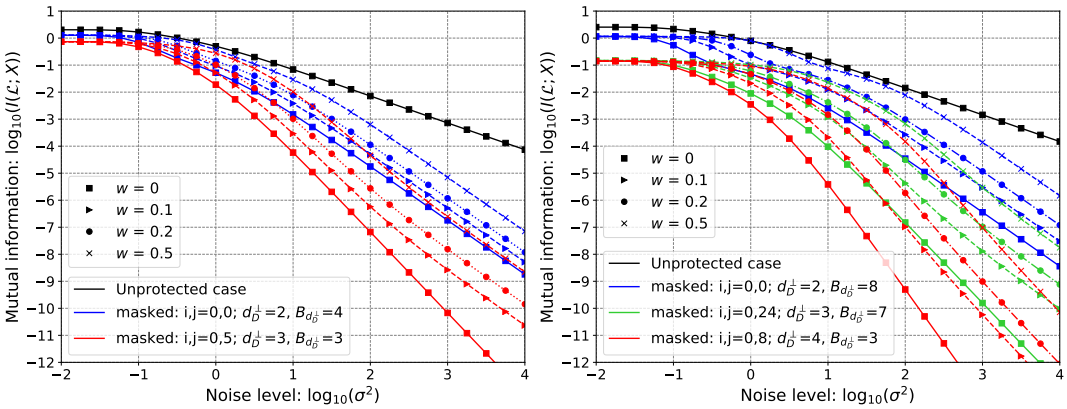
This function can glitch because of the term $Z_1 \times Z_2$. Indeed, if Z_1 changes, then the leading term still depends on Z_2 (derivative). Therefore, glitches are dreadful since they consist in combinations from within the chip, even before the measurement noise arrives.

An Information-Theoretic Evaluation of Intra-Share Independence. We consider the Hamming weight as leakage model in a perfect independent case and take the weighted square of Hamming weight as second-order (non-linear) leakages as follows:

$$\phi_P(Z_i) = \sum_{j=1}^{\ell} Z_{i,j} + w \sum_{j \neq k}^{\ell} Z_{i,j} Z_{i,k} = w_H(Z_i) + w \sum_{j \neq k}^{\ell} Z_{i,j} Z_{i,k} \tag{6.5}$$

where Z_i is an ℓ-bit share and w is the weight of second-order leakages. As a consequence, $P(Z) = \phi_P(Z)$ will be the same as Hamming weight model with $\deg(P) = 1$ if $w = 0$. Otherwise, there exists a different amount of second-order leakages indicated by w where the degree of P equals 2. The MI results on four candidates of w are shown in Fig. 6.2 for 4-bit and 8-bit variables, respectively. It is worthwhile to note that in 2-share settings with $n = 2$ and $t = 1$, the SSS-based masking can be transformed into IPM by changing the way of involving public parameters α_i for $1 \leq i \leq n$. Essentially, the two schemes are different because of the structure of $\mathbf{G}$ and $\mathbf{H}$ as in Tab. 5.1, but are comparable from a side-channel perspective.

The first observation from Fig. 6.2 is that MI increases along with the increasing amount of second-order leakages. More importantly, in the presence of second-order leakages, the security order under the bit-probing model [123] (indicated by the slope of MI curves when the noise level is high) decreases by one since the degree of ϕ_P is 2. Similarly, the security order will



(a) $\ell = 4$ for 4-bit case.

(b) $\ell = 8$ for 8-bit case.

Figure 6.2: The intra-share independence issue: the existence of higher-order leakages decreases the security of the corresponding masking scheme (two public parameters are $\alpha_1 = \alpha^i$, $\alpha_2 = \alpha^j$ as in Tab. 5.1). Note that the blue curves are for the Boolean masking.

reduce by two when the degree of ϕ_P equals 3 in the red curves of Fig. 6.2(b). However, the lowest security order under the bit-probing model is bounded by the Boolean masking under the word-probing model. More precisely, increasing the degree of ϕ_P only affects the intra-share independence and therefore decreases the security order under the bit-probing model, while the degree of φ_P (e.g., induced by couplings) affects the security order under the word-probing model.

6.4 Related Works

6.4.1 Differences with [37] in Detail

As summarized in Sec. 5.2, this work tackles GCM, which is a more general masking scheme than the one studied in [37]. In fact, we utilize the same notion of the numerical degree and a similar coding-theoretic approach as in [37], and also the same leakage assessment metrics like SNR and MI. However, generalizing [37] to this work is not trivial at all, we show hereafter the technical differences from [37].

We first highlight the different constructions of the generator matrices $\mathbf{G}$ and $\mathbf{H}$ in Tab. 5.1 for the codes $\mathcal{C}$ and $\mathcal{D}$, respectively. Indeed, $\mathcal{C}$ and $\mathcal{D}$ are not complementary in GCM, while they are complementary in IPM. In this respect, we show that Eqn. 5.7 is simplified as Eqn. 5.8 when $\mathcal{C}$ and $\mathcal{D}$ are complementary, thus we recover the main results in [37] (see Remark 5.3). As

a special case, the framework proposed in [37] is applicable when $\mathcal{C}$ and $\mathcal{D}$ are complementary, e.g., when $n = t + 1$ in SSS-based masking.

Moreover, we prove that GCM requires introducing a more general parameter B'_d (see Def. 5.2), which is a novel parameter for linear codes. Particularly, in [37] the parameter B_d only depends on $\mathcal{D}$. While B'_d depends on both $\mathcal{C}$ and $\mathcal{D}$, which indicates the importance of selecting appropriate candidates for both of them in practice. We also provide efficient magma scripts to evaluate this quantity [34].

Finally, we insist that the generalization in this work is a significant improvement that works for all GCMs. Since firstly, we show in Remark 5.4 that the security order can be greater than the dual distance minus one in GCM, which cannot be explained by the framework in [37], but can be explained perfectly by this work in a quantitative manner. Secondly, the redundancies in GCM allow detecting faults (e.g., for glitch-free designs [131]), which is currently an active research topic. We leave open the question on the construction of coding-theoretic countermeasures against both side-channel and fault injection attacks for future investigation.

6.4.2 Connections with [51]

The SSS-based masking is also the topic of a recent work [51], in which Costes et al. showed that the Boolean masking is a special case of SSS-based masking when $n \leq 6$. More interestingly, their simulation-based multivariate attacks [18] confirm our mathematical derivations, in particular, the information-theoretic evaluation in Fig. 6.1.

More generally, this work provides a unified framework for quantifying information leakage of all GCM instances. As a straightforward application, Theorems 5.2 and 5.3 in this chapter enable us to explain the empirical observations in practical attacks. For instance, the three codes for $(3, 1)$-SSS in Fig. 3 of [51] correspond to different $d^{\perp}_{\mathcal{D}}$ and/or $B_{d^{\perp}_{\mathcal{D}}}$. However, we stress that the three codes for $(2, 1)$-SSS in the same figure are not equivalent to each other but have the same $d^{\perp}_{\mathcal{D}}$ equal to 4 and closely distributed $B_{d^{\perp}_{\mathcal{D}}} \in \{11, 8, 8\}$. Moreover, this work presents a systematic way to select optimal codes for SSS-based masking and GCM, which is out of the scope of [51].

6.4.3 Efficient Implementations of GCM

In this chapter, we optimize security without touching the performances of GCM (there is no tradeoff between security and performance). Our coding-theoretic approach shows that both SNR and MI security metrics concur that dual distance and adjusted coefficient in weight

enumerator are the two drivers for security improvements. Essentially, we stick to the definition of GCM (recall the rightmost column in Tab. 5.1), and propose an effective way to tune the underlying codes.

In terms of performances, they are the same (with respect to memory and speed) as the generic GCM. A more detailed study could consist in attempting to represent the generator matrices $\mathbf{G}$ and $\mathbf{H}$ as compactly as possible (with as many zeros and ones in coefficients as possible, or with a specific structure, say "cyclic" for instance). Besides, Wang et al. [164] showed a complementary way to improve the overall performance of GCM implementations by an amortization technique. Both approaches would ease an efficient implementation of GCM, leaving an open problem for future study.

6.4.4 Further Application to Low Entropy Masking Schemes

Compared with the high cost of masking schemes, lower entropy masking scheme (LEMS) [78, 114, 170] provides a practical approach to reduce both randomness and implementation costs by only taking a small set of random masks. As a specific example, rotating S-Box masking (RSM) [12, 44, 110, 114] takes only 16 random masks which are elaborately chosen to achieve maximal protection. RSM is also the core protection used in DPA Contest v4.1 & v4.2 [154] for masked AES implementations.

In fact, RSM shall be represented in the form of code-based masking. Specifically, the two codes $\mathcal{C}$ and $\mathcal{D}$ in RSM are complementary. Let $X \in \mathbb{F}_2^8$ be the sensitive variable, then

$$Z = (X + YH,\, Y) \tag{6.6}$$

be the encoding in RSM [12] where $Z \in \mathbb{F}_2^{12}$ be the encoded variable, $Y \in \mathbb{F}_2^4$ is the mask and H is a 4×8 matrix with coefficients in $\mathbb{F}_2$. Indeed, $\mathcal{C}$ is spawn by the 8×12 matrix $\mathbf{G} = (I_8, 0)$ and $\mathcal{D}$ is spawn by the 4×12 matrix $\mathbf{H} = (H, I_4)$ where $H \in \mathbb{F}_2^{4\times 8}$ is the generator matrix of the code $[8, 4, 4]$ (which is known to be optimal and unique [24]). Therefore, we have

$$\begin{aligned} \mathbf{G} &= (I_8 \;\; 0_{8\times 4}), \\ \mathbf{H} &= (H \;\; I_4) = \begin{pmatrix} 1\,0\,0\,0 & 1\,1\,1\,0 & 1\,0\,0\,0 \\ 0\,1\,0\,0 & 1\,1\,0\,1 & 0\,1\,0\,0 \\ 0\,0\,1\,0 & 1\,0\,1\,1 & 0\,0\,1\,0 \\ 0\,0\,0\,1 & 0\,1\,1\,1 & 0\,0\,0\,1 \end{pmatrix}. \end{aligned} \tag{6.7}$$

As a consequence, the dual distance of $\mathcal{D}$ is $d_{\mathcal{D}}^{\perp} = 2$. Indeed, the dual code $\mathcal{D}^{\perp}$ has parameters $[12, 8, 2]$ and its weight distribution is: $[(0, 1), \mathbf{(2, 4)}, (4, 70), (6, 108), (8, 65), (10, 8)]$. Accordingly,

the security order of RSM with this $[8, 4, 4]$ code can only achieve a second-order side-channel resistance (under bit-probing model).

However, the above choice of the code $\mathcal{D}$ [12] is not optimal in the sense of $d_{\mathcal{D}}^{\perp}$. For instance, it can be improved by using the optimal code $[12, 4, 6]$ and its dual code has parameters $[12, 8, 3]$. The generator matrix of $\mathcal{D}$ is then as:

$$\mathbf{H} = (H' \quad I_4) = \begin{pmatrix} 0\,1\,1\,0 & 1\,1\,1\,0 & 1\,0\,0\,0 \\ 1\,0\,1\,0 & 1\,1\,0\,1 & 0\,1\,0\,0 \\ 1\,1\,1\,1 & 1\,0\,1\,1 & 0\,0\,1\,0 \\ 0\,0\,1\,1 & 0\,1\,1\,1 & 0\,0\,0\,1 \end{pmatrix}. \tag{6.8}$$

Note that the coordinates of $\mathbf{H}$ is permuted to have a systematic view. Accordingly, the weight distribution of the dual code $\mathcal{D}^{\perp}$ is: $[(0, 1), \mathbf{(3, 16)}, (4, 39), (5, 48), (6, 48), (7, 48), (8, 39), (9, 16), (12, 1)]$, which achieves a third-order side-channel resistance.

6.5 Conclusions and Perspectives

In this chapter, we investigate the side-channel resistance of redundant code-based masking and applications of our theoretical derivatives in SSS-based masking. In particular, we highlight the impact of the adjusted kissing number $B'_{d_{\mathcal{D}}^{\perp}}$ that depends on both codes $\mathcal{C}$ and $\mathcal{D}$. As applications, we present optimal codes for $(3, 1)$-SSS based masking as a first-order protection, and $(5, 2)$-SSS based masking as a second-order protection, respectively (over both $\mathbb{F}_{2^4}$ and $\mathbb{F}_{2^8}$).

However, the construction of optimal codes for a large number of shares is still an open problem. We launched an exhaustive study on $(3, 1)$-SSS based masking and presented some results on (5,2)-SSS in [34]. But the exhaustive enumeration would be computationally infeasible when n gets larger (e.g., $n > 8$) in SSS-based masking or, more generally, in GCM. A heuristic solution is to construct new (sub-) optimal codes by concatenating two optimal or sub-optimal codes, following a gradient descent idea. Alternatively, constructing the (sub-)optimal codes by an algebraic approach under certain constraints is a promising solution. We will explore both solutions for GCM in the future.

Part IV

Masked Cryptographic Implementations: Attacks and Information-Theoretic Bounds

CHAPTER 7

Optimal Attacks in the Presence of Code-based Masking

In this chapter, we present the results of higher-order optimal attacks against protected cryptographic implementations by the code-based masking. Part of work presented in this chapter is under submission (see [CGD21] in publication list).

Contents

7.1 Introduction

Side-channel analyses (SCAs) are among the most powerful attacks against cryptographic implementations. Since the seminal works [94, 95], a very large amount of SCAs have been proposed by exploiting various observable physical leakages in practice, like power consumption [46, 95], the electro-magnetic emanations [71, 132], etc. In essential, SCAs attempt to extract the sensitive information from noisy measurements containing unintended emissions or leakages, where the measurements are correlated with internal states or behaviors of a cryptographic device.

Along with a large body of attacks, numerous countermeasures have been proposed to protect practical implementations against SCAs. Relying on different strategies and principles, two major lines of countermeasures are hiding and masking [102]. Specifically, the hiding approach attempts to balance the leakage of different key-dependent operations or data, resulting in less informative signals in side-channel measurements [32, 156]. In contrast, the masking approach randomizes the internal states by splitting internal sensitive variables into several shares, which breaks the straightforward connection between the sensitive variables and the measurements. In particular, the latter is more preferable since it is featured with the provable security rather than engineering intuitions of designers. However, protecting cryptographic implementations against SCAs is usually not trivial and expensive in the sense of implementation cost [15, 30, 79, 128]. Furthermore, many proposals of protection are devised under abstract assumptions like independence assumption [7, 103], sufficiently noisy condition [30, 128], etc., which are not always fulfilled in real scenarios [5, 48].

Therefore, the evaluation of side-channel security of an implementation, especially in the presence of protection, plays a significant role in understanding its concrete security level and verifying the correctness and effectiveness of certain protections. In the following, we focus on the evaluation of masked implementations.

7.1.1 Evaluation of Side-Channel Security

According to different leakage models and the abstraction level of cryptographic implementations, evaluation tools are classified into four categories.

Firstly, the conformance-based leakage detection utilizes conformance testing to check whether there are significant differences in side-channel measurements of different key-dependent variables and/or operations [47, 107, 112, 144, 152]. It is intended to answer the following question at a high abstraction level: *does the device under test leak side-channel information?* Those

statistical tools include Welch's t-test, χ^2-test, etc. However, the conformance-based leakage detection only provides qualitative result, which is usually independent of the exploitation of the leakages, e.g., to launch a successful key-recovery attack. Furthermore, it might be difficult to interpret the detection result when the conformance testing gives a negative answer indicating no significant leakage. Therefore, other evaluations are necessary to further verify the leakage.

Secondly, the proof-based evaluation intends to prove the side-channel resistance of a masked design under abstract models like the probing model [86] and related variants [60, 62, 126, 128]. Typically, under independence assumption and noise condition, several leakage models are equivalent with certain forms of constants [126] by providing formal security guarantees of the masked implementation. However, physical defaults like couplings, glitches, etc. usually contradict assumptions behind the probing model [5, 103]. As a consequence, it is recommended to launch the attacked-based quantitative evaluation.

Thirdly, the information-theoretic evaluation aims at measuring side-channel leakages by utilizing information-theoretic measures [151, 166]. The frequently used measures include Shannon mutual information, Kullback-Leibler divergence, conditional entropy, etc. In fact, this category of evaluation measures the full distribution of leakages and provides insights on how much information an adversary can obtain. In essential, it usually provides information-theoretic bounds on the probability of success for any side-channel distinguishers given a set side-channel measurements [41, 57]. It is worth mentioning that not all distribution-based leakages can not be exploited by side-channel distinguishers. For instance, correlation power analysis is a typical non-profiling attack and it exploits only a few orders of moments of side-channel leakage. However, one of the major difficulties of using information-theoretic evaluation is how to estimate the leakage distribution accurately when the number of measurements is not sufficiently enough.

Finally, the attack-based evaluation is at the core of side-channel security evaluation, which aims at assessing the probability of success of a specific side-channel distinguisher. Relying on large variety of side-channel distinguishers like correlation power analysis [16, 82], mutual information analysis [55, 74, 163], template attacks [31, 119], stochastic attacks [75, 143], higher-order optimal distinguisher [18, 84], etc, the attack-based evaluation provides more accurate assessment of leakage, which captures device-specific features of side-channel leakage. However, it is infeasible to exhaust all distinguishers to launch attack-based evaluation provided a limited resources and time.

Summing up, the conformance-based leakage detection only provides qualitative assessment of side-channel security while other three evaluations give different levels of quantitative assessment.

To a large extent, four evaluation approaches are complementary to each other in practical application, varying with different evaluation requirements and necessary expertise on launching evaluations.

In this chapter, we complete the evaluation of side-channel resistance of masked implementations protected by the code-based masking, which is complementary to information-theoretic evaluation in Chap. 4 and 5 for IPM and GCM, respectively.

7.1.2 Metrics in Attack-based Evaluation

Considering a key-recovery attack in SCA, the ultimate metric is the success rate indicating the probability of an adversary can succeed in recovering the secret key [151][1]. In particular, two interrelated problems in attack-based evaluation are, on one hand, how many side-channel measurements are needed for a successful attack? Or on the other hand, what is the probability of success given a certain number of measurements? Therefore, it is preferable to show how the success rate evolve when the number of measurements increases.

Moreover, another attack metric is the guessing entropy (GE) [151], which measures the average rank of the correct key among all candidates based on distinguishing scores after an attack. GE is complementary to success rate as it indicates how wrongly guessed keys behave before a successful attack, and it converges to 1 when the success rate goes to 100% stably.

Therefore, we use both success rate and guessing entropy in evaluating the exploitability of information leakages in the code-based masking.

7.2 Contributions

We complete the evaluation of side-channel resistance of code-based masking by attacking results. In particular, our contributions are as follows.

HOOD-based Evaluation of GCM. We provide an extensive evaluation on the side-channel resistance of the generalized code-based masking. The attacks are based on the higher-order optimal distinguisher as it is the best attack strategy following the Maximum-Likelihood principle. We investigate both IPM and SSS-based masking, since they are representatives of non-redundant and redundant masking schemes, respectively. We highlight that the side-channel resistance

[1]There are different order of success rates when considering an adversary can launch key enumeration [124, 162] as a post-processing technique after side-channel attacks. However, we focus on the first-order success rate by convention as it is more straightforward.

of GCM is highly related to coding-theoretic properties wherein the dual distance and the adjusted kissing number are good indicators as we show in Chap. 5 from an information-theoretic perspective. Therefore, we verify our framework on quantifying the information leakage in GCM by HOOD-based attacks.

Redundancy in Code-based Masking Decreases Side-Channel Resistance. We leverage on both information-theoretic and attack-based evaluations to illustrate that the redundancy in sharing can only decrease the side-channel resistance of the corresponding masking schemes. Compared to the state-of-the-arts, our HOOD-based results challenge the evaluation launched in [29], but are in accordance with the ones in [51]. In particular, the authors showed in [29] that exploiting leakages from more shares does not always lead to more efficient attacks, whereas we show the improvements using leakages from more shares. Moreover, compared to [51], we extend the state-of-the-art in two directions: 1) we show the best cases of the linear codes, that are recommended to use, and 2) we give the worst cases of the linear codes that are not recommend for practical applications.

Challenges on Practical Use of Probing Model. Consider an (n, t)-sharing in redundant code-based masking, e.g., in (n, t)-SSS based masking, a sensitive variable is split into n shares while any $t+1$ shares among n are need to recover the sensitive. As a consequence of redundancy in sharing, increasing n can only decrease the concrete side-channel security given a fixed t. However, different sharings with the same t possess the same side-channel security order under the probing model. Therefore, only preserving security orders in proof-based evaluation of redundant code-based masking is not sufficient: it always has to be completed with the information-theoretic or attack-based evaluations. To verify this, we consider $(2, 1)$ and $(3, 1)$-SSS based masking, and show that exploiting leakages from all three shares always leads to more efficient attacks than using two shares.

7.3 Side-channel Distinguishers

We first recall the side-channel distinguishers in unprotected scenarios (without masking, etc). Let $X \in \mathbb{K}$ be the sensitive variable which depends on the secrets in the cryptographic implementations. For instance, the sensitive variable is usually $X = S(T \oplus K)$, the output of Sbox given a plaintext (or ciphertext) T and a subkey K, e.g., in AES or PRESENT, then we may use $X(k)$ in order to indicating a specific key guess k in generating X.

Considering simulated measurements, we adopt the common scenario in which the intermediates leak in Hamming weight model with independent additive white Gaussian noise (AWGN). Therefore, we have $\mathcal{L}^j = w_H(X^j) + N^j$, $1 \leq j \leq q$ for q traces, where w_H denotes the Hamming weight function and $N^j \sim \mathcal{N}(0, \sigma^2)$ is Gaussian noise with standard deviation σ. The basic setting of side-channel analysis seen as a communication channel is illustrated in Fig. 7.1.

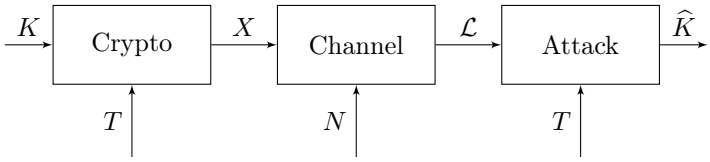


Figure 7.1: Side-channel seen as a communication channel.

7.3.1 Different Distinguishers

In SCA, a key-recovery attack intends to extract the secret key from q traces by exploiting certain side-channel distinguishers. In particular, a distinguisher takes maximization over all key hypothesis and gives the most possible candidate(s) by:

$$\widehat{k} = \underset{k \in \mathbb{K}}{\operatorname{argmax}}\, \widehat{\Delta}(k) = \underset{k \in \mathbb{K}}{\operatorname{argmax}}\, \widehat{\Delta}(\mathcal{L}, X(k)). \tag{7.1}$$

Formally, we define the side-channel distinguisher as follows.

Definition 7.1 (Side-Channel Distinguisher [80])**.** Given a set of side-channel measurements $\mathcal{L}$ and known cryptographic inputs (or outputs) T, a side-channel distinguisher returns a theoretical value

$$\Delta(k) = \Delta(\mathcal{L}, X(k)) \tag{7.2}$$

for any key guesses $k \in \mathbb{K}$ and the estimator $\widehat{\Delta}(k)$ converges to $\Delta(k)$ as $q \to \infty$, in the sense that the mean-squared error $\mathbb{E}\left[(\Delta(k) - \widehat{\Delta}(k))^2\right]$ approaches 0 when $q \to \infty$.

Note that we shall simplify $X(k)$ as X by implicitly indicating the link between the sensitive variable and the key hypothesis. In view of Def. 7.1, several classic side-channel distinguishers are presented as follows:

- Difference of Means (DoM): it is the original distinguisher proposed in the seminal work [95], known as Differential Power Analysis (DPA). Let $f_b(X)$ be the selection function which returns one specific bit of X, then we have

$$\begin{aligned} \Delta(k) &= |\mathbb{E}\left[\mathcal{L} | f_b(X) = 0\right] - \mathbb{E}\left[\mathcal{L} | f_b(X) = 1\right]|, \\ \widehat{\Delta}(k) &= \left| \frac{\sum_{j=1}^{q} (1 - f_b(X^j)) \mathcal{L}^j}{\sum_{j=1}^{q} (1 - f_b(X^j))} - \frac{\sum_{j=1}^{q} f_b(X^j) \mathcal{L}^j}{\sum_{j=1}^{q} f_b(X^j)} \right|, \end{aligned} \tag{7.3}$$

where the absolute value is always considered in maximization for each key hypothesis.

- Correlation Power Analysis (CPA) [16]: in which the distinguisher value is given by computing the Pearson correlation coefficient between the side-channel traces and the hypothetical leakages:

$$\begin{aligned}\Delta(k) &= |\rho(\mathcal{L}, f(X))| = \frac{|\mathrm{Cov}(\mathcal{L}, f(X))|}{\sigma_{\mathcal{L}}\sigma_{f(X)}} = \frac{|\mathbb{E}[\mathcal{L}f(X)] - \mathbb{E}[\mathcal{L}]\,\mathbb{E}[f(X)]|}{\sigma_{\mathcal{L}}\sigma_{f(X)}},\\ \widehat{\Delta}(k) &= \frac{\left|\frac{1}{q}\sum_{j=1}^{q}\mathcal{L}^j f(X^j) - \frac{1}{q}\sum_{j=1}^{q}\mathcal{L}^j \cdot \frac{1}{q}\sum_{j=1}^{q} f(X^j)\right|}{\sqrt{\frac{1}{q}\sum_{j=1}^{q}(\mathcal{L}^j)^2 - (\frac{1}{q}\sum_{j=1}^{q}\mathcal{L}^j)^2}\sqrt{\frac{1}{q}\sum_{j=1}^{q}(f(X^j))^2 - (\frac{1}{q}\sum_{j=1}^{q} f(X^j))^2}},\end{aligned} \tag{7.4}$$

  where $f(\cdot)$ denotes the leakage function, e.g., in the Hamming weight leakage model $f(X) = w_H(X)$. The absolute value is taken for each key hypothesis.

- Mutual Information Analysis (MIA) [74, 163]: the mutual information is used as a metric for assessing the dependency between the side-channel traces and the hypothetical leakages in an information-theoretic sense:

$$\begin{aligned}\Delta(k) &= I(\mathcal{L}, X) = H(L) - H(L|X),\\ \widehat{\Delta}(k) &= \sum_{l}\sum_{x}\widehat{\Pr}(\mathcal{L} = l, X = x)\log_2\frac{\widehat{\Pr}(\mathcal{L} = l, X = x)}{\widehat{\Pr}(\mathcal{L} = l)\widehat{\Pr}(X = x)},\end{aligned} \tag{7.5}$$

- Maximum Likelihood (ML)-based attack [31, 84]: when the leakage distribution is known, the optimal strategy for launching such attack is to use the maximum likelihood (ML) approach:

$$\begin{aligned}\Delta(k) &= \Pr(\mathcal{L}|X(k)),\\ \widehat{\Delta}(k) &= \widehat{\Pr}(\mathcal{L}, |X(k)) = \prod_{j=1}^{p}\widehat{\Pr}(\mathcal{L}^j, |X^j(k)),\end{aligned} \tag{7.6}$$

  where side-channel measurements are assumed to be i.i.d. Therefore, the best key guess is made by:

$$\widehat{k} = \underset{k\in\mathbb{K}}{\mathrm{argmax}}\,\widehat{\Delta}(k). \tag{7.7}$$

  Note that the ML rule is equivalent to Maximum A Posterior (MAP) rule with equiprobable keys. It is the case as commonly assumed that K is uniformly distributed over $\mathbb{K}$.

Given a side-channel distinguisher, a primary question arises: whether the attack utilizing the distinguisher will be succeed eventually? Therefore, we define the soundness of a distinguisher as follows.

Definition 7.2 (Soundness of a Distinguisher [80, 151])**.** A side-channel distinguisher $\widehat{\Delta(k)}$ is said to be sound if the theoretical distinguisher value is maximized at the correct key hypothesis, namely,

$$\Delta(k^*) > \Delta(k) \qquad \text{for any } k \neq k^*. \tag{7.8}$$

Apparently, if a distinguisher is sound, the attack tends to succeed with success rate equal to 100% eventually given enough number of traces (e.g., when $q \to \infty$).

Remark 7.1. For above classic distinguishers, CPA is sound [81], so as the DoM, since DoM can be seen as a special case of CPA [80] when $q \to \infty$. MIA is also proved to be sound under Gaussian noise [111, 127]. Moreover, ML-based distinguishers are sound by design, where the correct key guess will rank the first given enough amount of side-channel measurements.

7.3.2 Optimal Distinguisher in the Presence of Masking

We focus on code-based masking, which generalizes several existing masking schemes. The communication view in the presence of masking is depicted in Fig. 7.2. Let $X \in \mathbb{K}$ and $Y \in \mathbb{K}^t$ be respectively the sensitive variable and t random masks. Then the sharing in GCM writes:

$$Z = X\mathbf{G} + Y\mathbf{H} \in \mathbb{K}^n, \tag{7.9}$$

given that $t + 1 \leq n$, where $\mathbf{G}$ and $\mathbf{H}$ are generator matrices of two codes $\mathcal{C}$ and $\mathcal{D}$, respectively. For the sake of simplicity, we concentrate on scenarios in which the sensitive variable is a scalar. As assumed previously, the sensitive variable is $X = S(T \oplus K)$.

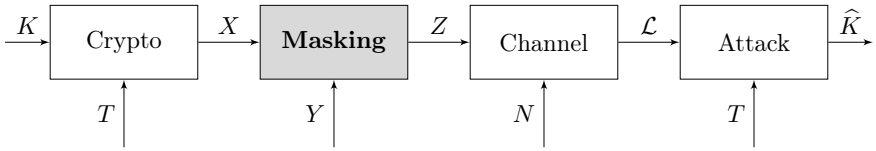


Figure 7.2: Side-channel seen as a communication channel in the presence of masking.

Regarding the simulated measurements, we utilize the Hamming weight model with independent AWGN. For each share Z_i, we have $\mathcal{L}_i = w_H(Z_i) + N_i$, $1 \leq i \leq n$ for n shares and $N_i \sim \mathcal{N}(0, \sigma^2)$ is Gaussian noise with standard deviation σ. Given a dataset of q traces, we further denote all traces as $\mathcal{L} = (\mathcal{L}_i^j)$ for $1 \leq i \leq n$ and $1 \leq j \leq q$.

In our scenario, as the leakage model is assumed to be known, the best strategy for performing key-recovery attacks is to utilize the ML-based approach. Following the principle of ML-based attack, the higher-order optimal distinguisher (HOOD) is known as follows.

Lemma 7.1 (Higher-Order Optimal Distinguisher [18]). *Given a set of q measurements $\mathcal{L} = (\mathcal{L}_i^j) = f(Z_i^j) + N_i^j$ for $1 \leq i \leq n$ and $1 \leq j \leq q$ such that N_i^j are i.i.d. across $1 \leq j \leq q$ and independent across $1 \leq i \leq n$. When the leakage distribution is known (both the leakage function and the noise distribution), the d-th order optimal distinguisher is:*

$$\Delta(k) = \prod_{j=1}^{q} \sum_{y \in \mathbb{K}^t} \Pr(Y = y) \prod_{i=1}^{d} \Pr(\mathcal{L}_i^j | Z_i^j), \tag{7.10}$$

where the calculation of Z_i^j implicitly involves $Y = y$. Therefore, the key hypothesis is given by

$$\widehat{k} = \underset{k \in \mathbb{K}}{\operatorname{argmax}}\, \Delta(k). \tag{7.11}$$

In the sequel, we focus on attack-based evaluation of the generalized code-based masking, particularly we target IPM with $n = 2$ and $n = 3$, and $(3, 1)$-SSS based masking.

7.4 Attacks against Non-Redundant Code-based Masking

Considering IPM as an instance of non-redundant code-based making, the generating matrices of $\mathcal{C}$ and $\mathcal{D}$ are:

$$\begin{aligned} \mathbf{G} &= \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \end{pmatrix} \\ \mathbf{H} &= \begin{pmatrix} \alpha_1 & 1 & 0 & \cdots & 0 \\ \alpha_2 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \alpha_t & 0 & 0 & \cdots & 1 \end{pmatrix} \end{aligned} \tag{7.12}$$

where $n = t + 1$ and $\alpha_i \in \mathbb{K} \backslash \{0\}$ for $1 \leq i \leq t$. In particular, by taking $\alpha_i = 1$ for $1 \leq i \leq t$ recovers the Boolean masking. As a result, the generator matrix of $\mathcal{D}^\perp$ is: $\mathbf{H}^\perp = (1 \ \alpha_1 \ \alpha_2 \ \cdots \ \alpha_t)$ with $d_{\mathcal{D}}^\perp = t + 1$, indicating that IPM with n shares has a security order equal to $n - 1$ under word-probing model [3, 123]. We denote $\alpha = (1, \alpha_1, \ldots, \alpha_t)$ the public parameters in IPM.

7.4.1 Optimal Distinguishers

Relying on Lemma 7.1, the HOOD is instantiated in the context of Hamming weight leakage with an AWGN as follows.

$$\begin{aligned} \Delta(k) &= \prod_{j=1}^{q} \sum_{y \in \mathbb{K}^{n-1}} \Pr(Y = y) \prod_{i=1}^{d} \Pr(\mathcal{L}_i^j | z_i^j) \\ &= \prod_{j=1}^{q} \sum_{y \in \mathbb{K}^{n-1}} \Pr(Y = y) \prod_{i=1}^{d} \mathcal{N}(\mathcal{L}_i^j | w_H(z_i^j), \sigma^2). \end{aligned} \tag{7.13}$$

Since Y is uniformly distributed ($\Pr(Y = y) = \frac{1}{|\mathbb{K}^{n-1}|}$) required by a sound masking scheme, it is independent of each key hypothesis and hence has no impact on $\Delta(k)$. Taking logarithms further eases the numerical computations (avoiding float overflows), the HOOD is equivalent to the following distinguisher score [29, 51]:

$$S(k) = \sum_{j=1}^{q} \log \sum_{y \in \mathbb{K}^{n-1}} \prod_{i=1}^{d} \mathcal{N}(\mathcal{L}_i^j | w_H(z_i^j), \sigma^2), \tag{7.14}$$

then the key guess is determined by maximizing $S(k)$.

Formally, thanks to masking, an adversary cannot obtain anything about the sensitive variable if the order d of a HOOD is not greater than the security order t. A prerequisite for launching a successful attack is $d > t$ in our scenario when targeting IPM, which is consistent with coding-theoretic conditions.

7.4.2 IPM with $n = 2$

Taking $n = 2$ gives $t = 1$, resulting that only one parameter in IPM is α_1 and $\mathbf{H} = (\alpha_1\ 1)$. There are 255 candidates for α_1 as it cannot be zero. In order to facilitate practical applications and fair comparison with the state-of-the-art, we aim at the irreducible polynomial $g(\mathsf{X}) = \mathsf{X}^8 + \mathsf{X}^4 + \mathsf{X}^3 + \mathsf{X} + 1$ that is used in AES to generate the field $\mathbb{K} = \mathbb{F}_{2^8}$.

As shown in Chaps. 4, the two coding-theoretic properties that indicate the side-channel resistance of IPM are the dual distance $d_{\mathcal{D}}^{\perp}$ and the kissing number $B_{d_{\mathcal{D}}^{\perp}}$. Herein, we first investigate the statistical properties of $d_{\mathcal{D}}^{\perp}$ and $B_{d_{\mathcal{D}}^{\perp}}$ among all linear code candidates. The distribution of $d_{\mathcal{D}}^{\perp}$ are enumerated in Tab. 7.1 and the corresponding choices of the codes with given $d_{\mathcal{D}}^{\perp}$ and $B_{d_{\mathcal{D}}^{\perp}}$ are in Tab. 7.2, while in the latter we are only interested in the linear codes with the maximal and minimal values of $B_{d_{\mathcal{D}}^{\perp}}$ for each $d_{\mathcal{D}}^{\perp}$.

Table 7.1: Distribution of $d_{\mathcal{D}}^{\perp}$ for IPM with $n = 2$.

$d_{\mathcal{D}}^{\perp} = d$	$\lvert\{\alpha_1\}\rvert$	$\max\{B_d\}$	$\min\{B_d\}$
$d = 2$	35	8	1
$d = 3$	146	6	1
$d = 4$	74	17	4

As shown in Tab. 7.2, there are only 12 optimal linear codes which maximize $d_{\mathcal{D}}^{\perp}$ and minimize $B_{d_{\mathcal{D}}^{\perp}}$ at the same time.

[1] We use $\overset{\text{eqv}}{\approx}$ to denote that two linear codes with the given parameters are equivalent over $\mathbb{F}_{2^\ell}$ or after the sub-field representation in $\mathbb{F}_2$, see [37, 42] for details.

Table 7.2: Choices of the codes for IPM with $n = 2$.

$d_{\mathcal{D}}^{\perp} = d$	B_d	$\lvert\{\alpha_1\}\rvert$	Candidates of α_1	Comments
$d = 2$	$B_d = 8$	1	$\{1\}$	Boolean masking
	$B_d = 1$	20	$\{16, 17, 34, 39, 60, 90, 115, 116, 119, 120, 133,$ $140, 180, 182, 201, 207, 215, 230, 234, 247\}$	
$d = 3$	$B_d = 6$	8	$\{3, 83, 101, 137, 158, 166, 202, 246\}$	
	$B_d = 1$	58	$\{14, 15, 19, 20, 40, 44, 48, 49, 52, 56, 61, 67$ $69, 75, 76, 80, 84, 94, 97, 99, 103, 112, 113, \ldots\}$	
$d = 4$	$B_d = 17$	2	$\{29, 64\}$	$29 \overset{\text{eqv}}{\approx} 64$ [1]
	$B_d = 4$	12	$\{23, 46, 51, 54, 81, 92, 95,$ $102, 108, 162, 165, 184\}$	12 optimal codes in total

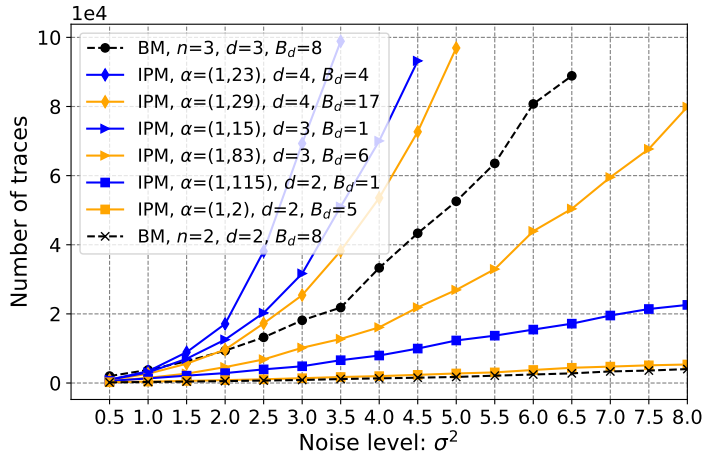
7.4.2.1 Experimental Results

As mentioned previously, the simulated traces are $\mathcal{L} = (\mathcal{L}_i^j)$ for $1 \leq i \leq n$ and $1 \leq j \leq q$ where $\mathcal{L}_i^j = w_H(Z_i^j) + N_i^j$ denotes the leakage of i-th share in j-th trace. The evaluation metric is the minimum number of traces achieving $P_s \geq 95\%$, which integrates the success rate along with different noise levels.

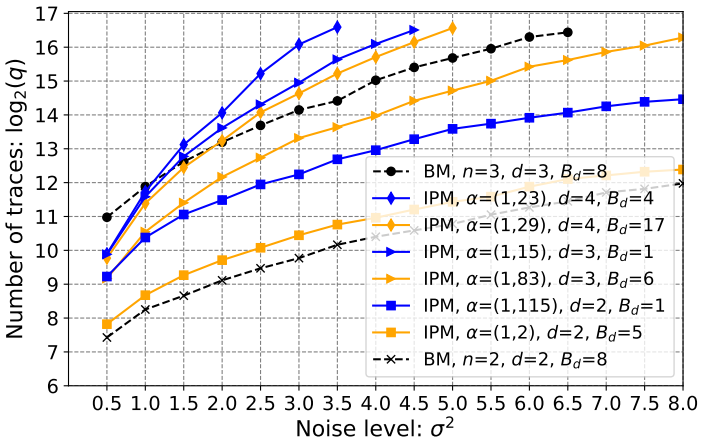
For linear codes of different $d_{\mathcal{D}}^{\perp}$ shown in Tab. 7.2, we choose both the minimum and the maximum of $B_{d_{\mathcal{D}}^{\perp}}$ excluding the Boolean one. The evaluation results of IPM with $n = 2$ are shown in Fig. 7.3 by using up to $q = 100,000$ traces. Moreover, we include Boolean masking (BM) with $n = 2$ and $n = 3$ shares in comparison.

The main takeaway point from Fig. 7.3 is that, IPM with the linear code of the maximized dual distance $d_{\mathcal{D}}^{\perp}$ and the minimized kissing number $B_{d_{\mathcal{D}}^{\perp}}$ indeed has the best achievable side-channel resistance. The attack-based evaluation also confirms: 1) all 2-share IPM are better than the first-order Boolean masking (with $n = 2$); 2) good choices of linear codes of 2-share IPM can even be better than the second-order Boolean masking (with $n = 3$) when the noise level is $\sigma^2 > 1.0$. The reason is that in IPM, the best cases of $d_{\mathcal{D}}^{\perp}$ is larger and $B_{d_{\mathcal{D}}^{\perp}}$ is smaller than that in the second-order Boolean masking, respectively; 3) it is also advantageous to adopt 2-share IPM rather than 3-share BM from a performance perspective. For instance, the clock cycles are 157,196 vs 160,357 as reported in [3] for an AES-128 implementations on an AVR architecture protected by the former and the latter, respectively.

Optimal Codes for 2-Share IPM. According to Tab. 7.2, there are only 12 optimal codes with the best coding-theoretic properties. For the sake of brevity, we present four cases of



(a) Number of traces in normal scale.

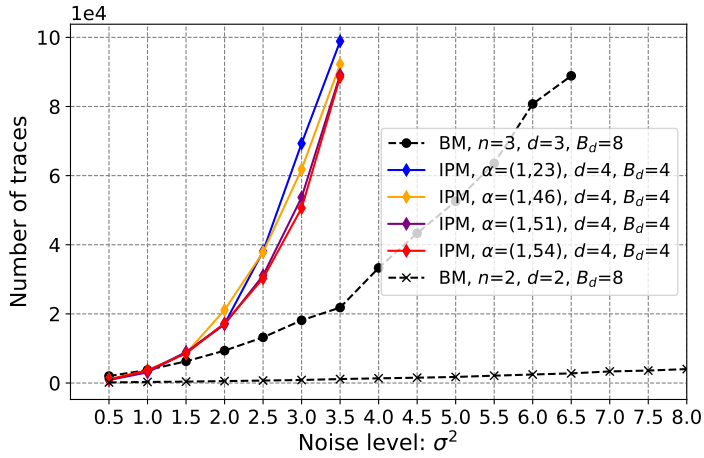


(b) Number of traces in $\log_2$ scale.

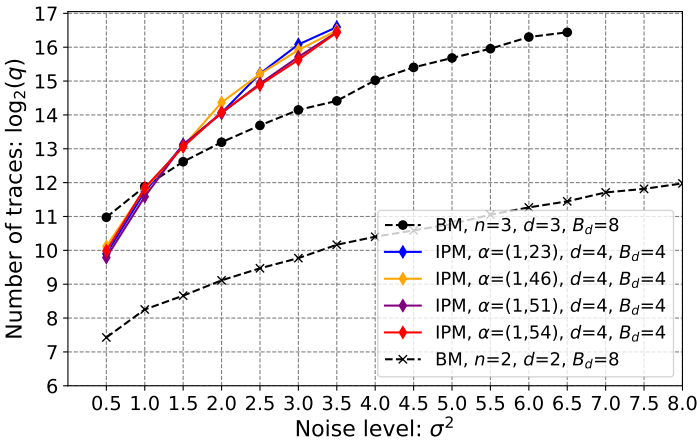
Figure 7.3: Attack-based evaluation of IPM with $n = 2$ shares. Taking two codes in each group with different $d^{\perp}_{\mathcal{D}}$ and/or $B_{d^{\perp}_{\mathcal{D}}}$.

optimal codes as in Fig. 7.4. The primary observation is that those four codes have similar side-channel resistance from an adversary perspective who launches HOOD-based attacks. Note that the fluctuations among those four curves are due to the nature of numerical simulation with certain random seeds. Overall, those four codes perform closely against HOOD-based attacks.

We assume that the leakage distribution is known when launching such attacks, this scenario allows a worst-case evaluation of side-channel resistance of IPM. However, this assumption is usually too radical in practice, where the leakage properties of the device is unknown and the acquisition environment might be various. Furthermore, when other distinguishers are adopted in carrying out attacks, ML-based analysis also provides an upper bound on the success rate. In



(a) Number of traces in normal scale.



(b) Number of traces in $\log_2$ scale.

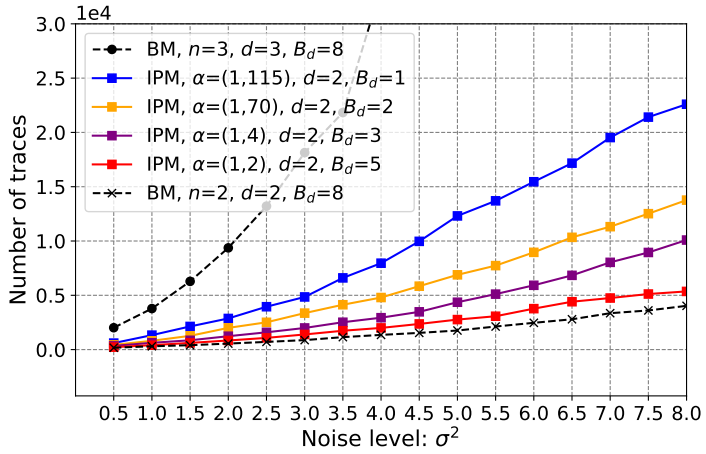
Figure 7.4: Comparison of four instances of optimal codes for 2-share IPM, according to the best coding-theoretic properties given by $d_{\mathcal{D}}^{\perp} = 4$ and $B_{d_{\mathcal{D}}^{\perp}} = 4$.

summary, our worst-case evaluation provides some insights on how successful can an attack be in practice and shows how to select optimal codes when applying IPM.

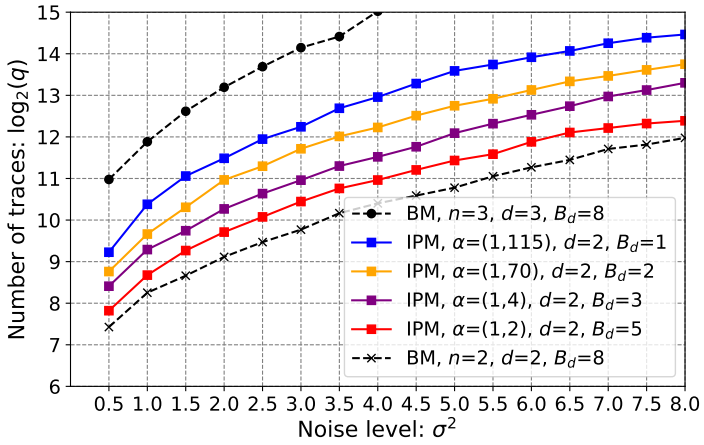
The Impact of $B_{d_{\mathcal{D}}^{\perp}}$. We have showed how to select optimal codes according to both $d_{\mathcal{D}}^{\perp}$ and $B_{d_{\mathcal{D}}^{\perp}}$, yet the solo role of $B_{d_{\mathcal{D}}^{\perp}}$ is not explicitly investigated. In the following, we compare several instances of the linear code in IPM with the same $d_{\mathcal{D}}^{\perp}$ while different $B_{d_{\mathcal{D}}^{\perp}}$.

As shown in Fig. 7.5, we set $d_{\mathcal{D}}^{\perp} = 2$ and $B_{d_{\mathcal{D}}^{\perp}} \in \{1, 2, 3, 5, 8\}$ where 2-share BM being a special case of IPM has $B_{d_{\mathcal{D}}^{\perp}} = 8$. Apparently, reducing $B_{d_{\mathcal{D}}^{\perp}}$ leads to a more difficult attack in the sense of the necessary number of traces to launch a successful attack. In addition, since we

choose $d_{\mathcal{D}}^{\perp} = 2$, all those codes will not outperform 3-share BM in full range of noise levels.



(a) Number of traces in normal scale.



(b) Number of traces in $\log_2$ scale.

Figure 7.5: Illustrating the impact of $B_{d_{\mathcal{D}}^{\perp}}$ given the same $d_{\mathcal{D}}^{\perp}$ in 2-share IPM.

Summing up, we demonstrate how $B_{d_{\mathcal{D}}^{\perp}}$ plays a significant role in indicating the side-channel resistance of IPM. More generally, it is integrated with the dual distance as indicators in evaluating side-channel security of non-redundant code-based maskings like DSM and LS, etc.

7.4.3 Linear Codes for IPM with $n = 3$

Herein we present the classification of the linear codes of 3-share IPM. Taking $n = 3$, resulting that $t = 2$ and two free parameters in $\alpha = (1, \alpha_1, \alpha_2)$ are $\alpha_1, \alpha_2 \in \mathbb{K}\backslash\{0\}$. There are $255 \times 255 = 65025$ candidates, where the number of candidates can be dramatically reduced by considering the

equivalence of the linear codes [35]. Therefore, we take $\alpha_1 \leq \alpha_2$, which reduces the number of the codes to 32640.

The distribution of $d_{\mathcal{D}}^{\perp}$ are enumerated in Tab. 7.3 and the choices of the codes under given $d_{\mathcal{D}}^{\perp}$ and $B_{d_{\mathcal{D}}^{\perp}}$ are in Tab. 7.4. Note again that in Tab. 7.4 we only focus on the maximal and minimal values of $B_{d_{\mathcal{D}}^{\perp}}$.

Table 7.3: Distribution of $d_{\mathcal{D}}^{\perp}$ for IPM with $n = 3$.

$d_{\mathcal{D}}^{\perp} = d$	$\lvert\{(\alpha_1, \alpha_2)\}\rvert$	$\max\{B_d\}$	$\min\{B_d\}$
$d = 3$	207	8	1
$d = 4$	1730	6	1
$d = 5$	7242	7	1
$d = 6$	15304	13	1
$d = 7$	7929	12	1
$d = 8$	228	20	6

As shown in Tab. 7.4, there are only 3 optimal codes which maximize $d_{\mathcal{D}}^{\perp}$ and minimize $B_{d_{\mathcal{D}}^{\perp}}$ at the same time. Since the maximized dual distance is $d_{\mathcal{D}}^{\perp} = 8$, IPM with those optimal codes should be comparable with the eighth-order BM, namely $n = 8$ given a certain level noise. In particular, considering the security order in the bit-probing model, the former and the latter share the same security order $t_b = d_{\mathcal{D}}^{\perp} - 1 = 7$. Therefore, it is recommended to apply IPM rather than BM with many more shares since as a rule of thumb, the implementation cost increases at least quadratically with n.

7.5 Attacks on Redundant Code-based Masking

In the sequel, we investigate the HOOD-based evaluation on the polynomial masking [77, 131], where in its central is Shamir's Secret Sharing (SSS) scheme. Taking SSS-based masking as an example of redundant code-based masking, the parameters are denoted as $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_n)$ and the condition for α_i is that $\alpha_i \neq \alpha_j$ for any $i \neq j$. Then we have the following generator matrices for the codes $\mathcal{C}$ and $\mathcal{D}$, respectively,

$$\mathbf{G} = \begin{pmatrix} 1 & 1 & \cdots & 1 \end{pmatrix}$$
$$\mathbf{H} = \begin{pmatrix} \alpha_1^1 & \alpha_2^1 & \cdots & \alpha_n^1 \\ \alpha_1^2 & \alpha_2^2 & \cdots & \alpha_n^2 \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_1^t & \alpha_2^t & \cdots & \alpha_n^t \end{pmatrix} \tag{7.15}$$

Table 7.4: Choices of the codes for IPM with $n = 3$.

$d^{\perp}_{\mathcal{D}} = d$	B_d	$\lvert\{(\alpha_1, \alpha_2)\}\rvert$	Candidates of (α_1, α_2)	Comments
$d = 3$	$B_d = 8$	1	$\{(1, 1)\}$	Boolean masking
	$B_d = 1$	151	$\{(1, 16), (1, 17), (1, 34), (1, 39), (1, 60), (1, 90), (1, 115), (1, 116), \ldots\}$	
$d = 4$	$B_d = 6$	3	$\{(2, 3), (140, 141), (246, 247)\}$	
	$B_d = 1$	1227	$\{(1, 14), (1, 18), (1, 19), (1, 20), (1, 21), (1, 30), (1, 41), (1, 42), \ldots\}$	
$d = 5$	$B_d = 7$	8	$\{(1, 176), (2, 164), (5, 143), (8, 64), (8, 232), (12, 12), (29, 232), (82, 141)\}$	
	$B_d = 1$	4586	$\{(1, 23), (1, 31), (1, 46), (1, 47), (1, 75), (1, 77), (1, 98), (1, 107), \ldots\}$	
$d = 6$	$B_d = 13$	2	$\{(1, 130), (127, 127)\}$	
	$B_d = 1$	7050	$\{(2, 184), (3, 45), (3, 46), (3, 47), (3, 59), (3, 65), (3, 77), (3, 81), \ldots\}$	
$d = 7$	$B_d = 12$	3	$\{(16, 185), (56, 142), (116, 242)\}$	
	$B_d = 1$	645	$\{(3, 53), (7, 45), (7, 49), (7, 77), (7, 99), (7, 106), (7, 107), (9, 154), \ldots\}$	
$d = 8$	$B_d = 20$	3	$\{(94, 109), (97, 124), (147, 161)\}$	
	$B_d = 6$	3	$\{(27, 196), (91, 204), (218, 240)\}$	Only three cases are optimal.

where α_i for $1 \leq i \leq n$ are also called public points in SSS-based masking. The corresponding scheme is also denoted as (n, t)-SSS based masking.

From a coding-theoretic perspective, the SSS scheme is connected to the Reed-Solomon (RS) code. Given the two generator matrices as in Eqn. 7.15, the rank of $\mathbf{H}$ equals t, so the dual distance of $\mathcal{D}$ is $t + 1$ [35]. Accordingly, the side-channel security order in the word-probing model is $t_w = t$.

7.5.1 Optimal Distinguishers

Recall the form of $\mathbf{H}$ in Eqn. 7.15 that, there are n public points to be determined in SSS-based masking. However, the masking itself is in the t-th order.

Similarly as in IPM, the optimal distinguisher is determined by applying the ML rule.

Considering the same assumption on leakage distribution, we have:

$$\begin{aligned}\Delta(k) &= \prod_{j=1}^{q} \sum_{y \in \mathbb{K}^t} \Pr(Y = y) \prod_{i=1}^{d} \Pr(\mathcal{L}_i^j | z_i^j) \\ &= \prod_{j=1}^{q} \sum_{y \in \mathbb{K}^t} \Pr(Y = y) \prod_{i=1}^{d} \mathcal{N}(\mathcal{L}_i^j | w_H(z_i^j), \sigma^2).\end{aligned} \tag{7.16}$$

Taking logarithms to ease the numerical computations, the HOOD is therefore equivalent to the following distinguisher score [51]:

$$S(k) = \sum_{j=1}^{q} \log \sum_{y \in \mathbb{K}^t} \prod_{i=1}^{d} \mathcal{N}(\mathcal{L}_i^j | w_H(z_i^j), \sigma^2). \tag{7.17}$$

Remark 7.2. Note that the distinguisher proposed in [29, Eqn. 13] is problematic, which leads to suspicious conclusions. In fact, the summation within the logarithm is over $y \in \mathbb{K}^t$ rather than over $y \in \mathbb{K}^{n-1}$ when $n > t + 1$, namely in redundant cases.

7.5.2 HOOD against $(3, 1)$-SSS based Masking

Considering $n = 3$ and $t = 1$, the generator matrices $\mathbf{G}$ and $\mathbf{H}$ are as follows.

$$\begin{aligned}\mathbf{G} &= \begin{pmatrix} 1 & 1 & 1 \end{pmatrix} \\ \mathbf{H} &= \begin{pmatrix} \alpha_1 & \alpha_2 & \alpha_3 \end{pmatrix},\end{aligned} \tag{7.18}$$

where α_i for $1 \leq i \leq 3$ are not equal to each other. We can fix $\alpha_1 = 1$ by utilizing the equivalence of the linear codes. Additionally, we set $\alpha_2 < \alpha_3$ as in [35] and resulting that there are 32131 candidates (instead of 2731135 codes for any pairwise different α_1, α_2 and α_3).

The distribution of $d_{\mathcal{D}}^{\perp}$ are exhausted in Tab. 7.5 and the choices of the codes under given $d_{\mathcal{D}}^{\perp}$ and $B_{d_{\mathcal{D}}^{\perp}}$ are in Tab. 7.6 in which we only focus on the maximal and minimal values of $B_{d_{\mathcal{D}}^{\perp}}$ as above.

Table 7.5: Distribution of $d_{\mathcal{D}}^{\perp}$ for $(3, 1)$-SSS based masking.

$d_{\mathcal{D}}^{\perp} = d$	$\lvert\{(\alpha_2, \alpha_3)\}\rvert$	$\max\{B_d\}$	$\min\{B_d\}$
$d = 2$	11460	13	1
$d = 3$	20581	19	1
$d = 4$	90	73	37

Remark 7.3. In SSS-based masking, we should use the adjusted kissing number $B'_{d_{\mathcal{D}}^{\perp}}$ instead of $B_{d_{\mathcal{D}}^{\perp}}$. Typically, we have $B'_{d_{\mathcal{D}}^{\perp}} \geq B_{d_{\mathcal{D}}^{\perp}}$ in SSS-based masking as pointed out in [35]. However, we use $B_{d_{\mathcal{D}}^{\perp}}$ here since it follows the same trend as $B'_{d_{\mathcal{D}}^{\perp}}$. Note that given a specific $\mathcal{C}$, different choices of $\mathcal{D}$ with the same $B_{d_{\mathcal{D}}^{\perp}}$ may lead to different $B'_{d_{\mathcal{D}}^{\perp}}$.

Table 7.6: Choices of the codes for $(3, 1)$-SSS.

$d_{\mathcal{D}}^{\perp} = d$	B_d	$\lvert\{(\alpha_2, \alpha_3)\}\rvert$	Candidates of (α_2, α_3)	Comments
$d = 2$	$B_d = 13$	3	$\{(2, 4), (2, 141), (141, 203)\}$	The worst cases
	$B_d = 1$	5976	$\{(3, 17), (3, 34), (3, 37), (3, 39), (3, 48), (3, 49), (3, 51), (5, 60), \ldots\}$	
$d = 3$	$B_d = 19$	3	$\{(6, 137), (71, 123), (105, 158)\}$	
	$B_d = 1$	435	$\{(7, 23), (7, 53), (7, 111), (7, 148), (7, 198), (11, 84), (11, 94), (11, 154), \ldots\}$	
$d = 4$	$B_d = 73$	3	$\{(29, 37), (64, 131), (77, 128)\}$	
	$B_d = 37$	3	$\{(51, 54), (102, 228), (108, 198)\}$	Only three cases are optimal

7.5.2.1 Experimental Results

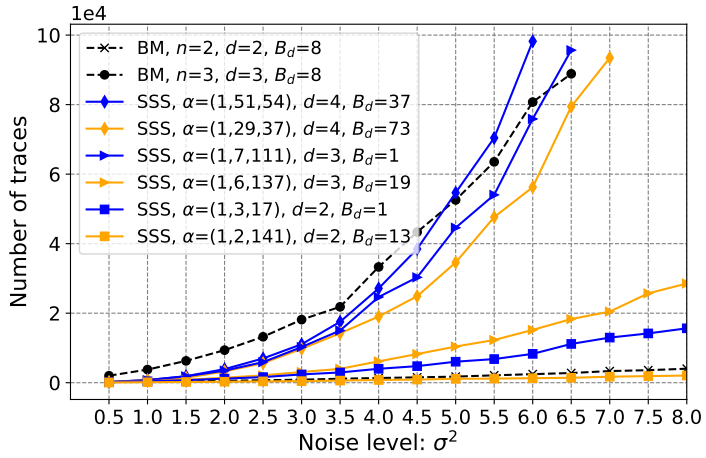
With the same setting as in evaluation of IPM, the simulated traces are $\mathcal{L} = (\mathcal{L}_i^j)$ for $1 \leq i \leq n$ and $1 \leq j \leq q$ where $\mathcal{L}_i^j = w_H(Z_i^j) + N_i^j$ denotes the leakage of i-th share in j-th trace.

For linear codes of different $d_{\mathcal{D}}^{\perp}$ shown in Tab. 7.6, we choose both the minimum and the maximum of $B_{d_{\mathcal{D}}^{\perp}}$. The evaluation results of $(3, 1)$-SSS based masking are shown in Fig. 7.6 by using up to $q = 100,000$ traces. Moreover, we include Boolean masking (BM) with $n = 2$ and $n = 3$ shares in comparison.

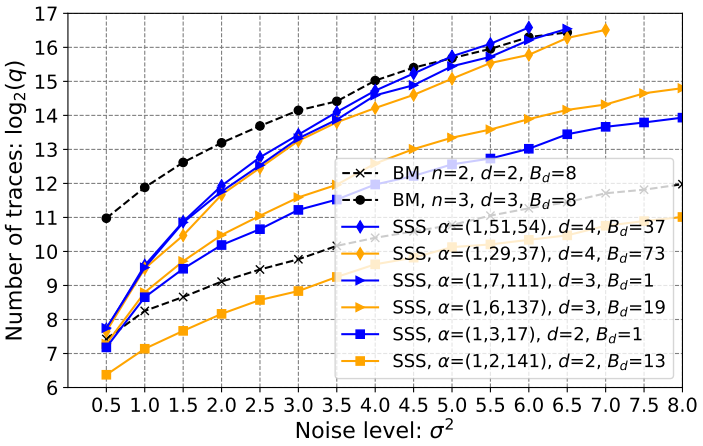
From Fig. 7.6, the most important takeaway point is that the public points in $(3, 1)$-SSS based masking make a significant difference in side-channel resistance of the corresponding masking scheme. Furthermore, we can observe that: 1) with dedicated selection of good linear codes, the side-channel resistance of the scheme can be improved significantly; 2) comparing with the attack-based evaluation on 2-share IPM, the side-channel security of $(3, 1)$-SSS based masking is degraded because of the redundancy, which is consistent with our information-theoretic evaluation in Chap. 6; 3) similarly as in 2-share IPM, the best codes can provide comparable security level as 3-share BM when the noise level is higher enough (e.g., $\sigma^2 \geq 5.0$); 4) for the first time, we show that with bad choices of the code, the security level of $(3, 1)$-SSS based masking can be continuously lower than 2-share BM.

In the sequel, we further leverage the last two points by providing more instances of the optimal and the worst codes for $(3, 1)$-SSS based masking, respectively.

Optimal Codes for $(3, 1)$-SSS based Masking. According to Tab. 7.6, there are only three cases of optimal codes. The evaluation results are depicted in Fig. 7.7. It is more obvious



(a) Number of traces in normal scale.



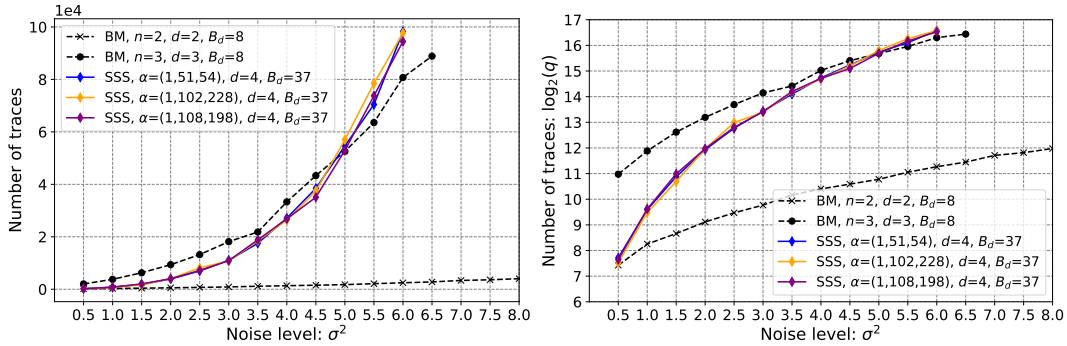
(b) Number of traces in $\log_2$ scale.

Figure 7.6: Attack-based evaluation of $(3, 1)$-SSS based masking. Taking two codes in each group with different $d_{\mathcal{D}}^{\perp}$ and/or $B_{d_{\mathcal{D}}^{\perp}}$.

in logarithmic view as in Fig. 7.7(b) that those three codes lead to very close side-channel resistance.

To sum up, the optimal choices of public points in SSS-based masking can significantly improve its side-channel resistance that is much higher than 2-share BM. In particular, those optimal codes can even provide comparable security as 3-share BM.

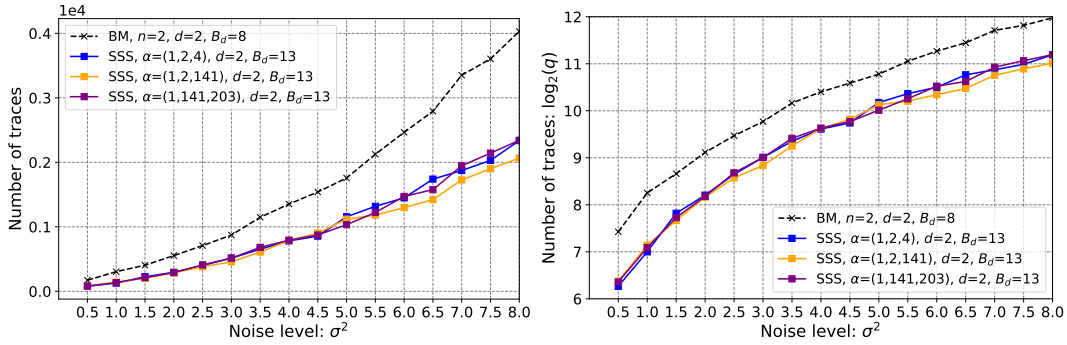
Worst Codes for $(3, 1)$-SSS based Masking. From Tab. 7.6, there are several classes of the linear codes that are worse than 2-share BM, including the three worst cases. The evaluation results are plotted in Fig. 7.8. Interestingly, those worst codes make the SSS-based masking



(a) Number of traces in normal scale. (b) Number of traces in $\log_2$ scale.

Figure 7.7: The optimal codes for $(3,1)$-SSS based masking, in which $d^{\perp}_{\mathcal{D}}$ is maximized and $B_{d^{\perp}_{\mathcal{D}}}$ is minimized given a specific $d^{\perp}_{\mathcal{D}} = 4$.

perform worse in full range of noise levels.



(a) Number of traces in normal scale. (b) Number of traces in $\log_2$ scale.

Figure 7.8: The worst codes for $(3,1)$-SSS based masking, where $d^{\perp}_{\mathcal{D}}$ is minimized and $B_{d^{\perp}_{\mathcal{D}}}$ is maximized given a specific $d^{\perp}_{\mathcal{D}} = 2$.

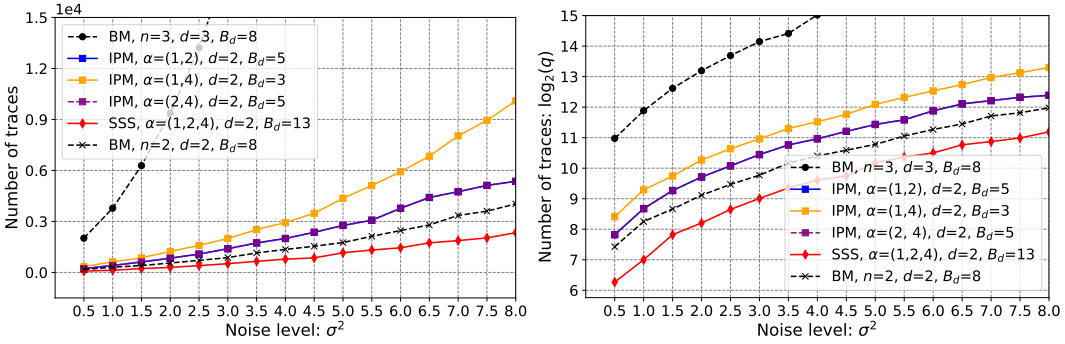
To the best of our knowledge, we identify, for the first time, the worst cases of public points in SSS-based masking or more generally in the context of secret sharing schemes, when each share leaks certain noisy information. Comparing with the state-of-the-art, our coding-theoretic approach not only provides the optimal cases, but also identifies the worst cases of the public points in SSS-based masking. Both of them are instructive in designing redundant code-based masking in protecting cryptographic implementations in practice.

7.6 Comparisons: How Redundancy Matters?

As shown in Chap. 6, the redundancy in code-based masking gives rise to more leakage from an information-theoretic sense when assessed by mutual information. However, more leakage detected by mutual information can not always be exploited by side-channel distinguishers. As a result, it is not clear how much impact can the redundancy have from an attacking perspective.

In this section, we demonstrate from an attack-based evaluation that, adding redundancy in code-based masking can only reduce the side-channel resistance of the corresponding masking scheme. To have a fair comparison, we consider two instances of $(3, 1)$-SSS based masking and reuse the parameters in 2-share IPM. Specifically, in $(3, 1)$-SSS based masking, the parameters are $\alpha = (1, \alpha_1, \alpha_2)$, while any 2-out-of-3 element in α gives an instance of IPM and there are three instances in total. Then those four instance of code-based masking are evaluated by HOOD-based attacks (e.g., refer to Eqn. 7.14 and Eqn. 7.17, respectively).

The first group of comparisons is shown in Fig. 7.9, where we have $\alpha = (1, 2, 4)$. The first observation is that adding one share of redundancy always reduces the concrete side-channel security of code-based masking. Secondly, given the same security order under the word-probing model, IPM always outperforms SSS-based masking. The more redundancy can only further reduce the security level. Interestingly, as three instances of IPM have the same security order under the bit-probing model, a major difference exists in $B_{d_{\mathcal{D}}^{\perp}}$. Put differently, given the same $d_{\mathcal{D}}^{\perp}$ over $\mathbb{F}_2$, more redundancy leads to a greater value of $B_{d_{\mathcal{D}}^{\perp}}$, indicating a lower security level.



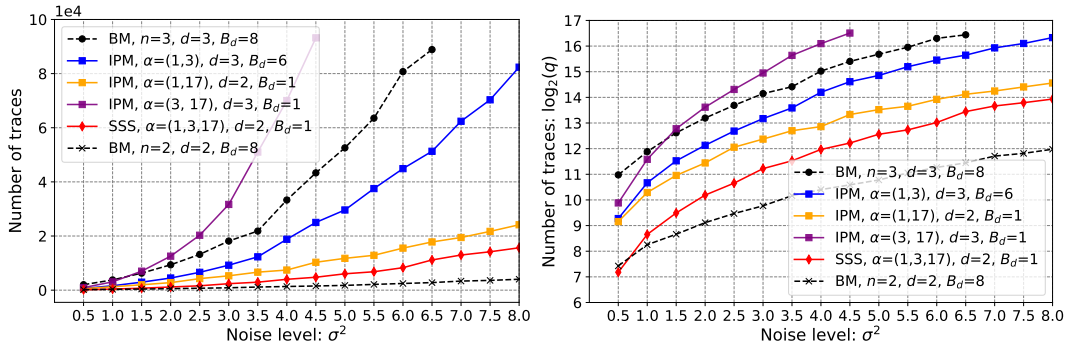
(a) Number of traces in normal scale.

(b) Number of traces in $\log_2$ scale.

Figure 7.9: Illustrating the impact of redundancy by comparing 2-share IPM with $(3, 1)$-SSS based masking, using $\alpha = (1, 2, 4)$ in the latter.

Another group of comparison is presented in Fig. 7.10 with $\alpha = (1, 3, 17)$. It is worth noting that, both coding-theoretic parameters are different in SSS-based masking and IPM. Although

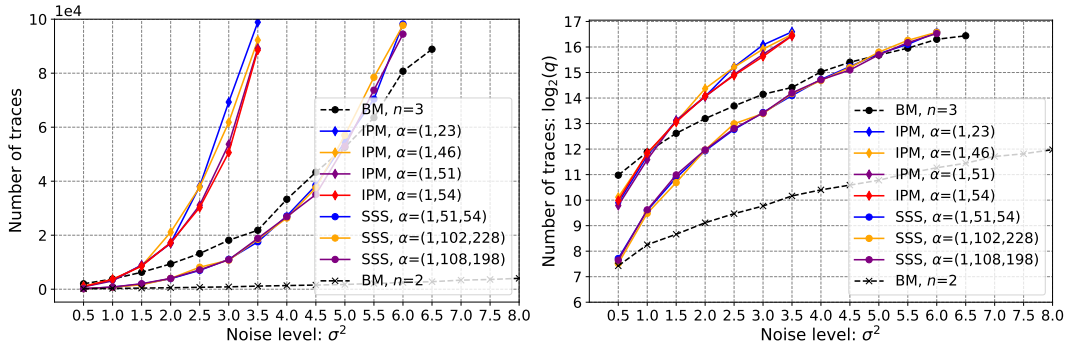
one instance of IPM is even better than the 3-share BM, the instance of SSS-based masking gets much worse with one share of redundancy. In particular, the latter is worse than the worst one among the three instances of IPM. Overall, the attack-based evaluation results verify the impact of redundancy on the concrete security level of code-based masking.



(a) Number of traces in normal scale. (b) Number of traces in $\log_2$ scale.

Figure 7.10: Illustrating the impact of redundancy by comparing 2-share IPM with $(3, 1)$-SSS based masking, using $\alpha = (1, 3, 17)$ in the latter.

At last, we illustrate the impact of redundancy by presenting a comparison between the optimal codes in IPM and SSS-based masking. Those optimal codes are visualized in Fig. 7.11. In particular, the four (out of twelve) optimal codes for 2-share IPM and three optimal codes for $(3, 1)$-SSS based masking are already shown in Fig. 7.4 and 7.7, respectively. Apparently, the redundancy can leverage an easier key-recovery attack in the sense of the necessary number of traces to succeed.



(a) Number of traces in normal scale. (b) Number of traces in $\log_2$ scale.

Figure 7.11: Illustrating the impact of redundancy by comparing 2-share IPM with $(3, 1)$-SSS based masking, using $\alpha = (1, 3, 17)$ in the latter.

Those observations made in above two groups of comparison invoke the need of a trade-off

between the amount of redundancy and the concrete security level in code-based masking. From a theoretical perspective, more redundancy can lead to more leakage, which is indicated by the two coding-theoretic properties. As a consequence, it is always advantageous to adopt non-redundant masking schemes rather than redundant ones when thwarting side-channel analysis. However, considering fault injection attacks (FIA) in real scenarios, a redundant masking scheme provides a combined countermeasure against both SCA and FIA.

More generally, above evaluation results pose a challenge on practical applications of the probing model when assessing the concrete security level of a protected cryptographic implementation. Specifically, given the same side-channel security order (irrespective to word-level or bit-level), adding redundancy will always facilitate the adversaries in recovering secrets, and lower the practical security in the sense of attacks. Therefore, we recommend to further assess the practical security of code-based masking by verifying both the dual distance and the kissing number in practice.

In summary, the attack-based evaluation confirms those theoretical findings in Parts II and III of this thesis. That is, we connect the dots in studying and improving code-based masking schemes. Particularly, we propose a unified framework to quantify the information leakage in code-based masking, and verify extensively by considering both IPM and SSS-based masking as instances of non-redundant and redundant code-based masking.

7.7 Revisiting All Codes in the State-of-the-Art

Regarding the state-of-the-art, various instances of code-based masking have been presented in literature, accompanied with specific linear codes (which are tuning parameters) used in them. We therefore revisit all linear codes in the literature for a thorough comparison.

For the purpose of a fair comparison, we focus on instances of code-based masking in which the codes are generated over $\mathbb{F}_{2^8}$ by using AES's irreducible polynomial (see Sec. 7.4.2). The results are detailed in Tab. 7.7. In particular, we present the best codes in several cases, along with the corresponding coding-theoretic properties.

The main takeaway point is that those optimal shall be used straightforwardly in practice, for instance, to protect AES implementations. We also provide instructive details for employing those codes in real circuits.

Table 7.7: Revisiting all linear codes used in literature over $\mathbb{F}_{2^8}$, with redundancy when $n > t+1$ while no redundancy when $n = t+1$.

Security Order t	Num. of Shares n	Parameters α in Sharing	Masking Scheme	Coding-Theoretic Properties			Comments
				$d_D^\perp$	$B_{d_D^\perp}$	$B'_{d_D^\perp}$	
$t=1$	$\boldsymbol{n=2}$ Non-redundant	$(1, 255)$	IPM	3	2	2	[2]
		$(3, 7)$	$(2,1)$-SSS	3	2	2	
		$(1, 17)$	IPM	2	1	1	[3] Three distinct codes
		$(1, 5)$		3	4	4	
		$(1, 7)$		4	8	8	
		$(221, 198)$, $(188, 189)$, $(237, 198)$	$(2,1)$-SSS	3, 3, 3	1, 1, 1	1, 1, 1	[51]. Note that $\alpha = (237, 175)$ is optimal
		$\mathbf{(237, 175)}$		4	4	4	
		$\mathbf{(1, 23)}$, $\mathbf{(1, 46)}$, $\mathbf{(1, 51)}$, ...	IPM	**4, 4, 4**, ...	**4, 4, 4**, ...	**4, 4, 4**, ...	**This work**. 12 optimal code in total, see Tab. 7.2
	$n=3$	$(5, 221, 198)$	$(3,1)$-SSS	3	1	1	[51]
		$(237, 175, 221)$		3	3	3	
		$(237, 221, 198)$		4	6	6	
		$\mathbf{(1, 51, 54)}$, $\mathbf{(1, 102, 228)}$, $\mathbf{(1, 108, 198)}$	$(3,1)$-SSS	**4, 4, 4**	**37, 37, 37**	**53, 53, 53**	**This work**. Only 3 optimal codes, see Tab. 7.6
	$n=4$	$(5, 237, 221, 198)$	$(4,1)$-SSS	3	10	10	[51]
		$(237, 175, 221, 198)$		3	12	12	
		$(12, 80, 176, 237)$		3	19	**53**	
	$n=5$	$(5, 237, 175, 221, 198)$	$(5,1)$-SSS	2	2	2	[51]
$t=2$	$\boldsymbol{n=3}$ Non-redundant	$(1, 15, 233)$	IPM	5	1	1	[2]
		$(13, 240, 163)$	$(3,2)$-SSS	6	2	2	
		$(1, 146, 147)$, $(1, 188, 189)$	$(3,2)$-SSS	3, 3	8, 8	8, 8	[51]. Both are equivalent to BM
		$\mathbf{(1, 27, 196)}$, $\mathbf{(1, 91, 204)}$, $\mathbf{(1, 218, 240)}$	IPM	**8, 8, 8**	**6, 6, 6**	**6, 6, 6**	**This work**. Only 3 optimal codes, see Tab. 7.4
	$n=5$	$(125, 246, 119, 104, 150)$, $(86, 23, 115, 107, 189)$	$(5,2)$-SSS	4, 4	1, 1	1, 1	[29]
		$(169, 63, 106, 49, 112)$		4	2	2	
		$(5, 237, 175, 221, 198)$		5	6	6	
		$\mathbf{(1, 23, 71, 167, 235)}$	$(5,2)$-SSS	**6**	**36**	**46**	**This work**. We find only one optimal code by fixing $\alpha_1 = 1$ and $\alpha_2 = 23$

7.8 Conclusions

In this chapter, we present an attack-based evaluation on two representative instances of code-based masking, namely IPM and SSS-based masking. The higher-order optimal distinguisher is employed in evaluation. As shown in previous chapters, we highlight that various encodings have significant impacts on the side-channel analysis of the corresponding scheme. Moreover, as an ultimate metric, the success rate of empirical attacks confirm the advantages of applying optimal instances of code-based masking.

Furthermore, our attack-based evaluation completes the assessment of code-based masking in known leakage model. However, this study shall be further verified on practical measurements from real circuits.

CHAPTER 8

Information-Theoretic Bounds on Attacks

Measuring the information leakage is critical for evaluating practical security of cryptographic devices against side-channel analysis. More straightforwardly, it is interesting to have an upper bound on success rate of any attack given a (fixed) number of side-channel measurements. Or conversely, we wish to derive a lower bound on the number of queries for a given success rate of optimal attacks. In this chapter, we derive several bounds in both directions by using information-theoretic tools, particularly for cryptographic implementations protected by masking schemes. We show that a generic upper bound on the probability of success, irrespective of specific attacks, is linked to mutual information between side-channel measurements and the secret. Moreover, our numerical evaluation confirms that, the success rate of optimal maximum likelihood distinguishers is tightly bounded given a fixed number of measurements.

Part of results shown in this chapter has been presented in [41] (preprint on ArXiv).

Contents

8.1 Introduction

Since the seminal work by Kocher et al. [95], side-channel analyses (SCAs) have been one of the most powerful attacks against cryptographic devices in practice. They exploit physical observable information leakages like instantaneous power consumption [95] or electromagnetic radiation [71] to recover the secrets used in cryptosystems. From an adversary's perspective, many attacks (distinguishers) have been proposed to exploit various leakages and there are several metrics to compare them in a fair way [151]. However, bounding how (any) side-channel attack succeeds is still an open problem. In other words, given a set of side-channel measurements, we seek a generic upper bound on the success rate of any attack. In this respect, Chérisey et al. [57] propose several bounds on the key extraction success rate by using information-theoretic tools, which are tight in assessing unprotected cryptographic implementations.

To counteract SCAs, many countermeasures are proposed wherein masking is a well-established one which provides provable security [86, 128]. Indeed, the number of measurements for a successful attack against masked implementations is exponential to the masking order (e.g., the number of random masks per sensitive variable) provided with a sufficient amount of noise [61]. However, the lower bounds proposed in [61, 128] are demonstrated by approximations and inequalities, resulting in loose bounds on the number of traces needed for a given success rate. Moreover, as we will show in this chapter, those bounds on success rate given in [57] by using mutual information (MI) for several measurements are also very loose when targeting protected cryptographic implementations.

In this chapter, we aim at providing tight bounds on the success rate of any SCA by leveraging information-theoretic tools. To do so, we consider a similar communication-channel framework

which has been developed in [57, 84] and adapt it to masking schemes. The overview of the framework is shown in Fig. 8.1 and notations are introduced in the following section.

8.1.1 Notations

In the sequel, uppercase letters (e.g. X) denote random variables where calligraphic letters (e.g., $\mathcal{X}$) are for sets, lowercase letters (e.g., x) are for realizations and bold letters are for vectors and matrices.

Therefore, as illustrated in Fig. 8.1, we have

- $K \in \mathbb{F}_{2^\ell}$ denotes the secret key (typically $\ell = 8$, e.g., in AES), and $\widehat{K}$ is the output of a side-channel attack
- $\mathbf{T} \in \mathbb{F}_{2^\ell}^q$ denotes plaintexts or ciphertexts of length q
- $\mathbf{U}$ is the sensitive variable, say $\mathbf{U} = S(\mathbf{T} \oplus K)$ where S denotes a cryptographic operations like the Sbox in AES
- without or with masking:
  - $\mathbf{V} = \mathbf{U}$ if no masking, which is the case for [57, 84]
  - $\mathbf{V} = (\mathbf{U} \oplus \mathbf{M}, \mathbf{M})$ if considering e.g., the 1st-order Boolean masking with a random mask $\mathbf{M} \in \mathbb{F}_{2^\ell}^q$
  - $\mathbf{V} = \mathbf{UG} + \mathbf{MH}$ if taking the code-based masking [35] where $\mathbf{G}$ and $\mathbf{H}$ are two generator matrices used in the masking
- $\mathbf{X} = f(\mathbf{V})$ is the noiseless leakage, say $\mathbf{X} = f(\mathbf{V})$ and $f = w_H$ in co-called Hamming weight model
- $\mathbf{Y}$ is the noisy leakage, which models q measurements (traces) in practice, say $\mathbf{Y} = \mathbf{X} + \mathbf{N}$ where $\mathbf{N}$ denotes the additive white Gaussian noise: $\mathbf{N} \sim \mathcal{N}(0, \sigma\mathbf{I})$. Additionally, the channel is assumed to be memoryless.

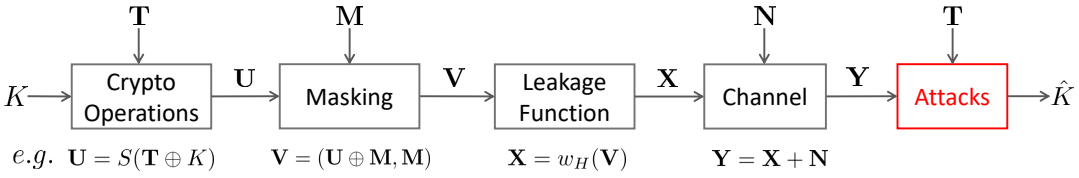


Figure 8.1: Representation of side-channel analysis of a masked cryptographic operation as a communication channel.

As a consequence, assuming $\mathbf{T}$ is known, other variables form a Markov chain: $K-\mathbf{U}-\mathbf{V}-\mathbf{X}-\mathbf{Y}-\widehat{K}$. By Markovity, when related to single-letter quantities, we have: $I(\mathbf{X};\mathbf{Y}|\mathbf{T}) \leq qI(X;Y|T)$ as in [57]. In particular,

- in [57, §3.1], one of the bounds is given by: $q \geq \frac{\text{cst}}{I(X;Y|T)}$. So $I(X;Y|T)$ comes naturally from bounds.
- in [37, Theorem 4], the leakage metric is: $I(K;Y|T) = I(U;Y|T)$, which is implicitly connected to q [24].

8.2 Contributions

In this work, we derive security bounds for side-channel attacks in the presence of countermeasures. First of all, instead of utilizing universal inequality-based bounds on mutual information as in [57], we use mutual information itself and derive bounds on the success rate by applying Fano's inequality [52]. Secondly, we suggest to use $I(\mathbf{U};\mathbf{Y}|\mathbf{T})$ instead of $I(\mathbf{X};\mathbf{Y}|\mathbf{T})$ in masked cases, since the bounds on success rate by the former is much tighter than by the latter. At last, we furnish numerical results in a commonly used side-channel setting which confirm that, our new bound provides more accurate security guarantees in the context of masked cryptographic implementations.

8.3 Applying MIs of Different Variables

8.3.1 Links between Different Pairs of MIs

With notations shown in Fig. 8.1, we have following equalities and inequalities with respect to MIs given different pairs of variables in the context of side-channel analysis.

Lemma 8.1. *By a side-channel setting as in Fig. 8.1, one has*

$$I(K;\mathbf{Y}|\mathbf{T}) = I(\mathbf{U};\mathbf{Y}|\mathbf{T}) \leq I(\mathbf{V};\mathbf{Y}|\mathbf{T}) = I(\mathbf{X};\mathbf{Y}|\mathbf{T}). \tag{8.1}$$

Proof. For $I(\mathbf{V};\mathbf{Y}|\mathbf{T}) = I(\mathbf{X};\mathbf{Y}|\mathbf{T})$, knowing $\mathbf{T}$ gives that $\mathbf{V}-\mathbf{X}-\mathbf{Y}$ forms a Markov chain and, since $\mathbf{X} = f(\mathbf{V})$ then $\mathbf{X}-\mathbf{V}-\mathbf{Y}$ also forms a Markov chain. Thus $I(\mathbf{V};\mathbf{Y}|\mathbf{T}) = I(\mathbf{X};\mathbf{Y}|\mathbf{T})$. Similarly, we get the first equality.

For $I(\mathbf{U};\mathbf{Y}|\mathbf{T}) \leq I(\mathbf{V};\mathbf{Y}|\mathbf{T})$, it is straightforward as $\mathbf{U}-\mathbf{V}-\mathbf{Y}$ is a Markov chain. Yet the converse is not true because of the random mask $\mathbf{M}$. □

As a consequence of Lemma 8.1, we only focus on two quantities $I(\mathbf{U};\mathbf{Y}|\mathbf{T})$ and $I(\mathbf{X};\mathbf{Y}|\mathbf{T})$, where intuitively, the former should give a better bound than the latter. Next, since the ML (Maximum Likelihood)-based distinguishers are optimal [84] in SCAs, we have the following lemma which works for any distinguisher.

Lemma 8.2. *Considering any distinguisher, including the optimal one (namely the ML-based one), one has*

$$I(K;\widehat{K}) \leq I(K;\widehat{K}|\mathbf{T}) = I(K;\mathbf{Y}|\mathbf{T}). \tag{8.2}$$

where $\widehat{K} = \varphi(\mathbf{y},\mathbf{t}) = \text{argmax}_k\, p(\mathbf{Y} = \mathbf{y}(k)|\mathbf{T} = \mathbf{t})$ *follows the distinguisher rule in SCAs.*

Proof. Since $H(K|\widehat{K}) \geq H(K|\widehat{K},\mathbf{T})$ and K is independent of $\mathbf{T}$, we have $I(K;\widehat{K}|\mathbf{T}) = H(K|\mathbf{T}) - H(K|\widehat{K},\mathbf{T}) = H(K) - H(K|\widehat{K},\mathbf{T}) \geq H(K) - H(K|\widehat{K}) = I(K;\widehat{K})$.

Secondly, knowing $\mathbf{T}$ implies that $\widehat{K}$ is a deterministic function of $\mathbf{Y}$ following the ML rule: $\widehat{K} = \varphi(\mathbf{Y},\mathbf{T}) = \text{argmax}_k\, p(\mathbf{y}(k)|\mathbf{T} = \mathbf{t})$, which proves equality. □

Remark 8.1. The ML rule coincides with MAP (Maximum A Posterior) rule assuming K is uniformly distributed (to maximize its entropy), which is a common setting in SCAs.

Interestingly, we can upper bound $I(K;\mathbf{Y}|\mathbf{T})$ as follows.

Lemma 8.3. *Given the same setting as in Fig. 8.1, one has*

$$I(K;\mathbf{Y}|\mathbf{T}) \leq H(K). \tag{8.3}$$

Proof. The inequality holds in the side-channel setting of Fig. 8.1, since $I(K;\mathbf{Y}|\mathbf{T}) = H(K|\mathbf{T}) - H(K|\mathbf{Y},\mathbf{T}) = H(K) - H(K|\mathbf{Y},\mathbf{T}) \leq H(K)$ where $H(K|\mathbf{Y},\mathbf{T}) \geq 0$. □

In practice, Lemma 8.3 reflects the fact that the total amount of information any adversary could extract cannot exceed the information carried by the secret key, while the latter is measured by the entropy $H(K)$.

8.3.2 Connecting to Capacity

Lemma 8.4. *Considering the same setting as in Fig. 8.1, one has*

$$I(\mathbf{X};\mathbf{Y}) - I(\mathbf{T};\mathbf{Y}) = I(\mathbf{X};\mathbf{Y}|\mathbf{T}) \geq 0. \tag{8.4}$$

Proof. Since $\mathbf{T} - \mathbf{X} - \mathbf{Y}$ forms a Markov chain, we have $H(\mathbf{Y}|\mathbf{X},\mathbf{T}) = H(\mathbf{Y}|\mathbf{X})$ [1]. Hence $I(\mathbf{X};\mathbf{Y}|\mathbf{T}) = H(\mathbf{Y}|\mathbf{T}) - H(\mathbf{Y}|\mathbf{X},\mathbf{T}) = H(\mathbf{Y}|\mathbf{T}) - H(\mathbf{Y}|\mathbf{X}) = H(\mathbf{Y}) - H(\mathbf{Y}|\mathbf{X}) - \big(H(\mathbf{Y}) - H(\mathbf{Y}|\mathbf{T})\big) = I(\mathbf{X};\mathbf{Y}) - I(\mathbf{T};\mathbf{Y})$. □

[1] We use H for entropy of both discrete and continuous variables, although h is used more frequently for differential entropy of a continuous variable.

In fact, $I(\mathbf{X};\mathbf{Y}) - I(\mathbf{T};\mathbf{Y}) \geq 0$ is also a direct consequence of the data processing inequality where $\mathbf{T} - \mathbf{X} - \mathbf{Y}$ forms a Markov chain.

This leads us to define the *capacity* of the side-channel as

$$\begin{aligned} C &= \max_{\mathbf{T}-\mathbf{X}-\mathbf{Y}} I(\mathbf{X};\mathbf{Y}) - I(\mathbf{T};\mathbf{Y}) \\ &= \max_{\mathbf{T}-\mathbf{X}-\mathbf{Y}} I(\mathbf{X};\mathbf{Y}|\mathbf{T}), \end{aligned} \tag{8.5}$$

where the maximum is taken over all distributions of $\mathbf{X}$ given $\mathbf{T}$ such that $\mathbf{T} - \mathbf{X} - \mathbf{Y}$ is a Markov chain. The capacity can be determined from the following lemma.

Lemma 8.5. *One has*

$$C = \max_{\mathbf{X}} I(\mathbf{X};\mathbf{Y}) \tag{8.6}$$

where the maximum is taken over all channel input distributions $\mathbf{X}$.

Proof. From (8.5), one has $I(\mathbf{X};\mathbf{Y}|\mathbf{T}) = \mathbb{E}_{\mathbf{T}} I(\mathbf{X};\mathbf{Y}|\mathbf{T}=\mathbf{t})$ and each $I(\mathbf{X};\mathbf{Y}|\mathbf{T}=\mathbf{t})$ is maximized taking $p(\mathbf{x}|\mathbf{t}) = p(\mathbf{x})$, so as to maximize $I(\mathbf{X};\mathbf{Y})$. Since the optimal distribution does not depend on $\mathbf{t}$, it also maximizes the expectation $\mathbb{E}_{\mathbf{T}} I(\mathbf{X};\mathbf{Y}|\mathbf{T}=\mathbf{t}) = I(\mathbf{X};\mathbf{Y}|\mathbf{T})$ and thus $\max_{\mathbf{T}-\mathbf{X}-\mathbf{Y}} I(\mathbf{X};\mathbf{Y}|\mathbf{T}) = \max_{\mathbf{X}} I(\mathbf{X};\mathbf{Y})$. □

Remark 8.2. We can consider the more general situation where the channel also depends on $\mathbf{T}$. In this case we would have $C = \mathbb{E}\{C_T\}$ where $C_{\mathbf{t}} = \max_{\mathbf{X}} I(\mathbf{X};\mathbf{Y}|\mathbf{T}=\mathbf{t})$.

8.4 Bounding Success Rates and Capacity

8.4.1 Upper Bounds on Success Rates

As shown in [57], the mutual information itself gives the tightest bound on the success rate of a side-channel attack. By combining Lemmas 8.1, 8.2 and 8.3, we have $I(K;\widehat{K}) \leq I(\mathbf{U};\mathbf{Y}|\mathbf{T}) \leq H(K)$.

The probability of success (say success rate) in SCA is defined as: $P_s = \mathbb{P}(\widehat{K} = K)$. Accordingly, the error rate is $P_e = 1 - P_s$. Now, we have the following upper bound on P_s.

Theorem 8.1. *Given a side-channel setting as in Fig. 8.1, we have*

$$f_P(P_s) \leq I(\mathbf{U};\mathbf{Y}|\mathbf{T}), \tag{8.7}$$

where $f_P(p) = H(K) - H_2(p) - (1-p)\log(2^\ell - 1)$ *and* $H_2(p) = -p\log p - (1-p)\log(1-p)$, *for* $p \in [2^{-\ell}, 1]$ *and* ℓ *denotes the number of bits in* $K = k$.

Proof. By Fano's inequality [52] and Lemma 8.2, we have: $f_P(P_s) = H(K) - H_2(P_s) - (1 - P_s)\log(2^\ell - 1) \leq I(K;\widehat{K}) \leq I(\mathbf{U};\mathbf{Y}|\mathbf{T})$. □

Since $f_P(p)$ is strictly increasing for $p \in [2^{-\ell}, 1]$ [56, §A], Theorem 8.1 not only provides an upper bound on P_s, but also gives a lower bound on the number of traces q to obtain a specific P_s in SCAs, where q is involved in $I(\mathbf{U}; \mathbf{Y}|\mathbf{T})$. Apparently, we have $I(\mathbf{U}; \mathbf{Y}|\mathbf{T}) \leq q \cdot I(U; Y|T)$.

Remark 8.3. It is trivial to have a much loose bound on P_s as: $f_P(P_s) \leq I(\mathbf{X}; \mathbf{Y}|\mathbf{T})$. However, as we will show later, this bound is too loose and is out of use in evaluating masked implementations. Furthermore, $I(\mathbf{X}; \mathbf{Y}|\mathbf{T})$ cannot be bounded by $H(K)$ (recall Lemma 8.3) and it increases linearly in q.

8.4.2 Bounding $I(\mathbf{X}; \mathbf{Y}|\mathbf{T})$ by Shannon's Channel Capacity

As mentioned in Remark 8.3, $I(\mathbf{X}; \mathbf{Y}|\mathbf{T})$ will not be bounded by $H(K)$ in protected cryptographic implementations. But still, it is upper bounded by the capacity defined in (8.5) as in the following lemma.

Lemma 8.6. *Given a side-channel setting in Fig. 8.1, we have*

$$I(\mathbf{X}; \mathbf{Y}|\mathbf{T}) \leq \frac{q}{2} \log(1 + SNR), \tag{8.8}$$

where SNR is the signal-to-noise ratio and σ^2 denotes the variation of noise.

Proof. $I(\mathbf{X}; \mathbf{Y}|\mathbf{T}) \leq q \cdot I(X; Y|T) = q \cdot (H(Y|T) - H(Y|X)) \leq q \cdot (H(Y) - H(Y|X)) \leq q \cdot C = \frac{q}{2} \log(1 + \mathrm{SNR})$. □

We will show in next section that this upper bound on $I(\mathbf{X}; \mathbf{Y}|\mathbf{T})$ is very tight in the presence of a Boolean masking.

8.5 Applying into Hamming Weight Leakages with Additive Gaussian Noise

By equalities in Lemma 8.1, the only two MIs that need to be evaluated are $I(\mathbf{X}; \mathbf{Y}|\mathbf{T})$ and $I(\mathbf{U}; \mathbf{Y}|\mathbf{T})$ [1]. Taking notations from Fig. 8.1, we calculate both MIs in a numerical manner. We have

$$\begin{aligned} I(\mathbf{X}; \mathbf{Y}|\mathbf{T}) &= H(\mathbf{Y}|\mathbf{T}) - H(\mathbf{Y}|\mathbf{X}, \mathbf{T}), \\ I(\mathbf{U}; \mathbf{Y}|\mathbf{T}) &= H(\mathbf{Y}|\mathbf{T}) - H(\mathbf{Y}|\mathbf{U}, \mathbf{T}), \end{aligned} \tag{8.9}$$

where $H(\mathbf{Y}|\mathbf{T})$ and $H(\mathbf{Y}|\mathbf{U}, \mathbf{T}) = H(\mathbf{Y}|\mathbf{U})$ can be estimated by Monte-Carlo simulations and,

$$\begin{aligned} H(\mathbf{Y}|\mathbf{X}, \mathbf{T}) &= H(\mathbf{Y}|\mathbf{X}) = H(\mathbf{N}) \\ &= q \cdot \frac{1}{2} \log\left(2\pi e \sigma^2\right). \end{aligned} \tag{8.10}$$

[1] We use $\log_2$ to have mutual information and entropy expressed in bits.

8.5.1 Importance Sampling in Monte-Carlo Simulation

Monte-Carlo simulation is a well-known method to estimate expectations of a function under certain distribution by repeated random sampling, where the importance sampling can be used to improve the efficiency and speedup the convergence procedure [100, Chap. 29].

By Monte-Carlo simulation, we can estimate the first term $H(\mathbf{Y}|\mathbf{T})$ in subtractions (8.9) by randomly drawing N_C samples. Particularly, equipped with importance sampling, we have

$$\begin{aligned} H(\mathbf{Y}|\mathbf{T}) &= \int_{\mathbf{y}} \sum_{\mathbf{t}} p(\mathbf{y},\mathbf{t}) \log \frac{1}{p(\mathbf{y}|\mathbf{t})} \, \mathrm{d}\mathbf{y} \\ &\approx \lim_{N_C \to \infty} -\frac{1}{N_C} \sum_{j=1}^{N_C} \log p(\mathbf{y}^j|\mathbf{t}^j), \end{aligned} \tag{8.11}$$

where each $(\mathbf{t}^j, \mathbf{y}^j)$, for $1 \le j \le N_C$, is drawn randomly. The estimation in (8.11) is sound based on the law of large numbers [52, Chap. 3] and it has been numerically verified in [57]. Similarly, $H(\mathbf{Y}|\mathbf{U})$ can be estimated using Monte-Carlo simulation by $H(\mathbf{Y}|\mathbf{U}) = -\frac{1}{N_C} \sum_{j=1}^{N_C} \log p(\mathbf{y}^j|\mathbf{u}^j)$.

Convergence in Monte-Carlo Simulation. Since the accuracy of Monte-Carlo simulation highly depends on the number of samples, we justify hereafter how we chose N_C. As shown in Fig. 8.2, the estimation of $I(\mathbf{X};\mathbf{Y}|\mathbf{T})$ (in unprotected case, cf Sec. 8.5.2) gets more accurate by using larger N_C. In particular, this estimation on $I(\mathbf{X};\mathbf{Y}|\mathbf{T})$ is accurate enough by using only $N_C = 100,000$ draws. However, we use $N_C = 1,000,000$ throughout this chapter (e.g., in Fig. 8.3, 8.4 and 8.5) to have a more stable estimation.

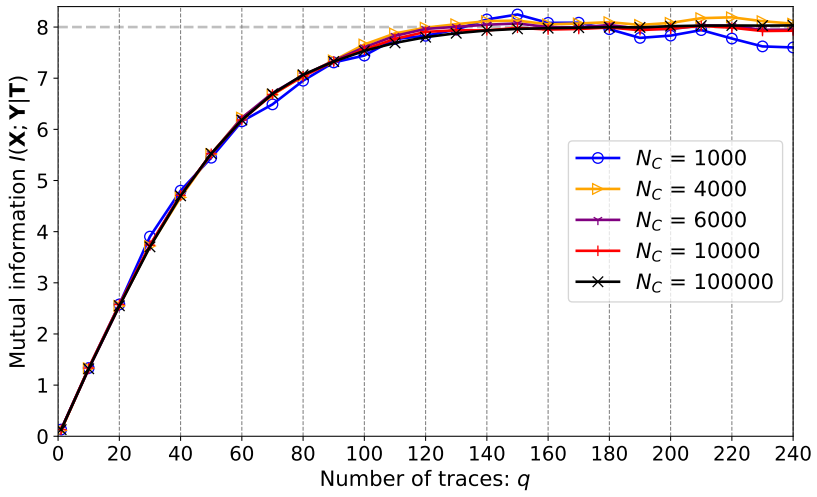


Figure 8.2: Monte-Carlo simulation with various N_C draws where $\sigma^2 = 10.00$.

In the following, we present different cases of $p(\mathbf{y}|\mathbf{t})$ of one draw in both unprotected and masked cases.

8.5.2 Without Masking

The unprotected case corresponds to the one considered in [57]. Here $(\mathbf{t}^j, \mathbf{y}^j)$, for $1 \leq j \leq N_C$, is drawn according to this process:

- $\mathbf{t}^j \sim \mathcal{U}(\mathbb{F}_{2^\ell}^q)$,
- $k^j \sim \mathcal{U}(\mathbb{F}_{2^\ell})$, and
- $\mathbf{y}^j \sim \mathcal{N}(w_H(S(\mathbf{t}^j \oplus k^j)), \sigma^2 \mathbf{I}_q) \in \mathbb{R}^q$.

We then have for one draw $(\mathbf{t}, \mathbf{y})$:

$$\begin{aligned} p(\mathbf{y}|\mathbf{t}) &= \sum_k p(k) p(\mathbf{y}|\mathbf{t}, k) = \sum_k p(k) \prod_{i=1}^{q} p(\mathbf{y}_i|\mathbf{t}_i, k) \\ &= \sum_k p(k) \prod_{i=1}^{q} \frac{1}{(2\pi\sigma^2)^{1/2}} e^{\frac{-(\mathbf{y}_i - w_H(S(\mathbf{t}_i \oplus k)))^2}{2\sigma^2}}. \end{aligned} \tag{8.12}$$

Since $K \in \mathbb{F}_{2^\ell}$ is uniformly distributed, (8.12) gives

$$\begin{aligned} \log p(\mathbf{y}|\mathbf{t}) &= \log \sum_k p(k) \prod_{i=1}^{q} \frac{1}{(2\pi\sigma^2)^{1/2}} e^{\frac{-(\mathbf{y}_i - w_H(S(\mathbf{t}_i \oplus k)))^2}{2\sigma^2}} \\ &= -\ell - \frac{q}{2} \log\left(2\pi\sigma^2\right) + \log \sum_k \prod_{i=1}^{q} e^{\frac{-(\mathbf{y}_i - w_H(S(\mathbf{t}_i \oplus k)))^2}{2\sigma^2}}. \end{aligned} \tag{8.13}$$

The numerical results of $I(\mathbf{X}; \mathbf{Y}|\mathbf{T})$ are shown in Fig. 8.3 with different levels of noise (σ^2). Note that we take $N_C = 1,000,000$ random draws in Monte-Carlo simulation. These bounds are the same as those already plotted in [57]. Here, plotting the bounds as a function of various values of σ^2 highlights that $I(\mathbf{X}; \mathbf{Y}|\mathbf{T})$ curves are about homothetic, in that they $I(\mathbf{X}; \mathbf{Y}|\mathbf{T})$ depends only on q/σ^2. Said differently, from an attacker perspective, the effort in terms of traces collection scales linearly in the noise variance (for a given value of mutual information).

8.5.3 With a First-order Boolean Masking

Here $(\mathbf{t}^j, \mathbf{y}^j)$, for $1 \leq j \leq N_C$, is drawn i.i.d. according to this process:

- $\mathbf{t}^j \sim \mathcal{U}(\mathbb{F}_{2^\ell}^q)$,
- $\mathbf{m}^j \sim \mathcal{U}(\mathbb{F}_{2^\ell}^q)$,

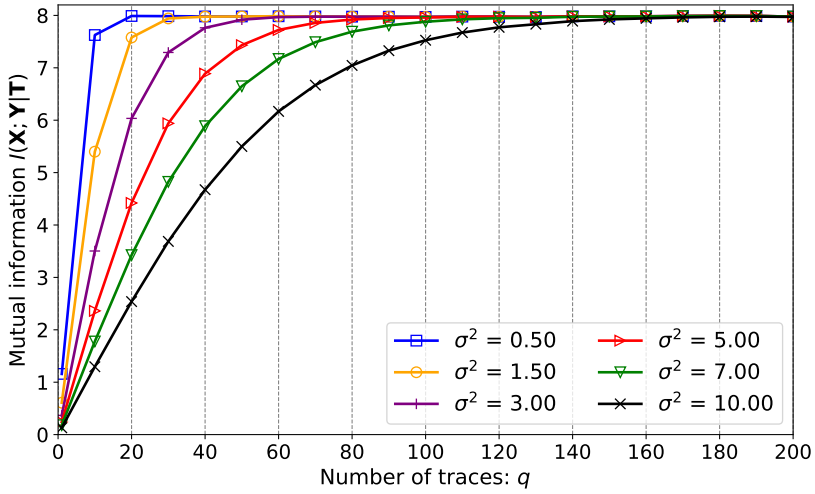


Figure 8.3: Evolution of $I(\mathbf{X};\mathbf{Y}|\mathbf{T})$ with the number of traces under different levels of noise in the unprotected case without masking, $N_C = 1,000,000$.

- $k^j \sim \mathcal{U}(\mathbb{F}_{2^\ell})$, and
- $\mathbf{y}^j \sim \mathcal{N}(w_H(S(\mathbf{t}^j \oplus k^j) \oplus \mathbf{m}^j) + w_H(\mathbf{m}^j), \sigma^2 \mathbf{I}_q) \in \mathbb{R}^q$.

Note that we consider the zero-offset leakage [24] where the leakages of each share are summed together (see the sum of two Hamming weights above).

We have for one draw $(\mathbf{t}, \mathbf{y})$:

$$\begin{aligned} p(\mathbf{y}|\mathbf{t}) &= \sum_k p(k) p(\mathbf{y}|\mathbf{t},k) = \sum_k p(k) \prod_{i=1}^{q} p(\mathbf{y}_i|\mathbf{t}_i,k) \\ &= \sum_k p(k) \prod_{i=1}^{q} \sum_{m_i} p(m_i) p(\mathbf{y}_i|\mathbf{t}_i,k,m_i) \\ &= \sum_k p(k) \prod_{i=1}^{q} \sum_{m_i} p(m_i) \frac{e^{\frac{-(\mathbf{y}_i - f(\mathbf{t}_i,k,m_i))^2}{2\sigma^2}}}{(2\pi\sigma^2)^{1/2}}, \end{aligned} \tag{8.14}$$

where $f(\mathbf{t}_i,k,m_i) = w_H(S(\mathbf{t}_i \oplus k) \oplus m_i) + w_H(m_i)$ is the zero-offset leakage under Hamming weight model. Again, taking $K \in \mathbb{F}_{2^\ell}$ uniformly, and considering that all masks are i.i.d. $\sim \mathcal{U}(\mathbb{F}_{2^\ell})$, we have

$$\log p(\mathbf{y}|\mathbf{t}) = -\ell(q+1) - \frac{q}{2}\log\left(2\pi\sigma^2\right) + \log \sum_k \prod_{i=1}^{q} \sum_m e^{\frac{-(\mathbf{y}_i - f(\mathbf{t}_i,k,m))^2}{2\sigma^2}}. \tag{8.15}$$

The numerical results of $I(\mathbf{X};\mathbf{Y}|\mathbf{T})$ are depicted in Fig. 8.4. It clearly appears that the effect of masking is to relax the values of $I(\mathbf{X};\mathbf{Y}|\mathbf{T})$, which motivates for the fact that a bound based

on $I(\mathbf{X};\mathbf{Y}|\mathbf{T})$ will be very loose. This motivates for the shifting to focus on $I(\mathbf{U};\mathbf{Y}|\mathbf{T})$. Back to Fig. 8.4, the dotted black lines show the upper bounds given by Lemma 8.6. The takeaway observation is that the bounds are all the tighter as the noise level increases.

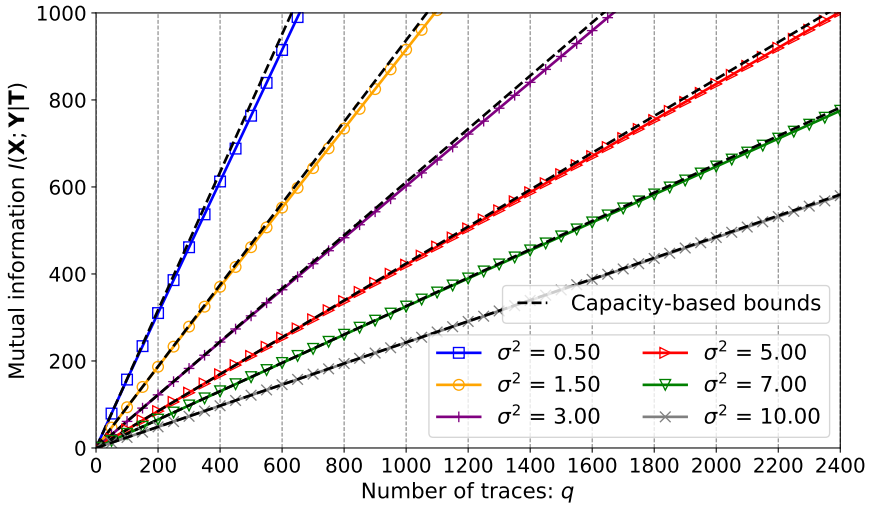


Figure 8.4: Bounding on $I(\mathbf{X};\mathbf{Y}|\mathbf{T})$ by Shannon's channel capacity in masked cases, $N_C = 1,000,000$.

Estimation of $I(\mathbf{U};\mathbf{Y}|\mathbf{T})$. Similarly as in (8.15), we have

$$\log p(\mathbf{y}|\mathbf{u}) = -q\ell - \frac{q}{2}\log\left(2\pi\sigma^2\right) + \log\prod_{i=1}^{q}\sum_{m} e^{\frac{-(\mathbf{y}_i - f'(\mathbf{u}_i,m))^2}{2\sigma^2}}. \tag{8.16}$$

where $f'(\mathbf{u}_i, m) = w_H(\mathbf{u}_i \oplus m) + w_H(m)$.

The numerical results of $I(\mathbf{U};\mathbf{Y}|\mathbf{T})$ are shown in Fig. 8.5 with different levels of noise (σ^2).

As shown in Fig. 8.5, $I(\mathbf{U};\mathbf{Y}|\mathbf{T})$ is bounded by $H(K)$ as expected (see Lemma 8.3). Particularly, given the same noise level, the number of traces needed to obtain $I(K;\mathbf{Y}|\mathbf{T}) = I(\mathbf{U};\mathbf{Y}|\mathbf{T}) = 8$ bits is much larger than in the unprotected case shown in Fig. 8.3. The curves $I(\mathbf{U};\mathbf{Y}|\mathbf{T})$ vs σ^2 also look homothetic with a scale of σ^2 (as was the case of curves $I(\mathbf{X};\mathbf{Y}|\mathbf{T})$ vs σ^2 without masking, cf. Fig. 8.4). This is justified by a simple scaling argument: if the number of traces for a given set of $(\mathbf{T},\mathbf{U})$ is doubled, then the mutual information is the same as with the nominal number of traces, but with SNR doubled as well.

8.5.4 Bounding Success Rate in Masked Implementations

Relying on Theorem 8.1, we have a upper bound on probability of success P_s, which also gives a lower bound on the minimum of q to get a specific P_s. Moreover, a linear bound on q is given

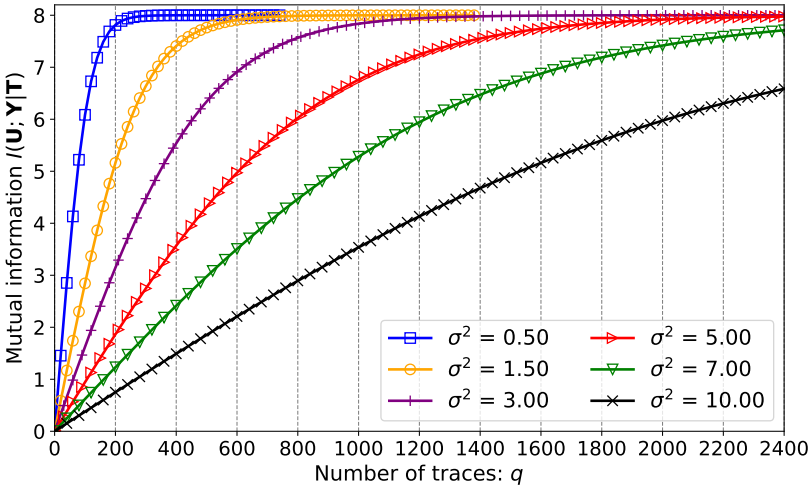


Figure 8.5: Evolution of mutual information $I(\mathbf{U};\mathbf{Y}|\mathbf{T})$ with the number of traces under different levels of noise in masked cases, $N_C = 1,000,000$. Note that $I(\mathbf{U};\mathbf{Y}|\mathbf{T})$ is upper bounded by $H(K) = 8$ bits.

by $I(\mathbf{U};\mathbf{Y}|\mathbf{T}) \leq q \cdot I(U;Y|T)$.

We apply Theorem 8.1 into the masked case. Numerical results are shown in Fig. 8.6 where we present several instances with different levels of Gaussian noises. In particular, the ML attacks utilize the higher-order distinguishers which have been demonstrated to be optimal in the presence of masking [18]. In order to evaluate P_s of ML attacks, each attack is repeated 200 times to have a more accurate success rate.

As shown in Fig. 8.6, the bound given by $I(\mathbf{U};\mathbf{Y}|\mathbf{T})$ is very tight. Indeed, a commonly used metric on attacks is the minimum number of traces to reach $P_s \geq 95\%$. Considering $\sigma^2 = 3.00$ in Fig. 8.6, we set $P_s = 95\%$ and the ML attack needs around $q = 800$ traces, where our new bound gives $q = 720$, while the bound proposed in [57] by using $I(\mathbf{X};\mathbf{Y}|\mathbf{T})$ only gives $q = 12$. Furthermore, the latter bound would be much looser when the noise level continues to increase. A more detailed comparison is depicted in Fig. 8.7, which shows the predicted minimum numbers of traces reaching $P_s \geq 95\%$ given by both $I(\mathbf{U};\mathbf{Y}|\mathbf{T})$ and $I(\mathbf{X};\mathbf{Y}|\mathbf{T})$. These curves show that our new bound is much tighter than the previous one from the state-of-the-art, as it captures the masking scheme (recall from Fig. 8.1 that the masking countermeasure step is between $\mathbf{U}$ and $\mathbf{Y}$ but not between $\mathbf{X}$ and $\mathbf{Y}$).

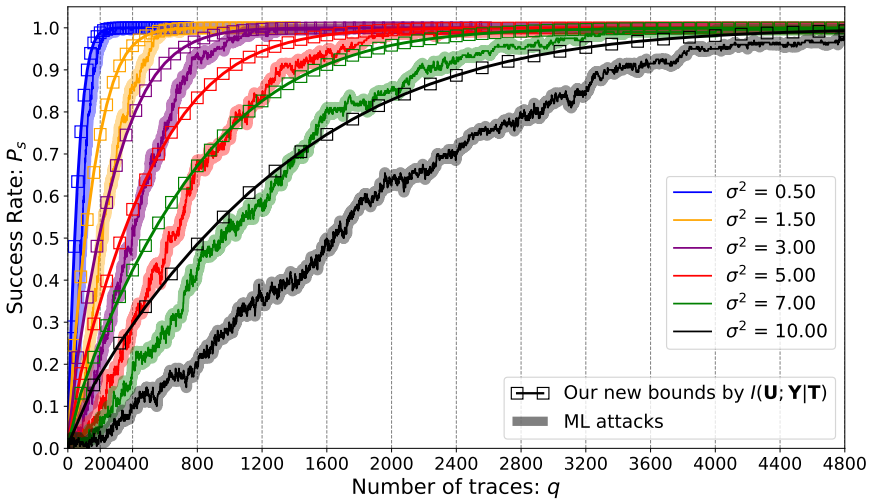
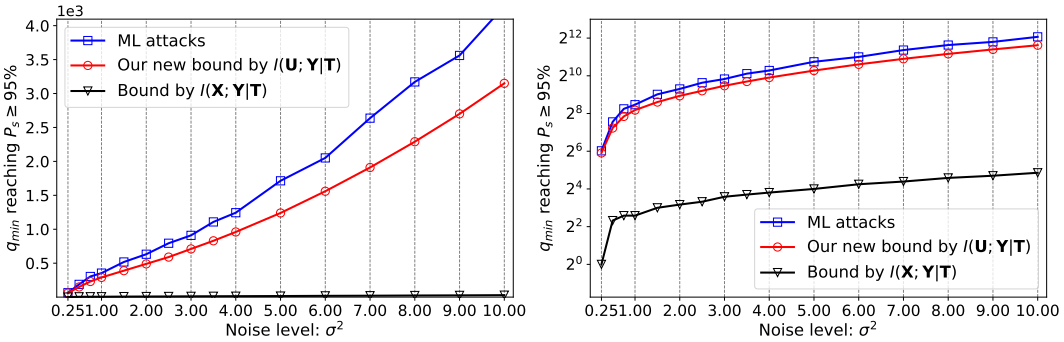


Figure 8.6: Application and comparison of bounds on success rate. We present six instances with different noise levels by using $q_{\max} = 4800$ traces. Note that we omit the bounds given by $I(\mathbf{X};\mathbf{Y}|\mathbf{T})$ as they are invisible when plotted together with bounds given by $I(\mathbf{U};\mathbf{Y}|\mathbf{T})$.



(a) Number of traces in normal scale.

(b) Number of traces in $\log_2$ scale.

Figure 8.7: Comparison of the minimum number of traces $q_{\min}$ to reach $P_s \geq 95\%$ predicted by our new bound, by $I(\mathbf{X};\mathbf{Y}|\mathbf{T})$ as in [57] and also the baseline given by an ML attack.

8.6 Extending to Code-based Masking

We have shown the advantages of the code-based masking against side-channel attacks in previous chapters. As the information-theoretic bounds in this chapter are generic, we therefore apply those evaluations into the code-based masking.

The general communication channel framework is the same as in Fig. 8.1, except that the Boolean masking is replaced by a more general masking scheme. Specifically, we say $\mathbf{V} = \mathbf{UG} + \mathbf{MH}$ where $\mathbf{G}$, $\mathbf{H}$ are the generator matrices of two codes $\mathcal{C}$ and $\mathcal{D}$ in code-based

masking, respectively. For the sake of simplicity, we consider IPM with $n = 2$, meaning that

$$\begin{aligned}\mathbf{V} &= \mathbf{UG} + \mathbf{MH} = \mathbf{U}\begin{pmatrix}1 & 0\end{pmatrix} + \mathbf{M}\begin{pmatrix}\alpha_1 & 1\end{pmatrix} \\ &= (\mathbf{U} + \alpha_1\mathbf{M},\ \mathbf{M}),\end{aligned} \tag{8.17}$$

where α_1 is the only public parameter in IPM. As a consequence, different values of α_1 lead to various linear codes in IPM.

8.6.1 Bounding Mutual Information

From a high abstract view of communication channel, utilizing code-based masking satisfies the same inequality as in Lemmas 8.1, 8.2 and 8.3. Therefore, we have that $I(K;\widehat{K}) \leq I(\mathbf{U};\mathbf{Y}|\mathbf{T}) \leq H(K)$, an information-theoretic upper bound for $I(K;\widehat{K})$ by the entropy of K, for instance, $H(K) = \ell$ under the uniform assumption.

Considering the Hamming weight leakages under AWGN, the simulation setting is the same as in the Boolean masking, except that: $\mathbf{y}^j \sim \mathcal{N}(w_H(S(\mathbf{t}^j \oplus k^j) \oplus (\alpha_1\mathbf{m}^j)) + w_H(\mathbf{m}^j), \sigma^2\mathbf{I}_q) \in \mathbb{R}^q$. Accordingly, Eqn. 8.16 is updated by replacing $f'(\mathbf{u}_i, m)$ by $f''(\mathbf{u}_i, m, \alpha_1) = w_H(\mathbf{u}_i \oplus \alpha_1 m) + w_H(m)$ for the zero-offset leakage:

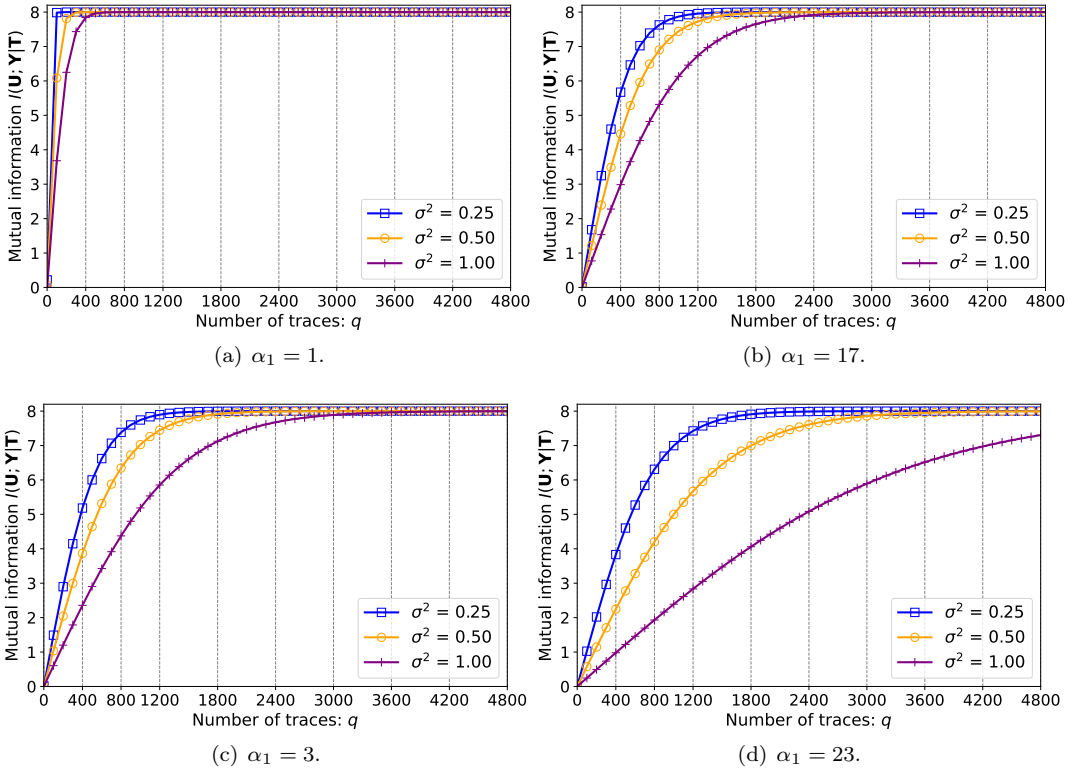
$$\log p(\mathbf{y}|\mathbf{u}) = -q\ell - \frac{q}{2}\log\left(2\pi\sigma^2\right) + \log\prod_{i=1}^{q}\sum_{m} e^{\frac{-\left(\mathbf{y}_i - w_H(\mathbf{u}_i \oplus f''(\mathbf{u}_i, m, \alpha_1))\right)^2}{2\sigma^2}}. \tag{8.18}$$

After inserting Eqn. 8.18 into Eqn. 8.11, we launch the Monte-Carlo simulation and the numerical results are shown in Fig. 8.8.

The takeaways from Fig. 8.8 are twofold. Firstly, $I(\mathbf{U};\mathbf{Y}|\mathbf{T})$ is well-bounded by $H(K) = 8$. The bound is tight, since given an enough amount of side-channel measures (q), we should have $I(\mathbf{U};\mathbf{Y}|\mathbf{T}) = H(K)$. Secondly, as already illustrated in Chaps. 4 and 7, different choices of the codes in IPM have distinct impact on the effectiveness of side-channel protection. We herein provide another argument such that $\alpha_1 = 23$ is one of the optimal codes for 2-share IPM. Additionally, compared with the state-of-the-art (e.g., [3, 37]), we complete the analysis of IPM by $I(\mathbf{U};\mathbf{Y}|\mathbf{T})$ rather than $I(U;Y|T)$, where the former enables concrete predication of number of traces to launch a successful attack.

8.6.2 Bounding the Probability of Success

In the presence of code-based masking, Fano's inequality also holds, meaning that $f_P(P_s) \leq I(\mathbf{U};\mathbf{Y}|\mathbf{T})$. By Theorem 8.1, the success rate P_s is straightforwardly linked to $I(\mathbf{U};\mathbf{Y}|\mathbf{T})$.



(a) $\alpha_1 = 1$. (b) $\alpha_1 = 17$.

(c) $\alpha_1 = 3$. (d) $\alpha_1 = 23$.

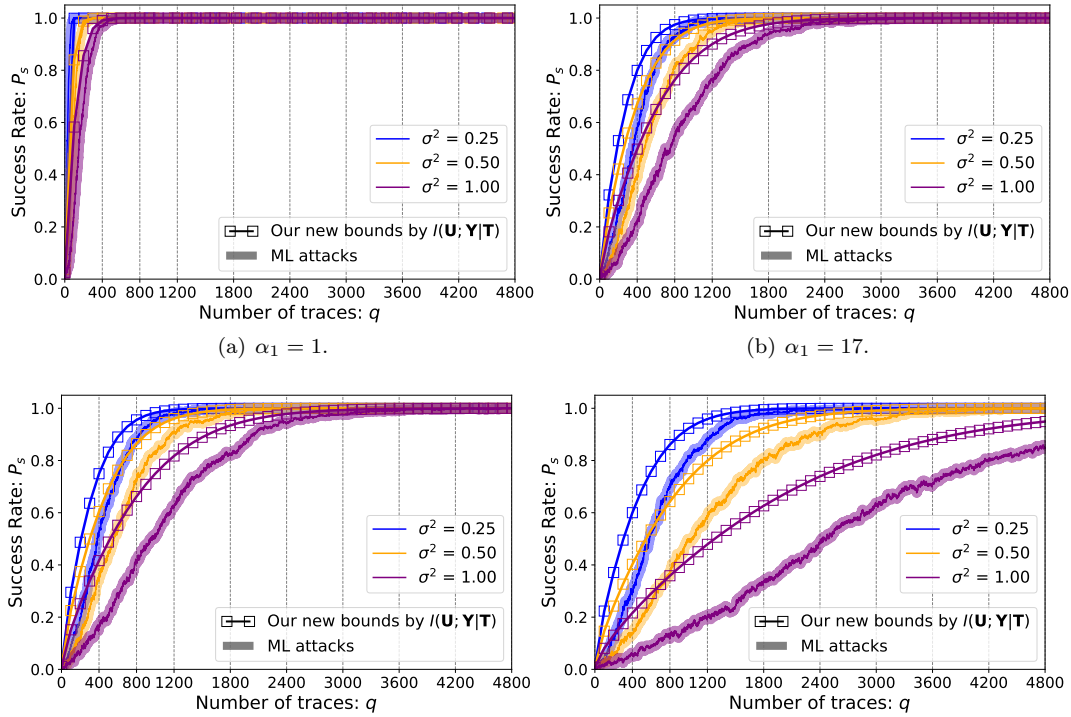
Figure 8.8: Numerical results of $I(\mathbf{U};\mathbf{Y}|\mathbf{T})$ under different choices of α_1 in IPM. Note that $\alpha_1 = 1$ corresponds to Boolean masking as shown in Fig. 8.5 for other levels of noise.

Therefore, we derive the same upper bound on P_s, the best probability of success an attack can obtain given a set of q measurements.

Similarly with the above setting, we can numerically verify the bound on success rate considering the Hamming weight leakage under AWGN. Those numerical results are depicted in Fig. 8.9, where the ML attack follows the maximum likelihood rule.

As shown in Fig. 8.9, first of all, the effectiveness of protection against SCA increases from $\alpha_1 = 1$ (corresponds to the Boolean masking) to $\alpha_1 = 23$. It can be explained by the coding-theoretic properties like the dual distance and the kissing number of the code $\mathcal{D}$ as IPM is non-redundant. Secondly, with the same noise level (σ^2), it is significantly more difficult to attack IPM with $\alpha_1 = 23$ than the Boolean one. The gap is much larger along with σ^2 increases, as frequently stated in demonstrating the security of masking schemes [128].

In above analysis, we derive the success rate of attack given a certain fixed number of traces. Conversely, given a certain value of P_s, we are able to predict the minimum number (lower



(a) $\alpha_1 = 1$. (b) $\alpha_1 = 17$.

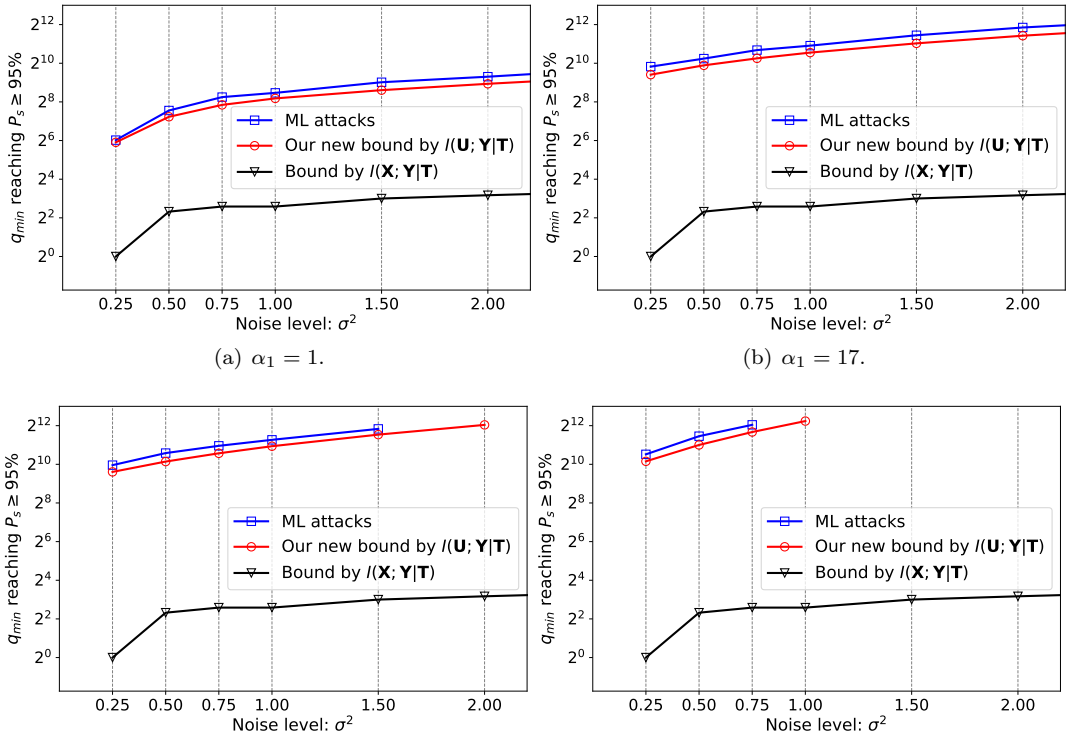
(c) $\alpha_1 = 3$. (d) $\alpha_1 = 23$.

Figure 8.9: Bounds on success rate P_s by $I(\mathbf{U};\mathbf{Y}|\mathbf{T})$ under different choices of α_1 in IPM.

bound) of traces needed to achieve this P_s. The twined problems are unified by Theorem 8.1, knowing the one direction gives the bound for the other one. Accordingly, inverse to Fig. 8.9, we can plot the predicted number of traces to achieve $P_s \geq 95\%$ as in Fig. 8.10. We also add the prediction by $I(\mathbf{X};\mathbf{Y}|\mathbf{T})$ [57, 58] in comparison. Note that several points in Fig. 8.10(c) and 8.10(d) are missing since $q_{\min}$ are already exceed 5000 in corresponding noise levels.

As shown in Fig. 8.10, the lower bound on $q_{\min}$ given by $I(\mathbf{U};\mathbf{Y}|\mathbf{T})$ is much tighter than that given by $I(\mathbf{X};\mathbf{Y}|\mathbf{T})$. Consequently, It is recommended to utilize the bound given by $I(\mathbf{U};\mathbf{Y}|\mathbf{T})$ rather than $I(\mathbf{X};\mathbf{Y}|\mathbf{T})$ in the presence of masking.

In order to show the exponential properties, we depict in Fig. 8.11 with the number of traces in normal scale. More apparently from Fig. 8.11, the inherent properties of higher algebraic complexity in IPM significantly improve the side-channel resistance when empowered by optimal linear codes. When moving to redundant code-based masking, for instance in SSS-base masking, those optimal codes also achieve the best protection against SCA, while bad codes may degrade the protection as already demonstrated in Chap. 7.



(a) $\alpha_1 = 1$. (b) $\alpha_1 = 17$.

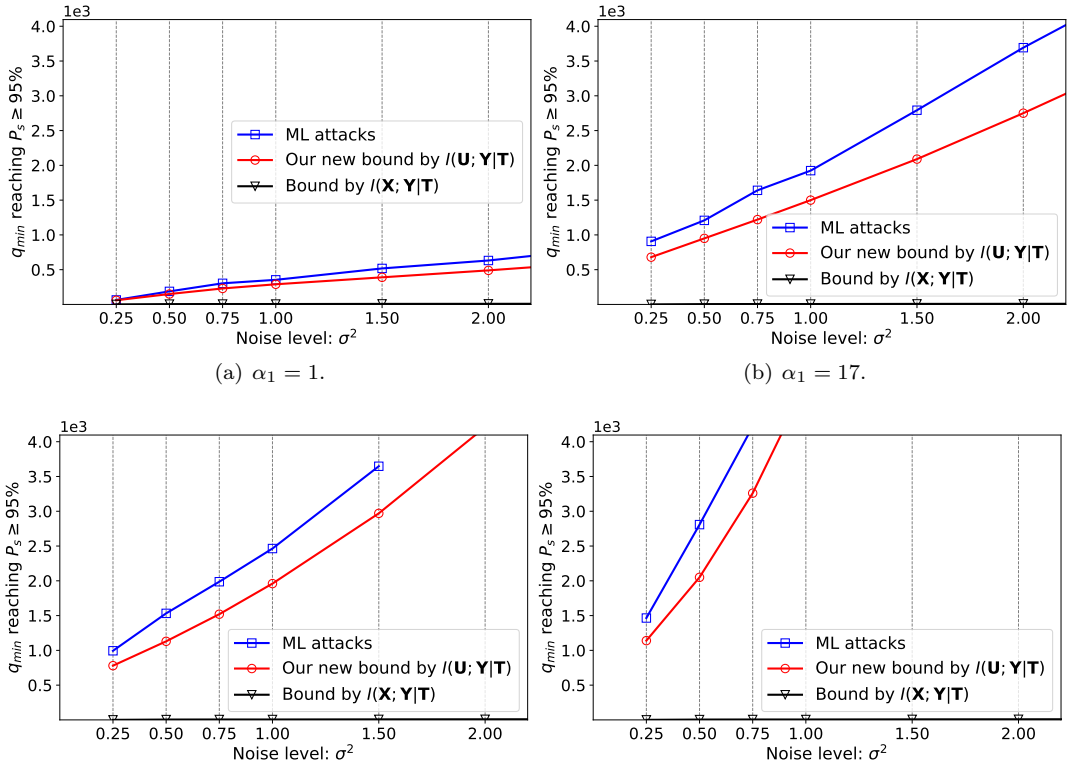
(c) $\alpha_1 = 3$. (d) $\alpha_1 = 23$.

Figure 8.10: Prediction of $q_{\min}$ achieving $P_s \geq 95\%$ under different choices of α_1 in IPM. Note that the number of traces are in $\log_2$ scale.

8.7 Conclusions

We derive security bounds for side-channel attacks in the presence of countermeasures. In this respect, we leverage the seminal framework from Chérisey et al. in TCHES 2019, and extend it to the case of a protection aiming at randomizing the leakage. Interestingly, the generalization allows to improve bounds compared to Chérisey et al's. Also, we improve on the computation method for the security metric, by resorting to a powerful probabilistic information estimation based on importance sampling.

Furthermore, we verify our information-theoretic bounds in the context of code-based masking. On the one hand, those bounds confirm again the advantages of code-based masking compared to the commonly used Boolean masking, when the former is equipped with optimal codes. On the other hand, those bounds also allow us to predict how successful can an ML-based distinguisher be: evaluated by either an upper bound on the success rate given a set of traces or a lower bound on the number of traces to succeed in recovering the secret key. In summary, our results provide



(a) $\alpha_1 = 1$. (b) $\alpha_1 = 17$.

(c) $\alpha_1 = 3$. (d) $\alpha_1 = 23$.

Figure 8.11: Prediction of $q_{\min}$ achieving $P_s \geq 95\%$ under different choices of α_1 in IPM.

quantitative bounds allowing for the theoretical (i.e., "pre-silicon") evaluation of protections applied on top of a cryptographic algorithm.

Part V

Generic Information-Theoretic Measures and Applications to Side-Channel Analysis

CHAPTER 9

Towards Exact Assessment of Side-Channel Leakage by α-Information

Measuring the leakages of sensitive variables is the key to evaluate the security level in many secrecy and privacy problems. In practice, an adversary can observe some "information" about the manipulation of secrets, e.g., she can get some physically observable leakages like cache timing variations, power consumption, electromagnetic radiations, etc. Those physical observations are particularly called side-channel leakages when targeting cryptographic systems. Therefore, the problem is how much information about a variable is carried in its side-channel leakages. In this chapter, we study this problem, instead of using Shannon information theory (Shannon entropy, mutual information, divergence, etc), in a more general sense by using Rényi entropies, Rényi divergences and alpha-information.

Part of results shown in this chapter has been presented in *IEEE Information Theory Workshop* (ITW 2021) [99].

Contents

9.1 Introduction

Since the seminal work [146] by Claude E. Shannon published in 1948, entropy and mutual information have been fundamental tools for measuring uncertainty and dependency in information theory. These information-theoretic tools have achieved great success in a large variety of domains and topics, including quantification, storage, and communication of digital information. Particularly, they have been widely used in cryptography [147], along with the concept of "perfect secrecy". Later on, several generalizations of entropy [1] and divergence have been proposed, wherein Alfred Rényi propose a parameterized one depending on α for $\alpha > 0$ and $\alpha \neq 1$ in 1961 [135], usually called Rényi entropies and divergences. Essentially, all these information-theoretic tools measure the intrinsic properties and connections between different distributions of variables (under certain assumptions) despite using various tools.

The generalization of Shannon information theory continues. On one hand, regarding the conditional version of Rényi entropies, at least six proposals have been come up with [1, 66, 148]. In particular, the one proposed by Arimoto [1], known as Arimoto-Rényi conditional entropy, is a good definition possessing several fundamental features. On the other hand, regarding the generalization of mutual information, the one proposed by Sibson [148], known as Sibson's mutual information, is perhaps the most preferred generalization of classical mutual information and has been applied in various scenarios [63, 64, 122, 137, 157, 160].

However, there is no widely accepted definition on conditional mutual information. Recently, two proposals are [63, 157], both of which do not have the uniform expansion property (UEP) as shown in [99]. In this respect, our proposal in [99] is featured with UEP and other good properties which are useful in the context of quantifying information leakage in side-channels.

[1] We call entropy, exclusively for Shannon entropy except stated otherwise.

9.1.1 Information-Theoretic Measures in Side-Channel Analysis

Mutual information has been extensively exploited in evaluating side-channel security of practical cryptographic implementations. Typically, as a theoretical tool, it can be applied as a side-channel distinguisher [9, 74, 84, 111, 163], or a leakage evaluation tool [41, 57, 151]. Interestingly, from a framework of communication channel model, an optimal side-channel distinguisher is derived in [84] in the sense that it can make the best use of leakage. Later on, the same framework is explored in universally upper bounding the success rate of any attacks [41, 57], or conversely giving lower bounds on the number of measurements to launch a successful attack (e.g., with a success probability $P_s \geq 95\%$).

However, there is a gap between the success rate of an optimal ML-based attack and the upper bounds given by mutual information, and the gap is even widened when other estimated bounds are adopted [57]. We will show how the conditional α-information closes the gap and provides an exact bound on attacks in this chapter.

9.2 Contributions

In this chapter, we aim at quantifying information leakage by utilizing generalized information-theoretic measures and provide numerical simulation results in the context of guessing the value of a discrete variable from its side-channel leakage. In this regard, we first show how conditional information-theoretic tools are applied to quantify the amount of information carried by the leakage in a Hamming weight leakage model. We present numerical results in both noiseless and noisy scenarios. The takeaway is that the conditional α-entropy brings different bounds on the information leakage when equipped with various α. In particular, the larger value of α gives a much tighter bound.

Secondly, we present a full spectrum of application α-information in side-channel analysis equipped with different α. In particular, we fully fill the gap between two worlds, namely information-theoretic measures and side-channel attacks by applying conditional α-information. Notably, it is the first time that we are able to predict exactly the success rate of the maximum-likelihood (ML)-based attack in SCA, or conversely to predict the minimum number of traces to launch a successful key-recovery attack. Furthermore, we also prove that when α approaches positive infinity, the ML-based attack converges to conditional α-information between the leakage and the sensitive variable. Therefore, this conditional α-information will provide the best upper bound on the success rate of any attacks.

9.3 Quantifying Hamming Weight Leakages by Rényi Entropy

Considering two random variables X and Y, where Y denotes the general leakage from the sensitive variable X. For instance, Y is the side-channel leakage of sensitive variables during its computation, storage, or even some micro-architectural caching information. An abstract overview of the leakage model is illustrated in Fig. 9.1, in which f can be any deterministic or probabilistic function on X and N denotes some additive noises.

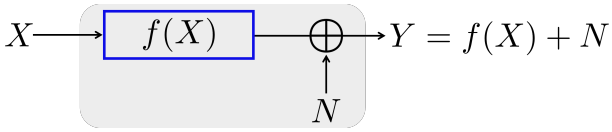


Figure 9.1: Leakage model: the sensitive variable X and the leakage Y with some noise N.

We first present several definitions. Let $p(x)$ and $p(y)$ be the probability distributions (e.g., in discrete case) or probability density functions (e.g., in continuous case) of X and Y, respectively. We recall the definition of Rényi entropy as follows.

Definition 9.1 (Rényi entropy [135])**.** The α-entropy, or Rényi entropy of order $\alpha \geq 0$ is defined for $0 < \alpha < +\infty$ and $\alpha \neq 1$ as

$$H_\alpha(X) = \frac{1}{1-\alpha} \log \mathbb{E}\big(p(X)^{\alpha-1}\big) = \frac{1}{1-\alpha} \log \sum\!\!\!\!\!\!\int p(x)^\alpha. \tag{9.1}$$

Remark 9.1. Considering different choices of α, we shall recover various entropies [137]:

- Hartley's entropy (max-entropy): taking $\alpha \to 0$ gives $H_0(X) = \log |\mathrm{Supp} X|$ where $|\mathrm{Supp} X| = \int\!\!\!\!\!\!\int_{p(x)>0} 1$ denotes the volume/cardinality of the support of the distribution.
- Shannon entropy: taking $\alpha \to 1$ recovers $H_1(X) = \mathbb{E} \log \frac{1}{p(X)} = \int\!\!\!\!\!\!\int p(x) \log \frac{1}{p(x)} = H(X)$.
- Min–entropy: taking $\alpha \to \infty$ leads to $H_\infty(X) = \log \frac{1}{\sup p}$ where the notation $\sup p$ denotes the ∞-norm $\|p\|_\infty$. In discrete case, we have $H_\infty(X) = \log \frac{1}{\max_x p(x)}$.

Then defining the conditional α-entropy $H_\alpha(X|Y)$ implies an expectation over Y. We use in this chapter the Arimoto's conditional α-entropy as it is attributed with some good properties:

Definition 9.2 (Conditional Arimoto-Rényi Entropies [1])**.** The conditional α-entropy or conditional Arimoto-Rényi entropy of order $\alpha \geq 0$ is defined for $0 < \alpha < +\infty$ and $\alpha \neq 1$ as

$$H_\alpha(X|Y) = \frac{\alpha}{1-\alpha} \log \mathbb{E}\|p(\cdot|Y)\|_\alpha = \frac{\alpha}{1-\alpha} \log \sum\!\!\!\!\!\!\int_y p(y) \Big(\sum\!\!\!\!\!\!\int_x p(x|y)^\alpha\Big)^{1/\alpha}. \tag{9.2}$$

Remark 9.2. Similarly, we shall recover conditional version of above entropies as follows.

- Conditional max-entropy (conditional 0-entropy): taking $\alpha \to 0$ gives $H_0(X|Y) = \sup_y H_0(X|Y=y) = \log\big(\sup_y |\mathrm{Supp}X|Y=y|\big)$ where the $\sup_y$ is the infinity norm as above.
- Conditional Shannon entropy (conditional 1-entropy): taking $\alpha \to 1$ gives Shannon's conditional entropy: $H_1(X|Y) = H(X|Y)$.
- Conditional min-entropy (conditional ∞-entropy): taking $\alpha \to \infty$ leads to $H_\infty(X|Y) = \log \frac{1}{\mathbb{E}_y \sup_x p(x|y)}$ where the $\sup_x$ is the infinity norm as above.

We aim at characterizing the leakages by utilizing the (conditional) guessing entropy and Rényi entropies [142, 159], then build quantitative connections and upper/lower bounds between the entropies and the success rates (SR). In particular, for quantifying the information that Y brings on X, we focus on following four metrics.

- Guessing entropy
- Conditional Shannon entropy
- Conditional Arimoto-Rényi entropies (the α-entropy [161])
- Success rate (or success probability)

Assume that the discrete variable $X \in \mathbb{F}_{2^\ell}$ is uniformly distributed over $\mathcal{X}$ with cardinality $|\mathcal{X}| = M$ and $N = 0$ in noiseless scenario, the marginal and joint distributions are:

$$\mathbb{P}(X=x) = 2^{-\ell} = \frac{1}{M}, \qquad \mathbb{P}(Y=y) = \frac{|f^{-1}(y)|}{M}, \qquad \mathbb{P}(X,Y) = \frac{1}{M} \cdot \mathbf{1}_{y=f(x)}. \tag{9.3}$$

The conditional probability distributions are as follows.

$$\mathbb{P}(Y|X) = \begin{cases} 1 & \text{if } y = f(x) \\ 0 & otherwise \end{cases}, \quad \mathbb{P}(X|Y) = \frac{\mathbb{P}(Y|X)\mathbb{P}(X)}{\mathbb{P}(Y)} = \begin{cases} \frac{1}{|f^{-1}(y)|} & \text{if } y = f(x) \\ 0 & otherwise \end{cases}. \tag{9.4}$$

Next, the maximum probability of success is defined as follows.

$$P_s(X) = \max_x p(x). \tag{9.5}$$

For any leakage function f, we have: $P_s(X|Y) = \mathbb{E}_y \max_x p(x|Y) = \sum_y p(y) \cdot \frac{1}{|f^{-1}(y)|} = \sum_y \frac{1}{M} = \frac{M'}{M} \geq \frac{1}{M}$, where $M' = |\mathcal{Y}|$ and $M = |\mathcal{X}|$.

In the following, we assume that the sensitive variable X leaks the Hamming weight model, which is the well-studied leakage model in side-channel analysis. The reason is that, hardware implementations leak bits in parallel, hence the leakage is the sum of the registers state bits, that is the Hamming weight of the register contents. The schematic is shown in Fig. 9.1, where we have $f = w_H$ for Hamming weight leakages.

9.3.1 Guessing with Noiseless Leakages

Let $f(X) = w_H(X)$ where $N = 0$ and $|\mathcal{X}| = M = 2^\ell$ for the sake of calculation. Hence,

$$\mathbb{P}(x) = \frac{1}{2^\ell}\ , \qquad \mathbb{P}(y) = \frac{\binom{\ell}{y}}{2^\ell}\ , \qquad \mathbb{P}(x|y) = \frac{\mathbb{P}(Y|X)\mathbb{P}(X)}{P(Y)} = \frac{\mathbf{1}_{y=w_H(x)}}{\binom{\ell}{y}}\ . \tag{9.6}$$

We focus on quantifying the reduction of uncertainty of X knowing its Hamming weight leakages Y. The four metrics are then calculated as follows.

- **Conditional guessing entropy**.

$$G(X|Y) = \sum_y \mathbb{P}(y) \sum_x x \cdot \mathbb{P}(x|y) = \sum_y \frac{\binom{\ell}{y}}{2^\ell} \left(\sum_{w_x = y} \frac{1}{\binom{\ell}{y}} \cdot x \right) = \frac{1}{2} + \frac{1}{2^{\ell+1}} \binom{2\ell}{\ell}. \tag{9.7}$$

- **Conditional Shannon entropy**.

$$H(X|Y) = -\sum_{x,\,y} p(x,y) \log p(x|y) = 2^{-\ell} \sum_y \binom{\ell}{y} \cdot \log \binom{\ell}{y}\ . \tag{9.8}$$

- **Conditional Arimoto-Rényi Entropies**.

$$H_\alpha(X|Y) = \frac{\alpha}{1-\alpha} \log \sum_y p(y) \left(\sum_x p(x|y)^\alpha \right)^{\frac{1}{\alpha}} = \frac{\alpha}{\alpha - 1} \left(\ell - \log \sum_y \binom{\ell}{y}^{\frac{1}{\alpha}} \right). \tag{9.9}$$

- **Conditional success probability**.

$$P_s(X|Y) = \mathbb{E}_Y \max_x p(x|Y) = \frac{M'}{M} = \frac{\ell + 1}{2^\ell}. \tag{9.10}$$

By using the upper bound from Fano's inequality [65] and the lower bound $H(X|Y) \geq \varphi^*(P_s(X|Y))$ where

$$\varphi^*(s) = \lfloor \tfrac{1}{s} \rfloor \big(s \lceil \tfrac{1}{s} \rceil - 1\big) \log \lfloor \tfrac{1}{s} \rfloor + \Big(1 - \lfloor \tfrac{1}{s} \rfloor \big(s \lceil \tfrac{1}{s} \rceil - 1\big)\Big) \log \lceil \tfrac{1}{s} \rceil, \tag{9.11}$$

and $H_\alpha(X|Y) \geq \frac{\alpha}{1-\alpha} \log \phi^*_\alpha(P_s(X|Y))$, where

$$\phi^*_\alpha(s) = \Big(\lceil \frac{1}{s} \rceil s - 1 \Big) \lfloor \frac{1}{s} \rfloor^{1/\alpha} + \Big(1 - \lfloor \frac{1}{s} \rfloor \Big(\lceil \frac{1}{s} \rceil s - 1 \Big)\Big) \lceil \frac{1}{s} \rceil^{\frac{1-\alpha}{\alpha}} \tag{9.12}$$

(proposed by Sason et al. [142]), we numerically show the conditional Shannon and Rényi entropies of X as Fig. 9.2 and Fig. 9.3. Specifically, the upper bound of Rényi entropy is highly dependent on the α. With α much larger than 1.0, the marked region is much smaller than the region with $\alpha < 1.0$.

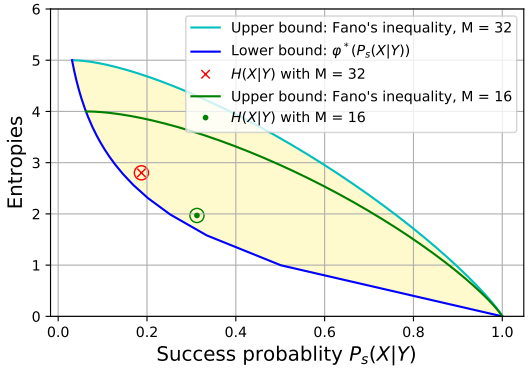
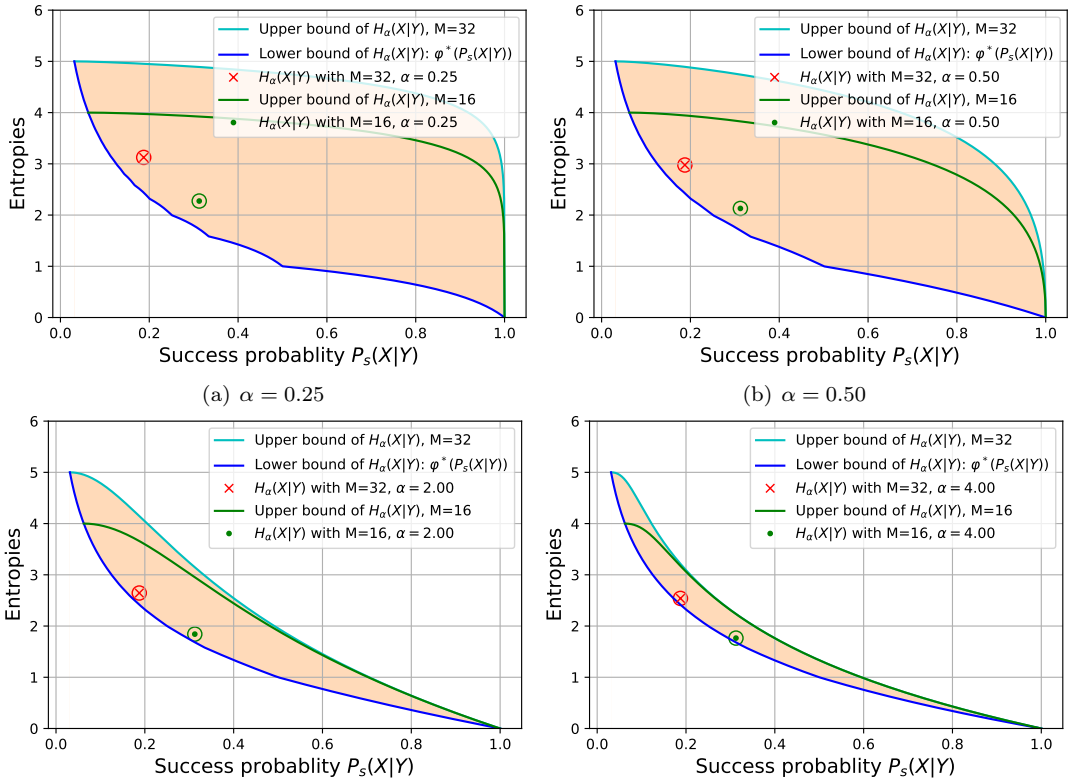


Figure 9.2: Conditional Shannon entropy of guessing X knowing Y.



(a) $\alpha = 0.25$ (b) $\alpha = 0.50$

(c) $\alpha = 2.00$ (d) $\alpha = 4.00$

Figure 9.3: Conditional Rényi entropies of guessing X knowing Y with different α.

9.3.2 Guessing with Noisy Leakages

In fact, noise is the intrinsic part in the side-channel leakages, like in the power consumption and electromagnetic radiations. Thus we consider the noisy leakages in a classic way by assuming the noise is the additive white Gaussian noise (AWGN), which is a common noise model to

mimic the effect of many random processes.

We assume that $N = \varphi(z) \sim \mathcal{N}(0, \sigma^2)$ and $\varphi(z) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{z^2}{2\sigma^2}}$ which is a nonincreasing function of $|z|$. Then, we have:

$$\begin{aligned} \mathbb{P}(X = x) &= \frac{1}{M} \ , \\ \mathbb{P}(Y = y) &= \sum_x p(x) \cdot p(y|x) = \frac{1}{M} \sum_x \varphi(y - f(x)), \\ \mathbb{P}(y|x) &= \varphi(y - f(x)) \ , \\ \mathbb{P}(x|y) &= \frac{p(y|x)p(x)}{p(y)} = \frac{\varphi(y - f(x))}{\sum_{x'} \varphi(y - f(x'))} \ . \end{aligned} \tag{9.13}$$

In addition, the maximum conditional probability of success is computed as follows.

$$\begin{aligned} P_s = \mathbb{E} \max_x p(x|Y) &= \int \left(\frac{1}{M} \sum_{x'} \varphi(y - f(x')) \right) \times \frac{\varphi(\min_x |y - f(x)|)}{\sum_{x'} \varphi(y - f(x'))} \, \mathrm{d}y \\ &= \frac{1}{M} \int \varphi(min_x |y - f(x)|) \, \mathrm{d}y \\ &= \frac{1}{M} \int \varphi(y - f(x^*(y))) \, \mathrm{d}y \qquad (\text{where } x^*(y) = \arg\min_x |y - f(x)|) \\ &= \frac{M'}{M} - 2\frac{M' - 1}{M} Q\left(\frac{\Delta/2}{\sigma} \right) \ , \end{aligned} \tag{9.14}$$

where M' is the cardinality of $f(x)$ and $Q(x) = \frac{1}{2} erfc\left(\frac{x}{\sqrt{2}}\right)$, Δ equals the regularly spaced distance of $f(x)$, for instance $\Delta = 1$ in the Hamming weight model.

Thus, we calculate the conditional Shannon and Arimoto-Rényi entropies as follows.

- **Conditional Shannon entropy**. Given $\mathsf{I}(X;Y) = H(X) - H(X|Y) = h(Y) - h(Y|X)$ and the differential entropy of Gaussian variable Y is $h(Y) = \frac{1}{2} \log(2\pi e\sigma^2)$, the conditional Shannon entropy is:

$$\begin{aligned} H(X|Y) &= H(X) - h(Y) + h(Y|X) \\ &= \log M + \frac{\log(2\pi e\sigma^2)}{2} - \int p(y) \log \frac{1}{p(y)} \, \mathrm{d}y. \end{aligned} \tag{9.15}$$

- **Conditional Arimoto-Rényi entropies**.

$$\begin{aligned} H_\alpha(X|Y) &= \frac{\alpha}{1-\alpha} \log \sum_y p(y) \big(\sum_{x \in \mathcal{X}} p(x|y)^\alpha \big)^{1/\alpha} \\ &= \frac{\alpha}{1-\alpha} \log \int_y p(y) \big(\sum_{x \in \mathcal{X}} p(x|y)^\alpha \big)^{1/\alpha} \, \mathrm{d}y \\ &= \frac{\alpha}{1-\alpha} \log \int_y \big(\sum_{x \in \mathcal{X}} p(x,y)^\alpha \big)^{1/\alpha} \, \mathrm{d}y \\ &= \frac{\alpha}{1-\alpha} \log \frac{1}{M} \int_y \big(\sum_{x \in \mathcal{X}} \varphi(y - f(x))^\alpha \big)^{1/\alpha} \, \mathrm{d}y \ . \end{aligned} \tag{9.16}$$

With function ***scipy.integrate.quad*** in `Python`, we numerically investigate the remaining information of X with knowing noisy Hamming weight leakages Y where the noise is the additive Gaussian noise. Specifically, for conditional Shannon entropy, its upper bound is given by Fano's inequality and lower bound is: $H(X|Y) \geq \varphi^*(P_s(X|Y))$ where

$$\varphi^*(s) = \lfloor \tfrac{1}{s} \rfloor \big(s\lceil \tfrac{1}{s} \rceil - 1\big) \log \lfloor \tfrac{1}{s} \rfloor + \Big(1 - \lfloor \tfrac{1}{s} \rfloor \big(s\lceil \tfrac{1}{s} \rceil - 1\big)\Big) \log \lceil \tfrac{1}{s} \rceil. \tag{9.17}$$

While for conditional Arimoto-Rényi entropies, the upper bound and lower bound are given by Sason et al. [142]. Particularly, $H_\alpha(X|Y) \geq \frac{\alpha}{1-\alpha} \log \phi^*_\alpha(P_s(X|Y))$, where

$$\phi^*_\alpha(s) = \Big(\lceil \frac{1}{s} \rceil s - 1\Big) \lfloor \frac{1}{s} \rfloor^{1/\alpha} + \Big(1 - \lfloor \frac{1}{s} \rfloor \Big(\lceil \frac{1}{s} \rceil s - 1\Big)\Big) \lceil \frac{1}{s} \rceil^{\frac{1-\alpha}{\alpha}}, \tag{9.18}$$

and it does not depend on M.

With noise level $\sigma \in [0.05, 5.00]$, the conditional Shannon entropies with different M are as in Fig. 9.4. Specially note that for $M = 4$ and noise level $\sigma = 0.05$, the lower bound is 0.5, while the conditional Shannon entropy is 0.5000000000000003, which is very close but greater than the lower bound.

From the Fig. 9.4, the conditional Shannon entropy is increasing along with M, which in fact impacts the success probability P_s. In the noisy scenarios, adding noise increases the difficulty of guessing, resulting in that the conditional entropy is increasing along with noise level and approaching $H(X)$. As a result, the numerical results are consistent with theoretical analysis.

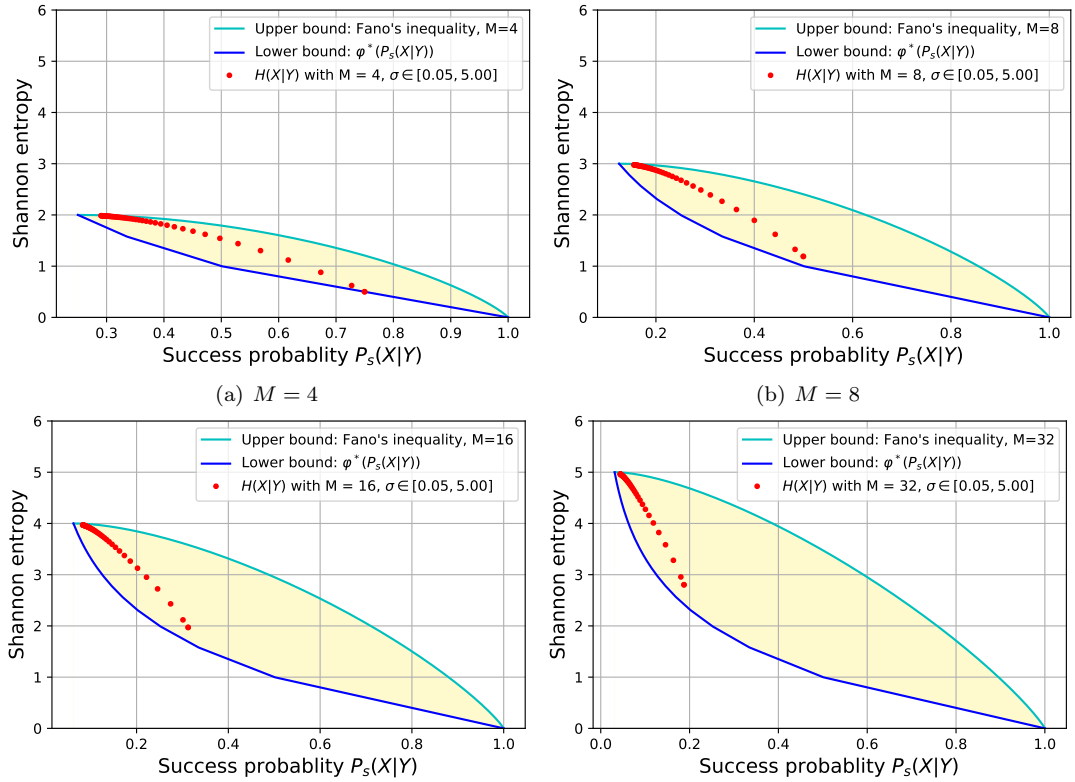
With the same setting of noise level, the conditional Arimoto-Rényi entropies with different α are plotted as in Fig. 9.5. It is interesting to show that with greater α, the upper bounds are tighter. This result is the same as in Fig. 9.3. The conditional Arimoto-Rényi entropies are increasing along with noise level σ as expected. But the shape of the conditional entropies curve changed from concave to convex, and approaching to lower bound with an increase of α.

With observations from Fig. 9.5, we recommend to use Arimoto-Rényi entropies with larger α under the Hamming weight leakages. It is worthy noting that $H_{1/2}$ is highly related to guessing entropy [45], and can be utilized to estimate the bounds for key ranking in a fast and scalable way.

9.4 Good Definition of α-Information

We extend our investigation to α-information [1]. It is well-known that α-information is related to α-divergence. Therefore, we first recall the definition of the latter.

[1] We remove "mutual" in α-information since in general: $I_\alpha(X;Y) \neq I_\alpha(Y;X)$.



(a) $M = 4$ (b) $M = 8$

(c) $M = 16$ (d) $M = 32$

Figure 9.4: Conditional Shannon entropies of guessing X knowing Y with different M, which indicating different probabilities of $P_s(X|Y)$.

Definition 9.3 (Rényi Divergence [135, 137, 159])**.** Let p, q be distributions such that $\frac{p}{q}$ is well defined. The α-divergence, or Rényi divergence of order $\alpha \geq 0$ is defined for $0 < \alpha < +\infty$ and $\alpha \neq 1$ as

$$D_\alpha(p\|q) = \frac{1}{\alpha - 1} \log \mathbb{E}\Big(\frac{q(X)}{p(X)}\Big)^{1-\alpha} = \frac{1}{\alpha - 1} \log \sum\!\!\!\!\!\!\int p^\alpha(x) q^{1-\alpha}(x), \tag{9.19}$$

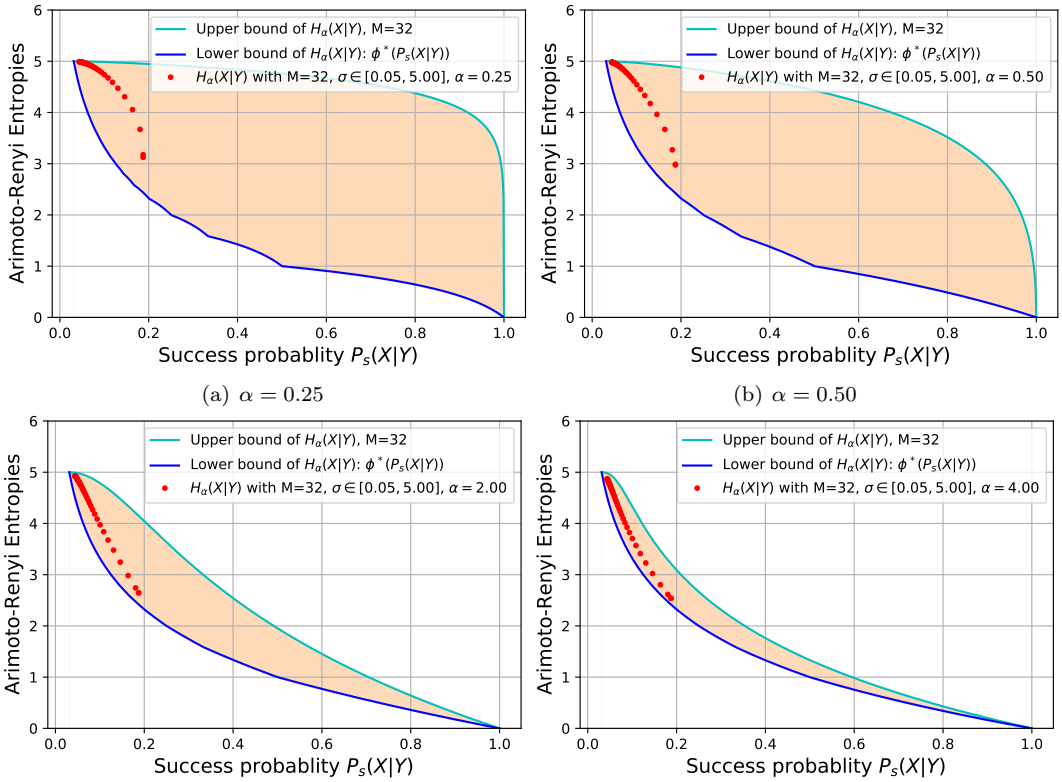
where $X \sim p(x)$. In particular, if p and q are binary distributions, say $(p, 1-p)$ and $(q, 1-q)$, respectively, then the binary α-divergence is:

$$d_\alpha(p \parallel q) = \frac{1}{\alpha - 1} \log\big(p^\alpha q^{1-\alpha} + (1-p)^\alpha (1-q)^{1-\alpha}\big). \tag{9.20}$$

Then we present Sibson's α-information and identity.

Definition 9.4 (Sibson's α-Information [53, 148, 160])**.** Let p, q be distributions such that $\frac{p}{q}$ is well defined as above. The Sibson's α-information of order $\alpha \geq 0$ is defined for $0 < \alpha < +\infty$ and $\alpha \neq 1$ as

$$I_\alpha(X;Y) = \min_{q_Y} D_\alpha(p_{Y|X}\|q_Y|p_X) = \min_{q_Y} D_\alpha(p_{X,Y}\|p_X q_Y). \tag{9.21}$$



(a) $\alpha = 0.25$ (b) $\alpha = 0.50$

(c) $\alpha = 2.00$ (d) $\alpha = 4.00$

Figure 9.5: Conditional Arimoto-Rényi entropies of guessing X knowing noisy Y with different $\alpha \in [0.25, 0.50, 2.00, 4.00]$, and also different noise level.

Explicitly, the closed-form formula for Sibson's α-information is:

$$I_\alpha(X;Y) = \frac{\alpha}{\alpha - 1} \log \sum_y \Big(\sum_x p(x) p^\alpha(y|x) \Big)^{1/\alpha}. \tag{9.22}$$

Definition 9.5 (Sibson's Identity [148, 160])**.** On the basis of above definition, we have Sibson's identity:

$$I_\alpha(X;Y) = D_\alpha(p_{Y|X} \| q_Y | p_X) - D_\alpha(q_Y^* \| q_Y) \tag{9.23}$$

for any probability distribution q_Y, where q_Y^* is the minimizing probability distribution such that $I_\alpha(X;Y) = \min_{q_Y} D_\alpha(p_{Y|X} \| q_Y | p_X) = D_\alpha(p_{Y|X} \| q_Y^* | p_X)$.

It is worth mentioning that Sibson's α-information satisfies the following basic properties [122, 137], which are seamlessly connected to Shannon mutual information.

- Shannon mutual information: taking $\alpha \to 1$ recovers Shannon's mutual information: $I_1(X;Y) = I(X;Y)$.
- Independence Characterization: $I_\alpha(X;Y) \geq 0$ with equality iff $X \perp\!\!\!\perp Y$.

- Uniform expansion property (UEP): if X is discrete uniformly distributed ($p(x) = \frac{1}{M}$), we obtain $I_\alpha(X;Y) = \log M - H_\alpha(X|Y)$. This links α-information to conditional α-entropy.

- Data processing inequality (DPI): if $X-Y-Z$ forms a Markov chain, we have the data "post-processing" inequality (post-processing cannot increase information): $I_\alpha(X;Z) \leq I_\alpha(X;Y)$, and the data "pre-processing inequality" (pre-processing cannot increase information): $I_\alpha(X;Z) \leq I_\alpha(Y;Z)$.

In summary, Sibson's α-information provides a continuous extension from Shannon mutual information to the parametric one by order α for $\alpha > 0$ and $\alpha \neq 1$. However, in the context of side-channel analysis, an adversary usually is allowed to know some public information, for instance plaintexts or ciphertexts when targeting cryptographic algorithms or implementations. Therefore, it is critical to define the conditional version of α-information.

9.4.1 Extending to Conditional α-Information

As a natural continuation of the definitions in the preceding section, 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 Eqn. 9.22, the expression of Sibson's (unconditional) α-information.

Rényi entropy and divergence are well-known generalizations of Shannon's entropy and Kullback-Leibler divergence:

Definition 9.6 (Compact Representation of Rényi Entropy and Divergence [99])**.** Assume that either $0 < \alpha < 1$ or $1 < \alpha < +\infty$. The α-entropy of a probability distribution P and α-divergence of P from Q are defined as

$$\begin{aligned} H_\alpha(P) &= \tfrac{\alpha}{1-\alpha} \log \|p\|_\alpha, \\ D_\alpha(P\|Q) &= \tfrac{1}{\alpha-1} \log \langle p\|q\rangle_\alpha^\alpha, \end{aligned} \tag{9.24}$$

where we have used the special notation:

$$\|p\|_\alpha = \left(\int p^\alpha d\mu\right)^{1/\alpha}, \qquad \langle p\|q\rangle_\alpha = \left(\int p^\alpha q^{1-\alpha} d\mu\right)^{1/\alpha}, \tag{9.25}$$

with the following convention: All considered probability distributions P, Q possess a dominating measure μ such that $P \ll \mu$ and $Q \ll \mu$, the corresponding lower-case letters p, q are densities of P, Q with respect to μ.

Therefore, we shall have the following definition for α-information.

Definition 9.7 (Conditional α-Information, Closed-Form Definition [99])**.** The α-information between random variables X and Y knowing Z is:

$$I_\alpha(X;Y|Z) = \frac{\alpha}{\alpha-1}\log \mathbb{E}_Z \mathbb{E}_{Y|Z}\langle p_{X|YZ}\|p_{X|Z}\rangle_\alpha = \frac{\alpha}{\alpha-1}\log \mathbb{E}_{YZ}\langle p_{X|YZ}\|p_{X|Z}\rangle_\alpha. \tag{9.26}$$

More explicitly, it is equivalent to:

$$I_\alpha(X;Y|Z) = \frac{\alpha}{\alpha-1}\log \iint p(z)\iint \left(\int p(x|z)p^\alpha(y|x,z)\,\mathrm{d}\mu_X(x)\right)^{\frac{1}{\alpha}}\mathrm{d}\mu_Y(y)\,\mathrm{d}\mu_Z(z). \tag{9.27}$$

To the best of our knowledge, this definition has not been considered elsewhere.

Interestingly, we have the following property for the impact of order α in conditional α-information.

Lemma 9.1. *Given fixed distributions of X, Y given Z, $I_\alpha(X,Y|Z)$ is non-decreasing in α, in particular,*

$$I_0(X;Y|Z) \le I_1(X;Y|Z) = I(X;Y|Z) \le I_2(X;Y|Z) \le I_\infty(X;Y|Z), \tag{9.28}$$

where $I(X;Y|Z)$ is the Shannon mutual information, and we call $I_2(X;Y|Z)$ the quadratic α-information as in convention. Additionally, we call $I_\infty(X;Y|Z)$ the (conditional) maximal information [1] *as $\alpha\to\infty$.*

Proof. Since α-divergence is non-decreasing in α, if $\alpha<\beta$, then $D_\alpha(P_{XYZ}\|P_{X|Z}Q_{YZ}) < D_\beta(P_{XYZ}\|P_{X|Z}Q_{YZ})$ given a distribution of X,Y and Z. By the conditional Sibson's identity, we have that $I_\alpha(X;Y|Z) \le D_\alpha(P_{XYZ}\|P_{X|Z}Q_{YZ}) \le D_\beta(P_{XYZ}\|P_{X|Z}Q_{YZ})$. Finally, taking Q^*_{YZ} for minimization gives $D_\beta(P_{XYZ}\|P_{X|Z}Q^*_{YZ}) = I_\beta(X;Y|Z)$. □

In the following, we present some important properties of this conditional α-information.

9.4.2 Basic Properties

The conditional α-information in Def. 9.7 enjoys three important properties, namely consistency, UEP and DPI. Note that we refer the interested reader to [99] for detailed proofs of above properties.

Property 1 (Consistency of Conditional α-Information w.r.t. α-Information [99])**.** *If Z is independent of (X,Y) then $I_\alpha(X;Y|Z) = I_\alpha(X;Y)$.*

Property 2 (UEP for Conditional α-Information [99])**.** *If $U\sim\mathcal{U}(M)$ is uniformly distributed independent of Z, then*

$$I_\alpha(U;Y|Z) = H_\alpha(U) - H_\alpha(U|YZ) = \log M - H_\alpha(U|YZ). \tag{9.29}$$

[1] We use term *maximal information* to highlight the essential meaning of "information" in reducing uncertainty, e.g., in guessing games, side-channel analysis, etc. Another similar notion proposed in [89] is called *maximal leakage*. In a nutshell, *maximal leakage* shall be larger than *maximal information* in conditional scenarios.

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

Property 3 (DPI for Conditional α-Information [99])**.** *If $W - X - Y - Z$ forms a conditional Markov chain given T, then*

$$I_\alpha(X;Y|T) \geq I_\alpha(W;Z|T). \tag{9.30}$$

It is noteworthy that, firstly, the consistency property provides a continuous connection to unconditional Sibson's α-information. Secondly, we are especially interested in UEP since in cryptographic applications, the sensitive variable is usually discrete and uniformly distributed, say $X \sim \mathcal{U}(M)$. Therefore, it enables us to derive some straightforward but non-trivial bounds on how much information an adversary could infer knowing certain leakages. At last, DPI allows us to apply the conditional α-information to several communication channel-based frameworks [57, 58, 84]. We shall therefore expect that conditional α-information will bring some insights in practical scenarios.

9.5 Applications in Side-channel Analysis

As we have demonstrated in previous chapters, side-channel analysis (SCA) is a very powerful attacks against cryptographic implementation. In 2009, Standaert et al. establish a connection between side-channel analysis and information theory for the first time [151]. Then it is exploited by [84] to derive the optimal distinguisher and by [57] to obtain generic and universal bounds on how successful an optimal distinguisher can be in context of SCA. In this section we will generalize the results of [57] by Rényi information measures, particularly the conditional α-information, and deduce new upper bounds for the probability of success of side-channel attacks.

9.5.1 Side-Channel in a Communication Channel View

Recall that the secret key is denoted as K, and the plaintext or ciphertext is T, which is the input or output of the cryptographic implementation. By cryptographic operations, K and T are "encoded", and processed by some leakage function, producing a sensitive variable X. Then X is leaked along with inherent noise N in the channel, denoted as the noisy leakage Y. From a perspective of an adversary, she exploits Y to recover the key by certain side-channel distinguishers, resulting in $\widehat{K}$ as a guess of K.

The communication channel view of side-channel analysis is shown in Fig. 9.6: In particular, we assume that,

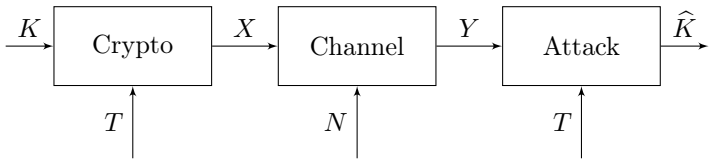


Figure 9.6: A communication channel view of side-channel analysis.

- K is uniformly distributed over $\mathcal{K} = \mathbb{F}_{2^\ell} = \{0, \ldots, 2^\ell - 1\}$. Denote $M = |\mathcal{K}| = 2^\ell$.
- T is independent with K, which is assumed to be available, e.g., in a key-recovering attack.
- the leakage function is a deterministic function, but not necessarily known to the attacker.

Based on this model, we have the following observation.

Lemma 9.2 (Conditional Markov Chains [57])**.** *The communication channel we described above admits the following Markov chains when given T:*

$$K \to Y \to \widehat{K}; \quad K \to X \to Y \tag{9.31}$$

Proof. When T is known, it is clear that $p(\widehat{k}|y,k) = p(\widehat{k}|y)$ and $p(y|x,k) = p(y|x)$. □

9.5.2 Upper Bounding the Success of Probability for Any Attacks

Considering the communication channel framework in Fig. 9.6, we have the following lemma.

Lemma 9.3.

$$I_\alpha(K, Y|T) = I_\alpha(X, Y|T) \tag{9.32}$$

Proof. Since $K \to X \to Y$ is a Markov chain given T, using Eqn. 9.30 we have $I_\alpha(K, Y|T) \leq I_\alpha(X, Y|T)$.

Conversely, when T is known, X is a deterministic function of K, which means $X \to K \to Y$ also forms an Markov chain given T. Again from Eqn. 9.30, we have $I_\alpha(K, Y|T) \geq I_\alpha(X, Y|T)$. □

In order to build a connection between α-information and the probability of success, we introduce the generalized Fano's inequality as follows.

Lemma 9.4 (Rioul's Generalized Fano Inequality [137, Thm. 1])**.**

$$I_\alpha(X;Y) \geq d_\alpha\big(P_s(X|Y) \| P_s(X)\big) \tag{9.33}$$

where

$$d_\alpha(p\|q) = \tfrac{1}{\alpha-1} \log\big(p^\alpha q^{1-\alpha} + (1-p)^\alpha (1-q)^{1-\alpha}\big) \tag{9.34}$$

denotes binary α-divergence.

Therefore, we derive the main result as follows.

Theorem 9.1 (Generic Upper Bound on Success Rate [99])**.** *Given a side-channel setting as in Fig. 9.6, we have*

$$d_\alpha(P_s \parallel \frac{1}{M}) \leq I_\alpha(K, Y|T). \tag{9.35}$$

Proof. Given Eqn. 9.32, it is enough to prove $I_\alpha(K, Y|T) \geq d_\alpha(P_s \parallel \frac{1}{M})$. Since $K - Y - \widehat{K}$ forms an Markov chain given T, by DPI in Eqn. 9.30, we have $I_\alpha(K; Y|T) \geq I_\alpha(K; \widehat{K}|T)$.

Using UEP in Eqn. 9.29, one has $I_\alpha(K; \widehat{K}|T) = \log M - H_\alpha(K|\widehat{K}, T) \geq \log M - H_\alpha(K|\widehat{K}) = I_\alpha(K; \widehat{K})$, where the inequality holds since conditioning reduces α-entropy [1, 66]. Then by Rioul's Fano inequality (Lemma 9.4), we have $I_\alpha(K; \widehat{K}) \geq d_\alpha(P_s \parallel \frac{1}{M})$. □

By applying binary α-divergence as in Eqn. 9.34, we have

$$\begin{aligned} d_\alpha(P_s \parallel \frac{1}{M}) &= \frac{1}{\alpha - 1} \log\Big(\frac{P_s^\alpha}{M^{1-\alpha}} + \frac{(1-P_s)^\alpha}{(\frac{M-1}{M})^{\alpha-1}}\Big) \\ &= \log\Big(\frac{M}{M-1}\Big) + \frac{1}{\alpha - 1} \log\big((M-1)^{\alpha-1} P_s^\alpha + (1-P_s)^\alpha\big), \end{aligned}$$

where $d_\alpha(P_s \parallel \frac{1}{M})$ is an increasing function of P_s when $P_s \geq \frac{1}{M}$, as illustrated in Fig. 9.7. Note that we have $P_s \geq \frac{1}{M}$ because if there is no leakage, e.g., when $q = 0$, then an adversary can only guess k randomly. Therefore, the probability of success is $\frac{1}{M}$.

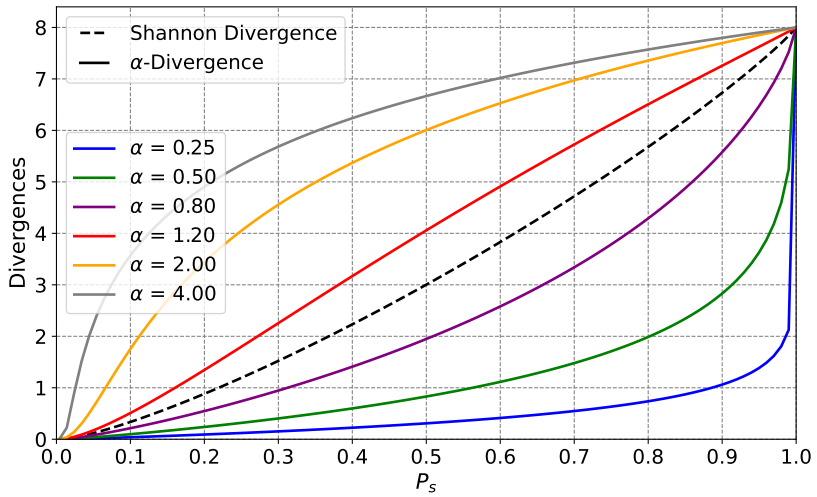


Figure 9.7: Illustration of $d_\alpha(P_s \parallel \frac{1}{M})$ as a function of P_s with different α, where $M = 2^8$.

Since $d_\alpha(P_s \parallel \frac{1}{M})$ is a monotonous function of P_s, it provides a lower bound on $I_\alpha(\mathbf{X}, \mathbf{Y}|\mathbf{T})$ by Theorem 9.1 given a specific P_s. As a result, it enables us to derive a lower bound on the number of traces to achieve that success rate P_s as $I_\alpha(\mathbf{X}, \mathbf{Y}|\mathbf{T}) \leq qI_\alpha(X, Y|T)$ where

$\mathbf{X} = (X_1, X_2, \ldots, X_q)$ and X_i are i.i.d., the same with $\mathbf{Y}$ and $\mathbf{T}$. More precisely, given a set of leakages with length q, $I_\alpha(\mathbf{X}, \mathbf{Y}|\mathbf{T})$ involves q itself, which should be more tighter than $q \geq \frac{d_\alpha(P_s \| \frac{1}{M})}{I_\alpha(X,Y|T)}$.

Inversely, given a fixed set of leakages with length q gives a fixed $I_\alpha(\mathbf{X}, \mathbf{Y}|\mathbf{T})$, then Theorem 9.1 leads to an upper bound on P_s. In particular, let f_d^{-1} be the inverse of $d_\alpha(P_s \parallel \frac{1}{M})$ of P_s, then we have $P_s \leq f_d^{-1}(I_\alpha(\mathbf{X}, \mathbf{Y}|\mathbf{T}))$. Note that f_d^{-1} is also monotonous in its valid domain, e.g., the range is determined by $\frac{1}{M} \leq P_s \leq 1.0$. The inverse function f_d^{-1} of $d_\alpha(P_s \parallel \frac{1}{M})$ is illustrated in Fig. 9.8.

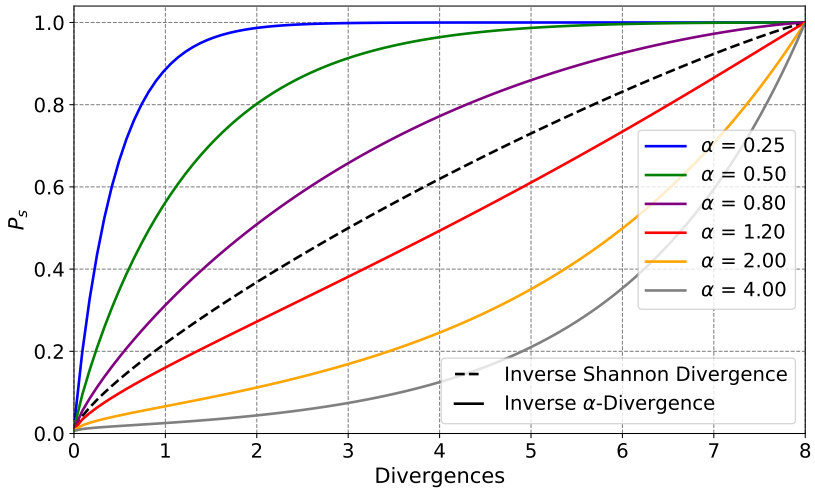


Figure 9.8: Illustration of the inverse of $d_\alpha(P_s \parallel \frac{1}{M})$ with different α where $M = 2^8$.

In summary, Theorem 9.1 allows us to derive twined bounds in two directions, namely an upper bound on P_s and a lower bound on q. In the sequel, we illustrate applications of this theorem by numerical experiments.

9.5.3 Maximal Information Meets ML-based Attacks

In view of Lemma 9.1, given a fixed distribution for corresponding variables, conditional α-information is non-decreasing with respect to the order α, and ∞-information with $\alpha \to \infty$ is the maximal one. Considering the communication channel model shown in Fig. 9.6, we shall apply this maximal information to side-channel analysis.

Recall that in the maximum-likelihood based distinguisher in Chap. 7 (see Eqn. 7.6 and 7.7), the best key guess is made by

$$\widehat{k} = \underset{k \in \mathbb{K}}{\operatorname{argmax}}\, \Delta(k) = \underset{k \in \mathbb{K}}{\operatorname{argmax}}\, p(Y|k, T). \tag{9.36}$$

It is worth mentioning that ML-based distinguisher is sound by design, implying the best key guess will be the true key used in the cipher with a success rate P_s given enough number of side-channel measurements. Therefore, we have the following theorem which bridges the maximum α-information to the ML-based distinguisher.

Theorem 9.2 (Exact Upper Bound on Success Rate by Maximal Information)**.** *Given a side-channel setting as in Fig. 9.6, we have*

$$d_\infty(P_{s|ML} \parallel \frac{1}{M}) = I_\infty(K, Y|T) \tag{9.37}$$

when $I_\infty(K, Y|T)$ denotes the (conditional) maximal information, and $P_{s|ML}$ is the success rate achieved by a maximum-likelihood based distinguisher.

Proof. By definition, we have

$$\begin{aligned}
I_\alpha(K;Y|T) &= \frac{\alpha}{\alpha-1}\log \oint p(t) \oint \Big(\sum_k p(k|t)p^\alpha(y|k,t)\,\mathrm{d}\mu_K(k)\Big)^{\frac{1}{\alpha}}\,\mathrm{d}\mu_Y(y)\,\mathrm{d}\mu_T(t) \\
&= \frac{\alpha}{\alpha-1}\log \sum_t p(t) \oint \Big(\sum_k p(k|t)p^\alpha(y|k,t)\Big)^{\frac{1}{\alpha}}\,\mathrm{d}\mu_Y(y) \\
&= \frac{\alpha}{\alpha-1}\log \sum_t p(t) \oint \Big(\sum_k p(k)p^\alpha(y|k,t)\Big)^{\frac{1}{\alpha}}\,\mathrm{d}\mu_Y(y) \\
&= \frac{\alpha}{\alpha-1}\log \mathbb{E}_{Y,T}\frac{\big(\sum_k p(k)p^\alpha(y|k,t)\big)^{\frac{1}{\alpha}}}{p(y|t)} \\
&= \frac{\alpha}{\alpha-1}\log \mathbb{E}_{Y,T}\frac{\big(\sum_k p(k)p^\alpha(y|k,t)\big)^{\frac{1}{\alpha}}}{\sum_k p(k)p(y|k,t)} \\
&= \log M + \frac{\alpha}{\alpha-1}\log \mathbb{E}_{Y,T}\frac{\big(\sum_k p^\alpha(y|k,t)\big)^{\frac{1}{\alpha}}}{\sum_k p(y|k,t)}.
\end{aligned} \tag{9.38}$$

The equality in $p(k|t) = p(k)$ holds since T is independent of K and $p(k) = \frac{1}{M}$ as it is uniformly distributed. Moreover, $\big(\sum_k p^\alpha(y|k,t)\big)^{\frac{1}{\alpha}}$ is the α-norm of $p(y|k,t)$. Therefore, we obtain $\big(\sum_k p^\alpha(y|k,t)\big)^{\frac{1}{\alpha}} = \max_k\{p(y|k,t)\} = P_{s|\mathrm{ML}}$ when $\alpha \to \infty$, resulting that

$$I_\infty(K;Y|T) = \log M + \log \mathbb{E}_{Y,T}\frac{P_{s|\mathrm{ML}}}{\sum_k p(y|k,t)}. \tag{9.39}$$

In other words, $P_{s|\mathrm{ML}}$ is the exact success rate (e.g., evaluated by hundreds or more repetitions) given by the ML-based distinguisher. Conversely, inserting the same $P_{s|\mathrm{ML}}$ and $\alpha \to \infty$ into d_α gives equality in Eqn. 9.37. □

Intuitively, Theorem 9.2 coincides with the fact that if the distinguisher is sound, then the probability of success for the correct key guess will exceed all other wrong key guesses. Indeed, the infinity norm in α-information when $\alpha \to \infty$ exactly fits with the ML-based distinguisher

by which it returns the most possible key guess. Still, it gives an upper bound for the success rate, implying that Theorem 9.2 leads to an exact bound achievable by using the optimal distinguishers (e.g., the ML-based one).

As we will show in the sequel, the derivatives in Eqn. 9.38 are easy to be implemented and evaluated by Monte-Carlo simulations. We shall simplify the notation $P_{s|\mathrm{ML}}$ to P_s when there is no ambiguity.

Remark 9.3. It is worth mentioning that Theorem 9.2 will recover the conditional maximal leakage proposed in [89, Def. 6]. However, the maximum is considered not only over $\Pr(K=k)$, but also over $\Pr(T=t)$, resulting in a larger value than our α-information when α tends to infinity.

9.6 Applications to Hamming Weight Leakage with AWGN

Let $K \in \mathbb{F}_{2^\ell}$ be the secret key and $T \in \mathbb{F}_{2^\ell}$ be the plaintext or ciphertext, where typically $\ell = 8$, e.g., in AES. Therefore, in side-channel analysis, an adversary aims to recover the secret key by exploiting several (many) side-channel measurements, say q traces. That is, considering the commonly used Hamming weight leakage model, the side-channel leakage can be generated by:

$$\mathbf{Y}_i = w_H(S(\mathbf{T}_i \oplus K)) + \mathbf{N}_i, \tag{9.40}$$

where w_H is the Hamming weight function, S denotes certain cryptographic operation within a cipher and $\mathbf{N}_i$ are i.i.d $\sim \mathcal{N}(0, \sigma^2)$ for $1 \le i \le q$.

Applying the definition of $I_\alpha(\mathbf{X}, \mathbf{Y}|\mathbf{T})$ as in Def. 9.7, we have:

$$\begin{aligned} I_\alpha(\mathbf{X}, \mathbf{Y}|\mathbf{T}) &= I_\alpha(K, \mathbf{Y}|\mathbf{T}) \\ &= \frac{\alpha}{\alpha-1} \log\Big(\sum_{\mathbf{t}} p(\mathbf{t}) \int_{\mathcal{Y}} \big(\sum_k p(k|\mathbf{t}) p^\alpha(\mathbf{y}|\mathbf{t},k) \big)^{\frac{1}{\alpha}} \mathrm{d}\mu_{\mathbf{Y}}(\mathbf{y}) \Big) \\ &= \frac{\alpha}{\alpha-1} \log\Big(\int_{\mathcal{Y}} \sum_{\mathbf{t}} p(\mathbf{y},\mathbf{t}) \frac{\big(\sum_k p(k|\mathbf{t}) p^\alpha(\mathbf{y}|\mathbf{t},k) \big)^{\frac{1}{\alpha}}}{p(\mathbf{y}|\mathbf{t})} \mathrm{d}\mu_{\mathbf{Y}}(\mathbf{y}) \Big). \end{aligned} \tag{9.41}$$

Next, Eqn. 9.41 can be estimated by using Monte-Carlo simulation by the law of large numbers. Indeed, we have

$$\begin{aligned} \exp\left(\frac{\alpha-1}{\alpha} I_\alpha(\mathbf{X}, \mathbf{Y}|\mathbf{T}) \right) &\approx \lim_{N_C \to \infty} \frac{1}{N_C} \sum_{j=1}^{N_C} \frac{\big(\sum_k p(k|\mathbf{t}^j) p^\alpha(\mathbf{y}^j|\mathbf{t}^j,k) \big)^{\frac{1}{\alpha}}}{p(\mathbf{y}^j|\mathbf{t}^j)} \\ &= \lim_{N_C \to \infty} \frac{1}{N_C} \sum_{j=1}^{N_C} \frac{\big(\sum_k p(k) p^\alpha(\mathbf{y}^j|\mathbf{t}^j,k) \big)^{\frac{1}{\alpha}}}{\sum_k p(k) p(\mathbf{y}^j|\mathbf{t}^j,k)}, \end{aligned} \tag{9.42}$$

where $\mathbf{t}^j \sim \mathcal{U}(\mathbb{F}_{2^\ell}^q)$ and $\mathbf{y}^j \sim \mathcal{N}(f(\mathbf{t}^j, k^j), \sigma^2 \mathbf{I}_q) \in \mathbb{R}^q$ by choosing $k^j \sim \mathcal{U}(\mathbb{F}_{2^\ell})$ and $f(\mathbf{t}^j, k^j) = w_H(S(\mathbf{t}^j \oplus k^j))$.

Considering independent Gaussian noise in each $\mathbf{y}^j$, we can simplify (9.42) and insert into (9.41), therefore,

$$\begin{aligned} I_\alpha(\mathbf{X}, \mathbf{Y}|\mathbf{T}) &\approx \ell + \frac{\alpha}{\alpha - 1} \log \frac{1}{N_C} \sum_{j=1}^{N_C} \frac{\left(\sum_k p^\alpha(\mathbf{y}^j|\mathbf{t}^j, k)\right)^{\frac{1}{\alpha}}}{\sum_k p(\mathbf{y}^j|\mathbf{t}^j, k)} \\ &= \ell + \frac{\alpha}{\alpha - 1} \log \frac{1}{N_C} \sum_{j=1}^{N_C} \frac{\left(\sum_k e^{-\frac{\alpha}{2\sigma^2}||\mathbf{y}^j - f(\mathbf{t}^j, k)||_2}\right)^{\frac{1}{\alpha}}}{\sum_k e^{-\frac{1}{2\sigma^2}||\mathbf{y}^j - f(\mathbf{t}^j, k)||_2}}, \end{aligned} \tag{9.43}$$

given a larger enough N_C.

Hereafter, we depict $I_\alpha(\mathbf{X}, \mathbf{Y}|\mathbf{T})$ with different choices of α.

9.6.1 Evaluation of $I_\alpha(\mathbf{X}, \mathbf{Y}|\mathbf{T})$ with Different α

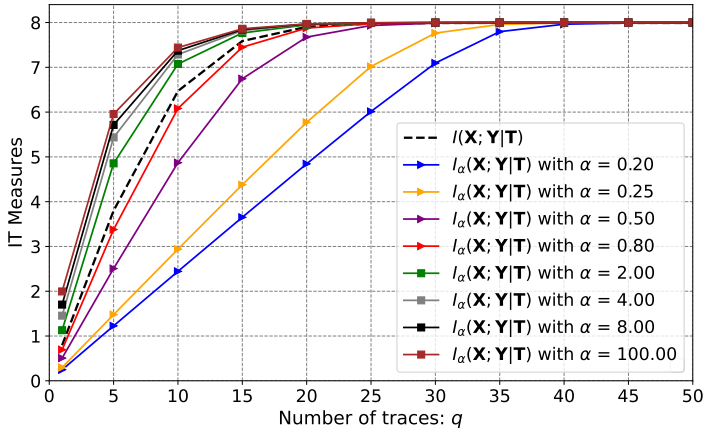
We first consider the lower level of Gaussian noise with $\sigma^2 = 1.00$ and 2.00. The results are depicted in Fig. 9.9 by using $q = 50$ side-channel traces in total. As the first observation, $I_\alpha(\mathbf{X}, \mathbf{Y}|\mathbf{T})$ is non-decreasing in α, which confirms Lemma 9.1. Note that for the sake of clarity on comparison at beginning, we ignore the first point when $q = 0$, which gives $I_\alpha(\mathbf{X}, \mathbf{Y}|\mathbf{T}) = 0$.

Secondly, we push forward our evaluation into scenarios with high noise levels as shown in Fig. 9.10. For instance, $\sigma^2 = 8.00$ corresponds to SNR $= 0.25$ in Fig. 9.10(b). Therefore, the second observation is that the gap between larger and smaller orders enlarges when the variance of noise increases. Intuitively, $I_\alpha(\mathbf{X}, \mathbf{Y}|\mathbf{T})$ with a larger order α should be more relevant in side-channel analysis. The reason is that a larger value of α makes this measure be more sensitive to key guesses that have higher probabilities, which is exactly the case when a sound distinguisher [80, 151] is adopted in corresponding attacks.

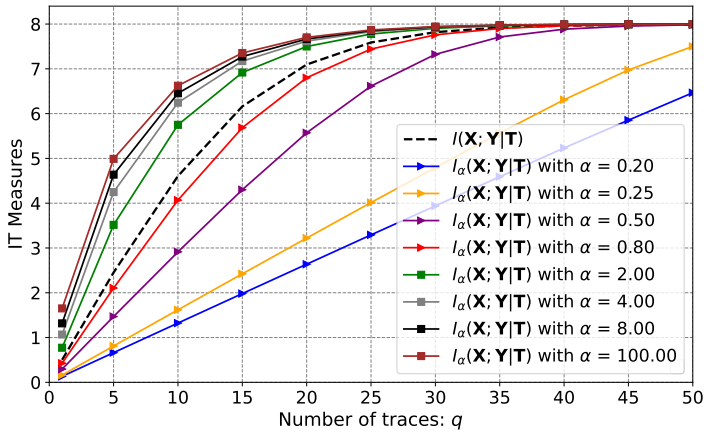
9.6.2 Bounding the Probability of Success

Relying on Theorem 9.1, we derive upper bounds on the probability of success P_s. For the sake of clarity, the bounds of P_s given by applying the generalized Fano's inequality is plotted in Fig. 9.11 and then the numerical results on the success rate are shown in Fig. 9.12 for different pairs of α in $I_\alpha(\mathbf{X}, \mathbf{Y}|\mathbf{T})$.

The main takeaway from Fig. 9.12 is that, larger orders in $I_\alpha(\mathbf{X}, \mathbf{Y}|\mathbf{T})$ enable to derive better bounds on the success rate. Specifically, taking $\alpha = 100.00$ as an example, it almost



(a) $\sigma^2 = 1.00$.

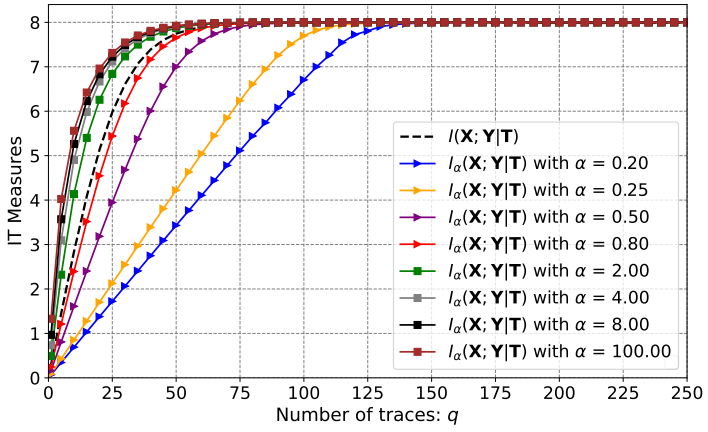


(b) $\sigma^2 = 2.00$.

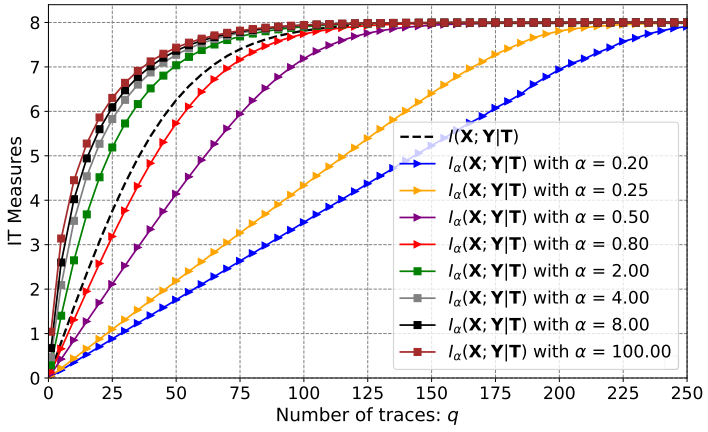
Figure 9.9: Numerical comparison of Shannon mutual information $I(\mathbf{X},\mathbf{Y}|\mathbf{T})$ and α-information $I_\alpha(\mathbf{X},\mathbf{Y}|\mathbf{T})$ with different α in a side-channel analysis context, with $q = 50$ traces.

pushes to the limit of supremum. However, smaller orders, e.g., taking $\alpha < 1.0$, only provide loose upper bounds on P_s.

Next, we compare those upper bounds given by $I_\alpha(\mathbf{X},\mathbf{Y}|\mathbf{T})$ with empirical success rate from ML-based attacks (using optimal distinguishers [84]). The numerical results under different noise levels are shown in Fig. 9.12 and 9.13 for $q_{\max} = 50$ and 200 traces, respectively. The most significant observation is that when the order is larger enough, the information-theoretic bounds provide exact upper bound on P_s, or conversely, the empirical success rate of the ML-based attack will converge to the upper bound by $I_\alpha(\mathbf{X},\mathbf{Y}|\mathbf{T})$. Compared with the state-of-the-art bound [57] given by Shannon mutual information $I(\mathbf{X},\mathbf{Y}|\mathbf{T})$, our new bounds with large orders α are significantly better in a sense of tight upper bounds.



(a) $\sigma^2 = 4.00$.

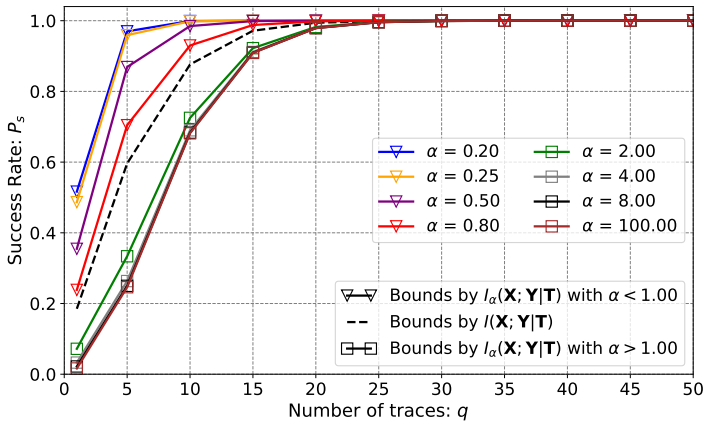


(b) $\sigma^2 = 8.00$.

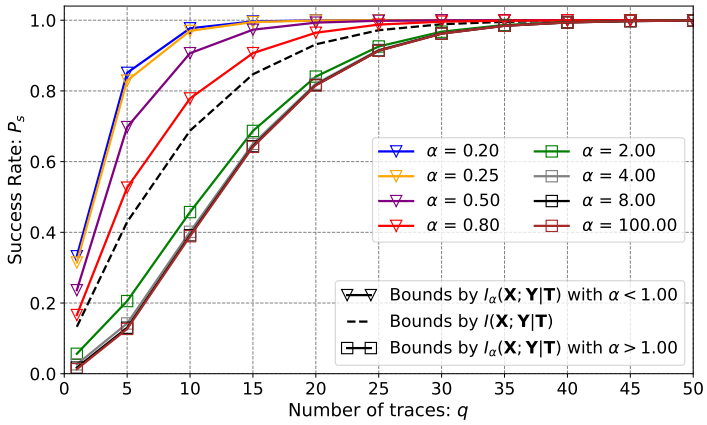
Figure 9.10: Numerical comparison of Shannon mutual information $I(\mathbf{X}, \mathbf{Y}|\mathbf{T})$ and α-information $I_\alpha(\mathbf{X}, \mathbf{Y}|\mathbf{T})$ with different α in a side-channel analysis context, with $q = 200$ traces.

To summing up, we present the full spectrum of upper bounds when applying $I_\alpha(\mathbf{X}, \mathbf{Y}|\mathbf{T})$ in bounding the success rate of the optimal attacks in SCA. Particularly, we shown that a larger order of $I_\alpha(\mathbf{X}, \mathbf{Y}|\mathbf{T})$ will give a tighter bound. When pushing to the limit, $I_\infty(\mathbf{X}, \mathbf{Y}|\mathbf{T})$ would lead to the best upper bound on P_s, since $I_\infty(\mathbf{X}, \mathbf{Y}|\mathbf{T})$ is increasing in α as proved in Lemma 9.1. At last, the optimality of the ML-based distinguisher indicates that there is no distinguisher better than it. In other words, the success rate of utilizing other distinguishers like CPA, DPA and MIA will not exceed that of ML-based distinguishers. As a consequence, the bound by $I_\infty(\mathbf{X}, \mathbf{Y}|\mathbf{T})$ is truly the supremum of success rate of any attacks, which confirms our theoretical derivatives shown in Theorem 9.2.

As applications, the bound given by $I_\infty(\mathbf{X}, \mathbf{Y}|\mathbf{T})$ bridges two worlds: the one is the



(a) $\sigma^2 = 1.00$.



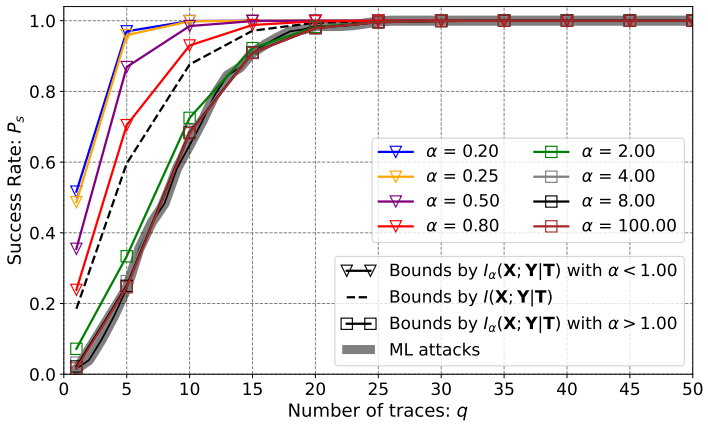
(b) $\sigma^2 = 2.00$.

Figure 9.11: Comparison of applying the Rioul's generalized Fano inequality on P_s in α-information $I_\alpha(\mathbf{X}, \mathbf{Y}|\mathbf{T})$ with different α in a side-channel analysis context, with $q = 50$ traces.

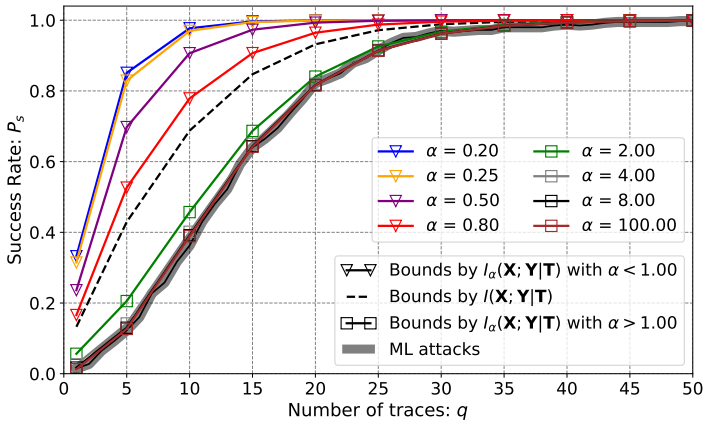
information-theoretic evaluation of side-channel leakage and the other is the exploitability of those leakage. Additionally, it should also bring us a deeper understanding on the precise security level of real devices in practice.

9.6.3 Predicting the Minimum Number of Traces for an Attack

As another application of Theorem 9.1, we shall derive lower bounds on the number of traces $q_{\min}$ to achieve a given success rate P_s, since q is implicitly involved in $I_\alpha(\mathbf{X}, \mathbf{Y}|\mathbf{T})$. The lower bounds on $q_{\min}$ are shown in Fig. 9.14, where in each figure, one pair of values of order α are added for comparison. Specifically, two groups of α are: 2.00 vs 0.50 and 100.00 vs 0.01. In particular, $\alpha = 2$ corresponds to *collision* information (or collision entropy $H_2(\mathbf{X}, \mathbf{Y}|\mathbf{T})$).



(a) $\sigma^2 = 1.00$.

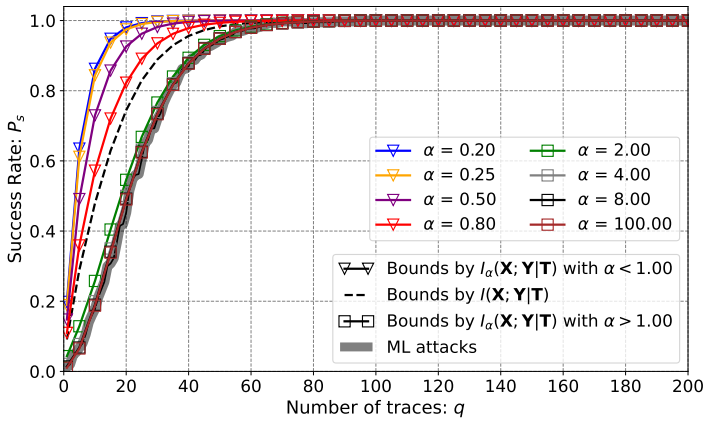


(b) $\sigma^2 = 2.00$.

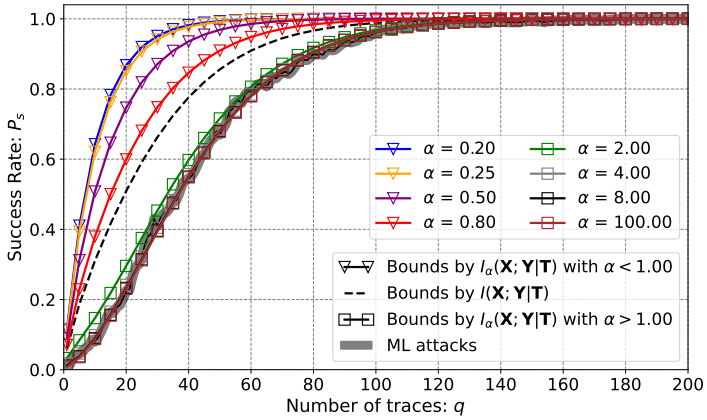
Figure 9.12: Comparison of upper bounds on success rate P_s given by Shannon mutual information $I(\mathbf{X}, \mathbf{Y}|\mathbf{T})$ and α-information $I_\alpha(\mathbf{X}, \mathbf{Y}|\mathbf{T})$ with different α in a side-channel analysis context, with $q = 50$ traces.

Moreover, we consider $\alpha = 100.00$ as a large enough order in α-information. Note than the empirical success rate given by ML-based attack is obtained by 400 repetitive experiments, while $N_C = 1,000,000$ random draws are used in Monte-Carlo evaluation of $I_\alpha(\mathbf{X}, \mathbf{Y}|\mathbf{T})$ to get a good convergence.

As expected, lower bounds given by $I_\alpha(\mathbf{X}, \mathbf{Y}|\mathbf{T})$ with larger values of α are tighter than those with smaller values of α. In particular, the order $\alpha = 100.00$ gives an almost exact prediction on the minimum number of traces to achieve $P_s \geq 95\%$. Additionally, we shall predict that $I_\infty(\mathbf{X}, \mathbf{Y}|\mathbf{T})$ will provide us the exact number of traces to achieve such a success rate, since the success rate of ML attack will converge to the bound given by the maximal information.



(a) $\sigma^2 = 4.00$.



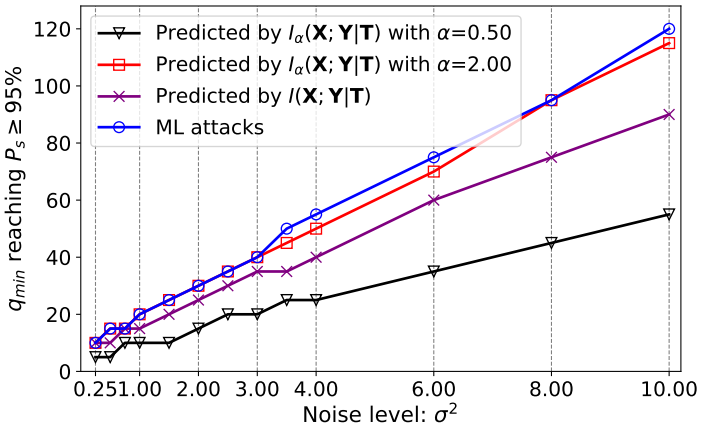
(b) $\sigma^2 = 8.00$.

Figure 9.13: Comparison of upper bounds on success rate P_s given by Shannon mutual information $I(\mathbf{X}, \mathbf{Y}|\mathbf{T})$ and α-information $I_\alpha(\mathbf{X}, \mathbf{Y}|\mathbf{T})$ with different α in a side-channel analysis context, with $q = 200$ traces.

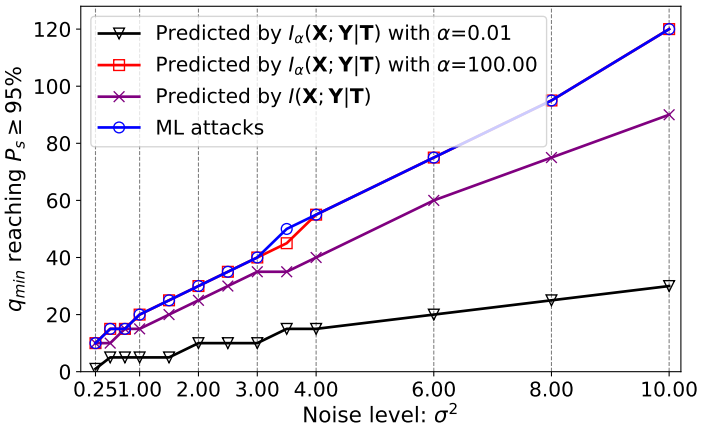
9.7 Conclusions

In this chapter, we aim at measuring information leakage by utilizing more general information-theoretic measures instead of Shannon information measures. We show first how the Hamming weight leakage is assessed by conditional α-entropy in both noiseless and noisy cases. In particular, we illustrate that the leakage quantification will be more accurate when the order α is larger enough.

More importantly, we present how the conditional α-information is applied in assessing the side-channel leakage. In this respect, we show a full spectrum of bounds given by conditional



(a) $\alpha = 2.00$ and 0.50.



(b) $\alpha = 100.00$ and 0.01.

Figure 9.14: Comparison of lower bounds on the number of traces $q_{\min}$ to reach $P_s \geq 95\%$.

α-information with different orders. The outputs are twofold. On one hand, the success rate of any key-recovering attack is upper bounded by the conditional α-information between the sensitive variable and the corresponding leakage. We therefore, for the first time, provide a supremum of empirical success rate of any attacks. On the other hand, the minimum number of traces that achieves a specific success rate is lower bounded by the conditional α-information. Again, the bound is tight when the order α is larger enough, meaning that optimal attacks can reach this bound, e.g., by utilizing maximum-likelihood based distinguishers.

Part VI

Conclusions and Perspectives

CHAPTER 10

Conclusions and Perspectives

10.1 Conclusion

Measuring the concrete side-channel security is in the center of designing and evaluating cryptographic implementations in practice, which is still an active and dynamic research area. This problem is all the more important for evaluating masked implementations to understand and enhance their practical security. In this thesis, we contribute to this problem in two aspects. On the one hand, we present a unified and generic information leakage quantification framework for the code-based masking, which allows us to assess the side-channel resistance of all instances of code-based masking. On the other hand, we explore possibilities of applying more general information-theoretic tools in side-channel analysis.

The first two parts of this thesis focus on quantifying information leakage in code-based masking. Because of the generalization, the leakage quantification framework works for all code-based masking instances like the simplest Boolean masking, inner product masking, direct sum masking, Shamir's secret sharing based masking, etc. Technically, our framework formally binds the coding-theoretic properties of the corresponding linear codes to two leakage metrics, namely signal-to-noise ratio (SNR) and mutual information (MI). Particularly in the case of the Hamming weight leakage model, we find that both SNR and MI depend exclusively on the dual distance and the kissing number of the linear codes used in the masking. Those theoretical

derivatives enable us to enhance the code-based masking by providing optimal linear codes for it in the sense of side-channel resistance.

Next, in the third part, we investigate the exploitability of those information leakages. We consider the higher-order optimal distinguisher as it is the most powerful one based on the maximum-likelihood rule. We first verify our theoretical framework from a perspective of attack-based evaluation. The experimental results fairly confirm those theoretical findings and demonstrate the advantages of employing code-based masking in practice. Another takeaway is that redundancy can only reduce the side-channel resistance as expected, implying that a trade-off must be considered in designing code-based masking against both side-channel analysis and fault injection attack simultaneously. Second, by utilizing traditional information-theoretic tools, we provide several theoretical bounds in attacking protected cryptographic implementations, in the presence of code-based masking.

Finally, in the fourth part, we devote ourselves to applying general information-theoretic measures for tighter universal bounds on how successful can any side-channel attack achieve. This is of special importance in understanding and defeating side-channel analysis in practice. In this respect, we propose to utilize α-information featured with an order α and investigate several relevant properties in the context of side-channel analysis. Interestingly, with various choices of α, we obtain a full spectrum of upper bounds on the success rate of the optimal attacks (distinguishers), from the loosest one to the tightest one. In particular, we prove that the success rate of the optimal attack converges to the upper bound given by α-information, indicating that our new bounds are exactly tight. As a straightforward application, we verify our theoretical bounds in side-channel analysis by considering the common Hamming weight leakages. The simulation results exactly match with theoretical predictions.

Relying on above progresses made in this study, we put forth the evaluation tools and pave the way to measurable side-channel security, especially in the presence of protections. From a perspective of protection, we also provide the best-practice guideline for applying code-based masking in practical cryptographic chips.

10.2 Further Perspectives

In view of topics studied and some progresses made in this thesis, we shall investigate the following aspects in the future.

Efficient Construction of Optimal Linear Codes. We showed in this thesis that different linear codes have significant impact on the side-channel resistance of a specific code-based masking. Thanks to our unified leakage quantification framework, we provided use-cases for several masking schemes with lower number of shares (e.g., $n \leq 5$) [35, 37]. However, it is still an open problem to construct those optimal codes rather than enumerate all possible candidates exhaustively, which would be infeasible soon when n increases [123]. A possible approach is to construct linear codes by adding small blocks recursively in a greedy fashion. This approach can be very efficient but the output of this approach might not be the global optimum. Moreover, it will be interesting to consider algebraic codes with certain good structures as well, which we shall explore more in the near future.

Generic Construction of Masked Gadgets against Both SCA and FIA. Devising combined countermeasures against both side-channel analysis and fault injection attacks (FIA) is always a very active topic in this field. Considering the intrinsic nature of a linear code, it can detect (or correct) errors provided that the number of erroneous "digits" is smaller than the minimum distance (or half of the minimum distance) of the code. Therefore, a question arising in code-based masking is, whether it can be extended to counteract both SCA and FIA. In this respect, an interesting construction of gadgets is proposed in [164], which presents several generic and efficient gadgets against SCA, while not all of them are applicable to thwart FIA. As a consequence, our interest particularly lies in constructing generic and efficient gadgets against both SCA and FIA for future study.

Practical Applications of Code-based Masking. We have demonstrated significant benefits of utilizing code-based masking from a security perspective and also provided evidence of its efficiency when implemented in practice. Our theoretical derivatives have been verified by numerical simulation experiments. However, it still remains to be validated in real devices. In particular, it is still non-trivial to devise a secure masked implementation, considering various physical defaults (like couplings, etc) and glitches in practical circuits (chips), which usually ruin the security guarantees provided by protections. In the case of code-based masking, we shall push forward the practical evaluation by considering various settings and platforms in practice. Moreover, it is also interesting to apply code-based masking in protecting implementations of post-quantum cryptographic schemes and algorithms.

Extended Applications of α-Information in Side-Channel Evaluations. As already shown in this thesis, the general α-information paves the way to seamlessly connect the information-theoretic evaluation and side-channel attacks. In particular, we presented a tight upper bound on the success rate of any side-channel distinguishers in unprotected scenarios. More generally, α-information is expected to provide tight bounds in the presence of masking or other protections. As perspectives, we will aim at applying α-information into side-channel security evaluations. Especially, we shall also explore possible construction of security proofs under the noisy leakage model, which may narrow down or even fill the gap between the theoretical proof-based security and the practical security of masked cryptographic implementations. Finally, we will investigate how α-information can be put into practice, where a tighter bound on success rate of any attacks implies an exact security guarantee against side-channel attacks. To summarize, those perspectives shall contribute to measurable side-channel security both in theory and in practice.

10.3 List of Publications

We list publications as follows during this thesis. Note that those with referred citations are more relevant to this thesis than that with increasing serial numbers.

Journal Papers

[35] **Wei Cheng**, Sylvain Guilley, Claude Carlet, Jean-Luc Danger, Sihem Mesnager. Information Leakages in Code-based Masking: A Unified Quantification Approach. IACR Trans. Cryptogr. Hardw. Embed. Syst. 2021(3): 465-495 (2021). https://doi.org/10.46586/tches.v2021.i3.465-495.

[40] **Wei Cheng**, Sylvain Guilley, Jean-Luc Danger. Categorizing all linear codes of IPM over $\mathbb{F}_{2^8}$. Cryptogr. Commun. 13(4): 527-542 (2021). https://doi.org/10.1007/s12095-021-00483-1.

[33] **Wei Cheng**, Claude Carlet, Kouassi Goli, Jean-Luc Danger, Sylvain Guilley. Detecting faults in inner product masking scheme. J. Cryptogr. Eng. 11(2): 119-133 (2021). https://doi.org/10.1007/s13389-020-00227-6.

[37] **Wei Cheng**, Sylvain Guilley, Claude Carlet, Sihem Mesnager, Jean-Luc Danger. Optimizing Inner Product Masking Scheme by a Coding Theory Approach. IEEE Trans. Inf. Forensics Secur. 16: 220-235 (2020). https://doi.org/10.1109/TIFS.2020.3009609.

[5] Trevor Kroeger, **Wei Cheng**, Sylvain Guilley, Jean-Luc Danger, Naghmeh Karimi. Assessment and Mitigation of Power Side-Channel based Cross-PUF Attacks on Arbiter-PUFs and their Derivatives. IEEE Trans. Very Large Scale Integr. Syst. 2021. (To appear)

[6] Jingdian Ming, Huizhong Li, Yongbin Zhou, **Wei Cheng**, Zehua Qiao. Revealing the Weakness of Addition Chain Based Masked SBox Implementations. IACR Trans. Cryptogr. Hardw. Embed. Syst. 2021(4): 326-350 (2021). https://doi.org/10.46586/tches.v2021.i4.326-350.

[7] Jingdian Ming, Yongbin Zhou, **Wei Cheng**, Huizhong Li, Guang Yang, Qian Zhang. Mind the Balance: Revealing the Vulnerabilities in Low Entropy Masking Schemes. IEEE Trans. Inf. Forensics Secur. 15: 3694-3708 (2020). https://doi.org/10.1109/TIFS.2020.2994775.

Conference & Workshop Papers

[42] **Wei Cheng**, Yi Liu, Sylvain Guilley, Olivier Rioul. Towards Finding Best Linear Codes for Side-Channel Protections. PROOFS 2021: 1-16 (2021). (To appear)

[138] Olivier Rioul, **Wei Cheng**, Sylvain Guilley. Cumulant Expansion of Mutual Information for Quantifying Leakage of a Protected Secret. ISIT 2021: 2596-2601 (2021). https://doi.org/10.1109/ISIT45174.2021.9517886.

[99] Yi Liu, **Wei Cheng**, Sylvain Guilley, Olivier Rioul. On Conditional α-Information and its Application to Side-Channel Analysis. ITW 2021: 1-6 (2021). https://doi.org/10.1109/ITW48936.2021.9611409.

[4] Jingdian Ming, **Wei Cheng**, Yongbin Zhou, Huizhong Li. APT: Efficient Side-Channel Analysis Framework against Inner Product Masking Scheme. ICCD 2021: 575-582 (2021). https://doi.org/10.1109/ICCD53106.2021.00093.

[5] Trevor Kroeger, **Wei Cheng**, Sylvain Guilley, Jean-Luc Danger and Naghmeh Karimi. Enhancing the Resiliency of Multi-Bit Parallel Arbiter-PUF and its Derivatives against Power Attacks. COSADE 2021:303-321 (2021). https://doi.org/10.1007/978-3-030-89915-8_14.

[6] Trevor Kroeger, **Wei Cheng**, Jean-Luc Danger, Sylvain Guilley and Naghmeh Karimi. Making Obfuscated PUFs Secure Against Power Side-Channel Based Modeling Attacks. DATE 2021:1000-1005 (2021). https://doi.org/10.23919/DATE51398.2021.9474137.

[7] Patrick Solé, **Wei Cheng**, Sylvain Guilley, Olivier Rioul. Bent Sequences over Hadamard Codes for Physically Unclonable Functions. ISIT 2021: 801-806 (2021). https://doi.org/10.1109/ISIT45174.2021.9517752.

[8] Patrick Solé, Yi Liu, **Wei Cheng**, Sylvain Guilley, Olivier Rioul. Linear Programming Bounds on the Kissing Number of q-ary Codes. ITW 2021:1-5 (2021). https://doi.org/10.1109/ITW48936.2021.9611478.

[9] Trevor Kroeger, **Wei Cheng**, Jean-Luc Danger, Sylvain Guilley and Naghmeh Karimi. Effect of Aging on PUF Modeling Attacks based on Power Side-Channel Observations. DATE 2020: 454–459 (2021). https://doi.org/10.23919/DATE48585.2020.9116428.

[10] Trevor Kroeger, **Wei Cheng**, Sylvain Guilley, Jean-Luc Danger, Naghmeh Karimi. Cross-PUF Attacks on Arbiter-PUFs through their Power Side-Channel. ITC 2020: 1-5 (2020). https://doi.org/10.1109/ITC44778.2020.9325241.

[11] **Wei Cheng**, Claude Carlet, Kouassi Goli, Sylvain Guilley, Jean-Luc Danger. Detecting Faults in Inner Product Masking Scheme - IPM-FD: IPM with Fault Detection. PROOFS 2019: 17-32 (2019). https://doi.org/10.29007/fv2n.

Pre-prints & Submissions

[41] **Wei Cheng**, Yi Liu, Sylvain Guilley, Olivier Rioul. Attacking Masked Cryptographic Implementations: Information-Theoretic Bounds. ArXiv.org/abs/2105.07436, 2021. (Pre-print on ArXiv)

[2] **Wei Cheng**, Sylvain Guilley, Jean-Luc Danger. Information Leakage in Code-based Masking: Another Look on Probing Model Security. (Submitted)

[3] Qianmei Wu, **Wei Cheng**, Sylvain Guilley, Fan Zhang. On Efficient and Secure Code-based Masking: A Pragmatic Evaluation. (Submitted)

Open Datasets of Optimal Codes for Code-based Masking

[1] **Wei Cheng**, Sylvain Guilley. Optimal linear codes for inner product masking (IPM) with 2 and 3 shares over both $\mathbb{F}_{2^4}$ and $\mathbb{F}_{2^8}$. https://github.com/Qomo-CHENG/OC-IPM.

[2] **Wei Cheng**, Sylvain Guilley. Optimal linear codes for generalized code-based masking (GCM): taking $(3,1)$ and $(5,2)$-SSS based masking over both $\mathbb{F}_{2^4}$ and $\mathbb{F}_{2^8}$. https://github.com/Qomo-CHENG/GeneralizedCM.

Part VII

Appendix

APPENDIX A

Further Proofs, Lemmas and Discussions

A.1 Detailed Proofs

Before presenting these proofs, we recall below two well-known properties of *Fourier transform*. We omit the proofs for the sake of brevity and refer to [22] for details.

Lemma 1.1 (Involution Property)**.** $\widehat{\widehat{P}}(z) = |\mathbb{K}^{n\ell}|P(z) = 2^{n\ell}P(z)$, $\forall z \in \mathbb{K}^{n\ell}$.

Lemma 1.2 (Inverse *Fourier Transform*)**.** $P(z) = 2^{-n\ell}\sum_{y\in\mathbb{K}^{n\ell}}\widehat{P}(y)(-1)^{y\cdot z}$, $\forall z \in \mathbb{K}^{n\ell}$.

A.1.1 Proof of Lemma 5.1

In order to demonstrate Lemma 5.1, we clarify the computations in $\mathbb{V}\left[\mathbb{E}\left[P(Z)|X\right]\right]$ as follows. Let us consider Eqn. 5.1 in basefield $\mathbb{F}_2$, and thus let $\mathcal{X} = \mathbb{F}_2^{\ell}$, $\mathcal{Y} = \mathbb{F}_2^{t\ell}$ and $\mathcal{Z} = \mathbb{F}_2^{n\ell}$. Moreover, the $\mathcal{C}$ and $\mathcal{D}$ are expanded into $\mathbb{F}_2$ by using code expansion (Def. 3.6):

- $\mathbb{E}\left[P(Z)|X = x\right]$ for a given $x \in \mathcal{X}$ is:

$$\mathbb{E}\left[P(x\mathbf{G} + Y\mathbf{H})\right] = \sum_{y\in\mathcal{Y}}\mathbb{P}(Y = y)P(x\mathbf{G} + y\mathbf{H}) = \frac{1}{|\mathcal{Y}|}\sum_{y\in\mathcal{Y}}P(x\mathbf{G} + y\mathbf{H})$$
$$= \frac{1}{|\mathcal{D}|}\sum_{d\in\mathcal{D}}P(x\mathbf{G} + d).$$

- For any variable X, we have that:

$$\mathbb{V}\left[\mathbb{E}\left[P(Z)|X\right]\right] = \mathbb{E}\left[\mathbb{E}\left[P(Z)|X\right]^2\right] - \mathbb{E}\left[\mathbb{E}\left[P(Z)|X\right]\right]^2.$$

Next, we derive formulas for both sub-terms $\mathbb{E}\left[\mathbb{E}\left[P(Z)|X\right]\right]$ and $\mathbb{E}\left[\mathbb{E}\left[P(Z)|X\right]^2\right]$ and their proofs are in Appendix A.1.2 and A.1.3, respectively.

Lemma 1.3. $\mathbb{E}\left[\mathbb{E}\left[P(Z)|X\right]\right] = \frac{1}{2^{n\ell}} \sum_{x \in (\mathcal{C} \oplus \mathcal{D})^\perp} \widehat{P}(x)$.

Lemma 1.4. $\mathbb{E}\left[\mathbb{E}\left[P(Z)|X\right]^2\right] = \frac{1}{2^{2n\ell}} \sum_{\substack{x,\, y \in \mathcal{D}^\perp \\ x+y \in \mathcal{C}^\perp}} \widehat{P}(x)\widehat{P}(y)$.

Particularly, when two codes $\mathcal{C}$ and $\mathcal{D}$ are complementary, we can simplify above two lemmas as follows, which is exactly the case for IPM (recall Lemmas 4.1 and 4.2).

Lemma 1.5. $\mathbb{E}\left[\mathbb{E}\left[P(Z)|X\right]\right] = \frac{1}{2^{n\ell}} \widehat{P}(0)$.

Proof. Given that $\mathcal{C} \oplus \mathcal{D} = \mathbb{F}_{2^{n\ell}}$, then we obtain $(\mathcal{C} \oplus \mathcal{D})^\perp = \{0\}$. Therefore, $\mathbb{E}\left[\mathbb{E}\left[P(Z)|X\right]\right] = \frac{1}{2^{n\ell}} \sum_{x \in (\mathcal{C} \oplus \mathcal{D})^\perp} \widehat{P}(x) = \frac{1}{2^{n\ell}} \widehat{P}(0)$. □

Lemma 1.6. $\mathbb{E}\left[\mathbb{E}\left[P(Z)|X\right]^2\right] = \frac{1}{2^{2n\ell}} \sum_{x \in \mathcal{D}^\perp} \left(\widehat{P}(x)\right)^2$.

Proof. Given that $\mathcal{C}$ and $\mathcal{D}$ are complementary, then $\mathcal{C} \cap \mathcal{D} = \{0\}$, so as $\mathcal{C}^\perp \cap \mathcal{D}^\perp = \{0\}$. Therefore, the conditional codewords $x,\, y \in \mathcal{D}^\perp$ and $x + y \in \mathcal{C}^\perp$ gives $x + y = 0$, or equivalently $x = y$. As a result, Lemma 1.4 becomes $\mathbb{E}\left[\mathbb{E}\left[P(Z)|X\right]^2\right] = \frac{1}{2^{2n\ell}} \sum_{x,\, y \in \mathcal{D}^\perp, x+y \in \mathcal{C}^\perp} \widehat{P}(x)\widehat{P}(y) = \frac{1}{2^{2n\ell}} \sum_{x \in \mathcal{D}^\perp} \left(\widehat{P}(x)\right)^2$. □

Therefore, relying on the two lemmas, the proof of Lemma 5.1 is as follows.

Proof of Lemma 5.1. From Lemma 4.1, we compute $\mathbb{E}\left[\mathbb{E}\left[P(Z)|X\right]\right]^2$ as follows:

$$\begin{aligned} \mathbb{E}\left[\mathbb{E}\left[P(Z)|X\right]\right]^2 &= \left(\frac{1}{2^{n\ell}} \sum_{x \in (\mathcal{C}^\perp \cap \mathcal{D}^\perp)} \widehat{P}(x)\right)^2 = \frac{1}{2^{2n\ell}} \left(\sum_{x \in (\mathcal{C}^\perp \cap \mathcal{D}^\perp)} \widehat{P}(x)\right)^2 \\ &= \frac{1}{2^{2n\ell}} \sum_{x,\, y \in (\mathcal{C}^\perp \cap \mathcal{D}^\perp)} \widehat{P}(x)\widehat{P}(y). \end{aligned} \tag{A.1}$$

Finally, we obtain $\mathbb{V}\left[\mathbb{E}\left[P(Z)|X\right]\right]$ by combining Lemma 4.2 and Eqn. A.1 as follows.

$$\begin{aligned} \mathbb{V}\left[\mathbb{E}\left[P(Z)|X\right]\right] &= \mathbb{E}\left[\mathbb{E}\left[P(Z)|X\right]^2\right] - \mathbb{E}\left[\mathbb{E}\left[P(Z)|X\right]\right]^2 \\ &= \frac{1}{2^{2n\ell}} \sum_{\substack{x,\, y \in \mathcal{D}^\perp; \\ x+y \in \mathcal{C}^\perp}} \widehat{P}(x)\widehat{P}(y) - \frac{1}{2^{2n\ell}} \sum_{x,\, y \in (\mathcal{C}^\perp \cap \mathcal{D}^\perp)} \widehat{P}(x)\widehat{P}(y) \\ &= \frac{1}{2^{2n\ell}} \left(\sum_{\substack{x,\, y \in \mathcal{D}^\perp; \\ x+y \in \mathcal{C}^\perp}} \widehat{P}(x)\widehat{P}(y) - \sum_{\substack{x,\, y \in \mathcal{D}^\perp; \\ x,\, y \in \mathcal{C}^\perp}} \widehat{P}(x)\widehat{P}(y) \right). \end{aligned} \tag{A.2}$$

Due to Lemma 3.3, we have $\mathcal{C}^\perp \cap \mathcal{D}^\perp = (\mathcal{C} \oplus \mathcal{D})^\perp$ in SSS-based polynomial masking, where $\oplus$ denotes the direct sum operation. Notice that $\{(x, y) \in \mathbb{K}^n \times \mathbb{K}^n | x, y \in \mathcal{D}^\perp,\ x + y \in \mathcal{C}^\perp\} \supseteq$

$\{(x,y) \in (\mathcal{D}^\perp \cap \mathcal{C}^\perp) \times (\mathcal{D}^\perp \cap \mathcal{C}^\perp)\}$. This means that in Eqn. A.2, the subtracted terms are already included in the first sum. Indeed, if $x \in \mathcal{D}^\perp$ also satisfies $x \in \mathcal{C}^\perp$, then $x+y \in \mathcal{C}^\perp$ in the first sum implies $y \in \mathcal{C}^\perp$. Therefore, Eqn. A.2 can be rewritten as follows:

$$\begin{aligned}\mathbb{V}\left[\mathbb{E}\left[P(Z)|X\right]\right] &= \mathbb{E}\left[\mathbb{E}\left[P(Z)|X\right]^2\right] - \mathbb{E}\left[\mathbb{E}\left[P(Z)|X\right]\right]^2 \\ &= \frac{1}{2^{2n\ell}} \sum_{x,\, y \in \mathcal{D}^\perp \setminus \mathcal{C}^\perp;\, x+y \in \mathcal{C}^\perp} \widehat{P}(x)\widehat{P}(y).\end{aligned} \tag{A.3}$$

□

A.1.2 Proof of Lemma 4.1

Proof. Note that $\mathcal{C} \cap \mathcal{D} = \{0\}$, while $(\mathcal{C} \oplus \mathcal{D})^\perp = (\mathcal{C}^\perp \cap \mathcal{D}^\perp) \supseteq \{0\}$. We have

$$\begin{aligned}\mathbb{E}\left[\mathbb{E}\left[P(Z)|X\right]\right] &= \frac{1}{|\mathcal{X}|}\sum_{x\in\mathcal{X}}\left(\frac{1}{|\mathcal{Y}|}\sum_{d\in\mathcal{D}} P(x\mathbf{G}+d)\right) = \frac{1}{|\mathcal{C}|}\sum_{c\in\mathcal{C}}\left(\frac{1}{|\mathcal{D}|}\sum_{d\in\mathcal{D}} P(c+d)\right) \\ &= \frac{1}{|\mathcal{C}||\mathcal{D}|}\sum_{c\in\mathcal{C},\, d\in\mathcal{D}} P(c+d) \\ &= \frac{1}{|\mathcal{C}||\mathcal{D}|}\cdot\frac{1}{2^{n\ell}}\sum_{c\in\mathcal{C},\, d\in\mathcal{D}}\sum_{x\in\mathbb{F}_2^{n\ell}} \widehat{P}(x)(-1)^{(c+d)\cdot x} \qquad \triangleright \textit{ By Lemma 1.2} \\ &= \frac{1}{|\mathcal{C}||\mathcal{D}|}\cdot\frac{1}{2^{n\ell}}\sum_{x\in\mathbb{F}_2^{n\ell}} \widehat{P}(x)\left(\sum_{c\in\mathcal{C}}(-1)^{c\cdot x}\right)\left(\sum_{d\in\mathcal{D}}(-1)^{d\cdot x}\right) \\ &= \frac{1}{2^{n\ell}}\sum_{x\in\mathbb{F}_2^{n\ell}} \widehat{P}(x)\mathbb{1}_{\mathcal{C}^\perp}(x)\mathbb{1}_{\mathcal{D}^\perp}(x) = \frac{1}{2^{n\ell}}\sum_{x\in\mathcal{C}^\perp,\, x\in\mathcal{D}^\perp} \widehat{P}(x) \\ &= \frac{1}{2^{n\ell}}\sum_{x\in(\mathcal{C}+\mathcal{D})^\perp} \widehat{P}(x)\ . \qquad \triangleright \textit{ By Lemma 3.3}\end{aligned} \tag{A.4}$$

□

A.1.3 Proof of Lemma 4.2

Proof. By definition,

$$\mathbb{E}\left[\mathbb{E}\left[P(Z)|X\right]^2\right] = \frac{1}{|\mathcal{C}|}\sum_{c\in\mathcal{C}}\left(\frac{1}{|\mathcal{D}|}\sum_{d\in\mathcal{D}} P(c+d)\right)^2 = \frac{1}{|\mathcal{C}||\mathcal{D}|^2}\sum_{c\in\mathcal{C}}\left(\sum_{d\in\mathcal{D}} P(c+d)\right)^2. \tag{A.5}$$

We have:

$$\sum_{c\in\mathcal{C}}\left(\sum_{d\in\mathcal{D}} P(c+d)\right)^2 = \frac{1}{2^{n\ell}}\cdot\frac{1}{2^{n\ell}}\sum_{\substack{c\in\mathcal{C},\, d,d'\in\mathcal{D}\\ x,y\in\mathbb{F}_2^{n\ell}}} \widehat{P}(x)\widehat{P}(y)(-1)^{x\cdot(c+d)+y\cdot(c+d')}, \tag{A.6}$$

since, according to the inverse *Fourier transform* (by using Lemma 1.2), we have:

$$P(a) = 2^{-n\ell}\sum_{x\in\mathbb{F}_2^{n\ell}} \widehat{P}(x)(-1)^{x\cdot a}\ .$$

Hence we obtain

$$\begin{aligned}\text{Eqn. A.6} &= \frac{1}{2^{n\ell}} \cdot \frac{1}{2^{n\ell}} \sum_{\substack{c\in\mathcal{C},\, d,d'\in\mathcal{D} \\ x,\, y\in\mathbb{F}_2^{n\ell}}} \widehat{P}(x)\widehat{P}(y)(-1)^{(x+y)\cdot c + x\cdot d + y\cdot d'} \\ &= \frac{1}{2^{n\ell}} \cdot \frac{1}{2^{n\ell}} \sum_{\substack{c\in\mathcal{C},\, d,d'\in\mathcal{D} \\ x,\, y\in\mathbb{F}_2^{n\ell}}} \widehat{P}(x)\widehat{P}(y)(-1)^{(x+y)\cdot c}(-1)^{x\cdot d}(-1)^{y\cdot d'} \\ &= \frac{1}{2^{2n\ell}} \cdot |\mathcal{C}| \cdot |\mathcal{D}|^2 \sum_{x,\, y\in\mathcal{D}^\perp;\, x+y\in\mathcal{C}^\perp} \widehat{P}(x)\widehat{P}(y),\end{aligned} \tag{A.7}$$

where $\mathcal{C}$, $\mathcal{D}$ are not necessary to be complementary codes and $|\mathcal{C}||\mathcal{D}| = 2^{t\ell} \leq 2^{n\ell}$. Indeed, since $\mathcal{C}$ is linear, $\sum_{c\in\mathcal{C}}(-1)^{(x+y)\cdot c}$ is null when $x+y$ does not belong to $\mathcal{C}^\perp$ and equals the size of $\mathcal{C}$ if it does, and the same with $\mathcal{D}$. Note that x, $y \in \mathcal{D}^\perp$ and $x+y \in \mathcal{C}^\perp$ which implies $x+y \in \mathcal{C}^\perp \cap \mathcal{D}^\perp$. In summary, we have the following result for $\mathbb{E}\left[\mathbb{E}\left[P(Z)|X\right]^2\right]$.

$$\begin{aligned}\mathbb{E}\left[\mathbb{E}\left[P(Z)|X\right]^2\right] &= \frac{1}{|\mathcal{C}||\mathcal{D}|^2} \cdot \frac{1}{2^{2n\ell}} \cdot |\mathcal{C}| \cdot |\mathcal{D}|^2 \sum_{x,\, y\in\mathcal{D}^\perp;\, x+y\in\mathcal{C}^\perp} \widehat{P}(x)\widehat{P}(y) \\ &= \frac{1}{2^{2n\ell}} \sum_{x,\, y\in\mathcal{D}^\perp;\, x+y\in\mathcal{C}^\perp} \widehat{P}(x)\widehat{P}(y).\end{aligned} \tag{A.8}$$

□

APPENDIX B

Generator Matrices for Some Optimal Linear Codes

B.1 Optimal Codes for IPM with $n = 2$

This appendix provides the details about the three non-equivalent optimal codes identified by Alg. 1 and reported in the last line of Tab. 4.2.

Extension of the first optimal code from $\mathbb{F}_{256}$ to $\mathbb{F}_2$. The generating matrix for the expanded code spanned by $\begin{pmatrix} 1 & \alpha^8 \end{pmatrix}$ from $\mathbb{F}_{256}$ on the base field $\mathbb{F}_2$ is:

$$\mathbf{H}_1^{\perp} = \begin{pmatrix} 1\,0\,0\,0\,0\,0\,0\,0 & 1\,0\,1\,1\,1\,0\,0\,0 \\ 0\,1\,0\,0\,0\,0\,0\,0 & 0\,1\,0\,1\,1\,1\,0\,0 \\ 0\,0\,1\,0\,0\,0\,0\,0 & 0\,0\,1\,0\,1\,1\,1\,0 \\ 0\,0\,0\,1\,0\,0\,0\,0 & 0\,0\,0\,1\,0\,1\,1\,1 \\ 0\,0\,0\,0\,1\,0\,0\,0 & 1\,0\,1\,1\,0\,0\,1\,1 \\ 0\,0\,0\,0\,0\,1\,0\,0 & 1\,1\,1\,0\,0\,0\,0\,1 \\ 0\,0\,0\,0\,0\,0\,1\,0 & 1\,1\,0\,0\,1\,0\,0\,0 \\ 0\,0\,0\,0\,0\,0\,0\,1 & 0\,1\,1\,0\,0\,1\,0\,0 \end{pmatrix} \in \mathbb{F}_2^{8\times 16}.$$

Extension of the second optimal code from $\mathbb{F}_{256}$ to $\mathbb{F}_2$. The generating matrix for the expanded code spanned by $\begin{pmatrix} 1 & \alpha^{126} \end{pmatrix}$ from $\mathbb{F}_{256}$ on the base field $\mathbb{F}_2$ is:

$$\mathbf{H}_2^{\perp} = \begin{pmatrix} 1\,0\,0\,0\,0\,0\,0\,0 & 1\,0\,1\,1\,1\,0\,0\,0 \\ 0\,1\,0\,0\,0\,0\,0\,0 & 0\,1\,0\,1\,1\,1\,0\,0 \\ 0\,0\,1\,0\,0\,0\,0\,0 & 0\,0\,1\,0\,1\,1\,1\,0 \\ 0\,0\,0\,1\,0\,0\,0\,0 & 0\,0\,0\,1\,0\,1\,1\,1 \\ 0\,0\,0\,0\,1\,0\,0\,0 & 1\,0\,1\,1\,0\,0\,1\,1 \\ 0\,0\,0\,0\,0\,1\,0\,0 & 1\,1\,1\,0\,0\,0\,0\,1 \\ 0\,0\,0\,0\,0\,0\,1\,0 & 1\,1\,0\,0\,1\,0\,0\,0 \\ 0\,0\,0\,0\,0\,0\,0\,1 & 0\,1\,1\,0\,0\,1\,0\,0 \end{pmatrix} \in \mathbb{F}_2^{8\times 16}.$$

Extension of the third optimal code from $\mathbb{F}_{256}$ to $\mathbb{F}_2$. The generating matrix for the expanded code spanned by $\begin{pmatrix}1 & \alpha^{127}\end{pmatrix}$ from $\mathbb{F}_{256}$ on the base field $\mathbb{F}_2$ is:

$$\mathbf{H}_3^{\perp} = \begin{pmatrix}
1\,0\,0\,0\,0\,0\,0\,0 & 0\,0\,1\,1\,0\,0\,1\,1 \\
0\,1\,0\,0\,0\,0\,0\,0 & 1\,0\,1\,0\,0\,0\,0\,1 \\
0\,0\,1\,0\,0\,0\,0\,0 & 1\,1\,1\,0\,1\,0\,0\,0 \\
0\,0\,0\,1\,0\,0\,0\,0 & 0\,1\,1\,1\,0\,1\,0\,0 \\
0\,0\,0\,0\,1\,0\,0\,0 & 0\,0\,1\,1\,1\,0\,1\,0 \\
0\,0\,0\,0\,0\,1\,0\,0 & 0\,0\,0\,1\,1\,1\,0\,1 \\
0\,0\,0\,0\,0\,0\,1\,0 & 1\,0\,1\,1\,0\,1\,1\,0 \\
0\,0\,0\,0\,0\,0\,0\,1 & 0\,1\,0\,1\,1\,0\,1\,1
\end{pmatrix} \in \mathbb{F}_2^{8\times 16}.$$

B.2 Optimal Codes for $(3,1)$-SSS based Masking

As shown in Tab. 5.1, the generator matrix of $\mathcal{D}$ is $\mathbf{H} = \begin{pmatrix}\alpha_1 & \alpha_2 & \alpha_3\end{pmatrix}$. From an exhaustive study on 32131 candidates, the three optimal codes for $(3,1)$-SSS based masking are: $(\alpha_1,\alpha_2,\alpha_3) \in \{(\alpha^0,\alpha^{72},\alpha^{80}), (\alpha^0,\alpha^{175},\alpha^{247}), (\alpha^0,\alpha^{8},\alpha^{183})\}$. Note that permutation on three public points does not change the codes due to equivalence.

The generator matrices of the three optimal codes are shown below.

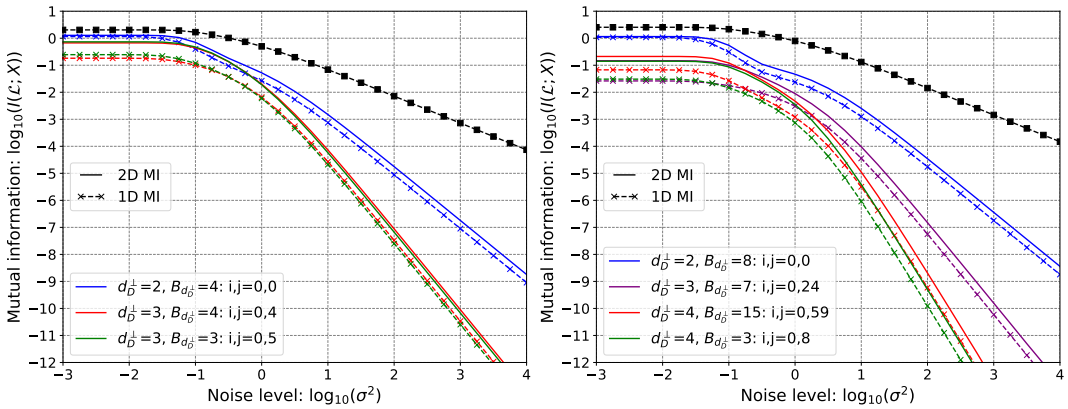
$$\mathbf{H}_1 = \begin{pmatrix}\alpha^0 & \alpha^{72} & \alpha^{80}\end{pmatrix} = \begin{pmatrix}
1\,0\,0\,0\,0\,0\,0\,0 & 1\,0\,1\,0\,0\,1\,1\,0 & 1\,0\,1\,1\,1\,1\,1\,1 \\
0\,1\,0\,0\,0\,0\,0\,0 & 0\,1\,0\,1\,0\,0\,1\,1 & 1\,1\,1\,0\,0\,1\,1\,1 \\
0\,0\,1\,0\,0\,0\,0\,0 & 1\,0\,0\,1\,0\,0\,0\,1 & 1\,1\,0\,0\,1\,0\,1\,1 \\
0\,0\,0\,1\,0\,0\,0\,0 & 1\,1\,1\,1\,0\,0\,0\,0 & 1\,1\,0\,1\,1\,1\,0\,1 \\
0\,0\,0\,0\,1\,0\,0\,0 & 0\,1\,1\,1\,1\,0\,0\,0 & 1\,1\,0\,1\,0\,1\,1\,0 \\
0\,0\,0\,0\,0\,1\,0\,0 & 0\,0\,1\,1\,1\,1\,0\,0 & 0\,1\,1\,0\,1\,0\,1\,1 \\
0\,0\,0\,0\,0\,0\,1\,0 & 0\,0\,0\,1\,1\,1\,1\,0 & 1\,0\,0\,0\,1\,1\,0\,1 \\
0\,0\,0\,0\,0\,0\,0\,1 & 0\,0\,0\,0\,1\,1\,1\,1 & 1\,1\,1\,1\,1\,1\,1\,0
\end{pmatrix} \in \mathbb{F}_2^{8\times 24}$$

$$\mathbf{H}_2 = \begin{pmatrix}\alpha^0 & \alpha^{175} & \alpha^{247}\end{pmatrix} = \begin{pmatrix}
1\,0\,0\,0\,0\,0\,0\,0 & 1\,1\,1\,1\,1\,1\,1\,1 & 1\,1\,0\,0\,0\,0\,0\,1 \\
0\,1\,0\,0\,0\,0\,0\,0 & 1\,1\,0\,0\,0\,1\,1\,1 & 1\,1\,0\,1\,1\,0\,0\,0 \\
0\,0\,1\,0\,0\,0\,0\,0 & 1\,1\,0\,1\,1\,0\,1\,1 & 0\,1\,1\,0\,1\,1\,0\,0 \\
0\,0\,0\,1\,0\,0\,0\,0 & 1\,1\,0\,1\,0\,1\,0\,1 & 0\,0\,1\,1\,0\,1\,1\,0 \\
0\,0\,0\,0\,1\,0\,0\,0 & 1\,1\,0\,1\,0\,0\,1\,0 & 0\,0\,0\,1\,1\,0\,1\,1 \\
0\,0\,0\,0\,0\,1\,0\,0 & 0\,1\,1\,0\,1\,0\,0\,1 & 1\,0\,1\,1\,0\,1\,0\,1 \\
0\,0\,0\,0\,0\,0\,1\,0 & 1\,0\,0\,0\,1\,1\,0\,0 & 1\,1\,1\,0\,0\,0\,1\,0 \\
0\,0\,0\,0\,0\,0\,0\,1 & 0\,1\,0\,0\,0\,1\,1\,0 & 0\,1\,1\,1\,0\,0\,0\,1
\end{pmatrix} \in \mathbb{F}_2^{8\times 24}$$

$$\mathbf{H}_3 = \begin{pmatrix} \alpha^0 & \alpha^8 & \alpha^{183} \end{pmatrix} = \begin{pmatrix} 1\,0\,0\,0\,0\,0\,0\,0 & 1\,0\,1\,1\,1\,0\,0\,0 & 0\,0\,1\,0\,0\,0\,1\,1 \\ 0\,1\,0\,0\,0\,0\,0\,0 & 0\,1\,0\,1\,1\,1\,0\,0 & 1\,0\,1\,0\,1\,0\,0\,1 \\ 0\,0\,1\,0\,0\,0\,0\,0 & 0\,0\,1\,0\,1\,1\,1\,0 & 1\,1\,1\,0\,1\,1\,0\,0 \\ 0\,0\,0\,1\,0\,0\,0\,0 & 0\,0\,0\,1\,0\,1\,1\,1 & 0\,1\,1\,1\,0\,1\,1\,0 \\ 0\,0\,0\,0\,1\,0\,0\,0 & 1\,0\,1\,1\,0\,0\,1\,1 & 0\,0\,1\,1\,1\,0\,1\,1 \\ 0\,0\,0\,0\,0\,1\,0\,0 & 1\,1\,1\,0\,0\,0\,0\,1 & 1\,0\,1\,0\,0\,1\,0\,1 \\ 0\,0\,0\,0\,0\,0\,1\,0 & 1\,1\,0\,0\,1\,0\,0\,0 & 1\,1\,1\,0\,1\,0\,1\,0 \\ 0\,0\,0\,0\,0\,0\,0\,1 & 0\,1\,1\,0\,0\,1\,0\,0 & 0\,1\,1\,1\,0\,1\,0\,1 \end{pmatrix} \in \mathbb{F}_2^{8\times 24}$$

B.3 Comparison of MI on 1-D and n-D Leakages

We add more results on MI to compare the efficiency of different combination functions φ_P in exploiting information leakages. In Fig. 5.3, we show the advantages to use joint distribution in trivariate leakages. In addition, we compare the two combination function in 2-share cases by plotting MI curves together. As shown in Fig. B.1, the combination by using joint distribution is more efficient than the one by using sum in bivariate leakages scenarios. Moreover, this is true for n-variate leakages.



(a) 1-D vs. 2-D MI on 4-bit case.

(b) 1-D vs. 2-D MI on 8-bit case.

Figure B.1: Comparing 1-D and 2-D MI on different linear codes where the sum and joint distribution are used to combine the bivariate leakages, respectively. Note that the blue curves are for the Boolean masking.

More importantly, the superiority of GCM can be fully unleashed by choosing appropriate codes. In this respect, our leakage quantitation approach is a simple, generic and effective way to choose the optimal codes for GCM.

APPENDIX C

The Impact of Encoding on Leakage Distributions

C.1 The Impact of Encoding on Leakage Distributions

In this section, we further show the distinct impact of different encoding in Boolean masking, IPM and DSM. Note that IPM is a special case of DSM, but not vice versa.

The leakage distribution in bivariate manner is shown in Fig. C.1, and two cases of IPM are in Fig. C.2 and C.1 for $\alpha_1 = \alpha$ and $\alpha_1 = \alpha^5$, respectively. At last, DSM equipped with the *BKLC* code $[8, 4, 4]$ is shown in Fig. C.4.

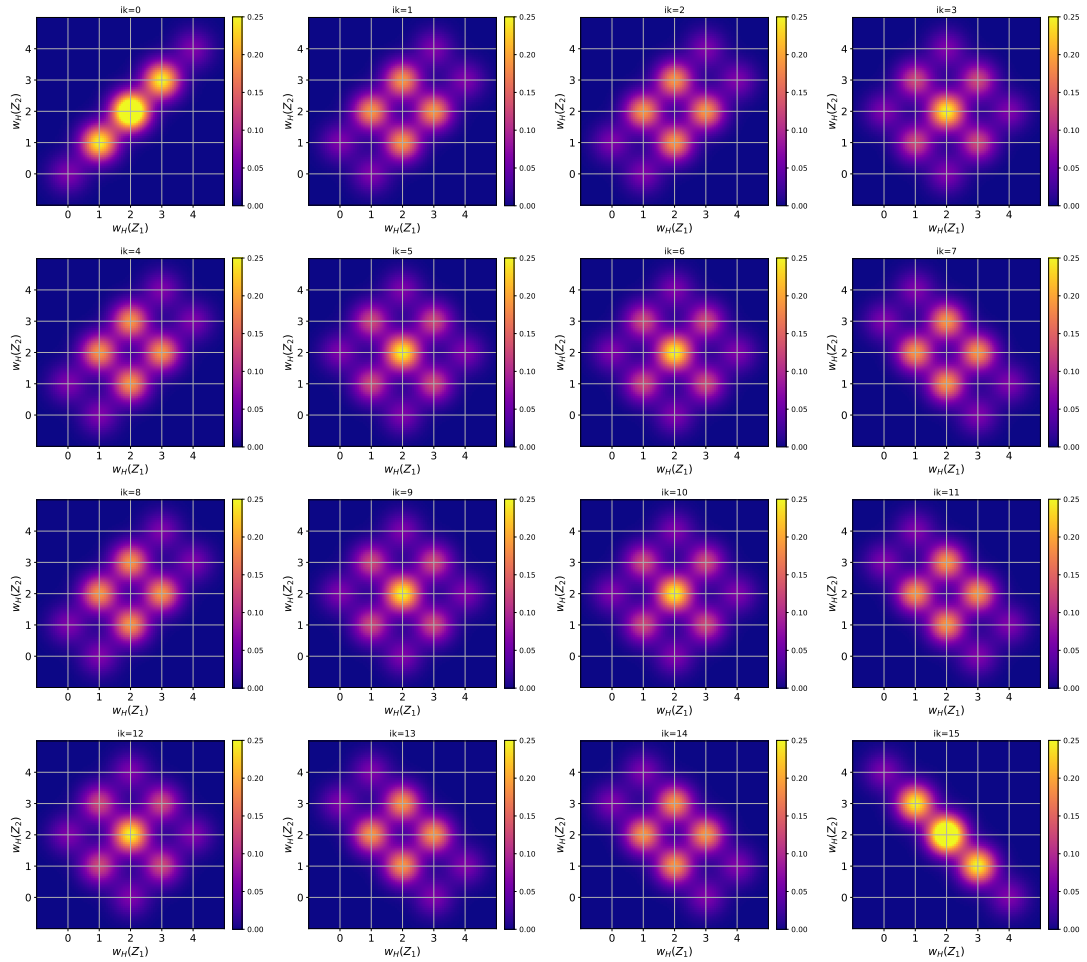


Figure C.1: Bivariate leakage distribution of 2-share Boolean masking under the Hamming weight model.

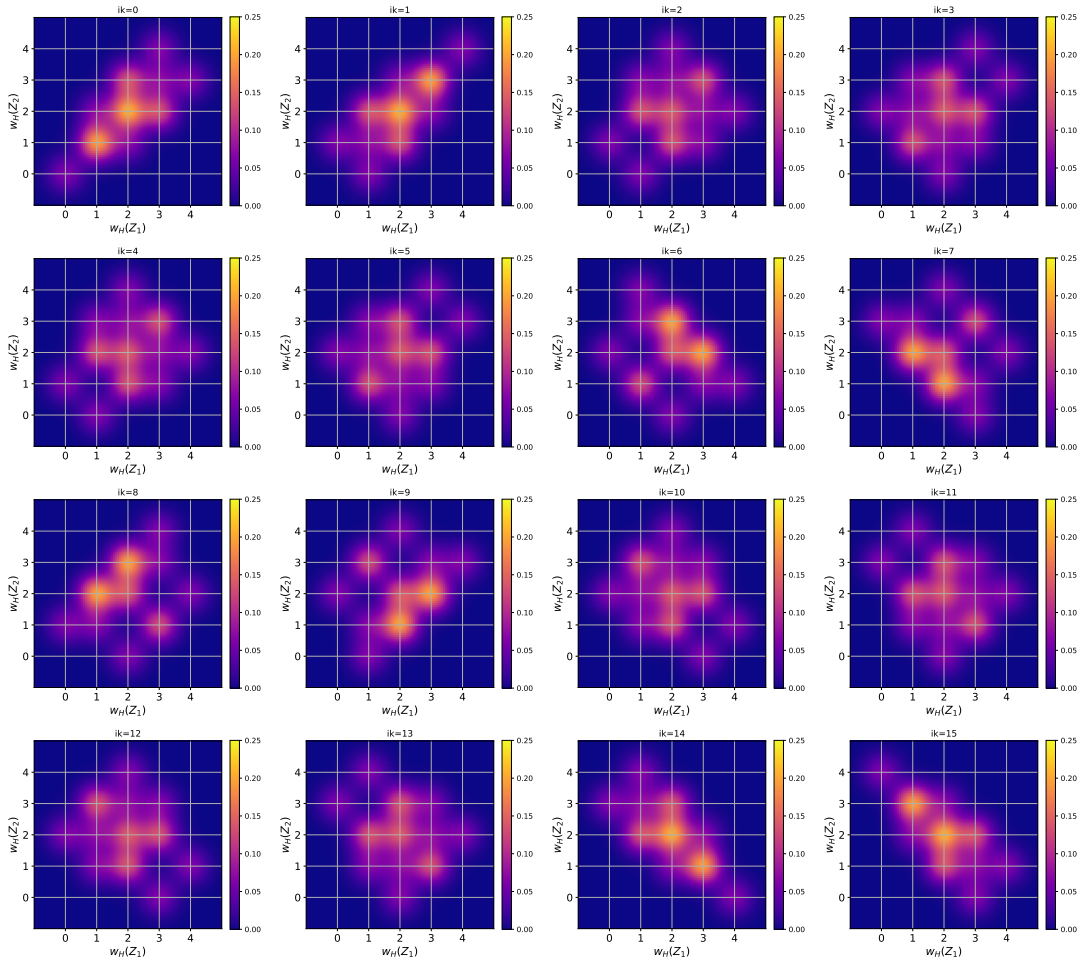


Figure C.2: Bivariate leakage distribution of 2-share IPM under the Hamming weight model, with $\alpha_1 = \alpha^1$.

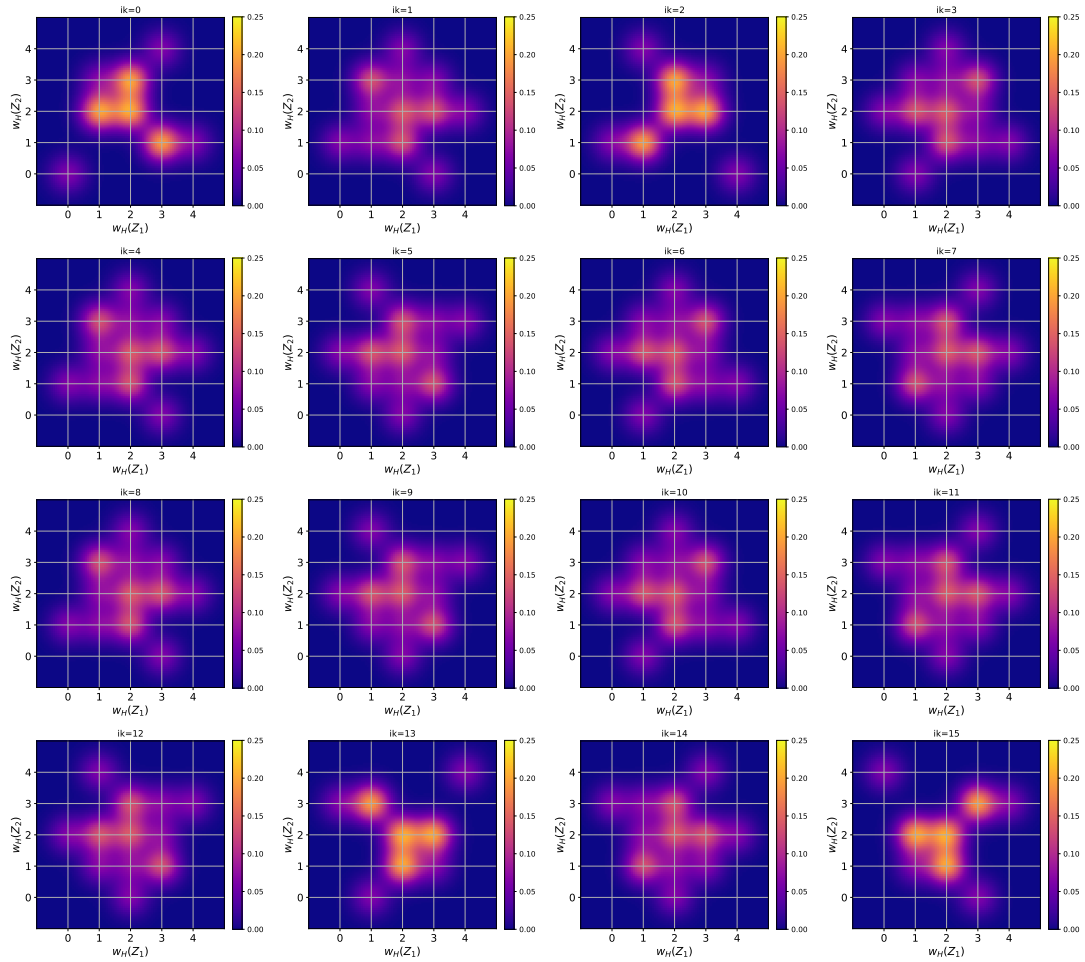


Figure C.3: Bivariate leakage distribution of 2-share IPM under the Hamming weight model, with $\alpha_1 = \alpha^5$, which is one of the optimal case.

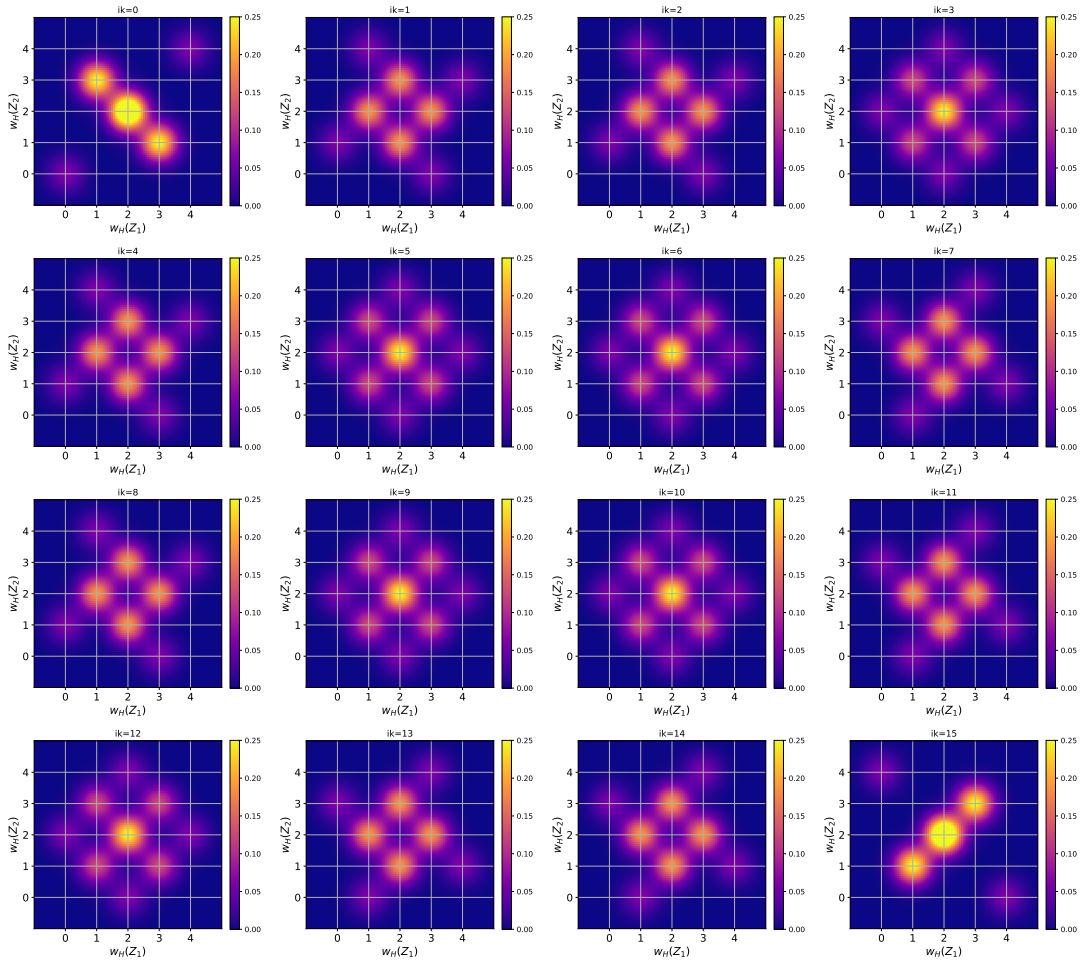


Figure C.4: Bivariate leakage distribution of 2-share DSM over bits under the Hamming weight model, with *BKLC* code $[8, 4, 4]$ involved.

Bibliography

[1] 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. 138, 140, 152

[2] Josep Balasch, Sebastian Faust, and Benedikt Gierlichs. Inner Product Masking Revisited. In Oswald and Fischlin [118], pages 486–510. 7, 28, 29, 30, 31, 41, 60, 61, 66, 116

[3] Josep Balasch, Sebastian Faust, Benedikt Gierlichs, Clara Paglialonga, and François-Xavier Standaert. Consolidating Inner Product Masking. In Tsuyoshi Takagi and Thomas Peyrin, editors, *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*, volume 10624 of *Lecture Notes in Computer Science*, pages 724–754. Springer, 2017. 7, 8, 28, 29, 30, 40, 41, 53, 60, 62, 66, 101, 103, 116, 130

[4] Josep Balasch, Sebastian Faust, Benedikt Gierlichs, and Ingrid Verbauwhede. Theory and Practice of a Leakage Resilient Masking Scheme. In Xiaoyun Wang and Kazue Sako, editors, *Advances in Cryptology - ASIACRYPT 2012 - 18th International Conference on the Theory and Application of Cryptology and Information Security, Beijing, China, December 2-6, 2012. Proceedings*, volume 7658 of *Lecture Notes in Computer Science*, pages 758–775. Springer, 2012. 7, 28, 29, 60

[5] Josep Balasch, Benedikt Gierlichs, Vincent Grosso, Oscar Reparaz, and François-Xavier Standaert. On the Cost of Lazy Engineering for Masked Software Implementations. In Marc Joye and Amir Moradi, editors, *Smart Card Research and Advanced Applications - 13th International Conference, CARDIS 2014, Paris, France, November 5-7, 2014. Revised Selected Papers*, volume 8968 of *Lecture Notes in Computer Science*, pages 64–81. Springer, 2014. 9, 94, 95

[6] Gilles Barthe, Sonia Belaïd, François Dupressoir, Pierre-Alain Fouque, Benjamin Grégoire, Pierre-Yves Strub, and Rébecca Zucchini. Strong non-interference and type-directed higher-order masking. In Edgar R. Weippl, Stefan Katzenbeisser, Christopher Kruegel, Andrew C. Myers, and Shai Halevi, editors, *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, Vienna, Austria, October 24-28, 2016*, pages 116–129. ACM, 2016. 60

[7] Gilles Barthe, François Dupressoir, Sebastian Faust, Benjamin Grégoire, François-Xavier Standaert, and Pierre-Yves Strub. Parallel Implementations of Masking Schemes and the Bounded Moment Leakage Model. In Jean-Sébastien Coron and Jesper Buus Nielsen, editors, *Advances in Cryptology - EUROCRYPT 2017 - 36th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Paris, France, April 30 - May 4, 2017, Proceedings, Part I*, volume 10210 of *Lecture Notes in Computer Science*, pages 535–566, 2017. 29, 94

[8] Gilles Barthe, François Dupressoir, Sebastian Faust, Benjamin Grégoire, François-Xavier Standaert, and Pierre-Yves Strub. Parallel Implementations of Masking Schemes and the Bounded Moment Leakage Model. In *Advances in Cryptology - EUROCRYPT 2017, Paris, France, April 30 - May 4, 2017, Proceedings, Part I*, pages 535–566, 2017. 63, 64, 67, 86

[9] Lejla Batina, Benedikt Gierlichs, Emmanuel Prouff, Matthieu Rivain, François-Xavier Standaert, and Nicolas Veyrat-Charvillon. Mutual Information Analysis: a Comprehensive Study. *J. Cryptology*, 24(2):269–291, 2011. 139

[10] Ryad Benadjila, Emmanuel Prouff, Rémi Strullu, Eleonora Cagli, and Cécile Dumas. Deep learning for side-channel analysis and introduction to ASCAD database. *J. Cryptogr. Eng.*, 10(2):163–188, 2020. 6

[11] Koichi Betsumiya and Masaaki Harada. Binary optimal odd formally self-dual codes. *Des. Codes Cryptography*, 23(1):11–22, 2001. http://www.math.nagoya-u.ac.jp/~koichi/paper/fsd-odd.pdf. 49

[12] Shivam Bhasin, Nicolas Bruneau, Jean-Luc Danger, Sylvain Guilley, and Zakaria Najm. Analysis and improvements of the DPA contest v4 implementation. In Rajat Subhra Chakraborty, Vashek Matyas, and Patrick Schaumont, editors, *Security, Privacy, and Applied Cryptography Engineering - 4th International Conference, SPACE 2014, Pune, India, October 18-22, 2014. Proceedings*, volume 8804 of *Lecture Notes in Computer Science*, pages 201–218. Springer, 2014. 89, 90

[13] Shivam Bhasin, Jean-Luc Danger, Sylvain Guilley, and Zakaria Najm. NICV: Normalized Inter-Class Variance for Detection of Side-Channel Leakage. In *International Symposium on Electromagnetic Compatibility (EMC '14 / Tokyo)*. IEEE, May 12-16 2014. Session OS09: EM Information Leakage. Hitotsubashi Hall (National Center of Sciences), Chiyoda, Tokyo, Japan. 9

[14] Shivam Bhasin, Jean-Luc Danger, Sylvain Guilley, and Zakaria Najm. Side-Channel Leakage and Trace Compression using Normalized Inter-Class Variance. *IACR Cryptology ePrint Archive*, 2014:1020, 2014. 10

[15] Johannes Blömer, Jorge Guajardo, and Volker Krummel. Provably Secure Masking of AES. In Helena Handschuh and M. Anwar Hasan, editors, *Selected Areas in Cryptography*, volume 3357 of *Lecture Notes in Computer Science*, pages 69–83. Springer, 2004. 94

[16] Éric Brier, Christophe Clavier, and Francis Olivier. Correlation power analysis with a leakage model. In Marc Joye and Jean-Jacques Quisquater, editors, *Cryptographic Hardware and Embedded Systems - CHES 2004: 6th International Workshop Cambridge, MA, USA, August 11-13, 2004. Proceedings*, volume 3156 of *Lecture Notes in Computer Science*, pages 16–29. Springer, 2004. 5, 10, 25, 70, 95, 99

[17] 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 David Naccache and Damien Sauveron, editors, *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, volume 8501 of *Lecture Notes in Computer Science*, pages 40–56. Springer, 2014. 7, 36, 49, 53, 60, 66, 72

[18] Nicolas Bruneau, Sylvain Guilley, Annelie Heuser, and Olivier Rioul. Masks Will Fall Off – Higher-Order Optimal Distinguishers. In Palash Sarkar and Tetsu Iwata, editors, *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*, volume 8874 of *Lecture Notes in Computer Science*, pages 344–365. Springer, 2014. 10, 30, 76, 88, 95, 101, 128

[19] Eleonora Cagli, Cécile Dumas, and Emmanuel Prouff. Convolutional Neural Networks with Data Augmentation Against Jitter-Based Countermeasures - Profiling Attacks Without Pre-processing. In Fischer and Homma [68], pages 45–68. 6

[20] Giovanni Camurati, Sebastian Poeplau, Marius Muench, Tom Hayes, and Aurélien Francillon. Screaming Channels: When Electromagnetic Side Channels Meet Radio Transceivers. In David Lie, Mohammad Mannan, Michael Backes, and XiaoFeng Wang, editors, *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security, CCS 2018, Toronto, ON, Canada, October 15-19, 2018*, pages 163–177. ACM, 2018. 5

[21] Jean-François Cardoso. Dependence, Correlation and Gaussianity in Independent Component Analysis. *Journal of Machine Learning Research*, 4:1177–1203, 2003. ISSN 1533-7928. 43, 74

[22] Claude Carlet. Boolean Functions for Cryptography and Error Correcting Codes: Chapter of the monography Boolean Models and Methods in Mathematics, Computer Science, and Engineering. pages 257–397. Cambridge University Press, Y. Crama and P. Hammer eds, 2010. Preliminary version is available at `http://www.math.univ-paris13.fr/~carlet/chap-fcts-Bool-corr.pdf`. 25, 29, 43, 74, 173

[23] Claude Carlet, Jean-Luc Danger, Sylvain Guilley, Houssem Maghrebi, and Emmanuel Prouff. Achieving side-channel high-order correlation immunity with leakage squeezing. *J. Cryptographic Engineering*, 4(2):107–121, 2014. 7, 43, 53, 60, 66, 74

[24] 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. 7, 29, 30, 46, 53, 60, 62, 89, 120, 126

[25] Claude Carlet and Philippe Guillot. A New Representation of Boolean Functions. In Marc P. C. Fossorier, Hideki Imai, Shu Lin, and Alain Poli, editors, *AAECC*, volume 1719 of *Lecture Notes in Computer Science*, pages 94–103. Springer, 1999. 25, 36, 70

[26] Claude Carlet, Cem Güneri, Sihem Mesnager, and Ferruh Özbudak. Construction of some codes suitable for both side channel and fault injection attacks. In Lilya Budaghyan and Francisco Rodríguez-Henríquez, editors, *Arithmetic of Finite Fields - 7th International Workshop, WAIFI 2018, Bergen, Norway, June 14-16, 2018, Revised Selected Papers*, volume 11321 of *Lecture Notes in Computer Science*, pages 95–107. Springer, 2018. 81

[27] Elad Carmon, Jean-Pierre Seifert, and Avishai Wool. Photonic side channel attacks against RSA. In *2017 IEEE International Symposium on Hardware Oriented Security and Trust, HOST 2017, McLean, VA, USA, May 1-5, 2017*, pages 74–78. IEEE Computer Society, 2017. 5

[28] Guilhem Castagnos, Soline Renner, and Gilles Zémor. High-order masking by using coding theory and its application to AES. In Martijn Stam, editor, *Cryptography and Coding - 14th IMA International Conference, IMACC 2013, Oxford, UK, December 17-19, 2013. Proceedings*, volume 8308 of *Lecture Notes in Computer Science*, pages 193–212. Springer, 2013. 66

[29] Hervé Chabanne, Houssem Maghrebi, and Emmanuel Prouff. Linear repairing codes and side-channel attacks. *IACR Trans. Cryptogr. Hardw. Embed. Syst.*, 2018(1):118–141, 2018. 8, 62, 64, 66, 67, 80, 97, 102, 109, 116

[30] Suresh Chari, Charanjit S. Jutla, Josyula R. Rao, and Pankaj Rohatgi. Towards Sound Approaches to Counteract Power-Analysis Attacks. In Wiener [167], pages 398–412. 6, 7, 60, 94

[31] Suresh Chari, Josyula R. Rao, and Pankaj Rohatgi. Template attacks. In Burton S. Kaliski, Jr., Çetin Kaya Koç, and Christof Paar, editors, *Cryptographic Hardware and Embedded Systems - CHES 2002, 4th International Workshop, Redwood Shores, CA, USA, August 13-15, 2002, Revised Papers*, volume 2523 of *Lecture Notes in Computer Science*, pages 13–28. Springer, 2002. 6, 10, 95, 99

[32] Zhimin Chen and Yujie Zhou. Dual-Rail Random Switching Logic: A Countermeasure to Reduce Side Channel Leakage. In *CHES*, volume 4249 of *LNCS*, pages 242–254. Springer, October 10-13 2006. Yokohama, Japan, http://dx.doi.org/10.1007/11894063_20. 94

[33] Wei Cheng, Claude Carlet, Kouassi Goli, Jean-Luc Danger, and Sylvain Guilley. Detecting faults in inner product masking scheme. *J. Cryptogr. Eng.*, 11(2):119–133, 2021. 54, 168

[34] Wei Cheng and Sylvain Guilley. Open-source: Quantifying Information Leakages in GCM, September 2020. http://github.com/Qomo-CHENG/GeneralizedCM. 64, 82, 88, 90

[35] 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. 7, 14, 15, 24, 59, 79, 107, 108, 109, 119, 167, 168

[36] Wei Cheng, Sylvain Guilley, Claude Carlet, Jean-Luc Danger, and Alexander Schaub. Optimal Codes for Inner Product Masking, June 24 - 25 2019. CRYPTARCHI, Pruhonice, Czech republic. https://labh-curien.univ-st-etienne.fr/cryptarchi/workshop19/abstracts/cheng.pdf. 13

[37] 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. xiii, 8, 13, 24, 27, 48, 62, 63, 65, 68, 69, 70, 79, 87, 88, 102, 120, 130, 167, 168

[38] Wei Cheng, Sylvain Guilley, and Jean-Luc Danger. Categorizing All Linear Codes of IPM over $\mathbb{F}_{2^8}$. *Cryptography and Communications and Discrete Structures*, 2020. Link to GitHub sources: https://github.com/Qomo-CHENG/OC-IPM. 27

[39] Wei Cheng, Sylvain Guilley, and Jean-Luc Danger. Optimal Linear Codes for IPM, January 2020. https://github.com/Qomo-CHENG/OC-IPM. xxiii, 40, 46, 52

[40] Wei Cheng, Sylvain Guilley, and Jean-Luc Danger. Categorizing All Linear Codes of IPM over $\mathbb{F}_{2^8}$. *Cryptogr. Commun.*, 13(4):527–542, 2021. 13, 168

[41] Wei Cheng, Yi Liu, Sylvain Guilley, and Olivier Rioul. Attacking masked cryptographic implementations: Information-theoretic bounds. *CoRR*, abs/2105.07436, 2021. 10, 15, 95, 117, 139, 170

[42] Wei Cheng, Yi Liu, Sylvain Guilley, and Olivier Rioul. Towards Finding Best Linear Codes for Side-Channel Protections, September 17 2021. 10th International Workshop on Security Proofs for Embedded Systems (PROOFS). Beijing, China. 59, 79, 102, 169

[43] Wei Cheng, Olivier Rioul, and Sylvain Guilley. Guessing a secret cryptographic key from side-channel leakages. In *2019 IEEE European School of Information Theory (ESIT'19), Sophia Antipolis, France, Apr. 15-19, 2019*, 2019. 15

[44] Wei Cheng, Chao Zheng, Yuchen Cao, Yongbin Zhou, Hailong Zhang, Sylvain Guilley, and Laurent Sauvage. How Far Can We Reach? Breaking RSM-Masked AES-128 Implementation Using Only One Trace. *IACR Cryptology ePrint Archive*, 2017:1144, 2017. 89

[45] Marios O. Choudary and P. G. Popescu. Back to massey: Impressively fast, scalable and tight security evaluation tools. In Fischer and Homma [68], pages 367–386. 145

[46] Christophe Clavier, Jean-Sébastien Coron, and Nora Dabbous. Differential Power Analysis in the Presence of Hardware Countermeasures. In Çetin Kaya Koç and Christof Paar, editors, *CHES*, volume 1965 of *Lecture Notes in Computer Science*, pages 252–263. Springer, 2000. 5, 6, 94

[47] Jeremy Cooper, Gilbert Goodwill, Joshua Jaffe, Gary Kenworthy, and Pankaj Rohatgi. Test Vector Leakage Assessment (TVLA) Methodology in Practice, Sept 24–26 2013. International Cryptographic Module Conference (ICMC), Holiday Inn Gaithersburg, MD, USA. 9, 94

[48] Jean-Sébastien Coron, Christophe Giraud, Emmanuel Prouff, Soline Renner, Matthieu Rivain, and Praveen Kumar Vadnala. Conversion of security proofs from one leakage model to another: A new issue. In Werner Schindler and Sorin A. Huss, editors, *Constructive Side-Channel Analysis and Secure Design - Third International Workshop, COSADE 2012, Darmstadt, Germany, May 3-4, 2012. Proceedings*, volume 7275 of *LNCS*, pages 69–81. Springer, 2012. 94

[49] Jean-Sébastien Coron, Emmanuel Prouff, and Matthieu Rivain. Side Channel Cryptanalysis of a Higher Order Masking Scheme. In Pascal Paillier and Ingrid Verbauwhede, editors, *CHES*, volume 4727 of *LNCS*, pages 28–44. Springer, 2007. 6

[50] Jean-Sébastien Coron and Lorenzo Spignoli. Secure wire shuffling in the probing model. In Tal Malkin and Chris Peikert, editors, *Advances in Cryptology - CRYPTO 2021 - 41st Annual International Cryptology Conference, CRYPTO 2021, Virtual Event, August 16-20,*

2021, Proceedings, Part III, volume 12827 of *Lecture Notes in Computer Science*, pages 215–244. Springer, 2021. 6

[51] Nicolas Costes and Martijn Stam. Redundant code-based masking revisited. *IACR Trans. Cryptogr. Hardw. Embed. Syst.*, 2021(1):426–450, 2021. xiii, 61, 65, 79, 83, 88, 97, 102, 109, 116

[52] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. Wiley-Interscience, July 18 2006. ISBN-10: ISBN-10: 0471241954, ISBN-13: 978-0471241959, 2nd edition. 120, 122, 124

[53] Imre Csiszár. Generalized cutoff rates and Rényi's information measures. *IEEE Trans. Inf. Theory*, 41(1):26–34, 1995. 146

[54] Jean-Luc Danger, Sylvain Guilley, Annelie Heuser, Axel Legay, and Ming Tang. Physical Security Versus Masking Schemes. In Çetin Kaya Koç, editor, *Cyber-Physical Systems Security.*, pages 269–284. Springer, 2018. 68

[55] Éloi de Chérisey, Sylvain Guilley, Annelie Heuser, and Olivier Rioul. On the optimality and practicability of mutual information analysis in some scenarios. *Cryptography and Communications*, 10(1):101–121, 2018. 95

[56] Éloi de Chérisey, Sylvain Guilley, Olivier Rioul, and Pablo Piantanida. Best information is most successful. Cryptology ePrint Archive, Report 2019/491, extended version of [57], 2019. https://eprint.iacr.org/2019/491. 123

[57] 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. xx, 10, 95, 118, 119, 120, 122, 124, 125, 128, 129, 132, 139, 150, 151, 157, 194

[58] Éloi de Chérisey, Sylvain Guilley, Olivier Rioul, and Pablo Piantanida. An information-theoretic model for side-channel attacks in embedded hardware. In *IEEE International Symposium on Information Theory, ISIT 2019, Paris, France, July 7-12, 2019*, pages 310–315. IEEE, 2019. 132, 150

[59] Jean-F. Dhem, F. Koeune, P.-A. Leroux, P. Mestre, J.-J. Quisquater, and J.-L. Willems. A Practical Implementation of the Timing Attack. In *CARDIS*, pages 167–182, 1998. http://citeseer.nj.nec.com/dhem98practical.html. 5

[60] Alexandre Duc, Stefan Dziembowski, and Sebastian Faust. Unifying Leakage Models: From Probing Attacks to Noisy Leakage. In Phong Q. Nguyen and Elisabeth Oswald, editors, *Advances in Cryptology - EUROCRYPT 2014 - 33rd Annual International Conference on the Theory and Applications of Cryptographic Techniques, Copenhagen, Denmark, May 11-15, 2014. Proceedings*, volume 8441 of *Lecture Notes in Computer Science*, pages 423–440. Springer, 2014. 9, 63, 95

[61] Alexandre Duc, Sebastian Faust, and François-Xavier Standaert. Making Masking Security Proofs Concrete - Or How to Evaluate the Security of Any Leaking Device. In Oswald and Fischlin [118], pages 401–429. 63, 118

[62] Stefan Dziembowski, Sebastian Faust, and Maciej Skorski. Noisy leakage revisited. 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 II*, volume 9057 of *Lecture Notes in Computer Science*, pages 159–188. Springer, 2015. 9, 95

[63] Amedeo Roberto Esposito, Michael Gastpar, and Ibrahim Issa. Generalization error bounds via rényi-, f-divergences and maximal leakage. 138

[64] Amedeo Roberto Esposito, Diyuan Wu, and Michael Gastpar. On conditional sibson's α-mutual information. *CoRR*, abs/2102.00720, 2021. 12, 138

[65] RM Fano. Class notes for transmission of information. In *Course 6.574*. MIT, Cambridge, 1952. 142

[66] Serge Fehr and Stefan Berens. On the conditional Rényi entropy. *IEEE Trans. Inf. Theory*, 60(11):6801–6810, 2014. 138, 152

[67] Julie Ferrigno and Martin Hlavác. When AES blinks: introducing optical side channel. *IET Inf. Secur.*, 2(3):94–98, 2008. 5

[68] Wieland Fischer and Naofumi Homma, editors. *Cryptographic Hardware and Embedded Systems - CHES 2017 - 19th International Conference, Taipei, Taiwan, September 25-28, 2017, Proceedings*, volume 10529 of *Lecture Notes in Computer Science*. Springer, 2017. 190, 193, 198

[69] Aurélien Francillon and Pankaj Rohatgi, editors. *Smart Card Research and Advanced Applications - 12th International Conference, CARDIS 2013, Berlin, Germany, November 27-29, 2013. Revised Selected Papers*, volume 8419 of *LNCS*. Springer, 2014. 197, 207

[70] Guillaume Fumaroli, Ange Martinelli, Emmanuel Prouff, and Matthieu Rivain. Affine Masking against Higher-Order Side Channel Analysis. In Alex Biryukov, Guang Gong, and Douglas R. Stinson, editors, *Selected Areas in Cryptography*, volume 6544 of *Lecture Notes in Computer Science*, pages 262–280. Springer, 2010. 28

[71] Karine Gandolfi, Christophe Mourtel, and Francis Olivier. Electromagnetic analysis: Concrete results. In *Proceedings of the Third International Workshop on Cryptographic Hardware and Embedded Systems*, CHES '01, pages 251–261, London, UK, UK, 2001. Springer-Verlag. 5, 94, 118

[72] Qian Ge, Yuval Yarom, David Cock, and Gernot Heiser. A survey of microarchitectural timing attacks and countermeasures on contemporary hardware. *J. Cryptographic Engineering*, 8(1):1–27, 2018. 5

[73] Daniel Genkin, Adi Shamir, and Eran Tromer. RSA Key Extraction via Low-Bandwidth Acoustic Cryptanalysis. In Juan A. Garay and Rosario Gennaro, editors, *Advances in Cryptology - CRYPTO 2014 - 34th Annual Cryptology Conference, Santa Barbara, CA, USA, August 17-21, 2014, Proceedings, Part I*, volume 8616 of *Lecture Notes in Computer Science*, pages 444–461. Springer, 2014. 5

[74] Benedikt Gierlichs, Lejla Batina, Pim Tuyls, and Bart Preneel. Mutual information analysis. In *CHES, 10th International Workshop*, volume 5154 of *Lecture Notes in Computer Science*, pages 426–442. Springer, August 10-13 2008. Washington, D.C., USA. 5, 95, 99, 139

[75] Benedikt Gierlichs, Kerstin Lemke-Rust, and Christof Paar. Templates vs. Stochastic Methods. In *CHES*, volume 4249 of *LNCS*, pages 15–29. Springer, October 10-13 2006. Yokohama, Japan. 95

[76] Jovan Dj. Golić and Christophe Tymen. Multiplicative Masking and Power Analysis of AES. In *CHES*, volume 2523 of *Lecture Notes in Computer Science*, pages 198–212. Springer, August 13-15 2002. San Francisco, USA. 28

[77] Louis Goubin and Ange Martinelli. Protecting AES with Shamir's Secret Sharing Scheme. In Preneel and Takagi [125], pages 79–94. 7, 28, 61, 66, 107

[78] Vincent Grosso, François-Xavier Standaert, and Emmanuel Prouff. Low Entropy Masking Schemes, Revisited. In Francillon and Rohatgi [69], pages 33–43. 89

[79] Vincent Grosso, François-Xavier Standaert, and Sebastian Faust. Masking vs. Multiparty Computation: How Large Is the Gap for AES? In Guido Bertoni and Jean-Sébastien Coron, editors, *Cryptographic Hardware and Embedded Systems - CHES 2013 - 15th International Workshop, Santa Barbara, CA, USA, August 20-23, 2013. Proceedings*, volume 8086 of *Lecture Notes in Computer Science*, pages 400–416. Springer, 2013. 6, 28, 60, 61, 94

[80] Sylvain Guilley, Annelie Heuser, and Olivier Rioul. A Key to Success - Success Exponents for Side-Channel Distinguishers. In Alex Biryukov and Vipul Goyal, editors, *Progress in Cryptology - INDOCRYPT 2015 - 16th International Conference on Cryptology in India, Bangalore, India, December 6-9, 2015, Proceedings*, volume 9462 of *Lecture Notes in Computer Science*, pages 270–290. Springer, 2015. 98, 100, 156

[81] Sylvain Guilley, Philippe Hoogvorst, Renaud Pacalet, and Johannes Schmidt. Improving Side-Channel Attacks by Exploiting Substitution Boxes Properties. In Presse Universitaire de Rouen et du Havre, editor, *BFCA*, pages 1–25, 2007. May 02–04, Paris, France, http://www.liafa.jussieu.fr/bfca/books/BFCA07.pdf. 100

[82] Neil Hanley, Robert McEvoy, Michael Tunstall, Claire Whelan, Colin Murphy, and William P. Marnane. Correlation Power Analysis of Large Word Sizes. In *ISSC (Irish Signals and System Conference)*, pages 145–150. IET, 13-14 Sept 2007. Edinburgh, Scotland, UK. 95

[83] Christoph Herbst, Elisabeth Oswald, and Stefan Mangard. An AES Smart Card Implementation Resistant to Power Analysis Attacks. In Jianying Zhou, Moti Yung, and Feng Bao, editors, *ACNS*, volume 3989 of *Lecture Notes in Computer Science*, pages 239–252, 2006. 6

[84] 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. 95, 99, 119, 121, 139, 150, 157

[85] Vincent Immler, Robert Specht, and Florian Unterstein. Your rails cannot hide from localized EM: how dual-rail logic fails on fpgas. In Fischer and Homma [68], pages 403–424. 6

[86] Yuval Ishai, Amit Sahai, and David Wagner. Private Circuits: Securing Hardware against Probing Attacks. In *CRYPTO*, volume 2729 of *Lecture Notes in Computer Science*, pages 463–481. Springer, August 17–21 2003. Santa Barbara, California, USA. 6, 9, 28, 32, 60, 95, 118

[87] ISO/IEC JTC 1/SC 27/WG 3. ISO/IEC 17825:2016: Information technology – Security techniques – Testing methods for the mitigation of non-invasive attack classes against cryptographic modules. https://www.iso.org/standard/60612.html. 9

[88] Ibrahim Issa and Aaron B. Wagner. Operational definitions for some common information leakage metrics. In *2017 IEEE International Symposium on Information Theory, ISIT 2017, Aachen, Germany, June 25-30, 2017*, pages 769–773. IEEE, 2017. 12

[89] Ibrahim Issa, Aaron B. Wagner, and Sudeep Kamath. An Operational Approach to Information Leakage. *CoRR*, abs/1807.07878, 2018. 12, 149, 155

[90] Auguste Kerckhoffs. La cryptographie militaire (1). *Journal des sciences militaires*, 9:5–38, January 1883. http://en.wikipedia.org/wiki/Kerckhoffs_law. 4

[91] Auguste Kerckhoffs. La cryptographie militaire (2). *Journal des sciences militaires*, 9:161–191, February 1883. http://en.wikipedia.org/wiki/Kerckhoffs_law. 4

[92] Neal Koblitz. Elliptic curve cryptosystems. *Mathematic of Computation*, 48:203–209, 1987. 4

[93] Paul Kocher, Daniel Genkin, Daniel Gruss, Werner Haas, Mike Hamburg, Moritz Lipp, Stefan Mangard, Thomas Prescher, Michael Schwarz, and Yuval Yarom. Spectre Attacks: Exploiting Speculative Execution. *CoRR*, abs/1801.01203, 2018. 5

[94] Paul C. Kocher. Timing Attacks on Implementations of Diffie-Hellman, RSA, DSS, and Other Systems. In Neal Koblitz, editor, *Advances in Cryptology - CRYPTO '96, 16th Annual International Cryptology Conference, Santa Barbara, California, USA, August 18-22, 1996, Proceedings*, volume 1109 of *Lecture Notes in Computer Science*, pages 104–113. Springer, 1996. 5, 94

[95] Paul C. Kocher, Joshua Jaffe, and Benjamin Jun. Differential power analysis. In Wiener [167], pages 388–397. 5, 25, 94, 98, 118

[96] Juliane Krämer, Dmitry Nedospasov, Alexander Schlösser, and Jean-Pierre Seifert. Differential Photonic Emission Analysis. In Emmanuel Prouff, editor, *Constructive Side-Channel Analysis and Secure Design - 4th International Workshop, COSADE 2013, Paris, France, March 6-8, 2013, Revised Selected Papers*, volume 7864 of *Lecture Notes in Computer Science*, pages 1–16. Springer, 2013. 5

[97] Thanh-Ha Le and Maël Berthier. Mutual Information Analysis under the View of Higher-Order Statistics. In Isao Echizen, Noboru Kunihiro, and Ryôichi Sasaki, editors, *IWSEC*, volume 6434 of *Lecture Notes in Computer Science*, pages 285–300. Springer, 2010. 38

[98] Moritz Lipp, Michael Schwarz, Daniel Gruss, Thomas Prescher, Werner Haas, Anders Fogh, Jann Horn, Stefan Mangard, Paul Kocher, Daniel Genkin, Yuval Yarom, and Mike Hamburg. Meltdown: Reading Kernel Memory from User Space. In William Enck and Adrienne Porter Felt, editors, *27th USENIX Security Symposium, USENIX Security 2018, Baltimore, MD, USA, August 15-17, 2018.*, pages 973–990. USENIX Association, 2018. 5

[99] Yi Liu, Wei Cheng, Sylvain Guilley, and Olivier Rioul. On conditional α-information and its application to side-channel analysis. *CoRR*, abs/2105.07167, 2021. 12, 15, 137, 138, 148, 149, 150, 152, 169

[100] David J. C. MacKay. *Information theory, inference, and learning algorithms*. Cambridge University Press, 2003. ISBN-13: 978-0521642989. 124

[101] F. Jessie MacWilliams and Neil J. A. Sloane. *The Theory of Error-Correcting Codes*. Elsevier, Amsterdam, North Holland, 1977. ISBN: 978-0-444-85193-2. 8, 21, 22, 23, 41, 81

[102] Stefan Mangard, Elisabeth Oswald, and Thomas Popp. *Power Analysis Attacks: Revealing the Secrets of Smart Cards*. Springer, December 2006. ISBN 0-387-30857-1, http://www.dpabook.org/. 6, 30, 34, 94

[103] Stefan Mangard, Norbert Pramstaller, and Elisabeth Oswald. Successfully Attacking Masked AES Hardware Implementations. In LNCS, editor, *Proceedings of CHES'05*, volume 3659 of *LNCS*, pages 157–171. Springer, August 29 – September 1 2005. Edinburgh, Scotland, UK. 9, 94, 95

[104] Ben Marshall, Dan Page, and James Webb. MIRACLE: micro-architectural leakage evaluation. *IACR Cryptol. ePrint Arch.*, page 261, 2021. 5

[105] James L Massey. Minimal codewords and secret sharing. In *Proceedings of the 6th joint Swedish-Russian international workshop on information theory*, pages 276–279. Citeseer, 1993. 62

[106] Loïc Masure, Cécile Dumas, and Emmanuel Prouff. A comprehensive study of deep learning for side-channel analysis. *IACR Trans. Cryptogr. Hardw. Embed. Syst.*, 2020(1):348–375, 2020. 6

[107] Luke Mather, Elisabeth Oswald, Joe Bandenburg, and Marcin Wójcik. Does my device leak information? an a priori statistical power analysis of leakage detection tests. In Kazue Sako and Palash Sarkar, editors, *Advances in Cryptology - ASIACRYPT 2013 - 19th International Conference on the Theory and Application of Cryptology and Information Security, Bengaluru, India, December 1-5, 2013, Proceedings, Part I*, volume 8269 of *Lecture Notes in Computer Science*, pages 486–505. Springer, 2013. 94

[108] Silvio Micali and Leonid Reyzin. Physically observable cryptography (extended abstract). In Moni Naor, editor, *Theory of Cryptography, First Theory of Cryptography Conference, TCC 2004, Cambridge, MA, USA, February 19-21, 2004, Proceedings*, volume 2951 of *Lecture Notes in Computer Science*, pages 278–296. Springer, 2004. 5

[109] Victor S. Miller. Use of Elliptic Curves in Cryptography. In Hugh C. Williams, editor, *Advances in Cryptology - CRYPTO '85, Santa Barbara, California, USA, August 18-22, 1985, Proceedings*, volume 218 of *Lecture Notes in Computer Science*, pages 417–426. Springer, 1985. 4

[110] Jingdian Ming, Yongbin Zhou, Wei Cheng, Huizhong Li, Guang Yang, and Qian Zhang. Mind the balance: Revealing the vulnerabilities in low entropy masking schemes. *IEEE Trans. Inf. Forensics Secur.*, 15:3694–3708, 2020. 89

[111] Amir Moradi, Nima Mousavi, Christof Paar, and Mahmoud Salmasizadeh. A Comparative Study of Mutual Information Analysis under a Gaussian Assumption. In *WISA (Information Security Applications, 10th International Workshop)*, volume 5932 of *Lecture Notes in Computer Science*, pages 193–205. Springer, August 25-27 2009. Busan, Korea. 100, 139

[112] Amir Moradi, Bastian Richter, Tobias Schneider, and François-Xavier Standaert. Leakage Detection with the χ^2-Test. *IACR Trans. Cryptogr. Hardw. Embed. Syst.*, 2018(1):209–237, 2018. 9, 94

[113] Gary L. Mullen and Daniel Panario. *Handbook of Finite Fields*. Chapman and Hall/CRC, June 17 2013. ISBN 9781439873786 - CAT# K13417. 23

[114] Maxime Nassar, Youssef Souissi, Sylvain Guilley, and Jean-Luc Danger. RSM: A small and fast countermeasure for AES, secure against 1st and 2nd-order zero-offset SCAs. In Wolfgang Rosenstiel and Lothar Thiele, editors, *2012 Design, Automation & Test in Europe Conference & Exhibition, DATE 2012, Dresden, Germany, March 12-16, 2012*, pages 1173–1178. IEEE, 2012. 89

[115] Noam Nisan and Mario Szegedy. On the degree of boolean functions as real polynomials. *Computational Complexity*, 4:301–313, 1994. 25

[116] NIST/ITL/CSD. Data Encryption Standard. FIPS PUB 46-3, Oct 1999. http://csrc.nist.gov/publications/fips/fips46-3/fips46-3.pdf. 4

[117] NIST/ITL/CSD. Advanced Encryption Standard (AES). FIPS PUB 197, Nov 2001. http://nvlpubs.nist.gov/nistpubs/FIPS/NIST.FIPS.197.pdf (also ISO/IEC 18033-3:2010). 4

[118] 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*. Springer, 2015. 187, 195

[119] Elisabeth Oswald and Stefan Mangard. Template Attacks on Masking — Resistance Is Futile. In Masayuki Abe, editor, *CT-RSA*, volume 4377 of *Lecture Notes in Computer Science*, pages 243–256. Springer, 2007. 95

[120] Athanasios Papoulis and S Unnikrishna Pillai. *Probability, random variables, and stochastic processes (fourth edition)*. Tata McGraw-Hill Education, 2002. 34

[121] Guilherme Perin, Lukasz Chmielewski, Lejla Batina, and Stjepan Picek. Keep it unsupervised: Horizontal attacks meet deep learning. *IACR Trans. Cryptogr. Hardw. Embed. Syst.*, 2021(1):343–372, 2021. 6

[122] 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. 138, 147

[123] Romain Poussier, Qian Guo, François-Xavier Standaert, Claude Carlet, and Sylvain Guilley. Connecting and Improving Direct Sum Masking and Inner Product Masking. In Thomas Eisenbarth and Yannick Teglia, editors, *Smart Card Research and Advanced Applications - 16th International Conference, CARDIS 2017, Lugano, Switzerland, November 13-15, 2017, Revised Selected Papers*, volume 10728 of *Lecture Notes in Computer Science*, pages 123–141. Springer, 2017. 7, 29, 30, 31, 32, 36, 37, 40, 41, 46, 60, 62, 66, 68, 86, 101, 167

[124] Romain Poussier, François-Xavier Standaert, and Vincent Grosso. Simple key enumeration (and rank estimation) using histograms: An integrated approach. In Benedikt Gierlichs and Axel Y. Poschmann, editors, *Cryptographic Hardware and Embedded Systems - CHES 2016 - 18th International Conference, Santa Barbara, CA, USA, August 17-19, 2016, Proceedings*, volume 9813 of *Lecture Notes in Computer Science*, pages 61–81. Springer, 2016. 96

[125] Bart Preneel and Tsuyoshi Takagi, editors. *Cryptographic Hardware and Embedded Systems - CHES 2011 - 13th International Workshop, Nara, Japan, September 28 – October 1, 2011. Proceedings*, volume 6917 of *LNCS*. Springer, 2011. 196, 203

[126] Thomas Prest, Dahmun Goudarzi, Ange Martinelli, and Alain Passelègue. Unifying leakage models on a rényi day. In Alexandra Boldyreva and Daniele Micciancio, editors, *Advances in Cryptology - CRYPTO 2019 - 39th Annual International Cryptology Conference, Santa Barbara, CA, USA, August 18-22, 2019, Proceedings, Part I*, volume 11692 of *Lecture Notes in Computer Science*, pages 683–712. Springer, 2019. 9, 95

[127] Emmanuel Prouff and Matthieu Rivain. Theoretical and practical aspects of mutual information-based side channel analysis. *International Journal of Applied Cryptography (IJACT)*, 2(2):121–138, 2010. 100

[128] Emmanuel Prouff and Matthieu Rivain. Masking against Side-Channel Attacks: A Formal Security Proof. In Thomas Johansson and Phong Q. Nguyen, editors, *Advances in Cryptology - EUROCRYPT 2013, 32nd Annual International Conference on the Theory and Applications of Cryptographic Techniques, Athens, Greece, May 26-30, 2013. Proceedings*,

volume 7881 of *Lecture Notes in Computer Science*, pages 142–159. Springer, 2013. 6, 9, 28, 60, 94, 95, 118, 131

[129] Emmanuel Prouff, Matthieu Rivain, and Régis Bevan. Statistical Analysis of Second Order Differential Power Analysis. *IEEE Trans. Computers*, 58(6):799–811, 2009. 33

[130] Emmanuel Prouff, Matthieu Rivain, and Thomas Roche. On the Practical Security of a Leakage Resilient Masking Scheme. In Josh Benaloh, editor, *Topics in Cryptology - CT-RSA 2014 - The Cryptographer's Track at the RSA Conference 2014, San Francisco, CA, USA, February 25-28, 2014. Proceedings*, volume 8366 of *Lecture Notes in Computer Science*, pages 169–182. Springer, 2014. 60

[131] Emmanuel Prouff and Thomas Roche. Higher-Order Glitches Free Implementation of the AES Using Secure Multi-party Computation Protocols. In Preneel and Takagi [125], pages 63–78. 7, 28, 61, 65, 66, 88, 107

[132] Jean-Jacques Quisquater and David Samyde. ElectroMagnetic Analysis (EMA): Measures and Counter-Measures for Smard Cards. In I. Attali and T. P. Jensen, editors, *Smart Card Programming and Security (E-smart 2001)*, volume 2140 of *LNCS*, pages 200–210. Springer-Verlag, September 2001. Nice, France. ISSN 0302-9743. 5, 94

[133] Keyvan Ramezanpour, Paul Ampadu, and William Diehl. SCAUL: power side-channel analysis with unsupervised learning. *IEEE Trans. Computers*, 69(11):1626–1638, 2020. 6

[134] Pablo Rauzy, Sylvain Guilley, and Zakaria Najm. Formally Proved Security of Assembly Code Against Leakage. *IACR Cryptology ePrint Archive*, 2013:554, 2013. (Also appears at PROOFS 2014, Busan, South Korea). 6

[135] 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. 12, 138, 140, 146

[136] Bastian Richter, David Knichel, and Amir Moradi. A comparison of χ^2-test and mutual information as distinguisher for side-channel analysis. In Sonia Belaïd and Tim Güneysu, editors, *Smart Card Research and Advanced Applications - 18th International Conference, CARDIS 2019, Prague, Czech Republic, November 11-13, 2019, Revised Selected Papers*, volume 11833 of *Lecture Notes in Computer Science*, pages 237–251. Springer, 2019. 9

[137] 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. 138, 140, 146, 147, 151

[138] Olivier Rioul, Wei Cheng, and Sylvain Guilley. Cumulant expansion of mutual information for quantifying leakage of a protected secret. In *IEEE International Symposium on Information Theory, ISIT 2021, Melbourne, Australia, July 12-20, 2021*, pages 2596–2601. IEEE, 2021. 15, 169

[139] Matthieu Rivain and Emmanuel Prouff. Provably Secure Higher-Order Masking of AES. In Stefan Mangard and François-Xavier Standaert, editors, *CHES*, volume 6225 of *LNCS*, pages 413–427. Springer, 2010. 6, 28, 60

[140] Matthieu Rivain, Emmanuel Prouff, and Julien Doget. Higher-Order Masking and Shuffling for Software Implementations of Block Ciphers. In *CHES*, volume 5747 of *Lecture Notes in Computer Science*, pages 171–188. Springer, September 6-9 2009. Lausanne, Switzerland. 6

[141] Ronald L. Rivest, Adi Shamir, and Leonard M. Adleman. A Method for Obtaining Digital Signatures and Public-Key Cryptosystems. *Communications of the ACM*, 21(2):120–126, 1978. 4

[142] Igal Sason and Sergio Verdú. Improved bounds on lossless source coding and guessing moments via Rényi measures. *IEEE Trans. Information Theory*, 64(6):4323–4346, 2018. 141, 142, 145

[143] Werner Schindler, Kerstin Lemke, and Christof Paar. A Stochastic Model for Differential Side Channel Cryptanalysis. In LNCS, editor, *CHES*, volume 3659 of *LNCS*, pages 30–46. Springer, Sept 2005. Edinburgh, Scotland, UK. 6, 10, 39, 95

[144] Tobias Schneider and Amir Moradi. Leakage assessment methodology - A clear roadmap for side-channel evaluations. In Tim Güneysu and Helena Handschuh, editors, *Cryptographic Hardware and Embedded Systems - CHES 2015 - 17th International Workshop, Saint-Malo, France, September 13-16, 2015, Proceedings*, volume 9293 of *Lecture Notes in Computer Science*, pages 495–513. Springer, 2015. 94

[145] Adi Shamir. How to share a secret. *Commun. ACM*, 22(11):612–613, 1979. 7, 61

[146] C. E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27(3):379–423, July 1948. 12, 138

[147] Claude E. Shannon. Communication theory of secrecy systems. *Bell System Technical Journal, Vol 28, pp. 656–715*, October 1949. 4, 12, 138

[148] Robin Sibson. Information radius. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 14(2):149–160, 1969. 12, 138, 146, 147

[149] Richard C. Singleton. Maximum distance q -nary codes. *IEEE Trans. Information Theory*, 10(2):116–118, 1964. 8

[150] Geoffrey Smith. On the foundations of quantitative information flow. In Luca de Alfaro, editor, *Foundations of Software Science and Computational Structures, 12th International Conference, FOSSACS 2009, Held as Part of the Joint European Conferences on Theory and Practice of Software, ETAPS 2009, York, UK, March 22-29, 2009. Proceedings*, volume 5504 of *Lecture Notes in Computer Science*, pages 288–302. Springer, 2009. 12

[151] François-Xavier Standaert, Tal Malkin, and Moti Yung. A Unified Framework for the Analysis of Side-Channel Key Recovery Attacks. In *EUROCRYPT*, volume 5479 of *LNCS*, pages 443–461. Springer, April 26-30 2009. Cologne, Germany. 10, 42, 95, 96, 100, 118, 139, 150, 156

[152] François-Xavier Standaert. How (not) to use welch's t-test in side-channel security evaluations. In Begül Bilgin and Jean-Bernard Fischer, editors, *Smart Card Research and Advanced Applications, 17th International Conference, CARDIS 2018, Montpellier, France, November 12-14, 2018, Revised Selected Papers*, volume 11389 of *Lecture Notes in Computer Science*, pages 65–79. Springer, 2018. 94

[153] François-Xavier Standaert, Nicolas Veyrat-Charvillon, Elisabeth Oswald, Benedikt Gierlichs, Marcel Medwed, Markus Kasper, and Stefan Mangard. The World Is Not Enough: Another Look on Second-Order DPA. In Masayuki Abe, editor, *Advances in Cryptology - ASIACRYPT 2010 - 16th International Conference on the Theory and Application of Cryptology and Information Security, Singapore, December 5-9, 2010. Proceedings*, volume 6477 of *Lecture Notes in Computer Science*, pages 112–129. Springer, 2010. 76

[154] TELECOM ParisTech SEN research group. DPA Contest (4^{th} edition), 2013–2014. http://www.DPAcontest.org/v4/. 89

[155] Benjamin Timon. Non-Profiled Deep Learning-Based Side-Channel Attacks. *IACR Cryptology ePrint Archive*, 2018:196, 2018. 6

[156] Kris Tiri and Ingrid Verbauwhede. A Logic Level Design Methodology for a Secure DPA Resistant ASIC or FPGA Implementation. In *2004 Design, Automation and Test in Europe Conference and Exposition (DATE 2004), 16-20 February 2004, Paris, France*, pages 246–251. IEEE Computer Society, 2004. 94

[157] 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. 12, 138

[158] University of Sydney (Australia). Magma Computational Algebra System. `http://magma.maths.usyd.edu.au/magma/`, Accessed on 2021-06-22. 30, 64, 82

[159] Tim van Erven and Peter Harremoës. Rényi divergence and Kullback-Leibler divergence. *IEEE Trans. Information Theory*, 60(7):3797–3820, 2014. 141, 146

[160] 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. 138, 146, 147

[161] 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, 2015. 141

[162] Nicolas Veyrat-Charvillon, Benoît Gérard, Mathieu Renauld, and François-Xavier Standaert. An Optimal Key Enumeration Algorithm and Its Application to Side-Channel Attacks. In Lars R. Knudsen and Huapeng Wu, editors, *Selected Areas in Cryptography*, volume 7707 of *Lecture Notes in Computer Science*, pages 390–406. Springer, 2012. 96

[163] Nicolas Veyrat-Charvillon and François-Xavier Standaert. Mutual Information Analysis: How, When and Why? In Christophe Clavier and Kris Gaj, editors, *Cryptographic Hardware and Embedded Systems - CHES 2009, 11th International Workshop, Lausanne, Switzerland, September 6-9, 2009, Proceedings*, volume 5747 of *Lecture Notes in Computer Science*, pages 429–443. Springer, 2009. 5, 95, 99, 139

[164] 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. 7, 24, 53, 60, 61, 66, 69, 70, 79, 89, 167

[165] Weijia Wang, François-Xavier Standaert, Yu Yu, Sihang Pu, Junrong Liu, Zheng Guo, and Dawu Gu. Inner Product Masking for Bitslice Ciphers and Security Order Amplification for Linear Leakages. In Kerstin Lemke-Rust and Michael Tunstall, editors, *Smart Card Research and Advanced Applications - 15th International Conference, CARDIS 2016, Cannes, France, November 7-9, 2016, Revised Selected Papers*, volume 10146 of *Lecture Notes in Computer Science*, pages 174–191. Springer, 2016. 29, 30, 36, 37, 40, 46

[166] Carolyn Whitnall and Elisabeth Oswald. A Comprehensive Evaluation of Mutual Information Analysis Using a Fair Evaluation Framework. In Phillip Rogaway, editor, *CRYPTO*, volume 6841 of *Lecture Notes in Computer Science*, pages 316–334. Springer, 2011. 10, 95

[167] Michael J. Wiener, editor. *Advances in Cryptology - CRYPTO '99, 19th Annual International Cryptology Conference, Santa Barbara, California, USA, August 15-19, 1999, Proceedings*, volume 1666 of *Lecture Notes in Computer Science*. Springer, 1999. 191, 199

[168] Lennert Wouters, Victor Arribas, Benedikt Gierlichs, and Bart Preneel. Revisiting a methodology for efficient CNN architectures in profiling attacks. *IACR Trans. Cryptogr. Hardw. Embed. Syst.*, 2020(3):147–168, 2020. 6

[169] Xiang Yang and James L Massey. The condition for a cyclic code to have a complementary dual. *Discrete Mathematics*, 126(1):391–393, 1994. 24

[170] Xin Ye and Thomas Eisenbarth. On the Vulnerability of Low Entropy Masking Schemes. In Francillon and Rohatgi [69], pages 44–60. 89

[171] Gabriel Zaid, Lilian Bossuet, Amaury Habrard, and Alexandre Venelli. Methodology for efficient CNN architectures in profiling attacks. *IACR Trans. Cryptogr. Hardw. Embed. Syst.*, 2020(1):1–36, 2020. 6

[172] Hui Zhao, Yongbin Zhou, François-Xavier Standaert, and Hailong Zhang. Systematic Construction and Comprehensive Evaluation of Kolmogorov-Smirnov Test Based Side-Channel Distinguishers. In Robert H. Deng and Tao Feng, editors, *ISPEC*, volume 7863 of *Lecture Notes in Computer Science*, pages 336–352. Springer, 2013. 39

ECOLE DOCTORALE

Titre : Qu'est ce que l'information permet de deviner ? Vers une quantification des fuites d'informations dans l'analyse de canaux auxiliaires

Mots clés : Analyse des canaux auxiliaires ; Contre-mesure ; Quantification des fuites ; L'informations alpha

Résumé : Les algorithmes cryptographiques jouent un rôle prédominant pour établir une connectivité sécurisée dans notre société numérique actuelle. Ces calculs traitent des informations sensibles telles que des clés de chiffrement, qui sont généralement très exposées lors de la manipulation. Dans le domaine de la sécurité des systèmes embarqués, l'analyse des canaux auxiliaires est l'une des techniques les plus puissantes contre les implémentations cryptographiques. Globalement, cette thèse se compose de deux sujets. L'un est la quantification des fuites de la forme la plus générale de masquage équipé des codes linéaires ; l'autre est l'exploration de l'application de mesures d'information plus génériques dans un contexte d'analyse de canaux auxiliaires.
Pour ce qui concerne le premier sujet, nous proposons un cadre théorique de codage unifié pour mesurer la fuite d'informations dans le masquage basé sur les codes. Plus précisément, notre cadre établit des connexions formelles entre les propriétés de codage et les métriques de fuite dans l'analyse des canaux auxiliaires. Ces connexions formelles nous permettent de faire avancer l'évaluation quantitative sur la façon dont les codes linéaires peuvent affecter la sécurité concrète de tous les schémas de masquage basés sur les codes.
Concernant le deuxième sujet, nous proposons d'utiliser une mesure plus générale du point de vue de la théorie de l'information, à savoir l'information alpha (alpha-information). Ce qui est remarquable, c'est qu'avec des choix appropriés, l'information alpha fournit des bornes très proches de la réalité ; en particulier, lorsque alpha tend vers l'infini (positif), les limites seront exactes. En fait, les distingueurs basés sur le maximum de vraisemblance convergeront vers les limites.
En résumé, notre étude dans cette thèse fait avancer l'évaluation et la consolidation de la sécurité des canaux auxiliaires des implémentations cryptographiques. Du point de vue de la protection, nous fournissons un guide des meilleures pratiques pour l'application du masquage basé sur le code. Du point de vue de l'évaluation, l'application de l'alpha-information permet aux évaluateurs et concepteurs (développeurs) d'avoir une estimation plus précise (voire exacte) du niveau de sécurité concret des canaux auxiliaires émanant de leurs puces cryptographiques.

Title : What Can Information Guess ? Towards Information Leakage Quantification in Side-Channel Analysis

Keywords : Side-Channel Analysis ; Code-based Masking ; Leakage Quantification ; Alpha-Information

Abstract : Cryptographic algorithms are nowadays prevalent in establishing secure connectivity in our digital society. Such computations handle sensitive information like encryption keys, which are usually very exposed during manipulation. In the field of embedded systems security, side-channel analysis is one of the most powerful techniques against cryptographic implementations. Overall, this thesis consists of two topics. One is the leakage quantification of the most general form of masking equipped with the linear codes, so-called code-based masking ; the other one is exploration of applying more generic information measures in a context of side-channel analysis.
Regarding the former, we propose a unified coding-theoretic framework for measuring the information leakage in code-based masking. Specifically, our framework builds formal connections between coding properties and leakage metrics in side-channel analysis. Those formal connections enable us to push forward the quantitative evaluation on how the linear codes can affect the concrete security of all code-based masking schemes.
Regarding the latter, we present a full spectrum of application of alpha-information, a generalization of mutual information, for assessing side-channel security. With proper choices, alpha-information provides very tight bounds, in particular, when alpha approaches to positive infinity, the bounds will be exact. As a matter of fact, maximum-likelihood based distinguishers will converge to the bounds when alpha approaches to positive infinity. Therefore, we demonstrate how the two world, information-theoretic measures (bounds) and maximum-likelihood based side-channel attacks, are seamlessly connected in side-channel analysis.
In summary, our study in this thesis pushes forward the evaluation and consolidation of side-channel security of cryptographic implementations. From a protection perspective, we provide a best-practice guideline for the application of code-based masking. From an evaluation perspective, the application of alpha-information enables practical evaluators and designers to have a more accurate (or even exact) estimation of concrete side-channel security level of their cryptographic chips.

Institut Polytechnique de Paris
91120 Palaiseau, France