Quantitative Security of Block Ciphers: Designs and Cryptanalysis Tools

THÈSE N^{O} 4208 (2008)

PRÉSENTÉE LE 14 NOVEMBRE 2008

À LA FACULTÉ INFORMATIQUE ET COMMUNICATIONS
LABORATOIRE DE SÉCURITÉ ET DE CRYPTOGRAPHIE
PROGRAMME DOCTORAL EN INFORMATIQUE, COMMUNICATIONS ET INFORMATION

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Thomas BAIGNÈRES

ingénieur en systèmes de communication EPF
de nationalités française et suisse et originaire de Vira (Gambarogno) (TI)

acceptée sur proposition du jury:

Prof. A. Lenstra, président du jury
Prof. S. Vaudenay, directeur de thèse
Dr H. Gilbert, rapporteur
Prof. S. Morgenthaler, rapporteur
Prof. J. Stern, rapporteur

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Suisse
2008

Contents

II On the (In)Security of Block Ciphers: Tools for Security Analysis 29

Abstract

Block ciphers probably figure in the list of the most important cryptographic primitives. Although they are used for many different purposes, their essential goal is to ensure confidentiality. This thesis is concerned by their *quantitative security*, that is, by *measurable attributes* that reflect their ability to guarantee this confidentiality.

The first part of this thesis deals with well know results. Starting with Shannon's Theory of Secrecy, we move to practical implications for block ciphers, recall the main schemes on which nowadays block ciphers are based, and introduce the Luby-Rackoff security model. We describe distinguishing attacks and key-recovery attacks against block ciphers and show how to turn the firsts into the seconds. As an illustration, we recall linear cryptanalysis which is a classical example of statistical cryptanalysis.

In the second part, we consider the (in)security of block ciphers against statistical cryptanalytic attacks and develop some tools to perform optimal attacks and quantify their efficiency. We start with a simple setting in which the adversary has to distinguish between two sources of randomness and show how an optimal strategy can be derived in certain cases. We proceed with the practical situation where the cardinality of the sample space is too large for the optimal strategy to be implemented and show how this naturally leads to the concept of *projection-based distinguishers*, which reduce the sample space by compressing the samples. Within this setting, we re-consider the particular case of linear distinguishers and generalize them to sets of arbitrary cardinality. We show how these distinguishers between random sources can be turned into distinguishers between random oracles (or block ciphers) and how, in this setting, one can generalize linear cryptanalysis to Abelian groups. As a proof of concept, we show how to break the block cipher TOY100, introduce the block cipher DEAN which encrypts blocks of decimal digits, and apply the theory to the SAFER block cipher family.

In the last part of this thesis, we introduce two new constructions. We start by recalling some essential notions about provable security for block ciphers and about Serge Vaudenay's Decorrelation Theory, and introduce new simple modules for which we prove essential properties that we will later use in our designs. We then present the block cipher C and prove that it is immune against a wide range of cryptanalytic attacks. In particular, we compute the *exact* advantage of the best distinguisher limited to two plaintext/ciphertext samples between C and the perfect cipher and use it to compute the exact value of the maximum expected linear probability (resp. differential probability) of C which is known to be inversely proportional to the number of samples required by the best possible linear (resp. differential) attack. We then introduce KFC a block cipher which builds upon the same foundations as C but for which we can prove results for higher order adversaries. We conclude both discussions about C and KFC by implementation considerations.

Keywords: Cryptography, block cipher, statistical cryptanalysis, linear cryptanalysis, hypothesis testing, SAFER, Decorrelation Theory

Résumé

Les algorithmes de chiffrement à clef secrète font très certainement partie des primitives cryptographiques les plus importantes. Bien qu'ils soient utilisés à des fins très diverses, leur principale fonction est d'assurer la confidentialité des données. Cette thèse s'intéresse à leur *sécurité quantitative*, c'est-à-dire aux *attributs mesurables* qui reflètent leur habilité à garantir cette confidentialité.

La première partie de cette thèse traite d'un certain nombre de résultats bien connus. En partant de la théorie du secret de Shannon, nous considérons les implications pratiques pour les algorithmes de chiffrement à clef secrète, nous rappelons les schémas élémentaires sur lesquels ces derniers sont conçus, et introduisons le modèle de Luby et Rackoff. Nous décrivons les attaques visant à distinguer une permutation aléatoire d'une autre puis les attaques dont l'objectif est de retrouver la clef secrète pour enfin montrer comment les premières peuvent entraîner les deuxièmes. En guise d'exemple, nous rappelons les concepts de la cryptanalyse linéaire qui est un exemple classique de cryptanalyse statistique.

Dans la deuxième partie, nous considérons l'(in)sécurité des algorithmes de chiffrement à clef secrète face au attaques cryptanalytiques statistiques et développons quelques outils pour exécuter certaines attaques et quantifier leur efficacité. Nous considérons un cadre initial très simple dans lequel un adversaire doit distinguer une source aléatoire d'une autre et montrons que, dans certains cas, une stratégie optimale peut être trouvée. Nous traitons ensuite le cas pratique dans lequel la cardinalité de l'espace échantillon est trop grande pour que la stratégie optimale puisse être utilisée telle quelle, ce qui entraîne naturellement la définition de *distingueurs basés sur des projections* qui réduisent l'espace en compressant chaque échantillon. Dans cette optique, nous reconsidérons le cas des distingueurs linéaires et les généralisons aux ensembles de cardinalité arbitraire. Nous montrons comment ces distingueurs entre des sources aléatoires peuvent être transformés en distingueurs entre des oracles aléatoires et comment, de cette façon, il est possible de généraliser la cryptanalyse linéaire aux groupes Abéliens. En guise de preuve de concept, nous montrons comment casser l'algorithme de chiffrement TOY100, introduisons l'algorithme DEAN qui permet de chiffrer des blocs de chiffres décimaux, et appliquons la théorie à la famille d'algorithmes SAFER.

Dans la dernière partie de cette thèse, nous proposons deux nouvelles constructions. Nous commençons par rappeler quelques notions essentielles concernant la sécurité prouvée des algorithmes de chiffrement à clef secrète et la Théorie de la Décorrélation développée par Serge Vaudenay. Nous introduisons de nouveaux modules pour lesquels un certain nombre de résultats de sécurité peuvent être prouvés et qui seront au coeur des deux constructions à suivre. Nous présentons ensuite l'algorithme de chiffrement C et prouvons sa sécurité contre une certain nombre d'attaques. En particulier, nous calculons l'avantage *exact* du meilleur distingueur limité à deux paires de textes clairs/chiffrés entre C et l'algorithme de chiffrement parfait et utilisons ce résultat pour calculer la valeur exacte de la valeur moyenne maximum de la probabilité linéaire (ainsi que celle de la valeur moyenne de la probabilité différentielle) de C que l'on sait être inversement proportionnelle au nombre d'échantillons nécessaires pour mener une attaque concluante. Nous introduisons ensuite KFC, un algorithme qui repose sur les mêmes bases que C mais pour lequel nous arrivons à prouver des résultats concernant des adversaires d'ordres plus élevés. Dans les deux cas, nous concluons la discussion par des considérations expérimentales.

Mots-clefs: Cryptographie, algorithme de chiffrement à clef secrète, cryptanalyse statistique, cryptanalyse linéaire, test d'hypothèse, SAFER, Théorie de la Décorrélation

Remerciements

Je tiens en premier lieu à remercier mon directeur de thèse, Serge Vaudenay, sans qui ce manuscrit n'aurait jamais vu le jour. Source d'inspiration permanente, exigeant et talentueux, parfois dur mais toujours juste, intègre et attentionné, il a souvent su aller au delà de mes espérances. J'espère que mon travail aura été à la hauteur des siennes.

Merci à tous les membres du jury de m'avoir accordé leur temps, leurs conseils et leurs encouragements pour l'avenir. En particulier, merci à Stephan Morgenthaler d'avoir pris la peine de s'écarter quelque peu de son domaine de prédilection, merci à Jacques Stern pour avoir non seulement accepté mon invitation mais aussi pour son aide regardant la généralisation de la cryptanalyse linéaire, merci à Henri Gilbert pour son aide, sa gentillesse, mais aussi pour ses conseils éclairés (déjà bien avant la défense !). Merci à Arjen Lenstra, président du jury, pour sa bonne humeur et son franc-parler !

Merci au Fond National Suisse d'avoir contribué à la plupart de mes travaux (bourse 200021-107982/1).

Le LASEC ne serait pas ce qu'il est aujourd'hui sans ceux qui m'ont précédés: merci à Pascal Junod pour m'avoir donné l'envie d'aller plus loin dans la recherche (et pour ce qui restera sans doute ma bière la plus inoubliable: Singapour, 30° aux petites heures du matin, assis à une table au beau milieu d'une rue déserte dans le quartier Indien), merci à Jean Monnerat pour les discussions sans fins (au café à Lausanne, dans un boui-boui à Shanghai, un bar high tech à Saint-Pétersbourg), merci à Gildas Avoine pour son humour (presque) toujours décapant (les amis de Pépin se reconnaîtront). Merci à Martine Corval pour son aide inestimable (et pour savoir toujours trouver les mots quand rien ne va plus !). Je n'oublie pas la relève: merci à Sylvain Pasini pour sa gentillesse sans égale, merci à Martin Vuagnoux de toujours savoir partager son énergie avec les autres ! Le "labo" a de beaux jours devant lui. *Last but not least*, merci à mon "collègue de bureau" Matthieu Finiasz d'avoir supporté mon sale caractère pendant deux ans ! Ceux qui le connaissent ne me contrediront pas, sa gentillesse et à la hauteur de son talent. Nous avons fait ensemble un travail que je crois formidable, et le réaliser m'a apporté un immense bonheur.

La crypto, c'est rigolo ! peut-on lire sur une plaquette du département de Mathématiques et d'Informatique de l'Ecole Normale Supérieure. La crypto, c'est aussi beaucoup de conférences et de voyages à l'autre bout du monde pour y assister. Merci à la communauté des cryptographes de m'y avoir fait passer de très bon moments. Merci en particulier à Frédéric Muller, Claude Barral, Thomas Peyrin, Raphael Overbeck, Khaled Ouafi, Rafik Chaabouni, Antoine Joux, Willi Meier, David Naccache, Pascal Paillier, Phong Q. Nguyen, Julien Stern, Olivier Billet, Emmanuel Bresson, David Pointcheval, Jean-Philippe Aumasson, Kaisa Nyberg, et Raphael Phan.

Merci à Chrissie et John Barlow pour leur amitié. Merci en particulier à John pour son aide, ses conseils et ses corrections !

J'ai eu la chance de refaire le monde plusieurs fois avec eux (et il en a bien besoin): Merci à Robert Bargmann, Numa Schmeder et Damien Tardieu de m'avoir fait partager cinq années inoubliables. Avec vous, travailler est redevenu un plaisir. Merci à Thibaut Davain pour plus de 20 ans d'amitié inaltérable.

Merci à Jacqueline de m'avoir fait (re)-découvrir la Rafraire et ceux qui la font. Merci à Danielle et René de m'y avoir accueilli comme si j'en avais toujours fait partie. Merci mille fois à Caroline et Jacques-André pour autant de soirées inoubliables. Merci à Jacques et Yolène pour leur soutient et leur bonne humeur !

Merci enfin à Elena et Bernard qui, il y a presque quinze ans, m'ont accueilli comme si je faisais déjà partie de la famille. Merci à Stefaan et Hélène pour tous les bons moments passés depuis. C'est aussi à votre soutien que je dois ma réussite.

Merci à ma grand-mère qui a su me transmettre tant de choses, y compris son amour des mathématiques. Merci à mon oncle pour son aide et son affection.

Merci à Yvonne et Mathias pour leur soutient de chaque instant. Merci à Mathias pour ses conseils toujours avisés et d'être là qu'en j'ai besoin de lui (merci pour les dés !).

Merci enfin à mes parents pour leur soutien inconditionnel et pour avoir toujours cru en moi. Je ne saurais exprimer ici l'amour et la gratitude immense que je leur porte. Ma réussite est aussi la votre.

Merci à Valérie de m'accompagner, de me soutenir, de partager, de croire en moi plus que je ne le fais moi-même depuis plus de quinze ans. Merci de m'apporter l'équilibre qui sinon me ferait défaut. Merci d'avoir enchanté ma vie.

Merci à Clara d'avoir donné tout son sens à mon existence.

A la mémoire de mon père.

So long as we live, he too shall live.
For he is now a part of us,
As we remember him.

Part I

An Introduction to Modern Cryptology and an Approach to the Design and Cryptanalysis of Block Ciphers

Chapter 1

Shannon's Theory of Secrecy

The oldest concern of cryptography is probably to find the most efficient and elegant technique to transmit confidential information (through time or space) to a recipient, and to this recipient *only*. The first known reference to this problem dates back to quite ancient times, a fact that David Kahn illustrates from the very beginning of "*The Code-Breakers*" [78] by entitling the second chapter "*The first 3,000 years*". During this period of time, cryptography fascinated not only the most important world leaders (Julius Caesar's cipher is one of the first encryption method taught in almost every lecture on cryptography) but also the greatest artists and scientists. It is not surprising that several books relate its story [78,100,141,142] for a reason which is very clearly and concisely summarized in a (by now) famous leitmotiv propagated in the 90's by a young cryptographer [149] of the "*Ecole Normale Supérieure*":

"*La crypto c'est rigolo*".[1]

Yet, the bases of cryptography as a scientific discipline were only formulated in 1946 by Claude E. Shannon in the confidential report (by now declassified) "A Mathematical Theory of Cryptography" [139]. Its mathematical analysis provides a formal statement of what defines a cryptographic system and what one should require from it.

Shannon's theory of secrecy is concerned with encryption methods which allow one to conceal information originally contained in a message (or plaintext) in a so-called ciphertext. Ideally, the ciphertext alone should not allow the recovery of information, so the fact that it is eavesdropped by some adversary cannot do any harm[2].

1.1 The Encryption Model: Preserving Confidentiality

Shannon defines a secrecy system (or a cryptographic system) as "*a set of transformations of one space (the set of possible messages) into a second space (the set*

[1] Crypto is fun.

[2] Shannon makes reference to the "*enemy*" since at that time, cryptography was essentially of military concern.

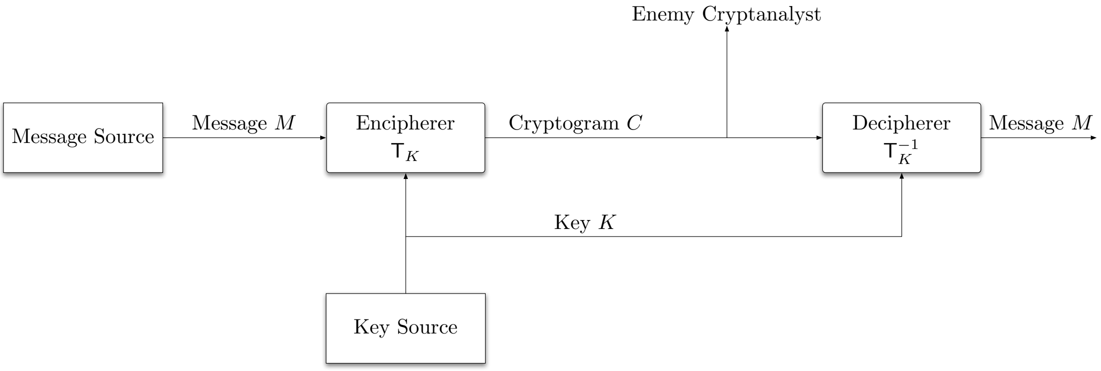


Figure 1.1: Symmetric Encryption

of possible cryptograms)." [139]. Each of the transformations is indexed by a *key* which shall only be shared by the sender and the recipient of the message. This situation is illustrated in Figure 1.1. Using the key K and the Encipherer T, the sender encrypts the message M and obtains the cryptogram (or ciphertext)

$$C = \mathsf{T}_K(M)$$

which is send over an insecure channel to the recipient, who recovers the original message M using the decipherer T^{-1} as

$$M = \mathsf{T}_K^{-1}(C).$$

According to this scenario, all the transformations defined by the system should be invertible in order to allow the recipient to recover only the original plaintext from the ciphertext. The channel on which the ciphertext is sent is assumed to be insecure in the sense that the enemy cryptanalyst (or adversary) can eavesdrop any message on that channel.

The key K is sampled by the key source in the finite space of all the possible keys allowed by the system. The key is usually considered as a random variable following some *a priori* distribution (which is known by the adversary). In most cases, this distribution is assumed to be uniform. Similarly, the message M is sampled by the message source according to some *a priori* distribution which is generally non-uniform. Once the adversary has intercepted the ciphertext C, the new distributions of M and K are referred to as the *a posteriori* distributions, since the adversary can benefit from any information that can be extracted from C. Intuitively, the level of security achieved by the system depends on how far the *a posteriori* distributions are from the *a priori* distributions.

In this scenario, the secret key must be transmitted to both parties over a *secure* (i.e., confidential) channel, which is obviously more "expensive" to use than the insecure one. This clearly makes sense when the encryption method is such that the message space from which M is chosen is larger than the key space. For example, modern encryption procedures (as block ciphers or stream ciphers) allow the encryption of several gigabytes of data with only one 128-bit key. But this model can also be meaningful when the message and the key are equal in length (which is mandatory when one aims at unconditional security, as we will see). In that case, one can *anticipate* any potential difficulty in transmitting confidential information at a certain time t by transmitting the key at a time $t' < t$ when such a transmission is easier.

Finally, this model assumes that the adversary knows the set from which the transformations T_K and T_K^{-1} are chosen from. In other words, the adversary knows the specifications of the cryptosystem that is used. Besides being conservative (which is often desirable from the point of view of security), this assumption has been proved correct in several situations and in particular when a large period of time is left to the adversary to break the system. This assumption corresponds to one of the most famous Kerckhoffs' principles [83], according to which the security of a cryptosystem should not rely on the secrecy of the cryptosystem itself (which is *not* to say that one should necessarily make it public in practice).

1.2 Perfect Secrecy and the Vernam Cipher

Ideally, no information about the plaintext should leak from the ciphertext C. In other words, the *a posteriori* distribution of the message should be identical to its *a priori* distribution so that an adversary with unlimited computational power cannot recover M (nor K) from C. We thus consider that the encryption system achieves perfect secrecy when

$$\Pr[M = m|C = c] = \Pr[M = m]$$

for any acceptable ciphertext c and message m, which also reads as

$$H(M|C) = H(M),$$

where $H(\cdot)$ denotes Shannon's entropy [139, 157].

The Vernam cipher [160] is a stream cipher developed by Gilbert Sandford Vernam in 1926 which achieves perfect secrecy when the *a priori* distribution of the key is uniform [139] and when its length (at least) corresponds to that of the plaintext. Assuming that the plaintext and the key are represented as bit strings and that they are of *equal length*, the Vernam cipher simply computes the ciphertext as

$$C = M \oplus K$$

where $\oplus$ corresponds the bit-wise exclusive-or operation. For several reasons, the Vernam cipher is impractical: not only a secret key cannot be used twice, but it has to be uniformly distributed, which is hard to achieve in practice. Yet, the problem of the secret key length is not inherent to the Vernam cipher but to the nature of perfect secrecy, as the following theorem shows.

Theorem 1.1 ***(Shannon, 1949)*** *Perfect secrecy implies* $H(K) \geq H(X)$.

1.3 Going Beyond Perfect Secrecy

Obviously, perfect secrecy is too expensive in many practical situations since the quantity of data to be sent over the insecure channel is necessarily (at least) equal to that of the data to be secured. Modern encryption methods are thus more concerned with *practical* security instead.

Essentially, a cryptographic system is considered to be practically secure when no *computationally bounded* adversary can recover meaningful information about M or K from the sole knowledge of the ciphertext C. Most of the currently widely used block ciphers (such as the Advanced Encryption Standard [41]) are assumed to be practically secure (although in almost all cases, no strong mathematical proof of this is provided).

Moreover, even in the case where perfect secrecy is not required, both ends still need to share the same secret key, which shall thus be transmitted to both end in a confidential way. This problem was solved with the invention of public key cryptography, as we will see in Chapter 2.

1.4 Thesis Outline

In the rest of Part I, we will recall several notions concerning Encipherers, which we rather call *symmetric encryption algorithms*. In particular we explain in Chapter 2 how to determine the secret key length by computing the complexity of black box attacks (i.e., generic attacks that apply to any block cipher) and show how the problem of sharing this secret key is solved by means of public key cryptography. Almost all practical block cipher constructions follow either a Feistel scheme [50] (or a generalization of it), a Lai-Massey scheme [96], or a substitution-permutation network (SPN). We recall these three schemes in Chapter 3 following a top-down approach, detailing various of the smallest building blocks used within these schemes together with some of the essential properties they should have. We recall in Chapter 4 the Luby-Rackoff security model [102]. We introduce the notion of perfect cipher together with statistical attacks against block ciphers. In particular, we explain the difference between distinguishing attacks and key-recovery attacks (and see how to turn the first ones into the seconds), and recall linear cryptanalysis [110] which is a classical example of statistical cryptanalysis. The notations used throughout the rest of this thesis are introduced in Chapter 5 as well as some elementary mathematical results.

In Part II we consider the (in)security of block ciphers against statistical cryptanalytic attacks and develop some tools to perform optimal attacks and quantify their efficiency. We do this step-by-step, starting by assuming in Chapter 6 a simple setting in which the adversary has to distinguish between two sources of randomness in a set of reasonable cardinality. Through the method of types, we show how to derive the *optimal distinguisher* limited to q samples and compute its advantage, which we proved to be linked to the Chernoff information between the two probability distributions. Our treatment is not only valid when both distributions are of full support[3] but also when their respective supports differ. Then we consider the case where both distributions are "close" to each other, which is a situation of practical interest in cryptography. We then turn to a more complex problem (from the point of view of the adversary) where one of the two hypotheses is *composite*. Finally, we study the case where the adversary has to decide whether or not the samples follow some known distribution, and we derive her advantage in this case also. In Chapter 7 we consider the case where the cardinality of the samples' set is too large to implement the optimal distinguisher. We introduce *projection-based distinguishers* which typically compress the samples before using them to decide between one hypothesis or another. Within this setting, we re-consider the particular case of linear distinguishers and generalize them to sets of arbitrary cardinality. We show how these distinguishers between random sources can be turned into distinguishers between random oracles (or block ciphers) in Chapter 8 and how, in this setting, one can generalize linear cryptanalysis to Abelian groups. Using these theoretical tools, we show how to break TOY100 and introduce the block cipher DEAN which encrypts blocks of decimal digits. We apply the theory to the SAFER block cipher family in Chapter 9. Most of the theoretical tools introduced in this part are published in [7, 10], except for the generalization of linear cryptanalysis which is published in [8], along with the attacks on SAFER.

We introduce two new block cipher designs in Part III. We start by recalling some essential notions about provable security for block ciphers and about Serge Vaudenay's Decorrelation Theory [155] in Chapter 10. Our contribution essentially relies on introducing new simple modules for which we prove essential properties that we will later use in our designs. In Chapter 11 we introduce C, a block cipher provably secure against a wide range of cryptanalytic attacks, including linear and differential cryptanalysis (taking into account the linear hull effect [125] and the differentials effects, which is unfortunately almost never done in so-called traditional block cipher security proofs). In particular, we compute the *exact* advantage of the best distinguisher limited to two plaintext/ciphertext samples between C and the perfect cipher and use it to compute the exact value of the maximum expected linear probability (resp. differential probability) of C which is known to be inversely proportional to the number of samples required by best possible linear (resp. differential) attack. We conclude the chapter by implementation considerations. Since we are unable to prove any security result on C concerning the best q-limited distinguisher for $q > 3$, we introduce the block cipher KFC in Chapter 12, for which we indeed manage to prove security results for higher order adversaries. The block cipher C is published in [6], based on previous security results we obtained in [9]. The development of KFC is published in [5].

[3] The support of a finite distribution is the set of points on which the probability is non-zero. A distribution is of full support when its support corresponds to the whole sample space.

Chapter 2

Computationally Bounded Adversaries

In this chapter we show how two address two questions raised by Shannon's encryption model, namely

- what should be the typical key length (in bits for example) of a secure block cipher,
- how one can transmit the secret key to both parties.

A block cipher on a finite set is a family of permutations on that set, indexed by a parameter called the key. More formally, let $\mathcal{T}$ and $\mathcal{K}$ be two finite sets, respectively called the *text space* and the *key space*. A block cipher C on the text space $\mathcal{T}$ and key space $\mathcal{K}$ is a set of $|\mathcal{K}|$ permutations on $\mathcal{T}$, i.e.,

$$\mathsf{C} = \{\mathsf{C}_k : \mathcal{T} \to \mathcal{T} : k \in \mathcal{K}\}.$$

To obtain a secure block cipher, it seems natural to require at least that the cardinality of $\mathcal{K}$ is large enough, for a reason that we will formalize here.

2.1 Black Box Attacks: Determining the Secret Key Length

Exhaustive Key Search

We first assume that the block cipher has no equivalent key, i.e., that $\mathsf{C}_k \neq \mathsf{C}_{k'}$ when $k \neq k'$ (otherwise, it suffices to keep in $\mathcal{K}$ exactly one representative of each equivalence class). We consider the scenario where the adversary is given a plaintext/ciphertext pair (P, C), such that $C = \mathsf{C}_{\widetilde{k}}(P)$ for some secret key $\widetilde{k} \in \mathcal{K}$. The objective of the adversary is to recover $\widetilde{k}$. Probably the most basic strategy is to exhaust all possible keys k and check whether $C = \mathsf{C}_k(P)$. If this is not the case, then k is certainly not the key $\widetilde{k}$. If the equality holds, then the algorithm outputs k and stops. This is illustrated in Algorithm 2.1. Assuming that $\widetilde{k} = k_{\sigma(j)}$ for some j and the permutation σ drawn on line 1, then it is clear that the algorithm succeeds if

$$C \neq \mathsf{C}_{k_{\sigma(i)}}(P)$$

```
Input: A plaintext/ciphertext pair $(P, C) \in \mathcal{T}^2$ such that $C = \mathsf{C}_{\widetilde{k}}(P)$ for some
       secret key $\widetilde{k} \in \mathcal{K} = \{k_1, k_2, \ldots, k_{|\mathcal{K}|}\}$
Output: A key $k$
Select $\sigma$ uniformly at random among all permutations of $\{1, 2, \ldots, |\mathcal{K}|\}$
for $i = 1, 2, \ldots, |\mathcal{K}|$ do
    if $C = \mathsf{C}_{k_{\sigma(i)}}(P)$ then return $k_{\sigma(i)}$
end
```

Algorithm 2.1: Exhaustive search for the secret key $\widetilde{k} \in \mathcal{K}$.

for all $i = 1, 2, \ldots, j-1$. We denote by p the probability of success. Since σ is uniformly distributed, we have

$$p = \frac{1}{N}$$

where N denotes the number of keys k in $\mathcal{K}$ such that $C = \mathsf{C}_k(P)$. We can approximate N by assuming that the $|\mathcal{K}|$ permutations defined by the block cipher are initially chosen (at the time of designing C) at random and in a uniform way so that, denoting $\mathsf{C}^\star : \mathcal{T} \to \mathcal{T}$ a uniformly distributed random permutation we have

$$N = \max(1, |\mathcal{K}| \Pr[\mathsf{C}^\star(P) = C]) = \max(1, |\mathcal{K}| \,/\, |\mathcal{T}|). \tag{2.1}$$

Since in practice $|\mathcal{K}|$ and $|\mathcal{T}|$ are close to each other, a few pairs are sufficient to obtain an overwhelming probability of success.

Assuming that the algorithm succeeds using only one pair, the time complexity is clearly equal to the position of $\widetilde{k}$ in the list $\{k_{\sigma(1)}, k_{\sigma(2)}, \ldots, k_{\sigma(|\mathcal{K}|)}\}$. In the worst case, the complexity is $|\mathcal{K}|$ encryptions while on average it is $(|\mathcal{K}|+1)/2$ since σ is uniformly distributed. In both cases this does not depend on the distribution of $\widetilde{k}$ (thanks to the random selection of σ). The memory complexity of Algorithm 2.1 is clearly negligible.

Codebook Attack

The exhaustive key search algorithm requires no memory but has a tremendous time complexity. One can rather imagine storing all possible $(\mathsf{C}_k(P), k)$ pairs in a huge table (for all possible k and one chosen P, sorted according to the first entry), request the encryption of P under the secret key $\widetilde{k}$, and perform one table look-up in order to recover k. The time complexity is now negligible (except for the table pre-computation time) and the memory requirement is in $O(|\mathcal{K}|)$.

Time-Memory Trade-offs

Martin Hellman showed in [66] how to obtain a trade-off between time and memory complexities (a concept that was further refined by in [127]). Essentially, the method allows the reduction of both time and memory complexities to $|\mathcal{K}|^{2/3}$.

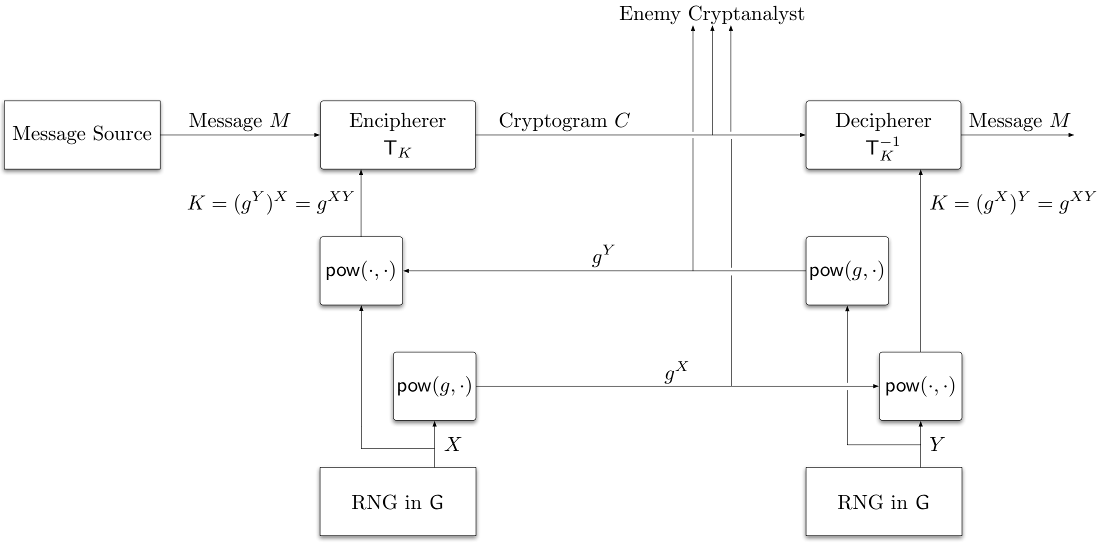


Figure 2.1: Secret key exchange by means of an authenticated channel

Conclusion

What the black box methods show is that $\mathcal{K}$ should be large enough in order for the time needed to encrypt $|\mathcal{K}|$ plaintext to be overwhelming. As a consequence, most of the current block ciphers use 128-bit or 256-bit keys whereas older ciphers used to have 64-bit (or even 56-bit) keys.

2.2 New Directions in Cryptography: reducing Confidentiality to Authenticity

In their seminal article "*New Directions in Cryptography*" [47], Diffie and Hellman explain how to build a confidential channel from an authentic channel. The way their construction integrates in Shannon's model of secrecy is illustrated in Figure 2.1. To simplify the description, we assume that the we are in the situation where Alice needs to send some confidential information to Bob. Let G be finite cyclic group and let $g \in \mathsf{G}$ be a (public) generator of this group. Alice and Bob respectively choose X and Y uniformly at random in G, send g^X and g^Y to each other through the insecure (but authenticated) channel and both compute $K = g^{XY}$. Without entering into the details (for which we refer to [47, 157]), the Diffie-Hellman key agreement protocol is assumed to be secure whenever the channels on which g, g^X, and g^Y are sent are *authenticated* and as soon as it is computationally hard to solve the Diffie-Hellman Problem (DHP) in G, that is, given two inputs $U, V \in \mathsf{G}$, compute $K = g^{XY}$ where $X = \log_g U$ and $Y = \log_g V$. In particular, this problem is assumed to be hard in $\mathbf{Z}_p^\star$, where p is a large prime number. We note that in practice, the secret key K will not be equal to g^{XY}

but rather to $h(g^{XY})$ where $h : \mathsf{G} \rightarrow \{0,1\}^n$ is a hash function and n is the secret key length.

Since Diffie and Hellman, various other means of exchanging a common secret key by means of an authenticated channel were invented. In particular, any public key cryptosystem (such as RSA [132], ElGamal [49], Paillier cryptosystem [128], the Naccache-Stern cryptosystem [117] or Cramer-Shoup [38], to cite only a few) can be used.

Chapter 3

Block Ciphers Design: a Top-Down Approach

In this chapter we introduce typical block cipher designs. We first consider iterated block ciphers, which encompass almost all block ciphers widely used today, key schedules, and then consider three particular cases of iterated block ciphers, namely Feistel ciphers, ciphers based on the Lai-Massey scheme, and substitution-permutation networks.

It will then become evident that, whatever the kind of scheme, the building blocks used within it must have certain desirable properties. Finally, we detail the design of the Advanced Encryption Standard (AES) since the block cipher C that we introduce in Chapter 11 is based on it.

3.1 Iterated Block Ciphers and Key Schedules

Let $\mathcal{T}$ and $\mathcal{K}$ respectively be the text space and the key space of a block cipher

$$\mathsf{C} = \{\mathsf{C}_k : \mathcal{T} \to \mathcal{T} : k \in \mathcal{K}\}.$$

Let $r > 0$ be a positive integer and let $\mathcal{K}_1, \mathcal{K}_2, \ldots \mathcal{K}_r$ be r finite sets. C is said to be an *r-round iterated block cipher* when it can be written as

$$\mathsf{C}_k = \mathsf{R}^{(r)}_{k_r} \circ \mathsf{R}^{(r-1)}_{k_{r-1}} \circ \cdots \circ \mathsf{R}^{(1)}_{k_1}, \tag{3.1}$$

for all $k \in \mathcal{K}$, where

$$\mathsf{R}^{(i)} = \{\mathsf{R}^{(i)}_{k_i} : \mathcal{T} \to \mathcal{T} : k_i \in \mathcal{K}_i\}$$

is called the ith round of C. Of course, this definition is not completely sound since, according on it, there is not a clear unique way of expressing an iterated cipher. Usually, the ith round of a block cipher is successively made of

- a key-mixing phase, where the key k_i is mixed to the data being encrypted,
- a confusion phase (in the sense of [139]),
- and a diffusion phase which dissipates the eventual redundancy.

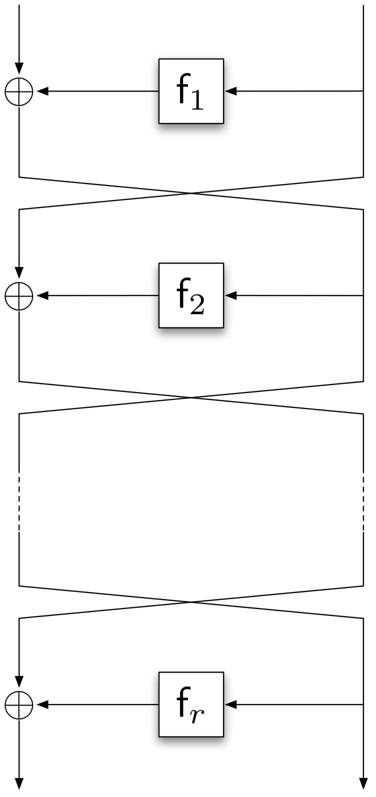


Figure 3.1: An r-round Feistel scheme $\Psi(\mathsf{f}_1, \mathsf{f}_2, \ldots, \mathsf{f}_r)$

The last round often restricts to the key-mixing phase. Finally, $k_1, k_2, \ldots, k_r$ are called the round keys of the block cipher and are derived from the main secret key k by means of a deterministic algorithm called the key schedule. We will see that in most cases, the length of each round key is comparable to that of the main secret key, so that when this secret key is considered as a random variable K, the round keys $K_1, K_2, \ldots, K_r$ cannot be independent.

3.2 Round Functions Based on Feistel Schemes

A Feistel scheme is a structure which allows to construct a permutation on $2n$-bit strings based on functions of n-bit strings. An $r > 0$ rounds Feistel scheme based on the functions

$$\mathsf{f}_1, \mathsf{f}_2, \ldots, \mathsf{f}_r : \{0,1\}^n \to \{0,1\}^n,$$

is denoted $\Psi(\mathsf{f}_1, \mathsf{f}_2, \ldots, \mathsf{f}_r)$ and is represented in Figure 3.1. It is easy to see that $\Psi(\mathsf{f}_1, \mathsf{f}_2, \ldots, \mathsf{f}_r)$ is invertible since

$$\Psi^{-1}(\mathsf{f}_1, \mathsf{f}_2, \ldots, \mathsf{f}_r) = \Psi(\mathsf{f}_r, \mathsf{f}_{r-1}, \ldots, \mathsf{f}_1).$$

To construct an r-round iterated block cipher $\mathsf{C} : \{0,1\}^{2n} \rightarrow \{0,1\}^{2n}$ (as in (3.1)) based on an r-round Feistel scheme, one typically defines a family of functions

$$\mathsf{f} = \left\{ \mathsf{f}_k : \{0,1\}^n \rightarrow \{0,1\}^n : k \in \mathcal{K}' \right\}$$

and then let for all $k \in \mathcal{K}'$

$$\mathsf{R}^{(i)}(x_{\mathtt{left}} \| x_{\mathtt{right}}) = (x_{\mathtt{right}} \| x_{\mathtt{left}} \oplus \mathsf{f}_{k_i}(x_{\mathtt{right}})),$$

where $x_{\mathtt{left}}$ (resp. $x_{\mathtt{right}}$) denotes the left-most (resp. right-most) n bits of the input x of the round. Usually, the last round does not permute the outputs (as in Figure 3.1). In this way, the construction of a family of permutations on $2n$ bits reduces to that of a family of functions on n bits. Moreover, Luby and Rackoff showed in [102] that from a secure family of functions, one only needs three rounds to obtain a secure block cipher (this is more formally stated in Chapter 10).

Practical examples of block ciphers based on a Feistel scheme include the Data Encryption Standard (DES) [122] and Blowfish [134]. The block cipher KFC that we introduce in Chapter 12 is based on a three rounds Feistel scheme.

3.3 Round Functions Based on Lai-Massey Schemes

Like the Feistel scheme, the Lai-Massey scheme enables us to construct a permutation from functions. An r rounds Lai-Massey scheme is represented in Figure 3.2. This scheme was developed by Xuejia Lai and James Massey during the design of the block cipher IDEA [94]. The particularity of the scheme is that it requires a commutative and associative law (which can be the exclusive-or operation or more complex group laws like in IDEA). As is, the Lai-Massey scheme is not secure even if the round functions are. The reason being that whatever the number of rounds, it is always true that

$$x_{\mathtt{left}} \boxminus x_{\mathtt{right}} = y_{\mathtt{left}} \boxminus y_{\mathtt{right}},$$

where $x = x_{\mathtt{left}} \| x_{\mathtt{right}}$ and $y = y_{\mathtt{left}} \| y_{\mathtt{right}}$ respectively denote the input and the output of the scheme. To break this undesirable property, Vaudenay demonstrates in [153] that introducing a special (fixed) permutation σ at the output of each round left branch allows one to obtain security results equivalent to those of the Feistel scheme. The permutation σ must be such that $z \mapsto \sigma(z) - z$ is also a permutation, in which case σ is called an orthomorphism.

Practical examples of block ciphers based on a Lai-Massey scheme include IDEA [94] and FOX [76].

3.4 Round Functions Based on Substitution-Permutation Networks

The last typical skeleton is probably the one which is closest to Shannon's conception of encryption [139] since it consists of a sequence wherein a substitution

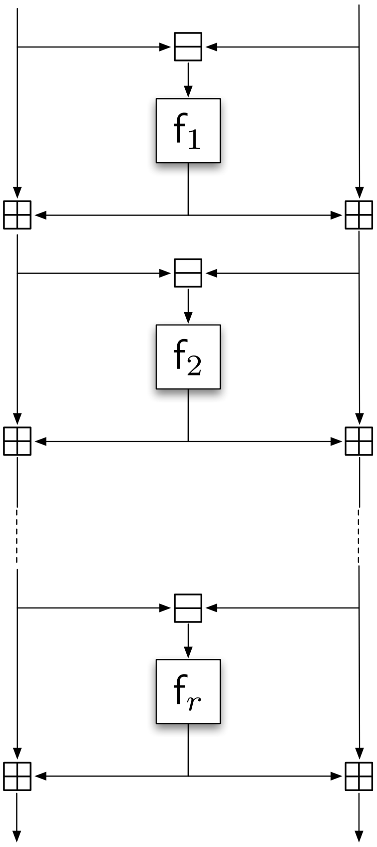


Figure 3.2: An r-round Lai-Massey scheme

layer producing confusion is followed by a confusion layer producing diffusion. Although any block cipher can be seen as a substitution-permutation network, those based on the Feistel or the Lai-Massey schemes are usually not considered to be part of this category.

The family of block ciphers SAFER [107, 109] (which we cryptanalyse in Chapter 9) and the Advanced Encryption Standard [41] (which we introduce in Section 3.7 and on which we base the design of the block cipher C in Chapter 11) are well known examples of substitution-permutation networks.

3.5 Providing Diffusion: on the Need for Multipermutations

According to Shannon, the diffusion process should "*dissipate* [the redundancy] into long range statistics" [139]. Yet, this definition leaves quite some space for interpretation. Schnorr and Vaudenay formalize in [137] the concept of multipermutation explaining what it technically means to provide good diffusion. Vaudenay further illustrates in [150] how fundamental this concept can be. In particular, he shows that if one replaces the substitution boxes of SAFER by other boxes then one obtains a weak block cipher in more than 6% of cases, the reason being that the diffusion of SAFER is *not* a multipermutation. This is also a feature we exploit in the generalized linear cryptanalysis that we propose in Chapter 9.

Definition 3.1 *A (r, n)-multipermutation over an alphabet $\mathcal{Z}$ is a function* f *from $\mathcal{Z}^r$ to $\mathcal{Z}^n$ such that two different $(r+n)$-tuples of the form $(x, \mathsf{f}(x))$ cannot collide in any r positions.*

Vaudenay notes that in the case where f is linear, Definition 3.1 corresponds to MDS codes. For example, one of the core transformations of the AES diffusion is based on a linear multipermutation (i.e., on an MDS code). In Chapter 11 we take advantage of the inherent properties of MDS codes to prove certain security results concerning the block cipher C.

3.6 Providing Confusion: Mixing key bits

Providing confusion is usually done by applying a (fixed) substitution box to a mixing of key bits and of text bits. This is the case for the DES, the AES and SAFER. Probably the most well known counter-example is IDEA. Sometimes the confusion is created by key-dependent substitution boxes, which is the case for Blowfish [134] for example, where the key bits have the particularity to be mixed with text bits in an non-linear way. It seems natural to look for substitution boxes as similar as possible to uniformly distributed random permutations (or functions, depending on the case), as indicated by several security results that we manage to prove for both C and KFC thanks to the ideal nature of the boxes we choose.

3.7 The Advanced Encryption Standard

As an example of substitution-permutation network, we introduce the encryption part of the Advanced Encryption Standard [41]. The AES is a 128-bit block cipher made of $r = 10$ rounds in the case where 128-bit keys are used[1], all identical in their structure (except the last one). Each round is parameterized by a *round-key* which is derived from the main 128 bits secret key using a so-called *key schedule algorithm.* The structure of each round is made of a (non-linear) substitution layer followed by a (linear) permutation layer.

A 128-bit plaintext p is considered as a 4×4 array of 8-bit elements $(p_{i,j})_{1 \leq i,j \leq 4}$ with

$$p = p_{1,1} \| p_{2,1} \| p_{3,1} \| p_{4,1} \| p_{1,2} \| \cdots \| p_{4,4}.$$

The first $r - 1$ first rounds successively apply to p the following transformations:

[1] The AES can also be used with 192 and 256-bit keys, in which cases the number of rounds are 12 and 14 respectively.

- `AddRoundKey` performs an exclusive-or operation between the bits of p and the bits of the round key k:

$$\begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} & a_{1,4} \\ a_{2,1} & a_{2,2} & a_{2,3} & a_{2,4} \\ a_{3,1} & a_{3,2} & a_{3,3} & a_{3,4} \\ a_{4,1} & a_{4,2} & a_{4,3} & a_{4,4} \end{bmatrix} = \begin{bmatrix} p_{1,1} & p_{1,2} & p_{1,3} & p_{1,4} \\ p_{2,1} & p_{2,2} & p_{2,3} & p_{2,4} \\ p_{3,1} & p_{3,2} & p_{3,3} & p_{3,4} \\ p_{4,1} & p_{4,2} & p_{4,3} & p_{4,4} \end{bmatrix} \bigoplus \begin{bmatrix} k_{1,1} & k_{1,2} & k_{1,3} & k_{1,4} \\ k_{2,1} & k_{2,2} & k_{2,3} & k_{2,4} \\ k_{3,1} & k_{3,2} & k_{3,3} & k_{3,4} \\ k_{4,1} & k_{4,2} & k_{4,3} & k_{4,4} \end{bmatrix}$$

- `SubBytes` applies to each 8-bit $a_{i,j}$ a fixed substitution box $\mathsf{S}[\cdot]$:

$$\begin{bmatrix} b_{1,1} & b_{1,2} & b_{1,3} & b_{1,4} \\ b_{2,1} & b_{2,2} & b_{2,3} & b_{2,4} \\ b_{3,1} & b_{3,2} & b_{3,3} & b_{3,4} \\ b_{4,1} & b_{4,2} & b_{4,3} & b_{4,4} \end{bmatrix} = \begin{bmatrix} \mathsf{S}[a_{1,1}] & \mathsf{S}[a_{1,2}] & \mathsf{S}[a_{1,3}] & \mathsf{S}[a_{1,4}] \\ \mathsf{S}[a_{2,1}] & \mathsf{S}[a_{2,2}] & \mathsf{S}[a_{2,3}] & \mathsf{S}[a_{2,4}] \\ \mathsf{S}[a_{3,1}] & \mathsf{S}[a_{3,2}] & \mathsf{S}[a_{3,3}] & \mathsf{S}[a_{3,4}] \\ \mathsf{S}[a_{4,1}] & \mathsf{S}[a_{4,2}] & \mathsf{S}[a_{4,3}] & \mathsf{S}[a_{4,4}] \end{bmatrix}$$

- `ShiftRows` shifts each row of the 4 × 4 array by an offset which depends on the row number:

$$\begin{bmatrix} c_{1,1} & c_{1,2} & c_{1,3} & c_{1,4} \\ c_{2,1} & c_{2,2} & c_{2,3} & c_{2,4} \\ c_{3,1} & c_{3,2} & c_{3,3} & c_{3,4} \\ c_{4,1} & c_{4,2} & c_{4,3} & c_{4,4} \end{bmatrix} = \begin{bmatrix} b_{1,1} & b_{1,2} & b_{1,3} & b_{1,4} \\ b_{2,2} & b_{2,3} & b_{2,4} & b_{2,1} \\ b_{3,3} & b_{3,4} & b_{3,1} & b_{3,2} \\ b_{4,4} & b_{4,1} & b_{4,2} & b_{4,3} \end{bmatrix}$$

- `MixColumns` applies a linear multipermutation to each column of $(c_{i,j})_{1\leq i,j\leq 4}$. Each 8-bit element is considered as a member of the finite field with 256 elements $\mathrm{GF}(2^8)$. The elements of this finite field are represented by polynomials of degree less than 8 with coefficients in $\mathrm{GF}(2)$, standard operations are performed modulo the irreducible polynomial $x^8+x^4+x^3+x+1$, and any 8-bit element $b = b_7b_6\dots b_0$ corresponds to the polynomial $b_7x^7 + b_6x^6 + \cdots + b_0$. With these notations, the `MixColumns` operation on the jth column of $(c_{i,j})_{i,j}$ is:

$$\begin{bmatrix} d_{1,j} \\ d_{2,j} \\ d_{3,j} \\ d_{4,j} \end{bmatrix} = \begin{bmatrix} \mathtt{0x02} & \mathtt{0x03} & \mathtt{0x01} & \mathtt{0x01} \\ \mathtt{0x01} & \mathtt{0x02} & \mathtt{0x03} & \mathtt{0x01} \\ \mathtt{0x01} & \mathtt{0x01} & \mathtt{0x02} & \mathtt{0x03} \\ \mathtt{0x03} & \mathtt{0x01} & \mathtt{0x01} & \mathtt{0x02} \end{bmatrix} \times \begin{bmatrix} c_{1,j} \\ c_{2,j} \\ c_{3,j} \\ c_{4,j} \end{bmatrix}$$

The last round of AES is identical to the $r - 1$ previous ones, except that there no `MixColumns` operation. Finally, a last `AddRoundKey` completes the algorithm.

Chapter 4

The Luby-Rackoff Model: Statistical Attacks against Block Ciphers

In Chapter 2 we considered computationally bounded adversaries and used them to determine the length of the secret key of a typical block cipher. We also showed that public-key cryptography can be used to turn an authentic channel into an (expensive) confidential channel that can be used to exchange this secret key.

Conversely, an adversary in the Luby-Rackoff Model is assumed to be computationally unbounded[1] and only limited by the number of plaintext and/or ciphertext samples she has access to.

4.1 The Perfect Cipher and Security Models

Let $\mathcal{T}$ and $\mathcal{K}$ respectively be the text space and the key space of a block cipher

$$\mathsf{C} = \{\mathsf{C}_k : \mathcal{T} \to \mathcal{T} : k \in \mathcal{K}\}.$$

The block cipher C can be considered as a random permutation by simply considering the key $K \in \mathcal{K}$ is a random variable. Intuitively the perfect cipher should have no particular property common to each permutation that it defines. As a consequence, the perfect cipher

$$\mathsf{C}^\star : \mathcal{T} \to \mathcal{T}$$

is defined as a uniformly distributed random permutation on $\mathcal{T}$. Obviously, the perfect cipher cannot be implemented for realistic block sizes, since the key length is proportional to $\log(|\mathcal{T}|!)$. When studying the security of a block cipher C in the Luby-Rackoff model, one is essentially concerned with how easy it is to distinguish C from $\mathsf{C}^\star$.

More formally, we consider an algorithm, called a *distinguisher* and denoted by A, that queries an oracle $\mathcal{O}$ which implements either a random instance of the block cipher C (an hypothesis that we denote $\mathsf{H}_1 : \mathcal{O} = \mathsf{C}$) or a random instance of the perfect cipher (an hypothesis that we denote $\mathsf{H}_0 : \mathcal{O} = \mathsf{C}^\star$). The distinguisher

[1] So that we can assume without loss of generality that it is deterministic, see [157].

eventually outputs a bit to indicate which hypothesis between H_0 and H_1 is more likely to be correct. The ability to distinguish between these two hypotheses is defined as the *advantage* of the distinguisher and is defined by

$$\mathrm{Adv}_{\mathsf{A}}(\mathsf{H}_0, \mathsf{H}_1) = \left|\Pr_{\mathsf{H}_1}[\mathsf{A} = 1] - \Pr_{\mathsf{H}_0}[\mathsf{A} = 1]\right|,$$

which we also denote by $\mathrm{Adv}_{\mathsf{A}}(\mathsf{C}, \mathsf{C}^\star)$. The distinguisher is essentially limited by the number q of queries it can make to the oracle, so that A is usually referred to as a q-limited distinguisher. Furthermore, a distinguisher that can actually choose the ith query made to the oracle based on the answers of the $i-1$ previous ones is said to be *adaptive.* A distinguisher which asks the q queries at once is said to be *non-adaptive.* Obviously, adaptive distinguishers are more powerful than non-adaptive distinguishers. We say that the block cipher C is resistant to q-limited (non-)adaptive distinguishers if any q-limited (non-)adaptive distinguisher A has a negligible advantage.

This security model is the one used by Michael Luby and Charles Rackoff in [102] to study the security of the Feistel scheme (on which the DES is based). This is also the model in which we prove security results for the block ciphers C and KFC that we introduce in chapters 11 and 12 respectively.

4.2 From Distinguishing to Key Recovery

Most of the concrete statistical cryptanalytic attacks against block ciphers implicitly assume the Luby-Rackoff model. Moreover, most of the well known attack categories (if not all) are non-adaptive. Within these, cryptanalysts generally distinguish between *known-plaintext attacks* (KPA), in which the adversary has no control on which queries are made to the oracle, and *chosen-plaintext attacks* (CPA), in which the queries follow a certain distribution chosen by the adversary. For example, linear cryptanalysis [110, 147] is a known-plaintext attack and differential cryptanalysis [21] is a chosen-plaintext attack.

The objective of a cryptanalytic attack can either be to distinguish between the two hypothesis mentioned in the previous section, namely $\mathsf{H}_0 : \mathcal{O} = \mathsf{C}^\star$ and $\mathsf{H}_1 : \mathcal{O} = \mathsf{C}$, or to recover the key that is used to encrypt the plaintext/ciphertext pairs that are available. In the rest of this section, we introduce a formalism close to Wagner's unified view of block cipher cryptanalysis [163], which is based on Vaudenay's model of statistical cryptanalysis [151]. We apply it within the scope of iterated ciphers and show why distinguishing attacks often lead to key recovery attacks.

Cryptanalytic attacks can be formalized using the notion of *projection* and *commutative diagrams.* Consider an adversary performing a known plaintext attack against $r+1$ rounds of an iterated block cipher

$$\mathsf{C} = \{\mathsf{C}_k : \mathcal{T} \to \mathcal{T} : k \in \mathcal{K}\}.$$

To emphasize the fact that C is made of $r+1$ rounds, we denote it $\mathsf{C}^{(r+1)}$ and denote the ith round by $\mathsf{R}^{(i)}$. Recursively, we let $\mathsf{C}^{(r+1)} = \mathsf{R}^{(r+1)} \circ \mathsf{C}^{(r)}$ and denote by k_i the ith

$$\begin{tikzcd}
\mathcal{T} \arrow[r, "\rho"] \arrow[d, "\mathsf{C}^{(r)}"'] & \mathcal{X} \arrow[d, "g"] \\
\mathcal{T} \arrow[r, "\phi"] & \mathcal{Y}
\end{tikzcd}$$

Figure 4.1: A commutative diagram representing a distinguishing property on C^r

round key (computed from the main key k by means of the key schedule). To simplify the notations, we assume that all the rounds have the same structure, so that we simply denote any round by R.

In an ideal scenario, the adversary is able to find a *distinguishing property* for the r first rounds of the cipher. More formally, we assume that the adversary has discovered two *projections*

$$\rho : \mathcal{T} \to \mathcal{X} \quad \text{and} \quad \phi : \mathcal{T} \to \mathcal{Y}$$

(where $\mathcal{X}$ and $\mathcal{Y}$ typically are sets of small cardinality) and some function $g : \mathcal{X} \to \mathcal{Y}$ such that

$$g \circ \rho = \phi \circ \mathsf{C}_k^{(r)} \tag{4.1}$$

holds for all keys $k \in \mathcal{K}$. Assume also that this property is not *trivial*, i.e., not true in general if we replace C_k by $\mathsf{C}^\star$. This can be represented by means of a *commutative diagram* as shown on Figure 4.1, in which the facts that (4.1) holds and that the diagram commutes are equivalent. In such a case, the adversary can often mount a key recovery attack against $r+1$ rounds of the block cipher by first guessing the last round key k_{r+1}, decrypting one round of the cipher for all the ciphertexts made available to her using her guess $\widetilde{k}$ of k_{r+1}, and finally checking whether

$$g \circ \rho = \phi \circ \mathsf{C}_k^{(r+1)} \circ \mathsf{R}_{\widetilde{k}}^{-1} \tag{4.2}$$

holds. When her guess is correct, i.e., when $\widetilde{k} = k_{r+1}$, then (4.2) is equivalent to (4.1), so that it will always hold. When $\widetilde{k} \neq k_{r+1}$ then we can consider that the adversary is actually performing an additional one-round encryption of all the ciphertexts. Consequently, we can (abusively) consider that the adversary checks whether

$$g \circ \rho = \phi \circ \mathsf{C}_k^{(r+2)} \tag{4.3}$$

holds in this case. As the distinguishing property was assumed to be non trivial, there is no particular reason why (4.3) should hold, so that the adversary will easily check that her guess is incorrect as (4.3) is likely to be false for several plaintext/ciphertext pairs.

In practical attacks, it is usually only necessary to guess some bits of the last round key in order to check the distinguishing property, the remaining bits being recovered by exhaustive search. Once k_{r+1} is recovered the adversary can peel-off an entire round of the block cipher and iterate the whole process (usually, once a distinguishing property can be found for a certain number of rounds, a distinguishing property on fewer rounds is easy to find). In certain cases, recovering the last round key can be sufficient to recover the key k.

As a distinguishing attack on $\mathsf{C}_k^{(r)}$ often leads to a key recovery on $\mathsf{C}_k^{(r+1)}$, from now on we only consider distinguishing attacks, i.e., attacks aiming at finding some non trivial distinguishing property on the block cipher. We illustrate these notions by introducing a concrete example, namely linear cryptanalysis.

4.3 Linear Cryptanalysis

Linear cryptanalysis is a known-plaintext attack proposed by Matsui in [110] to break the DES [122], based on concepts introduced by Tardy-Corfdir and Gilbert in [147]. It assumes that the plaintexts are independent and uniformly distributed in the text space $\mathcal{T} = \{0,1\}^n$, and consider linear (in the sense of GF(2)) binary projections of the form

$$\rho(P) = a \bullet P = a_0P_0 \oplus a_1P_1 \oplus \cdots \oplus a_{n-1}P_{n-1} \in \{0,1\},$$

where $a \in \{0,1\}^n$ is called a *mask*. Essentially, linear cryptanalysis aims at finding an input mask a and an output mask b on r rounds of an iterated cipher C, such that

$$(a \bullet P) \oplus (b \bullet \mathsf{C}_k^{(r)}(P)) = 0 \tag{4.4}$$

holds with a probability far distant from $\frac{1}{2}$ for all keys $k \in \mathcal{K}$. More precisely, if one let $\frac{1}{2} + \epsilon$ be the probability that the linear relation (4.4) holds, then the efficiency of the cryptanalysis based on it is known to depend on the *linear probability* coefficient [32]

$$\mathrm{LP}_{a,b}(\mathsf{C}_k) = \mathrm{LP}\left((a \bullet P) \oplus (b \bullet \mathsf{C}_k^{(r)}(P))\right) = 4\epsilon^2$$

where the *linear probability* of a random bit B is defined by

$$\mathrm{LP}(B) = (2\Pr[B=0]-1)^2 = \left(\mathrm{E}\left((-1)^B\right)\right)^2.$$

The linear probability is often assumed to be close to the expected linear probability

$$\mathrm{ELP}_{a,b}(\mathsf{C}) = \mathrm{E}_K\left(\mathrm{LP}_{a,b}(\mathsf{C}_K)\right),$$

an hypothesis referred to as the *hypothesis of stochastic equivalence* (a concept formalized by Lai [94, 97]).

In practice, to derive a linear relation such as (4.4) on an iterated cipher made of r rounds, the cryptanalyst first derives adequate linear relations on each round of the block cipher, such that the output mask of round $i-1$ is equal to the input mask of round i. This forms a so called characteristic $(a_0, a_1, \ldots, a_r)$. Using Matsui's piling-up lemma, which states that for two independent random bits B_1 and B_2 we have

$$\mathrm{LP}(B_1 \oplus B_2) = \mathrm{LP}(B_1)\mathrm{LP}(B_2),$$

the cryptanalyst then usually assumes that

$$\mathrm{ELP}_{a_0,a_r}(\mathsf{C}) \approx \prod_{i=1}^{r} \mathrm{ELP}_{a_{i-1},a_i}(\mathsf{R}). \tag{4.5}$$

This strategy is the one adopted by Matsui in his cryptanalysis of the DES. In that particular case, the experiments justify the approximations [72, 111].

Yet, Nyberg shows in [125] that the right hand-side of (4.5) essentially underestimates the true expected linear probability since, in the case of Markov ciphers [97] (see Definition 8.8) we actually have

$$\mathrm{ELP}_{a_0,a_r}(\mathsf{C}) = \sum_{a_1,\ldots,a_{r-1}} \prod_{i=1}^{r} \mathrm{ELP}_{a_{i-1},a_i}(\mathsf{R}),$$

a property which is often referred to as the *linear hull effect.* We emphasize the fact that since the approximation (4.5) underestimates the true value of the expected linear probability, it also underestimates the efficiency of the cryptanalysis. Whereas this is perfectly acceptable from the point of view of the adversary (since the attack can only perform better than expected), it is unfortunate to see the same approximation made in so-called security proofs of block ciphers. For example, the maximum value (over all non-zero input/output masks) of the expected linear probability over 8 rounds of the AES was initially assumed to be less than 2^{-300} [41, pp.30–31], which is obviously wrong: since for any input mask a, the sum over all the 2^{128} values of $\mathrm{ELP}_{a,b}(\mathsf{AES})$ is equal to 1, at least one must be greater than 2^{-128}. Yet, in that particular case, Keliher proves that the maximum value of the ELP's can be bounded by $1.778 \cdot 2^{-107}$ for 8 or more rounds [79–81]. In the cases of the block ciphers C and KFC that we introduce in chapters 11 and 12 respectively, we manage to compute the exact value of the expected linear probability (taking the linear hull effect into account) for various number of rounds.

Other examples of statistical cryptanalytic attacks include differential cryptanalysis (which is a chosen plaintext attack introduced by Biham and Shamir in [23]), several of its variants (such as truncated differentials [88], impossible differentials [18] or higher order differentials [88, 95]), Vaudenay's χ^2 cryptanalysis [62, 151], and integral attacks [69, 93]. An exhaustive review is provided by Junod in [73].

Chapter 5

Notations and Elementary Results

We introduce in this last chapter the notations that will be used throughout as well as some elementary results.

5.1 Random Variables, Probabilities, Strings, etc.

If $\mathcal{Z}$ is a finite set, we denote by $|\mathcal{Z}|$ its cardinality. Let P denote a probability distribution over the finite set $\mathcal{Z}$. We denote the fact that a random variable X is drawn according to the distribution P by $X \sim \mathsf{P}$ or $X \overset{\mathsf{P}}{\leftarrow} \mathcal{Z}$ in the case of an algorithm. The probability that X takes a particular value $a \in \mathcal{Z}$ is either denoted by $\Pr_{\mathsf{P}}[a]$, $\Pr[X = a]$, or $\mathsf{P}[a]$, where in the last case the probability distribution is simply seen as a vector in $[0,1]^{|\mathcal{Z}|}$. The *support* of P is the subset of $\mathcal{Z}$ made of all elements a such that $\mathsf{P}[a] \neq 0$ and is denoted by $\mathrm{supp}(\mathsf{P})$. The distribution P is said to be of *full support* if $\mathrm{supp}(\mathsf{P}) = \mathcal{Z}$. If A and B denote some random events such that $\Pr[A] > 0$, we will denote $\Pr[B|A]$ or $\Pr_A[B]$ the probability of the event B given the occurrence of the event A.

If $z_1, z_2, \ldots, z_q \in \mathcal{Z}$ are q elements of $\mathcal{Z}$, we denote by $\mathbf{z}^q = (z_1, z_2, \ldots, z_q) \in \mathcal{Z}^q$ the vector of $\mathcal{Z}$ having z_i as its ith component. We adopt a similar notation for random variables. If $Z_1, Z_2, \ldots, Z_q \in \mathcal{Z}$ are q independent and identically-distributed (i.i.d.) random variables drawn according to distribution P, we denote by P^d the distribution of $\mathbf{Z}^q = (Z_1, Z_2, \ldots, Z_q)$ so that

$$\Pr[Z_1 = z_1, Z_2 = z_2, \ldots, Z_q = z_q] = \Pr[\mathbf{Z}^q = \mathbf{z}^q] = \Pr_{\mathsf{P}^q}[\mathbf{z}^q] = \mathsf{P}^q[\mathbf{z}^q].$$

We note that since the random variables are assumed to be independent, it always holds that $\mathsf{P}^q[\mathbf{z}^q] = \prod_{i=1}^{q} \mathsf{P}[z_i]$.

The set of all functions $\mathcal{F}$ from the finite set $\mathcal{Z}$ to R is a vector space of finite dimension thus all norms $\|\cdot\|$ on this set define the same topology. The *open ball* of radius $\epsilon > 0$ around $f_0 \in \mathcal{F}$ is the set $\mathcal{B}_\epsilon(f_0) = \{f \in \mathcal{F} : \|f - f_0\| < \epsilon\}$. An *open set* is a union of open balls. The *interior* of a set Π is the union of all open sets included in Π and is denoted by $\overset{\circ}{\Pi}$. The *closed ball* of radius $\epsilon > 0$ around $f_0 \in \mathcal{F}$ is the set

$\overline{\mathcal{B}}_\epsilon(f_0) = \{f \in \mathcal{F} : \|f - f_0\| \leq \epsilon\}$. A *closed set* is an intersection of closed balls. The *closure* of a set Π is the intersection of all closed sets containing Π and is denoted by $\overline{\Pi}$.

5.2 Vector Norms and Fundamental Inequalities

In this section we recall several fundamental inequalities valid for the specific norms that we use throughout this thesis.

Definition 5.1 *Let $z_1, z_2, \ldots, z_n \in \mathsf{R}$ and $\mathbf{z}^n = (z_1, z_2, \ldots, z_n)$. Let r be a positive integer. The r-norm of the vector $\mathbf{z}^n$ is denoted $\|\mathbf{z}^n\|_r$ and defined by*

$$\|\mathbf{z}^n\|_r = \left(\sum_{i=1}^{n} |z_i|^r\right)^{1/r}.$$

When $r = 2$, we obtain the Euclidean norm. The infinity norm of $\mathbf{z}^n$ is denoted $\|\mathbf{z}^n\|_\infty$ and is defined by

$$\|\mathbf{z}^n\|_\infty = \max_{i=1,\ldots,n} |z_i|.$$

Theorem 5.1 ***(Cauchy's Inequality)*** *Let $\mathbf{a}, \mathbf{b} \in \mathsf{R}^n$ with $\mathbf{a} = a_1, a_2, \ldots, a_n$ and $\mathbf{b} = b_1, b_2, \ldots, b_n$. We have*

$$\left|\sum_{i=1}^{n} a_i b_i\right| \leq \|\mathbf{a}\|_2 \cdot \|\mathbf{b}\|_2$$

with equality if and only if $\mathbf{a}$ and $\mathbf{b}$ are proportional (i.e., if there exists some non-zero real value k such that $a_i = k \cdot b_i$ for all $i \in \{1, 2, \ldots, n\}$.

Proof. We have

$$\begin{aligned} 2\|\mathbf{a}\|_2^2 \cdot \|\mathbf{b}\|_2^2 - 2\left|\sum_{i=1}^{n} a_i b_i\right|^2 &= \sum_{i,j} a_i^2 b_j^2 + \sum_{i,j} a_j^2 b_i^2 - 2\sum_{i,j}^{n} a_i b_i a_j b_j \\ &= \sum_{i,j} (a_i b_j - a_j b_i)^2 \geq 0 \end{aligned}$$

with equality if and only if $\mathbf{a}$ and $\mathbf{b}$ are proportional. □

Corollary 5.1 *Let $\mathbf{z} = (z_1, z_2, \ldots, z_n) \in \mathsf{R}^n$ and let r be a positive integer. We have*

$$\|z\|_r \leq n^{\frac{1}{2r}} \cdot \|z\|_{2r}$$

with equality when $z_i = z_j$ for all $i, j = 1, 2, \ldots, n$.

Proof. Using Cauchy's inequality with $a_i = |z_i|^r$ and $b_i = 1$ allows us to conclude the proof. □

5.3 Asymptotic Notations

Most of the results obtained in Part II concern the asymptotic behaviours of different types of distinguishers. We introduce in this section the main notation used to express most of our results, together with its basics properties.

Definition 5.2 *The fact that two strictly positive sequences $(a_q)_{q\in\mathbb{N}}$ and $(b_q)_{q\in\mathbb{N}}$ are equal to the first order in the exponent, i.e., are such that*

$$\lim_{q\to\infty} \frac{1}{q} \log \frac{a_q}{b_q} = 0,$$

is denoted $a_q \doteq b_q$.

The fact that $a_q \doteq b_q$ is equivalent to $a_q = b_q e^{o(q)}$. This notation is *multiplicative*, in the sense that if a_q, b_q, c_q, d_q are strictly positive sequences such that $a_q \doteq b_q$ and $c_q \doteq d_q$ then we have $a_q c_q \doteq b_q d_q$.

Lemma 5.1 *Let $0 < \alpha \leq \beta$ and let $a_q \doteq 2^{-\alpha q}$ and $b_q \doteq 2^{-\beta q}$. Then $\lim_{q\to\infty} a_q = \lim_{q\to\infty} b_q = 0$. Moreover,*

$$\frac{b_q}{a_q} \doteq 2^{-(\beta-\alpha)q} \quad \textit{and} \quad a_q + b_q \doteq 2^{-\alpha q}.$$

Proof. We have

$$a_q = 2^{-\alpha q + o(q)} \xrightarrow{q\to\infty} 0.$$

The fact that $\frac{b_q}{a_q} \doteq 2^{-(\beta-\alpha)q}$ comes from multiplicativity. The last result comes from

$$a_q + b_q = a_q \left(1 + \frac{b_q}{a_q}\right),$$

where $a_q \doteq 2^{-\alpha q}$ so that showing that $1 + \frac{b_q}{a_q} \doteq 1$ would suffice to conclude by multiplicativity. When $\alpha = \beta$ we have $1 + \frac{b_q}{a_q} = 2 \doteq 1$. When $\beta > \alpha$ we let $\gamma = \beta - \alpha > 0$ and have $\frac{b_q}{a_q} \doteq 2^{-\gamma q}$, so that $\frac{b_q}{a_q} \to 0$, which easily leads to $1 + \frac{b_q}{a_q} \doteq 1$. □

Part II

On the (In)Security of Block Ciphers: Tools for Security Analysis

Chapter 6

Distinguishers Between Two Sources

6.1 A Typical Introduction to Simple Hypothesis Testing

In this section, we consider a simple game between an oracle, called the *source*, generating independent and identically-distributed (i.i.d.) random values in some given finite set, and an algorithm, called the *distinguisher*, that aims at determining the distribution followed by the source.

More precisely, let $\mathcal{Z}$ be a finite set and let P_0 and P_1 be two probability distributions over $\mathcal{Z}$. Consider an oracle S, the source, which generates q samples according to a distribution $\mathsf{P} \in \{\mathsf{P}_0, \mathsf{P}_1\}$. We denote the values of the sample members by $Z_1, Z_2, \ldots, Z_q$, the Z_i's being i.i.d. random variables following distribution P. These values are the inputs of an algorithm A_q, the distinguisher, the objective of which is to guess whether $\mathsf{P} = \mathsf{P}_0$ (hypothesis H_0, often referred to as the *null hypothesis*) or $\mathsf{P} = \mathsf{P}_1$ (hypothesis H_1, often referred to as the *alternate hypothesis*) on the basis of these q values (and of the knowledge of both P_0 and P_1). A_q is called a *q-limited distinguisher* as we assume that it is computationally unbounded and only limited by the sample size. This algorithm eventually outputs 0 (respectively 1) if its guess is that H_0 (respectively H_1) holds. The distinguisher A_q can be defined by an *acceptance region* $\mathcal{A}_q \subset \mathcal{Z}^q$ such that A_q outputs 1 when $(Z_1, Z_2, \ldots, Z_q) \in \mathcal{A}_q$ and 0 otherwise. Finally, we note that since the distinguisher is computationally unbounded (and only restricted by the sample size q), we can assume without loss of generality that it is fully *deterministic*.

The situation just described is commonly referred to as the *simple hypothesis testing problem* since both alternatives fully determine the distribution. A more complex situation arises when one of the two hypotheses is *composite*, i.e., when the distinguisher has to guess whether the distribution followed by the source is one particular distribution ($\mathsf{H}_0 : \mathsf{P} = \mathsf{P}_0$) or if it belongs to a set of several distributions ($\mathsf{H}_1 : \mathsf{P} \in \{\mathsf{P}_1, \ldots, \mathsf{P}_d\}$, where $\mathsf{P}_i \neq \mathsf{P}_0$ for $i = 1, \ldots, d$), which corresponds to the composite hypothesis. Finally, the difficulty of the game can be increased from the point of view of the distinguisher if the exact description of the composite hypothesis is not available. In that case, it shall guess whether the source follows a specific (known) distribution ($\mathsf{H}_0 : \mathsf{P} = \mathsf{P}_0$) or not

```
$b \xleftarrow{\mathsf{u}} \{0,1\}$                    /* Random choice between $\mathsf{P}_0$ and $\mathsf{P}_1$ */
view $\leftarrow \{\mathsf{P}_0, \mathsf{P}_1\}$
for $i = 1, \ldots, q$ do
    $Z_i \xleftarrow{\mathsf{P}_b} \mathcal{Z}$
    view $\leftarrow$ view $\cup \{Z_i\}$
end
$\widehat{b} \leftarrow \mathsf{A}_q$(view)
if $\widehat{b} = b$ then return 1 else return 0
```

Algorithm 6.1: Game played by a q-limited distinguisher A_q between two probability distributions P_0 and P_1 over a finite set $\mathcal{Z}$.

$(\mathsf{H}_1 : \mathsf{P} \neq \mathsf{P}_0)$.

Definition 6.1 *Let $\mathbf{Z}^q = Z_1, Z_2, \ldots, Z_q$ be q i.i.d. random variables sampled in a finite set $\mathcal{Z}$ according to a distribution P. Let H_0 and H_1 be two hypotheses on P such that one is true. A q-limited distinguisher A_q between H_0 and H_1 is an algorithm which (at least) takes as an input the q samples and eventually outputs a bit $b \in \{0,1\}$ to indicate that its guess is that H_b is true.*

In all cases, the ability to distinguish the null hypothesis H_0 from the alternate hypothesis H_1 is called the *advantage* of the distinguisher. We will first give a general definition of this notion and detail the particular cases that we will consider in this chapter.

Definition 6.2 *Let $\mathbf{Z}^q = Z_1, Z_2, \ldots, Z_q$ be q i.i.d. random variables sampled in a finite set $\mathcal{Z}$ according to a distribution P. Let H_0 and H_1 be two hypotheses on P. The advantage of a q-limited distinguisher A_q between H_0 and H_1 is defined by*

$$\mathrm{Adv}_{\mathsf{A}_q}(\mathsf{H}_0, \mathsf{H}_1) = \left|\Pr_{\mathsf{H}_0}[\mathsf{A}_q(\mathbf{Z}^q) = 1] - \Pr_{\mathsf{H}_1}[\mathsf{A}_q(\mathbf{Z}^q) = 1]\right| .$$

In the *simple hypothesis testing problem*, we consider two distributions P_0 and P_1 and try to distinguish between $\mathsf{H}_0 : \mathsf{P} = \mathsf{P}_0$ and $\mathsf{H}_1 : \mathsf{P} = \mathsf{P}_1$. In that specific case, we also denote the advantage of a q-limited distinguisher A_q by $\mathrm{Adv}_{\mathsf{A}_q}(\mathsf{P}_0, \mathsf{P}_1)$ instead of $\mathrm{Adv}_{\mathsf{A}_q}(\mathsf{H}_0, \mathsf{H}_1)$, and refer to A_q as a distinguisher between P_0 and P_1. One can formalize this simple scenario as a *game* played by the distinguisher, described in Algorithm 6.1. In such a case, the advantage of the distinguisher can be defined in a different (yet equivalent) way.

Definition 6.3 ***(alternative when both hypotheses are simple)*** *Let P_0 and P_1 be two probability distributions over a finite set $\mathcal{Z}$. Let A_q be a q-limited distinguisher* *between* P_0 *and* P_1 *playing the game described in Algorithm* 6.*1 and let* B *denote the event that the algorithm outputs 1. The advantage of* A_q *is*

$$|2\Pr[B]-1|\,,$$

where the probabilities hold over the random coins of the game.

Proposition 6.1 *When both hypotheses are simple, definitions 6.2 and 6.3 are equivalent.*

Proof. Using the notations of Algorithm 6.1 we have

$$\begin{aligned}\Pr[P] &= \Pr[\widehat{b}=b]\\ &= \Pr[\widehat{b}=b|b=0]\Pr[b=0]+\Pr[\widehat{b}=b|b=1]\Pr[b=1]\\ &= \tfrac{1}{2}(\Pr[\widehat{b}=b|b=0]+\Pr[\widehat{b}=b|b=1])\\ &= \tfrac{1}{2}(\Pr[\widehat{b}=0|b=0]+\Pr[\widehat{b}=1|b=1])\end{aligned}$$

where the third equality comes from the fact that b is uniformly distributed. Thus $|2\Pr[P]-1|$ is equal to $\left|\Pr[\widehat{b}=1|b=0]-\Pr[\widehat{b}=1|b=1]\right|$, which is how the advantage in Definition 6.2 should be written when using the notations of the algorithm. □

In the general case, the ability of A_q to distinguish between both hypotheses can be expressed in another (equivalent) way. Obviously, A_q can make two kinds of mistakes:

- it can either reject H_0 whereas it is true, which is often called a type I error,
- or reject H_1 whereas it is true, which is often called a type II error.

By definition, a type I error occurs with a probability $\Pr_{\mathsf{H}_0}[\mathsf{A}(\mathbf{Z}^q)=1]$ and a type II error with probability $\Pr_{\mathsf{H}_1}[\mathsf{A}(\mathbf{Z}^q)=0]$. Respectively denoting these probabilities by α and β, one can define an *overall probability of error* $\mathrm{P_e}$ such that

$$\mathrm{P_e}=\frac{1}{2}(\alpha+\beta).$$

It is easy to see that the advantage of a distinguisher is related to its overall probability of error by

$$\mathrm{Adv}_{\mathsf{A}_q}(\mathsf{H}_0,\mathsf{H}_1)=|1-2\mathrm{P_e}|\,.$$

6.2 An Alternate View through the Method of Types

Since the q samples $Z_1,Z_2,\ldots,Z_q$ generated by the source are independent, their particular *order* must be irrelevant. On the other hand, what matters is the number of occurrences of each symbol of $\mathcal{Z}$ in the string $Z_1,Z_2,\ldots,Z_q$ or, equivalently, the *relative* number of occurrence (frequency) of each symbol. We introduce here the notion of *type* (or empirical probability distribution) of an i.i.d. sequence, and show in Lemma 6.1 that the probability of occurrence of a given string is uniquely determined by its type. Finally, we recall a fundamental theorem due to Sanov [133], that we will later use to compute the asymptotic complexity of two distinguishers (namely the best distinguisher and the χ^2 distinguisher).

Definition 6.4 *Let $\mathcal{Z}$ be a finite set. The* type *(or* empirical probability distribution*) $\mathsf{P}_{\mathbf{z}^q}$ of a sequence $\mathbf{z}^q \in \mathcal{Z}^q$ is the relative proportion of occurrences of each symbol of $\mathcal{Z}$, i.e.,*

$$\mathsf{P}_{\mathbf{z}^q}[a] = \frac{\mathrm{N}(a|\mathbf{z}^q)}{q}$$

for all $a \in \mathcal{Z}$, where $\mathrm{N}(a|\mathbf{z}^q)$ is the number of times the symbol a occurs in the sequence $\mathbf{z}^q$.

Definition 6.5 *Given a finite set $\mathcal{Z}$, we denote by $\mathcal{P}(\mathcal{Z})$ (or simply by $\mathcal{P}$) the set of all probability distributions defined over the finite set $\mathcal{Z}$, and by $\mathcal{P}_q(\mathcal{Z})$ (or simply by $\mathcal{P}_q$) the set of probability distributions over $\mathcal{Z}$ in which probabilities are integral fractions of q. Finally, we let $\mathcal{P}_\infty(\mathcal{Z}) = \cup_{q>0}\mathcal{P}_q(\mathcal{Z})$ (or simply $\mathcal{P}_\infty$) be the set of rational distributions.*

Definition 6.6 *For $\mathsf{P} \in \mathcal{P}$, we denote by $T_q(\mathsf{P})$ the set of sequences of length q whose type is equal to P, i.e.,*

$$T_q(\mathsf{P}) = \{\mathbf{z}^q \in \mathcal{Z}^q \ : \ \mathsf{P}_{\mathbf{z}^q} = \mathsf{P}\}.$$

$T_q(\mathsf{P})$ is usually called the *type class* of P and is more commonly defined over $\mathcal{P}_q$ (see [37, p.280]). Here we extend the definition to $\mathcal{P}$ and simply have $T_q(\mathsf{P}) = \emptyset$ when $\mathsf{P} \in \mathcal{P} \setminus \mathcal{P}_q$.

Definition 6.7 *Let P_0 and P_1 be two probability distributions over a finite set $\mathcal{Z}$. The* relative entropy *or* Kullback-Leibler distance *between P_0 and P_1 is defined as*

$$\mathrm{D}(\mathsf{P}_0\|\mathsf{P}_1) = \sum_{z\in\mathcal{Z}} \mathsf{P}_0[z]\log\frac{\mathsf{P}_0[z]}{\mathsf{P}_1[z]} = \sum_{z\in\mathsf{supp}(\mathsf{P}_0)} \mathsf{P}_0[z]\log\frac{\mathsf{P}_0[z]}{\mathsf{P}_1[z]}$$

with the convention that $0\log\frac{0}{p} = 0$, that $p\log\frac{p}{0} = +\infty$ for $p > 0$, and that $0\log\frac{0}{0} = 0$.

We note that $\mathrm{D}(\mathsf{P}_0\|\mathsf{P}_1) < +\infty$ if and only if the support of P_0 is included in the support of P_1. In what follows, we will refer to this notion using the term *relative* *entropy* as, being non-symmetric, it is not exactly a distance. Nevertheless, it is always positive since $-\log$ is convex.

Lemma 6.1 *Let $\mathcal{Z}$ be a finite set and P be a probability distribution over $\mathcal{Z}$. For all $z_1, z_2, \ldots, z_q \in \mathcal{Z}$ we have*

$$\mathsf{P}^q[\mathbf{z}^q] = 2^{-q(H(\mathsf{P}_{\mathbf{z}^q})+\mathrm{D}(\mathsf{P}_{\mathbf{z}^q}\|\mathsf{P}))}.$$

Proof. We have

$$\mathsf{P}^q[\mathbf{z}^q] = \prod_{i=1}^{q} \mathsf{P}[z_i] = \prod_{a\in\mathcal{Z}} \mathsf{P}[a]^{\mathrm{N}(a\|\mathbf{z}^q)} = \prod_{a\in\mathcal{Z}} \mathsf{P}[a]^{q\mathsf{P}_{\mathbf{z}^q}[a]}.$$

As for each $a \in \mathcal{Z}$ we have

$$\mathsf{P}[a]^{q\mathsf{P}_{\mathbf{z}^q}[a]} = 2^{q\mathsf{P}_{\mathbf{z}^q}[a]\log\mathsf{P}[a]} = 2^{q\left(\mathsf{P}_{\mathbf{z}^q}[a]\log\mathsf{P}_{\mathbf{z}^q}[a]-\mathsf{P}_{\mathbf{z}^q}[a]\log\frac{\mathsf{P}_{\mathbf{z}^q}[a]}{\mathsf{P}[a]}\right)},$$

we obtain that

$$\mathsf{P}^q[\mathbf{z}^q] = 2^{q\left(\sum_a \mathsf{P}_{\mathbf{z}^q}[a]\log\mathsf{P}_{\mathbf{z}^q}[a]-\sum_a \mathsf{P}_{\mathbf{z}^q}[a]\log\frac{\mathsf{P}_{\mathbf{z}^q}[a]}{\mathsf{P}[a]}\right)} = 2^{-q(H(\mathsf{P}_{\mathbf{z}^q})+\mathrm{D}(\mathsf{P}_{\mathbf{z}^q}\|\mathsf{P}))}.$$

□

According to Lemma 6.1, the probability of occurrence of a particular sequence only depends on its type. As a consequence, from now on we will make the following assumption about the acceptance region of the distinguishers we will consider.

Assumption 6.1 Let $\mathcal{A}_q \in \mathcal{Z}^q$ be the acceptance region of a q-limited distinguisher. If $\mathbf{z}^q$ belongs to $\mathcal{A}_q$, then it is also the case for all the strings of $T_q(\mathsf{P}_{\mathbf{z}^q})$, i.e.,

$$\mathbf{z}^q \in \mathcal{A}_q \;\Leftrightarrow\; T_q(\mathsf{P}_{\mathbf{z}^q}) \subset \mathcal{A}_q.$$

Under the previous assumption, for each acceptance region $\mathcal{A}_q \subset \mathcal{Z}^q$ of a q-limited distinguisher, there exists a subset $\Pi_q \subset \mathcal{P}_q(\mathcal{Z})$ such that

$$P \in \Pi_q \;\Leftrightarrow\; T_q(P) \subset \mathcal{A}_q.$$

The advantage of the distinguisher can then be written as

$$\begin{aligned}
\mathrm{Adv}_{\mathsf{A}_q}(\mathsf{P}_0,\mathsf{P}_1) &= \left|\sum_{\mathbf{z}^q\in\mathcal{A}_q}\mathsf{P}_0^q[\mathbf{z}^q] - \sum_{\mathbf{z}^q\in\mathcal{A}_q}\mathsf{P}_1^q[\mathbf{z}^q]\right| \\
&= \left|\sum_{\mathbf{z}^q\in\mathcal{A}_q} 2^{-q(H(\mathsf{P}_{\mathbf{z}^q})+\mathrm{D}(\mathsf{P}_{\mathbf{z}^q}\|\mathsf{P}_0))} - \sum_{\mathbf{z}^q\in\mathcal{A}_q} 2^{-q(H(\mathsf{P}_{\mathbf{z}^q})+\mathrm{D}(\mathsf{P}_{\mathbf{z}^q}\|\mathsf{P}_1))}\right| \\
&= \left|\sum_{P\in\Pi_q} |T_q(P)|\,2^{-q(H(P)+\mathrm{D}(P\|\mathsf{P}_0))} - \sum_{P\in\Pi_q} |T_q(P)|\,2^{-q(H(P)+\mathrm{D}(P\|\mathsf{P}_1))}\right|.
\end{aligned}$$

We summarize the assumptions and some of the results obtained in this section in the following definition.

Definition 6.8 *(Acceptance Region) Let $\mathcal{Z}$ be a finite set. The* sample acceptance region *of a q-limited distinguisher A_q is a subset $\mathcal{A} \subset \mathcal{Z}^q$ such that for any string $\mathbf{z}^q$ of q symbols of $\mathcal{Z}$ we have*

$$\mathsf{A}_q(\mathbf{z}^q) = 1 \quad \Leftrightarrow \quad \mathbf{z}^q \in \mathcal{A}.$$

The type acceptance region *of A_q is a subset $\Pi_q \subset \mathcal{P}_q(\mathcal{Z})$ such that for any distribution $P \in \mathcal{P}_q(\mathcal{Z})$ we have*

$$P \in \Pi_q \quad \Leftrightarrow \quad T_q(P) \subset \mathcal{A}.$$

When clear from the context, we might simply call *acceptance region* either sample or type acceptance regions.

To conclude this section, we recall a fundamental result that will be the basis of the data complexity analysis of specific distinguishers, such as the best distinguisher in subsection 6.4. We first note that, from a topological viewpoint, $\mathcal{P}$ is a compact subset of the vector space $\mathbf{R}^{|\mathcal{Z}|}$ which is of finite dimension. Its topology can therefore by defined by any norm, e.g., the infinite norm $\|P\|_\infty = \max_z |P(z)|$.

Theorem 6.1 *(Sanov's theorem [133]) Let P be a probability distribution over a finite set $\mathcal{Z}$, $\mathcal{Z}'$ be a non-empty subset of $\mathcal{Z}$, and Π be a set of probability distributions of full support over $\mathcal{Z}'$. If $Z_1, Z_2, \ldots, Z_q$ are q i.i.d. random variables drawn according to the distribution P, we have*

$$\Pr[\mathsf{P}_{\mathbf{Z}^q} \in \Pi] \leq (q+1)^{|\mathcal{Z}|} 2^{-q\mathrm{D}(\Pi\|\mathsf{P})},$$

where $\mathrm{D}(\Pi\|\mathsf{P}) = \inf_{\mathsf{P}' \in \Pi} \mathrm{D}(\mathsf{P}'\|\mathsf{P})$. Moreover, if the closure of $\Pi \subset \mathcal{P}(\mathcal{Z}')$ is equal to the closure of its interior, i.e., if $\overline{\Pi} = \overline{\overset{\circ}{\Pi}}$ under the topology of probability distributions over $\mathcal{Z}'$, then

$$\Pr[\mathsf{P}_{\mathbf{Z}^q} \in \Pi] \doteq 2^{-q\mathrm{D}(\Pi\|\mathsf{P})}.$$

Proof. A proof of this result is given in Appendix A. □

6.3 The Best Distinguisher: an Optimal Solution

Since $\mathcal{P}_q$ is finite for any given fixed q, this is also the case for the number of sample acceptance regions of q-limited distinguishers. Consequently, there exists one distinguisher $\mathsf{A}_q^\star$ which maximizes the advantage, i.e., such that for all distinguisher A_q,

$$\mathrm{Adv}_{\mathsf{A}_q^\star}(\mathsf{P}_0, \mathsf{P}_1) \geq \mathrm{Adv}_{\mathsf{A}_q}(\mathsf{P}_0, \mathsf{P}_1).$$

In what follows, we denote by $\mathrm{BestAdv}^q(\mathsf{P}_0, \mathsf{P}_1)$ the advantage of $\mathsf{A}_q^\star$, i.e.,

$$\mathrm{BestAdv}^q(\mathsf{P}_0, \mathsf{P}_1) = \max_{\mathsf{A}_q} \mathrm{Adv}_{\mathsf{A}_q}(\mathsf{P}_0, \mathsf{P}_1),$$

where the maximum is taken over all q-limited distinguishers. It is possible to give an explicit description of $\mathsf{A}_q^\star$ by deriving an adequate acceptance region[1].

Definition 6.9 *Let* P_0 *and* P_1 *be two probability distributions over a finite set* $\mathcal{Z}$. *Let* q *be a positive integer. The function*

$$\begin{array}{rcl} \mathrm{LR} : \mathcal{Z}^q & \longrightarrow & \mathbf{R}^+ \cup \{+\infty, \mathtt{NaN}\} \\ \mathbf{z}^q & \longmapsto & \mathrm{LR}(\mathbf{z}^q) = \frac{\mathsf{P}_0^q[\mathbf{z}^q]}{\mathsf{P}_1^q[\mathbf{z}^q]} \end{array}$$

is the q-limited likelihood ratio *(or simply likelihood ratio), with the convention that* $\frac{p}{0} = +\infty$ *for* $p > 0$ *and that* $\frac{0}{0} = \mathtt{NaN}$. *The function*

$$\begin{array}{rcl} \mathrm{LLR} : \mathcal{Z}^q & \longrightarrow & \mathbf{R} \cup \{-\infty, +\infty, \mathtt{NaN}\} \\ \mathbf{z}^q & \longmapsto & \mathrm{LLR}(\mathbf{z}^q) = \log\left(\mathrm{LR}(\mathbf{z}^q)\right) \end{array}$$

is the q-limited logarithmic likelihood ratio *(or simply logarithmic likelihood ratio), with the convention that* $\log \frac{0}{p} = -\infty$, *that* $\log \frac{p}{0} = +\infty$ *for* $p > 0$, *and that* $\log \frac{0}{0} = \mathtt{NaN}$.

Proposition 6.2 *Let* P_0 *and* P_1 *be two probability distributions over a finite set* $\mathcal{Z}$. *The q-limited distinguisher* $\mathsf{A}_q^\star$ *defined by the sample acceptance region*

$$\mathcal{A}_q^\star = \{\mathbf{z}^q \in \mathcal{Z}^q \ : \ \mathrm{LR}(\mathbf{z}^q) \leq 1\} = \{\mathbf{z}^q \in \mathcal{Z}^q \ : \ \mathrm{LLR}(\mathbf{z}^q) \leq 0\}$$

is optimal in the sense that its advantage is $\mathrm{BestAdv}^q(\mathsf{P}_0, \mathsf{P}_1)$.

Proof. To prove this proposition, we show that an arbitrary distinguisher has a smaller advantage than $\mathsf{A}_q^\star$. Let A_q be an arbitrary q-limited distinguisher and $\mathcal{A}_q$ be its acceptance region. Without loss of generality, we can assume that $\Pr_{\mathsf{H}_1}[\mathsf{A}_q(\mathbf{Z}^q) = 1] \geq \Pr_{\mathsf{H}_0}[\mathsf{A}_q(\mathbf{Z}^q) = 1]$ holds[2]. By definition we thus have

$$\mathrm{Adv}_{\mathsf{A}_q}(\mathsf{P}_0, \mathsf{P}_1) = \Pr_{\mathsf{H}_1}[\mathsf{A}_q(\mathbf{Z}^q) = 1] - \Pr_{\mathsf{H}_0}[\mathsf{A}_q(\mathbf{Z}^q) = 1].$$

Since the distinguisher outputs 1 if the sample $\mathbf{z}^q$ that it receives belongs to its acceptance region $\mathcal{A}_q$, we have

$$\Pr_{\mathsf{H}_i}[\mathsf{A}_q(\mathbf{Z}^q) = 1] = \sum_{\mathbf{z}^q \in \mathcal{A}_q} \mathsf{P}_i^q[\mathbf{z}^q]$$

[1] The method described here is similar to the proof of the Neyman-Pearson lemma [123] that was used by Junod and Vaudenay in order to derive optimal key ranking procedures for block cipher cryptanalysis [75].

[2] Otherwise we replace $\mathcal{A}_q$ by $\mathcal{A}_q^c$, and obtain a distinguisher with the exact same advantage as A_q but such that the inequality is true.

for $i \in \{0,1\}$. Thus,

$$\mathrm{Adv}_{\mathsf{A}_q}(\mathsf{P}_0,\mathsf{P}_1) = \sum_{\mathbf{z}^q\in\mathcal{A}_q} \left(\mathsf{P}_1^q[\mathbf{z}^q] - \mathsf{P}_0^q[\mathbf{z}^q]\right) \leq \sum_{\mathbf{z}^q\in\mathcal{A}_q^+} \left(\mathsf{P}_1^q[\mathbf{z}^q] - \mathsf{P}_0^q[\mathbf{z}^q]\right),$$

where $\mathcal{A}_q^+ \subset \mathcal{A}_q$ is the set of all $\mathbf{z}^q \in \mathcal{A}_q$ such that $\mathsf{P}_1^q[\mathbf{z}^q] - \mathsf{P}_0^q[\mathbf{z}^q] \geq 0$. Since $\mathcal{A}_q^\star$ is by definition the set of all $\mathbf{z}^q \in \mathcal{Z}^q$ such that $\mathsf{P}_1^q[\mathbf{z}^q] - \mathsf{P}_0^q[\mathbf{z}^q] \geq 0$, we have $\mathcal{A}_q^+ \subset \mathcal{A}_q^\star$ and

$$\sum_{\mathbf{z}^q\in\mathcal{A}_q^+} \left(\mathsf{P}_1^q[\mathbf{z}^q] - \mathsf{P}_0^q[\mathbf{z}^q]\right) \leq \sum_{\mathbf{z}^q\in\mathcal{A}_q^\star} \left(\mathsf{P}_1^q[\mathbf{z}^q] - \mathsf{P}_0^q[\mathbf{z}^q]\right) = \mathrm{Adv}_{\mathsf{A}_q^\star}(\mathsf{P}_0,\mathsf{P}_1),$$

which allows to conclude that $\mathrm{Adv}_{\mathsf{A}_q}(\mathsf{P}_0,\mathsf{P}_1) \leq \mathrm{Adv}_{\mathsf{A}_q^\star}(\mathsf{P}_0,\mathsf{P}_1)$. □

We note that if $\mathrm{supp}(\mathsf{P}_0) \neq \mathrm{supp}(\mathsf{P}_1)$, the LLR might become infinite. However, it never occurs that $\mathrm{LLR}(\mathbf{z}^q) = \mathtt{NaN}$ in practice.

We conclude this subsection with a link between the advantage of the best distinguisher and (simplified) distribution vectors from Vaudenay's Decorrelation Theory [155]. Given the sample size q, let $[\mathsf{P}_0]^q$ and $[\mathsf{P}_1]^q$ be the vectors defined by

$$[\mathsf{P}_j]^q_{z_1,\ldots,z_q} = \mathsf{P}_j^q[\mathbf{z}^q] \quad \text{for} \quad j \in \{0,1\}.$$

Using the notations of Proposition 6.2, the probability that $\mathsf{A}_q^\star$ outputs 1 in the case where the source follows distribution P_j is $\sum_{\mathbf{z}^q\in\mathcal{A}_q^\star}[\mathsf{P}_j]^q_{\mathbf{z}^q}$ for $j \in \{0,1\}$. From Definition 6.2 and from the definition of $\mathcal{A}_q^\star$ we have

$$\mathrm{BestAdv}^q(\mathsf{P}_0,\mathsf{P}_1) = \left|\sum_{\mathbf{z}^q\in\mathcal{A}_q^\star}\left([\mathsf{P}_0]^q_{\mathbf{z}^q} - [\mathsf{P}_1]^q_{\mathbf{z}^q}\right)\right| = \sum_{\mathbf{z}^q\in\mathcal{A}_q^\star}\left([\mathsf{P}_1]^q_{\mathbf{z}^q} - [\mathsf{P}_0]^q_{\mathbf{z}^q}\right).$$

Since

$$\sum_{\mathbf{z}^q\in\mathcal{A}_q^\star}\left([\mathsf{P}_1]^q_{\mathbf{z}^q} - [\mathsf{P}_0]^q_{\mathbf{z}^q}\right) + \sum_{\mathbf{z}^q\in\overline{\mathcal{A}_q^\star}}\left([\mathsf{P}_1]^q_{\mathbf{z}^q} - [\mathsf{P}_0]^q_{\mathbf{z}^q}\right) = \sum_{\mathbf{z}^q\in\mathcal{Z}^q}[\mathsf{P}_1]^q_{\mathbf{z}^q} - \sum_{\mathbf{z}^q\in\mathcal{Z}^q}[\mathsf{P}_0]^q_{\mathbf{z}^q} = 0,$$

this gives

$$\begin{aligned}2\mathrm{BestAdv}^q(\mathsf{P}_0,\mathsf{P}_1) &= \sum_{\mathbf{z}^q\in\mathcal{A}_q^\star}\left([\mathsf{P}_1]^q_{\mathbf{z}^q} - [\mathsf{P}_0]^q_{\mathbf{z}^q}\right) - \sum_{\mathbf{z}^q\in\overline{\mathcal{A}_q^\star}}\left([\mathsf{P}_1]^q_{\mathbf{z}^q} - [\mathsf{P}_0]^q_{\mathbf{z}^q}\right)\\ &= \sum_{\mathbf{z}^q\in\mathcal{A}_q^\star}\left|[\mathsf{P}_1]^q_{\mathbf{z}^q} - [\mathsf{P}_0]^q_{\mathbf{z}^q}\right| + \sum_{\mathbf{z}^q\in\overline{\mathcal{A}_q^\star}}\left|[\mathsf{P}_1]^q_{\mathbf{z}^q} - [\mathsf{P}_0]^q_{\mathbf{z}^q}\right|\\ &= \sum_{\mathbf{z}^q\in\mathcal{Z}^q}\left|[\mathsf{P}_1]^q_{\mathbf{z}^q} - [\mathsf{P}_0]^q_{\mathbf{z}^q}\right|.\end{aligned}$$

We summarize this result in the following proposition.

Proposition 6.3 *Let P_0 and P_1 be two probability distributions over a finite set $\mathcal{Z}$. The advantage of the best q-limited distinguisher between P_0 and P_1 is*

$$\mathrm{BestAdv}^q(\mathsf{P}_0,\mathsf{P}_1) = \frac{1}{2}\left\|[\mathsf{P}_0]^q - [\mathsf{P}_1]^q\right\|_1,$$

where the $\|\cdot\|_1$ *norm of a vector* $\mathbf{A} = \{A_i\}_i$ *is defined by* $\|\mathbf{A}\|_1 = \sum_i |A_i|$.

The statistical framework proposed in [35] by Coppersmith et al. is based on this norm.

6.4 The Best Distinguisher: Data Complexity Analysis

In the previous section we showed how to compute the advantage of the best q-limited distinguisher for some given q. Here we consider the case where the advantage is fixed and wonder how large the sample size q must be, so that the best distinguisher achieves this advantage.

We first note that the logarithmic likelihood ratio can be expressed in terms of the relative entropy and of the type of the sample received by the distinguisher [37].

Lemma 6.2 *Let* $\mathcal{Z}$ *be a finite set and* $z_1, \ldots, z_q$ *be a sequence of* q *elements in* $\mathcal{Z}$. *If* $\mathrm{LLR}(\mathbf{z}^q) = \texttt{NaN}$ *we have* $\mathrm{supp}(\mathsf{P}_{\mathbf{z}^q}) \not\subset \mathrm{supp}(\mathsf{P}_0)$ *and* $\mathrm{supp}(\mathsf{P}_{\mathbf{z}^q}) \not\subset \mathrm{supp}(\mathsf{P}_1)$. *Otherwise,*

$$\mathrm{LLR}(\mathbf{z}^q) = q\left(\mathrm{D}(\mathsf{P}_{\mathbf{z}^q}\|\mathsf{P}_1) - \mathrm{D}(\mathsf{P}_{\mathbf{z}^q}\|\mathsf{P}_0)\right).$$

Proof. If $\mathrm{LLR}(\mathbf{z}^q) = -\infty$ it means that $\mathrm{P}_1^q[\mathbf{z}^q] > 0$ and $\mathrm{P}_0^q[\mathbf{z}^q] = 0$, so that $\mathrm{supp}(\mathsf{P}_{\mathbf{z}^q}) \subset \mathrm{supp}(\mathsf{P}_1)$ and $\mathrm{supp}(\mathsf{P}_{\mathbf{z}^q}) \not\subset \mathrm{supp}(\mathsf{P}_0)$. As a consequence, $\mathrm{D}(\mathsf{P}_{\mathbf{z}^q}\|\mathsf{P}_1)$ is finite while $\mathrm{D}(\mathsf{P}_{\mathbf{z}^q}\|\mathsf{P}_0) = +\infty$, so that we indeed have

$$q\left(\mathrm{D}(\mathsf{P}_{\mathbf{z}^q}\|\mathsf{P}_1) - \mathrm{D}(\mathsf{P}_{\mathbf{z}^q}\|\mathsf{P}_0)\right) = -\infty,$$

and the equality is verified. The case where $\mathrm{LLR}(\mathbf{z}^q) = +\infty$ can be treated similarly. When $\mathrm{LLR}(\mathbf{z}^q)$ is finite, using the notations of Definition 6.4 we have

$$\mathrm{LLR}(\mathbf{z}^q) = \log \prod_{i=1}^{q} \frac{\mathsf{P}_0[z_i]}{\mathsf{P}_1[z_i]} = \sum_{i=1}^{q} \log \frac{\mathsf{P}_0[z_i]}{\mathsf{P}_1[z_i]} = \sum_{z\in\mathcal{Z}} \mathrm{N}(z|\mathbf{z}^q) \log \frac{\mathsf{P}_0[z]}{\mathsf{P}_1[z]}.$$

Introducing the empirical probability distribution of the sequence $\mathbf{z}^q$, we obtain

$$\begin{aligned}
\mathrm{LLR}(\mathbf{z}^q) &= q \sum_{z\in\mathcal{Z}} \mathsf{P}_{\mathbf{z}^q}[z] \log \frac{\mathsf{P}_0[z]}{\mathsf{P}_1[z]} \frac{\mathsf{P}_{\mathbf{z}^q}[z]}{\mathsf{P}_{\mathbf{z}^q}[z]} \\
&= q \sum_{z\in\mathcal{Z}} \left(\mathsf{P}_{\mathbf{z}^q}[z] \log \frac{\mathsf{P}_{\mathbf{z}^q}[z]}{\mathsf{P}_1[z]}\right) - q \sum_{z\in\mathcal{Z}} \left(\mathsf{P}_{\mathbf{z}^q}[z] \log \frac{\mathsf{P}_{\mathbf{z}^q}[z]}{\mathsf{P}_0[z]}\right) \\
&= q\left(\mathrm{D}(\mathsf{P}_{\mathbf{z}^q}\|\mathsf{P}_1) - \mathrm{D}(\mathsf{P}_{\mathbf{z}^q}\|\mathsf{P}_0)\right).
\end{aligned}$$

□

Following Proposition 6.2 and Lemma 6.2, we can easily derive the distribution acceptance region of the best distinguisher.

Proposition 6.4 *Let P_0 and P_1 be two probability distributions over a finite set $\mathcal{Z}$. Let*

$$\begin{array}{rccl} \mathrm{L}: & \mathcal{P} & \longrightarrow & \mathbf{R} \cup \{+\infty, -\infty, \mathtt{NaN}\} \\ & \mathsf{P} & \longmapsto & \mathrm{L}(\mathsf{P}) = \sum_{z \in \mathrm{supp}(\mathsf{P})} \mathsf{P}[z] \log \frac{\mathsf{P}_0[z]}{\mathsf{P}_1[z]} \end{array}$$

with the convention that $p \log \frac{q}{0} = +\infty$, $p \log \frac{0}{q} = -\infty$, and that $p \log \frac{0}{0} = \pm\infty \mp \infty = \mathtt{NaN}$ (for $p, q > 0$). Let

$$\Pi^\star = \{\mathsf{P} \in \mathcal{P} \ : \ \mathrm{D}(\mathsf{P}\|\mathsf{P}_1) - \mathrm{D}(\mathsf{P}\|\mathsf{P}_0) \leq 0\} = \{\mathsf{P} \in \mathcal{P} \ : \ \mathrm{L}(\mathsf{P}) \leq 0\}.$$

The q-limited distinguisher $\mathsf{A}_q^\star$ defined by the distribution acceptance region $\Pi_q^\star = \Pi^\star \cap \mathcal{P}_q$ is optimal in the sense that its advantage is $\mathrm{BestAdv}^q(\mathsf{P}_0, \mathsf{P}_1)$.

The previous proposition shows that the best distinguisher shall choose the distribution P_b which is the closest one (in the sense of the relative entropy) from the type of the sequence.

In practice, the output of L cannot be `NaN`. Indeed, this situation occurs when there exists z such that $\mathsf{P}[z] > 0$ and $\mathsf{P}_0[z] = \mathsf{P}_1[z] = 0$, or when there exists distinct $z, z' \in \mathrm{supp}(\mathsf{P})$ such that $\mathsf{P}_0[z] > 0$, $\mathsf{P}_1[z] = 0$, $\mathsf{P}_1[z'] > 0$, and $\mathsf{P}_0[z'] = 0$. Since L is evaluated in some $\mathsf{P}_{\mathbf{z}^q}$ where the z_i's where sampled according to either P_0 or P_1, it is always the case that $\mathrm{supp}(\mathsf{P}_{\mathbf{z}^q}) \subset \mathrm{supp}(\mathsf{P}_0)$ or that $\mathrm{supp}(\mathsf{P}_{\mathbf{z}^q}) \subset \mathrm{supp}(\mathsf{P}_1)$, so that neither of the two previous situations can occur.

Lemma 6.3 *The set $\Pi^\star$ defined in Proposition 6.4 is convex.*

Proof. Let $\mathsf{P}, \mathsf{P}' \in \Pi^\star$ and $t \in [0, 1]$. Since L is linear, we have

$$\mathrm{L}(t\mathsf{P} + (1-t)\mathsf{P}') = t\mathrm{L}(\mathsf{P}) + (1-t)\mathrm{L}(\mathsf{P}') \leq 0.$$

□

The following theorem (which actually is another way of formulating a result from Chernoff [34]) gives the number of samples the best distinguisher $\mathsf{A}_q^\star$ needs to achieve a *given* probability of error of type I (see page 33), in the case where both distributions are of full support.

Theorem 6.2 *Let P_0 and P_1 be two probability distributions of full support over a finite set $\mathcal{Z}$. Let $\mathrm{F} : [0, 1] \to \mathbf{R}$ be the function defined by*

$$\mathrm{F}(\lambda) = \sum_{z \in \mathcal{Z}} \mathsf{P}_0[z]^{1-\lambda} \mathsf{P}_1[z]^\lambda$$

and let

$$\mathrm{C}(\mathsf{P}_0, \mathsf{P}_1) = -\inf_{0<\lambda<1} \log \mathrm{F}(\lambda) = -\min_{0\leq\lambda\leq1} \log \mathrm{F}(\lambda)$$

be the Chernoff information between P_0 *and* P_1*. The probability of error of type I of the best* q*-limited distinguisher* $\mathsf{A}_q^\star$ *is such that*

$$\Pr_{\mathsf{H}_0}[\mathsf{A}_q^\star(\mathbf{Z}^q) = 1] \doteq 2^{-q\mathrm{C}(\mathsf{P}_0,\mathsf{P}_1)} = \left(\min_{\lambda\in[0,1]} \mathrm{F}(\lambda)\right)^q . \tag{6.1}$$

Proof. For $\mathsf{P}_0 = \mathsf{P}_1$ the result is trivial. We now assume that P_0 and P_1 are distinct which ensures that there exists $z \in \mathcal{Z}$ such that $0 < \mathsf{P}_0[z] < \mathsf{P}_1[z]$ and $z' \in \mathcal{Z}$ such that $z' \neq z$ and $0 < \mathsf{P}_1[z'] < \mathsf{P}_0[z']$.

According to Proposition 6.4, the best q-limited distinguisher $\mathsf{A}_q^\star$ is defined by an acceptance region $\Pi_q^\star = \Pi^\star \cap \mathcal{P}_q$ where

$$\Pi^\star = \{\mathsf{P} \in \mathcal{P} \,:\, \mathrm{D}(\mathsf{P}\|\mathsf{P}_1) - \mathrm{D}(\mathsf{P}\|\mathsf{P}_0) \leq 0\} = \{\mathsf{P} \in \mathcal{P} \,:\, \mathrm{L}(\mathsf{P}) \leq 0\},$$

using the notations of the proposition. If z is such that $0 < \mathsf{P}_0[z] < \mathsf{P}_1[z]$, the distribution $\mathsf{P} \in \mathcal{P}$ such that $\mathsf{P}[z] = 1$ belongs to $\Pi^\star$ since $\mathrm{L}(\mathsf{P}) = \log \frac{\mathsf{P}_0[z]}{\mathsf{P}_1[z]} < 0$. This ensures that $\Pi^\star$ is nonempty.

Considering the topology of $\mathcal{P}$ (as discussed on page 36), we note that L is continuous. Since there exists $\mathsf{P} \in \Pi^\star$ such that $\mathrm{L}(\mathsf{P}) < 0$, then, for a sufficiently small $\epsilon > 0$, all distributions within a distance to P smaller than ϵ are in $\Pi^\star$ as well. This means that the interior of $\Pi^\star$ is nonempty.

Since $\Pi^\star$ is a nonempty and convex set (see Lemma 6.3), we have $\overline{\overset{\circ}{\Pi^\star}} = \overline{\Pi^\star}$ so that we can apply Theorem 6.1 and obtain

$$\Pr_{\mathsf{H}_0}[\mathsf{A}_q^\star(\mathbf{Z}^q) = 1] \doteq 2^{-q\mathrm{D}(\Pi^\star\|\mathsf{P}_0)}.$$

We now show that $\mathrm{D}(\Pi^\star\|\mathsf{P}_0)$ is actually equal to $\mathrm{C}(\mathsf{P}_0, \mathsf{P}_1)$.

The set $\Pi^\star$ is *topologically closed*: Since both P_0 and P_1 are of full support, $\mathrm{L}(\mathsf{P}) \in \mathbf{R}$ for all $\mathsf{P} \in \mathcal{P}$. Consequently, $(\Pi^\star)^c = \{\mathsf{P} \in \mathcal{P} \,:\, \mathrm{L}(P) > 0\}$. It is easy to see that there exists $\mathsf{P} \in \mathcal{P}$ such that $\mathrm{L}(\mathsf{P}) > 0$ (so that $(\Pi^\star)^c$ is nonempty) and that, since L is continuous, all distributions at a sufficiently small distance of P are in $(\Pi^\star)^c$ as well. This shows that any $\mathsf{P} \in (\Pi^\star)^c$ is the center of an open ball included in $(\Pi^\star)^c$, which makes $(\Pi^\star)^c$ an open set and thus, $\Pi^\star$ is closed.

Since $\Pi^\star$ is closed and bounded in the Euclidean space $\mathbf{R}^{|\mathcal{Z}|}$, it is *compact*. Since $\mathsf{P} \mapsto \mathrm{D}(\mathsf{P}\|\mathsf{P}_0)$ is continuous, the extreme value theorem states $\mathrm{D}(\Pi^\star\|\mathsf{P}_0) = \mathrm{D}(\mathsf{P}\|\mathsf{P}_0)$ for some $\mathsf{P} \in \Pi^\star$: there exists a global minimum for this function in $\Pi^\star$. Furthermore, since the function $\mathsf{P} \mapsto \mathrm{D}(\mathsf{P}\|\mathsf{P}_0)$ is convex, the set of P's such that $\mathrm{D}(\mathsf{P}\|\mathsf{P}_0) < r$ is a convex set for any radius $r > 0$. As a consequence, there is no local minimum in $\Pi^\star$ which is not global as well. Finally, if P reaches a minimum, then the segment between P_0 and P (excluding P itself) contains distributions closer to P_0, and thus, must be outside of $\Pi^\star$: the value of L on these points must be non-negative. So, either the segment is reduced to P_0 (meaning that $\mathrm{L}(\mathsf{P}_0) \leq 0$) or we must have $\mathrm{L}(\mathsf{P}) = 0$ for the closest $\mathsf{P} \in \Pi^\star$ of P_0, due to the continuity of L. Since the former case is impossible (as $\mathrm{L}(\mathsf{P}_0) = \mathrm{D}(\mathsf{P}_0\|\mathsf{P}_1) > 0$ since $\mathsf{P}_0 \neq \mathsf{P}_1$), then only the latter case can be true.

The problem now reduces to an optimization problem under constraints since we need to minimize $\mathsf{P} \mapsto \mathrm{D}(\mathsf{P}\|\mathsf{P}_0)$ under the conditions that $\mathrm{L}(\mathsf{P}) = \mathsf{c}$ (where c is a constant) and that $\mathrm{N}(\mathsf{P}) = \sum_{z\in\mathcal{Z}} \mathsf{P}[z] = 1$. According to the method of Lagrange multipliers, a minimum can only be obtained in a point P such that

$$\nabla\mathrm{D}(\mathsf{P}\|\mathsf{P}_0) = \lambda\nabla\mathrm{L}(\mathsf{P}) + \mu\nabla\mathrm{N}(\mathsf{P})$$

for some $\lambda, \mu \in \mathbf{R}$. Solving the previous equation under the two constraints leads to a solution of the form

$$P_\lambda[z] = \frac{\mathsf{P}_0[z]^{1-\lambda}\mathsf{P}_1[z]^\lambda}{\sum_{a\in\mathcal{Z}} \mathsf{P}_0[a]^{1-\lambda}\mathsf{P}_1[a]^\lambda}.$$

Moreover, it is easy to check that for distinct a and b we have

$$\frac{\partial^2\mathrm{D}(\mathsf{P}\|\mathsf{P}_0)}{\partial\mathsf{P}[a]\partial\mathsf{P}[b]} = 0 \quad\text{and}\quad \frac{\partial^2\mathrm{D}(\mathsf{P}\|\mathsf{P}_0)}{\partial\mathsf{P}[a]^2} > 0$$

so that P_λ is indeed a *minimum.*

Finally, we look for a P_λ such that $\mathsf{c} = 0$, i.e., such that $\mathrm{L}(P_\lambda) = 0$. Letting $f : \mathbf{R} \to \mathbf{R}$ be the function defined by

$$f(\lambda) = \log \sum_{a\in\mathcal{Z}} \mathsf{P}_0[a]^{1-\lambda}\mathsf{P}_1[a]^\lambda$$

we note that

$$\mathrm{D}(P_\lambda\|\mathsf{P}_0) = -\lambda\mathrm{L}(P_\lambda) - f(\lambda) \tag{6.2}$$

and that $f'(\lambda) = -\mathrm{L}(P_\lambda)$. Since $f'(0) = -\mathrm{L}(\mathsf{P}_0) < 0$ and $f(0) = f(1) = 0$, there must exist $\lambda^\star \in [0, 1]$ such that $f'(\lambda^\star) = 0$ and such that $f(\lambda^\star)$ is minimal on $[0, 1]$. This minimum is clearly $-\mathrm{C}(\mathsf{P}_0, \mathsf{P}_1)$. We deduce that $\mathrm{L}(P_{\lambda^\star}) = 0$ so that $P_{\lambda^\star}$ is the closest distribution to P_0 in $\Pi^\star$. From (6.2) we deduce that $\mathrm{D}(\Pi^\star\|\mathsf{P}_0) = \mathrm{D}(P_{\lambda^\star}\|\mathsf{P}_0) = -f(\lambda^\star) = \mathrm{C}(\mathsf{P}_0, \mathsf{P}_1)$, which concludes the proof. □

We stress the fact that the validity of the previous theorem is indeed limited to distributions that have the same support. Indeed the result is wrong in the general case as the following example shows.

Example 6.1 Let $\mathcal{Z} = \{0, 1, 2, \ldots, n\}$ for some $n > 0$ and define the distributions P_0 and P_1 over $\mathcal{Z}$ by

$$\mathsf{P}_0 = \left(0, \frac{1}{n}, \ldots, \frac{1}{n}\right) \quad\text{and}\quad \mathsf{P}_1 = \left(\frac{1}{n+1}, \frac{1}{n+1}, \ldots, \frac{1}{n+1}\right).$$

The common support is $\mathcal{Z}' = \{1, 2, \ldots, n\}$. Letting L be as in Proposition 6.4, we have $\mathrm{L}(\mathsf{P}) > 0$ for any distribution P such that $\mathrm{supp}(\mathsf{P}) \subset \mathcal{Z}'$. Consequently, we always have

$$\Pr_{\mathsf{H}_0}[\mathsf{A}_q^\star(\mathbf{Z}^q) = 1] = 0.$$

On the other hand, letting F be as in Theorem 6.2, we have

$$F(\lambda) = \left(\frac{n}{n+1}\right)^{\lambda}.$$

This function is decreasing so that $\inf_{0<\lambda<1} F(\lambda) = F(1) = \frac{n}{n+1}$, and thus $\mathrm{C}(\mathsf{P}_0, \mathsf{P}_1) = \log(1 + 1/n)$, which shows that if the assumption made on the common support in Theorem 6.2 does not hold, then the conclusions of the theorem do not hold either. We further note that by exchanging P_0 and P_1 we obtain the same Chernoff information (which is not surprising since the expression is symmetric) but an error probability of type II equal to $(1 - \frac{1}{n+1})^q$, which is correct. □

We will now extend the validity of the previous theorem to arbitrary distributions, that is, to distributions with arbitrary supports. We will actually show (in Theorem 6.3) that the expression obtained for the asymptotic behavior of the type I probability of error is actually too restrictive when considering arbitrary distributions, namely, λ should be free to take values greater than 1. We first prove a lemma that we then use to prove Theorem 6.3.

Lemma 6.4 *Let P_0 be a probability distribution of full support over a finite set $\mathcal{Z}$ and let $g : \mathcal{Z} \to]0, +\infty[$ be a strictly positive function over $\mathcal{Z}$. Given a distribution P over $\mathcal{Z}$ we define*

$$\mathrm{L}(\mathsf{P}) = \sum_{z\in\mathcal{Z}} \mathsf{P}[z] \log \frac{\mathsf{P}_0[z]}{g(z)}$$

and let $\Pi = \{\mathsf{P} \in \mathcal{P} \ : \ \mathrm{L}(\mathsf{P}) \leq 0\}$. Let $\mathrm{F} : \mathbf{R} \to \mathbf{R}$ be the function defined by

$$\mathrm{F}(\lambda) = \sum_{z\in\mathcal{Z}} \mathsf{P}_0[z]^{1-\lambda} g(z)^{\lambda}.$$

The probability of error of type I of the q-limited distinguisher A_q defined by the set of acceptance $\Pi_q = \Pi \cap \mathcal{P}_q$ is such that

$$\Pr_{\mathsf{H}_0}[\mathsf{A}_q(\mathbf{Z}^q) = 1] \doteq \left(\inf_{\lambda>0} \mathrm{F}(\lambda)\right)^q.$$

Proof. If $\mathsf{P}_0[z] > g(z)$ for all $z \in \mathcal{Z}$, the type I error probability is equal to $\mathsf{P}_0[\mathcal{E}]^q$ where $\mathcal{E}$ is the set of all $z \in \mathcal{Z}$ such that $\mathsf{P}_0[z] = g(z)$. Since $\inf_{\lambda>0} \mathrm{F}(\lambda) = \lim_{\lambda\to 0} \mathrm{F}(\lambda) = \mathsf{P}_0[\mathcal{E}]$ in this case, the result holds.

We now assume that there exists $z \in \mathcal{Z}$ such that $0 < \mathsf{P}_0[z] < g(z)$. Clearly, the distribution P such that $\mathsf{P}[z] = 1$ verifies $\mathrm{L}(\mathsf{P}) < 0$, so that Π is nonempty. Considering the topology of $\mathcal{P}$ (as discussed on page 36), we note that L is continuous. Since there exists $\mathsf{P} \in \Pi$ such that $\mathrm{L}(\mathsf{P}) < 0$, then, for a sufficiently small $\epsilon > 0$, all distributions within a distance to P smaller than ϵ are in Π as well. This means that the interior of Π is nonempty.

Since Π is a nonempty and convex set (since L is linear), we have $\overline{\overset{\circ}{\Pi}} = \overline{\Pi}$ so that we can apply Theorem 6.1 and obtain

$$\Pr_{\mathsf{H}_0}[\mathrm{A}_q(\mathbf{Z}^q) = 1] \doteq 2^{-q\mathrm{D}(\Pi\|\mathsf{P}_0)}. \tag{6.3}$$

We now show that $\mathrm{D}(\Pi\|\mathsf{P}_0)$ is actually equal to $-\inf_{\lambda>0}\log\mathrm{F}(\lambda)$.

The set Π is *topologically closed*: Since P_0 is of full support and $g(z) > 0$ for all $z \in \mathcal{Z}$, $\mathrm{L}(\mathsf{P}) \in \mathbf{R}$ for all $\mathsf{P} \in \mathcal{P}$. Consequently, $\Pi^c = \{\mathsf{P} \in \mathcal{P} \ : \ \mathrm{L}(P) > 0\}$. It is easy to see that there exists $\mathsf{P} \in \mathcal{P}$ such that $\mathrm{L}(\mathsf{P}) > 0$ (so that Π^c is nonempty) and that, since L is continuous, all distributions at a sufficiently small distance of P are in Π^c as well. This shows that any $\mathsf{P} \in \Pi^c$ is the center of an open ball included in Π^c, which makes Π^c an open set and thus, Π is closed.

Since Π is closed and bounded in the Euclidean space $\mathbf{R}^{|\mathcal{Z}|}$, it is *compact*. Since $\mathsf{P} \mapsto \mathrm{D}(\mathsf{P}\|\mathsf{P}_0)$ is continuous, the extreme value theorem states $\mathrm{D}(\Pi\|\mathsf{P}_0) = \mathrm{D}(\mathsf{P}\|\mathsf{P}_0)$ for some $\mathsf{P} \in \Pi$: there exists a global minimum for this function in Π. Furthermore, since the function $\mathsf{P} \mapsto \mathrm{D}(\mathsf{P}\|\mathsf{P}_0)$ is convex, the set of P's such that $\mathrm{D}(\mathsf{P}\|\mathsf{P}_0) < r$ is a convex set for any radius $r > 0$. As a consequence, there is no local minimum in Π which is not global as well. Finally, if P reaches a minimum, then the segment between P_0 and P (excluding P itself) contains distributions closer to P_0, and thus, must be outside of Π: the value of L on these points must be non-negative. So, either the segment is reduced to P_0 (meaning that $\mathrm{L}(\mathsf{P}_0) \leq 0$) or we must have $\mathrm{L}(\mathsf{P}) = 0$ for the closest $\mathsf{P} \in \Pi^\star$ of P_0, due to the continuity of L.

The problem now reduces to an optimization problem under constraints since we need to minimize $\mathsf{P} \mapsto \mathrm{D}(\mathsf{P}\|\mathsf{P}_0)$ under the conditions that $\mathrm{L}(\mathsf{P}) = \mathsf{c}$ (where c is a constant) and that $\mathrm{N}(\mathsf{P}) = \sum_{z\in\mathcal{Z}}\mathsf{P}[z] = 1$. According to the method of Lagrange multipliers, a minimum can only be obtained in a point P such that

$$\nabla\mathrm{D}(\mathsf{P}\|\mathsf{P}_0) = \lambda\nabla\mathrm{L}(\mathsf{P}) + \mu\nabla\mathrm{N}(\mathsf{P})$$

for some $\lambda, \mu \in \mathbf{R}$. Solving the previous equation under the two constraints leads to a solution of the form

$$P_\lambda[z] = \frac{\mathsf{P}_0[z]^{1-\lambda}g(z)^\lambda}{\sum_{a\in\mathcal{Z}}\mathsf{P}_0[a]^{1-\lambda}g(a)^\lambda}.$$

Moreover, it is easy to check that for distinct a and b we have

$$\frac{\partial^2\mathrm{D}(\mathsf{P}\|\mathsf{P}_0)}{\partial\mathsf{P}[a]\partial\mathsf{P}[b]} = 0 \quad \text{and} \quad \frac{\partial^2\mathrm{D}(\mathsf{P}\|\mathsf{P}_0)}{\partial\mathsf{P}[a]^2} > 0$$

so that P_λ is indeed a *minimum*.

Finally, we look for a P_λ such that $\mathsf{c} = 0$, i.e., such that $\mathrm{L}(P_\lambda) = 0$. Letting $f : \mathbf{R} \to \mathbf{R}$ be the function defined by

$$f(\lambda) = \log\sum_{a\in\mathcal{Z}}\mathsf{P}_0[a]^{1-\lambda}g(a)^\lambda$$

we note that

$$\mathrm{D}(P_\lambda\|\mathsf{P}_0) = -\lambda\mathrm{L}(P_\lambda) - f(\lambda) \tag{6.4}$$

and that $f'(\lambda) = -\mathrm{L}(P_\lambda)$. As previously noted, there are two eventualities: either $\mathrm{L}(\mathsf{P}_0) \leq 0$ (i.e., P_0 is in Π) or the closest $\mathsf{P} \in \Pi$ to P_0 verifies $\mathrm{L}(\mathsf{P}) = 0$:

- If $\mathrm{L}(\mathsf{P}_0) \leq 0$, then $f'(0) \geq 0$. We can see that

  $$f'(\lambda) = \frac{\sum_{z \in \mathcal{Z}} \mathsf{P}_0[z]^{1-\lambda} g(z)^\lambda \log \frac{g(z)}{\mathsf{P}_0[z]}}{\sum_{a \in \mathcal{Z}} \mathsf{P}_0[a]^{1-\lambda} g(a)^\lambda}.$$

  The denominator is always positive and it is easy to show that the nominator is an increasing function of λ. Since $f'(0) \geq 0$, this means that $f'(\lambda) \geq 0$ for all $\lambda \geq 0$, so that the minimum of f over $[0, +\infty[$ is $f(0) = 0$. Since we also have $\mathrm{D}(\Pi \| \mathsf{P}_0) = 0$, the lemma is correct in this case.

- If $\mathrm{L}(\mathsf{P}_0) > 0$, then $f'(0) < 0$. Moreover, since we assumed that there exists $z \in \mathcal{Z}$ such that $0 < \mathsf{P}_0[z] < g(z)$, then $\lim_{\lambda \to +\infty} f(\lambda) = +\infty$. Consequently, there must exists $\lambda^\star > 0$ such that $f'(\lambda^\star) = 0$ (and thus such that $\mathrm{L}(P_{\lambda^\star}) = 0$) and for which f is minimal. This minimum is clearly $\inf_{\lambda > 0} \log \mathrm{F}(\lambda)$. Combining this with (6.3) and (6.4) concludes the proof.

□

Theorem 6.3 *Let P_0 and P_1 be two probability distribution of finite supports with union $\mathcal{Z}$ and intersection $\mathcal{Z}'$. Given a distribution P over $\mathcal{Z}$ we define*

$$\mathrm{L}(\mathsf{P}) = \sum_{z \in \mathcal{Z}} \mathsf{P}[z] \log \frac{\mathsf{P}_0[z]}{\mathsf{P}_1[z]},$$

where $\mathrm{L}(\mathsf{P})$ can be infinite or undefined, and let $\Pi = \{\mathsf{P} \in \mathcal{P} : \mathrm{L}(\mathsf{P}) \leq 0\}$. Let $\mathrm{F} : \mathbf{R} \to \mathbf{R}$ be the function defined by

$$\mathrm{F}(\lambda) = \sum_{z \in \mathcal{Z}'} \mathsf{P}_0[z]^{1-\lambda} \mathsf{P}_1[z]^\lambda.$$

The probability of error of type I of the q-limited distinguisher A_q defined by the set of acceptance $\Pi_q = \Pi \cap \mathcal{P}_q$ is such that

$$\Pr_{\mathsf{H}_0}[\mathsf{A}_q(\mathbf{Z}^q) = 1] \doteq \left(\inf_{\lambda > 0} \mathrm{F}(\lambda) \right)^q. \tag{6.5}$$

Proof. Let P'_0 be the distribution of full support over $\mathcal{Z}' = \mathrm{supp}(\mathsf{P}_0) \cap \mathrm{supp}(\mathsf{P}_1)$, defined by

$$\mathsf{P}'_0[z] = \begin{cases} \frac{\mathsf{P}_0[z]}{\mathsf{P}_0[\mathcal{Z}']} & \text{if } z \in \mathcal{Z}' \\ 0 & \text{otherwise.} \end{cases}$$

Let $g : \mathcal{Z} \to \mathbf{R}$ be the function defined by

$$g(z) = \begin{cases} \frac{\mathsf{P}_1[z]}{\mathsf{P}_0[\mathcal{Z}']} & \text{if } z \in \mathcal{Z}' \\ 0 & \text{otherwise.} \end{cases}$$

Similarly to Lemma 6.4, for any distribution P over $\mathcal{Z}'$ we define

$$\mathrm{L}'(\mathsf{P}) = \sum_{z \in \mathcal{Z}'} \mathsf{P}[z] \log \frac{\mathsf{P}'_0[z]}{g(z)}$$

and $\Pi' = \{\mathsf{P} \in \mathcal{P}(\mathcal{Z}') \ : \ \mathrm{L}'(\mathsf{P}) \leq 0\}$. Clearly, we have $\mathrm{L}(\mathsf{P}) = \mathrm{L}'(\mathsf{P})$ for any distribution P over $\mathcal{Z}'$. Consequently, Π consists of Π' together with the distributions which support is included in that of P_1 but *not* in that of P_0 (in which case the value obtained for L is $-\infty$). Since the probability of reaching one of the latter distributions is 0 when sampling according to distribution P_0, we have

$$\Pr_{\mathsf{H}_0}[\mathrm{A}_q(\mathbf{Z}^q) = 1] = \Pr_{\mathsf{H}_0}[\mathsf{P}_{\mathbf{Z}^q} \in \Pi'],$$

where

$$\Pr_{\mathsf{H}_0}[\mathsf{P}_{\mathbf{Z}^q} \in \Pi'] = \Pr_{\mathsf{H}_0}[\mathsf{P}_{\mathbf{Z}^q} \in \Pi' | Z_1, \ldots, Z_q \in \mathcal{Z}'] \Pr_{\mathsf{H}_0}[Z_1, \ldots, Z_q \in \mathcal{Z}']$$

since $\Pr_{\mathsf{H}_0}[\mathsf{P}_{\mathbf{Z}^q} \in \Pi' | Z_1, \ldots, Z_q \notin \mathcal{Z}'] = 0$. We have $\Pr_{\mathsf{H}_0}[Z_1, \ldots, Z_q \in \mathcal{Z}'] = (\mathsf{P}_0[\mathcal{Z}'])^q$ and it is easy to show that

$$\Pr_{\mathsf{H}_0}[\mathsf{P}_{\mathbf{Z}^q} \in \Pi' | Z_1, \ldots, Z_q \in \mathcal{Z}'] = \Pr_{\mathsf{P}=\mathsf{P}'_0}[\mathsf{P}_{\mathbf{Z}^q} \in \Pi'].$$

From the previous equations we finally obtain

$$\Pr_{\mathsf{H}_0}[\mathrm{A}_q(\mathbf{Z}^q) = 1] = (\mathsf{P}_0[\mathcal{Z}'])^q \cdot \Pr_{\mathsf{P}=\mathsf{P}'_0}[\mathsf{P}_{\mathbf{Z}^q} \in \Pi'].$$

Applying Lemma 6.4 to the right-hand side of the previous equation, we get

$$\Pr_{\mathsf{P}=\mathsf{P}'_0}[\mathsf{P}_{\mathbf{Z}^q} \in \Pi'] \doteq \left(\inf_{\lambda > 0} \mathrm{G}(\lambda) \right)^q$$

where

$$\mathrm{G}(\lambda) = \sum_{z \in \mathcal{Z}'} \mathsf{P}'_0[z]^{1-\lambda} g(z)^\lambda = \frac{1}{\mathsf{P}_0[\mathcal{Z}']} \sum_{z \in \mathcal{Z}'} \mathsf{P}_0[z]^{1-\lambda} \mathsf{P}_1[z]^\lambda.$$

Combining the three previous equations allows to conclude. □

We will illustrate Theorem 6.3 on practical examples in section 6.5. Before that, we show that the previous theorem allows to easily deduce the asymptotic behavior of the advantage of the best distinguisher in the general case.

Corollary 6.1 *Let* P_0 *and* P_1 *be two probability distribution of finite supports with union* $\mathcal{Z}$ *and intersection* $\mathcal{Z}'$. *Let* $\mathrm{F} : \mathbf{R} \to \mathbf{R}$ *be the function defined by*

$$\mathrm{F}(\lambda) = \sum_{z \in \mathcal{Z}'} \mathsf{P}_0[z]^{1-\lambda} \mathsf{P}_1[z]^\lambda$$

and let

$$\mathrm{C}(\mathsf{P}_0,\mathsf{P}_1) = -\inf_{0<\lambda<1} \log \mathrm{F}(\lambda) \tag{6.6}$$

be the Chernoff information between P_0 *and* P_1*. The advantage of the best* q*-limited distinguisher between* P_0 *and* P_1 *is such that*

$$1 - \mathrm{BestAdv}^q(\mathsf{P}_0,\mathsf{P}_1) \doteq 2^{-q\mathrm{C}(\mathsf{P}_0,\mathsf{P}_1)} = \left(\inf_{0<\lambda<1} \mathrm{F}(\lambda)\right)^q .$$

Proof. The advantage of the best q-limited distinguisher $\mathsf{A}_q^\star$ is such that can be written as

$$1 - \mathrm{BestAdv}^q(\mathsf{P}_0,\mathsf{P}_1) = \Pr_{\mathsf{H}_0}[\mathsf{A}_q^\star(\mathbf{Z}^q) = 1] + \Pr_{\mathsf{H}_1}[\mathsf{A}_q^\star(\mathbf{Z}^q) = 0].$$

The acceptance region of $\mathsf{A}_q^\star$ being $\Pi^\star = \{\mathsf{P} \in \mathcal{P} \;:\; \mathrm{L}(\mathsf{P}) \le 0\}$ where

$$\mathrm{L}(\mathsf{P}) = \sum_{z\in\mathrm{supp}(\mathsf{P})} \mathsf{P}[z] \log \frac{\mathsf{P}_0[z]}{\mathsf{P}_1[z]}$$

(according to Proposition 6.4 and using the notations of the proposition), we can apply Theorem 6.3 and obtain

$$\Pr_{\mathsf{H}_0}[\mathsf{A}_q^\star(\mathbf{Z}^q) = 1] \doteq \left(\inf_{\lambda>0} \mathrm{F}(\lambda)\right)^q$$

where $\mathrm{F}(\lambda) = \sum_{z\in\mathcal{Z}'} \mathsf{P}_0[z]^{1-\lambda}\mathsf{P}_1[z]^\lambda$.

On the other hand, we can see that by symmetry one can get

$$\begin{aligned}
\Pr_{\mathsf{H}_1}[\mathsf{A}_q^\star(\mathbf{Z}^q) = 0] &= \Pr_{\mathsf{H}_1}[\mathrm{L}(\mathsf{P}_{\mathbf{Z}^q}) > 0] \\
&\le \Pr_{\mathsf{H}_1}[\mathrm{L}(\mathsf{P}_{\mathbf{Z}^q}) \ge 0] \\
&= \Pr_{\mathsf{H}_1}[-\mathrm{L}(\mathsf{P}_{\mathbf{Z}^q}) \le 0] \\
&\doteq \left(\inf_{\lambda>0} \mathrm{G}(\lambda)\right)^q
\end{aligned}$$

where $\mathrm{G}(\lambda) = \sum_{z\in\mathcal{Z}'} \mathsf{P}_1[z]^{1-\lambda}\mathsf{P}_0[z]^\lambda = \mathrm{F}(1-\lambda)$.

Consequently, if $\inf_{\lambda>0}\mathrm{F}(\lambda) > \inf_{\lambda<1}\mathrm{F}(\lambda)$, we obtain $1 - \mathrm{BestAdv}_q(\mathsf{P}_0,\mathsf{P}_1) \doteq (\inf_{\lambda>0}\mathrm{F}(\lambda))^q$. If $\inf_{\lambda>0}\mathrm{F}(\lambda) < \inf_{\lambda<1}\mathrm{F}(\lambda)$, we can exchange the roles of P_0 and P_1 and similarly obtain $1 - \mathrm{BestAdv}_q(\mathsf{P}_0,\mathsf{P}_1) \doteq (\inf_{\lambda<0}\mathrm{F}(\lambda))^q$. In all cases we get

$$1 - \mathrm{BestAdv}_q(\mathsf{P}_0,\mathsf{P}_1) \doteq \left(\inf_{\lambda>0} \mathrm{F}(\lambda)\right)^q + \left(\inf_{\lambda<1} \mathrm{F}(\lambda)\right)^q .$$

If the minimum of F is reached for some $\lambda \in \,]0,1[$ we are done. Otherwise, we first note that $\mathrm{F}'(\lambda) = \sum_{z\in\mathcal{Z}'} \mathsf{P}_0[z]^{1-\lambda}\mathsf{P}_1[z]^\lambda \log\frac{\mathsf{P}_0[z]}{\mathsf{P}_1[z]}$ and

$$\mathrm{F}''(\lambda) = \sum_{z\in\mathcal{Z}'} \mathsf{P}_0[z]^{1-\lambda}\mathsf{P}_1[z]^\lambda \left(\log\frac{\mathsf{P}_0[z]}{\mathsf{P}_1[z]}\right)^2 \ge 0,$$

so that F' is an increasing function. If the minimum of F is reached for $\lambda \leq 0$, then it must be the case that $\inf_{\lambda>0} \mathrm{F}(\lambda) = \mathrm{F}(0)$ and that $\inf_{\lambda<1} \mathrm{F}(\lambda) \leq \mathrm{F}(0)$, so that we obtain $1 - \mathrm{BestAdv}_q(\mathsf{P}_0, \mathsf{P}_1) \doteq \mathrm{F}(0)^q$. Similarly, if the minimum is reached for $\lambda \geq 1$, we necessarily have $\inf_{\lambda<1} \mathrm{F}(\lambda) = \mathrm{F}(1)$ and $\inf_{\lambda>0} \mathrm{F}(\lambda) \leq \mathrm{F}(1)$, so that in this case $1 - \mathrm{BestAdv}_q(\mathsf{P}_0, \mathsf{P}_1) \doteq \mathrm{F}(1)^q$. Hence, in all cases we can write

$$1 - \mathrm{BestAdv}_q(\mathsf{P}_0, \mathsf{P}_1) \doteq \left(\inf_{0<\lambda<1} \mathrm{F}(\lambda) \right)^q$$

which concludes the proof. □

6.5 The Best Distinguisher: Examples and Pathological Distributions

In this subsection we will consider several practical pairs of distributions and compute both the *exact* value of the advantage of the best distinguisher and its asymptotic value using Corollary 6.1. We will also illustrate the fact that (6.1) in Theorem 6.2 can prove to be *wrong* in the case where both distributions do not have the same support, the correct value of the error probability being obtained using (6.5). Unless otherwise stated, in these examples P_0 and P_1 are two probability distributions of finite supports with union $\mathcal{Z} = \{0, 1, \ldots, n\}$ for some positive integer n and intersection $\mathcal{Z}'$.

Example 6.2 We consider the trivial case where $\mathcal{Z}' = \emptyset$. Obviously, after one query the logarithmic likelihood ration will already take its final value (either $-\infty$ or $+\infty$) and we have $\mathrm{BestAdv}_q(\mathsf{P}_0, \mathsf{P}_1) = 1$ for all $q > 0$. This is coherent with Corollary 6.1 since we have $\mathrm{C}(\mathsf{P}_0, \mathsf{P}_1) = +\infty$ and thus $2^{-q\mathrm{C}(\mathsf{P}_0,\mathsf{P}_1)} = 0$. We similarly see that Theorem 6.3 holds since $\mathrm{F}(\lambda) = 0$ and since we indeed have that the probability of error is 0 in this case. □

Example 6.3 We reconsider here Example 6.1. Let $\mathsf{P}_0 = (0, \frac{1}{n}, \ldots, \frac{1}{n})$ and $\mathsf{P}_1 = (\frac{1}{n+1}, \ldots, \frac{1}{n+1})$. Using Proposition 6.3 is is easy to show that

$$\mathrm{BestAdv}_q(\mathsf{P}_0, \mathsf{P}_1) = 1 - \left(1 - \frac{1}{n}\right)^q .$$

The Chernoff information between P_0 and P_1 is given by

$$\mathrm{C}(\mathsf{P}_0, \mathsf{P}_1) = - \inf_{0<\lambda<1} \log \sum_{z \in \mathcal{Z}'} \mathsf{P}_0[z]^{1-\lambda} \mathsf{P}_1[z]^{\lambda} = - \log \left(1 - \frac{1}{n}\right),$$

so that Corollary 6.1 states that

$$1 - \mathrm{BestAdv}_q(\mathsf{P}_0, \mathsf{P}_1) \doteq \left(1 - \frac{1}{n}\right)^q .$$

We note that in this particular case, the corollary gives an *exact* value of the advantage, and thus tells more than just the asymptotic behavior. Concerning the type I error probability, using the notations of Theorem 6.3 we have

$$\mathrm{F}(\lambda) = \left(1 - \frac{1}{n}\right)^{\lambda}$$

so that $\inf_{\lambda>0} \mathrm{F}(\lambda) = \lim_{\lambda\to\infty} \mathrm{F}(\lambda) = 0$. Theorem 6.3 then states that $\Pr_{\mathsf{H}_0}[\mathsf{A}_q^\star(\mathbf{Z}^q) = 1] \doteq 0$, which is correct since the best distinguisher never makes a wrong guess under hypothesis H_0 as we have

$$\mathrm{L}(\mathsf{P}_{\mathbf{Z}^q}) = \sum_{z\in\mathcal{Z}'} \mathsf{P}_{\mathbf{Z}^q}[z] \log\frac{n+1}{n} > 0$$

in this case. □

Example 6.4 Let $\mathsf{P}_0 = (0, \frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n})$ and $\mathsf{P}_1 = (\frac{1}{n}, 0, \frac{1}{n}, \ldots, \frac{1}{n})$. We have $\mathcal{Z}' = \{2, 3, \ldots, n\}$. According to Theorem 6.3, and since $\mathrm{F}(\lambda) = 1 - \frac{1}{n}$ for all values of λ, we have

$$\Pr_{\mathsf{H}_0}[\mathsf{A}_q^\star(\mathbf{Z}^q) = 1] \doteq \left(1 - \frac{1}{n}\right)^q.$$

Indeed, according to the definition of the acceptance region of the best distinguisher given in Proposition 6.4, we clearly have that

$$\Pr_{\mathsf{H}_0}[\mathsf{A}_q^\star(\mathbf{Z}^q) = 1] = \Pr_{\mathsf{H}_0}[Z_1 \neq 1, \ldots, Z_q \neq 1] = (1 - \mathsf{P}_0[1])^q = \left(1 - \frac{1}{n}\right)^q,$$

which shows that the theorem is correct and quite precise in this case also. It is easy to see that we have $\Pr_{\mathsf{H}_1}[\mathsf{A}_q^\star(\mathbf{Z}^q) = 0] = 0$, and thus that

$$\mathrm{BestAdv}_q(\mathsf{P}_0, \mathsf{P}_1) = 1 - \left(1 - \frac{1}{n}\right)^q.$$

Corollary 6.1 thus gives us more than the asymptotic behavior of the advantage of the best distinguisher since we have in this case $\mathrm{C}(\mathsf{P}_0, \mathsf{P}_1) = -\log\left(1 - \frac{1}{n}\right)$, which leads to

$$1 - \mathrm{BestAdv}_q(\mathsf{P}_0, \mathsf{P}_1) \doteq \left(1 - \frac{1}{n}\right)^q.$$

□

Example 6.5 Let $\mathcal{Z} = \{0, 1, 2, 3\}$ and $\mathsf{P}_0 = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3}, 0)$ and $\mathsf{P}_0 = (\frac{1}{4}, \frac{1}{2}, 0, \frac{1}{4})$. We have $\mathcal{Z}' = \{0, 1\}$ and in this case the smallest value of

$$\mathrm{F}(\lambda) = \frac{1}{3}\left(\frac{3}{4}\right)^{\lambda} + \frac{1}{3}\left(\frac{3}{2}\right)^{\lambda}$$

over $\lambda > 0$ is achieved for $\lambda \to 0$, the limit being equal to $\frac{2}{3}$. We deduce that

$$\mathrm{C}(\mathsf{P}_0, \mathsf{P}_1) = -\log \frac{2}{3}$$

and that

$$1 - \mathrm{BestAdv}_q(\mathsf{P}_0, \mathsf{P}_1) \doteq \left(\frac{2}{3}\right)^q .$$

□

6.6 The Best Distinguisher: Case where the Distributions are Close to Each Other

In order to obtain an expression of the advantage of the best distinguisher easier to deal with in practice than that given in Corollary 6.1, we will derive a simple expression approximating the Chernoff information. We consider the case where P_0 and P_1 are both of full support over $\mathcal{Z}$ by letting

$$\epsilon_z = \frac{\mathsf{P}_1[z] - \mathsf{P}_0[z]}{\mathsf{P}_0[z]}$$

and assuming that $\epsilon_z = o(1)$ for all $z \in \mathcal{Z}$. Here, we do not consider that both distributions are fixed but rather consider that this is only the case for P_0 and that P_1 converges towards P_0. Although not realistic from a cryptographic point of view, this approach allows to obtain results which are mathematically sound, and that can lead to good approximates in practical situations.

Lemma 6.5 *Let P_0 and P_1 be two distributions of support $\mathcal{Z}$ such that P_1 tends towards the fixed distribution P_0. Let q be the number of queries available to the best distinguisher $\mathsf{A}_q^\star$ between P_0 and P_1. We have*

$$\mathrm{C}(\mathsf{P}_0, \mathsf{P}_1) = \frac{1}{8 \ln 2} \sum_{z \in \mathcal{Z}} \mathsf{P}_0[z] \epsilon_z^2 + o\left(\|\epsilon\|_2^2\right)$$

where for all $z \in \mathcal{Z}$ we define

$$\epsilon_z = \frac{\mathsf{P}_1[z] - \mathsf{P}_0[z]}{\mathsf{P}_0[z]} \quad \textit{and} \quad \epsilon = (\epsilon_z)_{z \in \mathcal{Z}}.$$

Proof. We let

$$\mathrm{F}(\lambda, x) = \sum_{z \in \mathcal{Z}} \mathsf{P}_0[z](1+\epsilon_z)^\lambda \quad \text{and} \quad g(\lambda, x) = \sum_{z \in \mathcal{Z}} \mathsf{P}_0[z](1+\epsilon_z)^\lambda \ln(1+\epsilon_z)$$

so that $\mathrm{C}(\mathsf{P}_0, \mathsf{P}_1) = -\min_{0\le\lambda\le 1} \log \mathrm{F}(\lambda, \epsilon) = -\log \mathrm{F}(\lambda^\star, \epsilon)$ and $g(\lambda^\star, \epsilon) = 0$ (since $\frac{\partial \mathrm{F}}{\partial \lambda}(\lambda, \epsilon) = g(\lambda, \epsilon)$). We will approximate $\mathrm{F}(\lambda^\star, \epsilon)$ when ϵ is small and subject to $\sum_z \mathsf{P}_0[z]\epsilon_z = 0$. We first have

$$\begin{aligned} g(\lambda,\epsilon) &= \sum_z \mathsf{P}_0[z](1+\lambda\epsilon_z + o(\epsilon_z))\left(\epsilon_z - \frac{\epsilon_z^2}{2} + o(\epsilon_z^2)\right) \\ &= \sum_z \mathsf{P}_0[z]\left(\lambda - \frac{1}{2}\right)\epsilon_z^2 + o\left(\|\epsilon\|_2^2\right) \end{aligned}$$

since $\sum_z \mathsf{P}_0[z]\epsilon_z$ is zero. As $g(\lambda^*, \epsilon) = 0$ we deduce that λ^* tends towards $\frac{1}{2}$ as ϵ tends towards 0. Taylor's theorem used with Lagrange's form of the remainder gives

$$\mathrm{F}(\lambda^*, \epsilon) = \mathrm{F}\left(\frac{1}{2}, \epsilon\right) + \left(\lambda^* - \frac{1}{2}\right)\frac{\partial \mathrm{F}}{\partial \lambda}\left(\frac{1}{2}, \epsilon\right) + \frac{1}{2}\left(\lambda^* - \frac{1}{2}\right)^2 R$$

with $|R| \le \max_\lambda \frac{\partial^2 \mathrm{F}}{\partial \lambda^2}(\lambda, \epsilon)$ for $\lambda \in [0,1]$. As $\frac{\partial \mathrm{F}}{\partial \lambda}(\lambda, \epsilon) = g(\lambda, \epsilon)$, previous computations immediately lead to $\frac{\partial \mathrm{F}}{\partial \lambda}(\frac{1}{2}, \epsilon) = g(\frac{1}{2}, \epsilon) = o(\|\epsilon\|_2^2)$. Similarly we have

$$\begin{aligned} \frac{\partial^2 \mathrm{F}}{\partial \lambda^2}(\lambda, \epsilon) &= \sum_{z\in\mathcal{Z}} \mathsf{P}_0[z](1+\epsilon_z)^\lambda \left(\ln(1+\epsilon_z)\right)^2 \\ &= \sum_{z\in\mathcal{Z}} \mathsf{P}_0[z](1+o(1))\left(\epsilon_z + o(\epsilon_z)\right)^2 \\ &= \sum_{z\in\mathcal{Z}} \mathsf{P}_0[z]\epsilon_z^2 + o(\|\epsilon\|^2) \end{aligned}$$

which is a $O(\|\epsilon\|^2)$, hence

$$\mathrm{F}(\lambda^*, \epsilon) = \mathrm{F}\left(\frac{1}{2}, \epsilon\right) + o(\|\epsilon\|^2) + \frac{1}{2}\left(\lambda^* - \frac{1}{2}\right)^2 O(\|\epsilon\|^2).$$

Since $\lambda^* - \frac{1}{2} = o(1)$ the previous equation can be reduced to

$$\mathrm{F}(\lambda^*, \epsilon) = \mathrm{F}\left(\frac{1}{2}, \epsilon\right) + o(\|\epsilon\|^2).$$

Now, we have

$$\begin{aligned} \mathrm{F}\left(\frac{1}{2}, \epsilon\right) &= \sum_{z\in\mathcal{Z}} \mathsf{P}_0[z]\sqrt{1+\epsilon_z} \\ &= \sum_{z\in\mathcal{Z}} \mathsf{P}_0[z]\left(1 + \frac{1}{2}\epsilon_z - \frac{1}{8}\epsilon_z^2 + o(\epsilon_z^2)\right) \\ &= 1 - \frac{1}{8}\sum_{z\in\mathcal{Z}} \mathsf{P}_0[z]\epsilon_z^2 + o(\|\epsilon\|_2^2) \end{aligned}$$

and therefore

$$\mathrm{F}(\lambda^*, x) = 1 - \frac{1}{8}\sum_{z\in\mathcal{Z}} \mathsf{P}_0[z]\epsilon_z^2 + o(\|\epsilon\|_2^2).$$

Since $\mathrm{C}(\mathsf{P}_0, \mathsf{P}_1) = -\log \mathrm{F}(\lambda^\star, \epsilon)$ this immediately leads to

$$\mathrm{C}(\mathsf{P}_0, \mathsf{P}_1) = \frac{1}{8\ln 2}\sum_{z\in\mathcal{Z}} \mathsf{P}_0[z]\epsilon_z^2 + o(\|\epsilon\|_2^2).$$

□

Based on Corollary 6.1, it is tempting in practice to make the following heuristic assumption.

Heuristic 6.1 Let P_0 and P_1 be two probability distributions of finite support with union $\mathcal{Z}$ and intersection $\mathcal{Z}'$. Let

$$\mathrm{C}(\mathsf{P}_0, \mathsf{P}_1) = -\inf_{0<\lambda<1} \log \sum_{z\in\mathcal{Z}'} \mathsf{P}_0[z]^{1-\lambda}\mathsf{P}_1[z]^{\lambda}$$

be the Chernoff information between P_0 and P_1. The best q-limited distinguisher between P_0 and P_1 reaches a non-negligible advantage when

$$q = \frac{1}{\mathrm{C}(\mathsf{P}_0, \mathsf{P}_1)}.$$

□

One can note that all the examples of Section 6.5 are in favor of the previous heuristic, since for all of them we have $\mathrm{BestAdv}_q(\mathsf{P}_0, \mathsf{P}_1) = 1 - 2^{-q\mathrm{C}(\mathsf{P}_0,\mathsf{P}_1)}$. Taking $q = 1/\mathrm{C}(\mathsf{P}_0, \mathsf{P}_1)$ would lead to an advantage equal to $\frac{1}{2}$ (in those specific cases). In practice though, it might be more comfortable to work with the approximation of the Chernoff information that we obtained in Lemma 6.5 than with the Chernoff information itself.

Heuristic 6.2 Let P_0 and P_1 be two distributions of support $\mathcal{Z}$ and let $\epsilon = (\epsilon_z)_{z\in\mathcal{Z}}$ be such that $\epsilon_z = \frac{\mathsf{P}_1[z]-\mathsf{P}_0[z]}{\mathsf{P}_0[z]}$ for all $z \in \mathcal{Z}$. Assuming that $\|\epsilon\|_2^2 \ll 1$, the best q-limited distinguisher reaches a non-negligible advantage when

$$q = \frac{8\ln 2}{\sum_{z\in\mathcal{Z}} \mathsf{P}_0[z]\epsilon_z^2}.$$

□

Whereas we cannot formally justify Heuristic 6.1, we can easily show that if it holds then Heuristic 6.2 does also. For this we use the following result.

Lemma 6.6 *Let P_0 and P_1 be two distributions of support $\mathcal{Z}$. For all $z \in \mathcal{Z}$ we define*

$$\epsilon_z = \frac{\mathsf{P}_1[z] - \mathsf{P}_0[z]}{\mathsf{P}_0[z]} \quad \textit{and} \quad \epsilon = (\epsilon_z)_{z\in\mathcal{Z}}.$$

Letting $\mathrm{B}(\mathsf{P}_0, \mathsf{P}_1) = -\log \sum_z \sqrt{\mathsf{P}_0[z]\mathsf{P}_1[z]}$ *we have* $\mathrm{C}(\mathsf{P}_0, \mathsf{P}_1) \geq \mathrm{B}(\mathsf{P}_0, \mathsf{P}_1)$ *and, assuming that* $\|\epsilon\|_\infty \leq \frac{1}{2}$*,*

$$\left| \mathrm{B}(\mathsf{P}_0, \mathsf{P}_1) - \frac{1}{8\ln 2} \sum_{z \in Z} \mathsf{P}_0[z]\epsilon_z^2 \right| \leq \frac{\sqrt{|\mathcal{Z}|}}{8\ln 2} \|\mathsf{P}_0\|_\infty \|\epsilon\|_2^3 + \frac{5}{96\ln 2} \|\mathsf{P}_0\|_\infty^2 \|\epsilon\|_2^4. \tag{6.7}$$

Proof. The fact that $\mathrm{B}(\mathsf{P}_0, \mathsf{P}_1) \leq \mathrm{C}(\mathsf{P}_0, \mathsf{P}_1)$ is trivial. The rest is proved in Appendix B. □

We see that if Heuristic 6.1 holds, then

$$q = \frac{1}{\mathrm{B}(\mathsf{P}_0, \mathsf{P}_1)}$$

samples are obviously sufficient to distinguish P_0 from P_1 since $\mathrm{B}(\mathsf{P}_0, \mathsf{P}_1) \leq \mathrm{C}(\mathsf{P}_0, \mathsf{P}_1)$. In that case it is easy to check whether the approximation

$$\mathrm{B}(\mathsf{P}_0, \mathsf{P}_1) \approx \frac{1}{8\ln 2} \sum_{z \in Z} \frac{(\mathsf{P}_1[z] - \mathsf{P}_0[z])^2}{\mathsf{P}_0[z]}$$

is acceptable by showing that the right-hand side of (6.7) is negligible. If it is so, then we can assume that Heuristic 6.2 is correct.

6.7 The Best Distinguisher: Case where one of the Distributions is Uniform

In this section, we assume that P_0 is the uniform distribution. This situation is quite typical when studying certain cryptographic devices, like for example pseudo-random generators which should generate a sequence of symbols indistinguishable from a uniformly distributed random string. To simplify the notations, we let $\mathsf{U} = \mathsf{P}_0$ denote the uniform distribution and $\mathsf{P} = \mathsf{P}_1$ be the biased distribution of full support.

Definition 6.10 *Let* P *be an arbitrary distribution over a finite set* $\mathcal{Z}$*, let* $\delta_z = \mathsf{P}[z] - \frac{1}{|\mathcal{Z}|}$*, and let* $\epsilon_z = |\mathcal{Z}|\,\delta_z$ *for all* $z \in \mathcal{Z}$*. The Squared Euclidean Imbalance*[3] *(SEI)* $\Delta(\mathsf{P})$ *of the distribution* P *is defined by*

$$\Delta(\mathsf{P}) = |\mathcal{Z}| \sum_{z \in \mathcal{Z}} \delta_z^2 = \frac{1}{|\mathcal{Z}|} \sum_{z \in \mathcal{Z}} \epsilon_z^2 = |\mathcal{Z}|\, \|\mathsf{P} - \mathsf{U}\|_2^2.$$

[3] Although this appellation coincides with the one of Harpes, Kramer, and Massey in [65], note that the definitions slightly differ.

Using this notation, we can re-write Lemma 6.6 as follows.

Lemma 6.7 *Let* P *be a probability distribution over a finite set* $\mathcal{Z}$ *and let* $\mathrm{B}(\mathsf{U},\mathsf{P}) = -\log\left(|\mathcal{Z}|^{-1/2}\sum_{z\in\mathcal{Z}}\sqrt{\mathsf{P}[z]}\right)$. *Assuming that* $\|\mathsf{P}-\mathsf{U}\|_\infty \leq \frac{1}{2|\mathcal{Z}|}$, *we have*

$$\left|\mathrm{B}(\mathsf{U},\mathsf{P}) - \frac{\Delta(\mathsf{P})}{8\ln 2}\right| \leq \frac{|\mathcal{Z}|}{8\ln 2}\Delta(\mathsf{P})^{3/2} + \frac{5}{96\ln 2}\Delta(\mathsf{P})^2.$$

This result validates the rule of thumb that states that in order to reach a non-negligible advantage, the best distinguisher between P and U needs a sample of size at least $1/\Delta(\mathsf{P})$. In particular, assuming that Heuristic 6.1 is correct and that P is close enough to the uniform distribution, it tells us that the number of samples required by the best distinguisher in order to reach an advantage close to $\frac{1}{2}$ is $\frac{8\ln 2}{\Delta(\mathsf{P})}$. If we consider, for example, the particular case where $\mathcal{Z} = \{0,1\}$ and denote $\epsilon = \epsilon_0 = -\epsilon_1$, this rule of thumb leads to the conclusion that the sample size should be close to $8/\epsilon^2$. This result is very similar to classical results coming from linear cryptanalysis [110].

6.8 The Best Distinguisher: Case where one Hypothesis is Composite

So far, we considered the problem of testing the null hypothesis $\mathsf{H}_0 : \mathsf{P} = \mathsf{P}_0$ against the simple alternate hypothesis $\mathsf{H}_1 : \mathsf{P} = \mathsf{P}_1$ where P_0 and P_1 were fully specified. This situation is usually referred to as the *simple hypothesis testing problem*. A more complex situation arises when one of the two hypotheses is *composite*, i.e., when the distribution might belong to a set of distributions. In this subsection, we will consider the latter situation and extend the results to the case where H_0 is still simple ($\mathsf{H}_0 : \mathsf{P} = \mathsf{P}_0$) but where H_1 is composite ($\mathsf{H}_1 : \mathsf{P} \in \mathcal{D}$, where $\mathcal{D} = \{\mathsf{P}_1,\ldots,\mathsf{P}_d\}$). Under H_1 we assume that the selection of P_i is taken with an *a priori* weight of π_i to define the advantage for distinguishing H_0 from H_1. For simplicity we assume that all distributions have the same support $\mathcal{Z}$.

Theorem 6.4 *Let* P_0 *be a distribution of support* $\mathcal{Z}$ *and* $\mathcal{D} = \{\mathsf{P}_1,\ldots,\mathsf{P}_d\}$ *be a finite set of distributions of support* $\mathcal{Z}$. *Let* $\mathsf{H}_0 : \mathsf{P} = \mathsf{P}_0$ *be the null hypothesis and* $\mathsf{H}_1 : \mathsf{P} \in \mathcal{D}$ *be the alternate hypothesis, in which* P_i *is chosen with an a priori weight of* π_i. *The q-limited distinguisher* $\mathsf{A}_q^\star$ *between* H_0 *and* H_1 *defined by the distribution acceptance region* $\Pi_q^\star = \Pi^\star \cap \mathcal{P}_q$, *where*

$$\Pi^\star = \left\{\mathsf{P}\in\mathcal{P} \;:\; \min_{1\leq i\leq d}\mathrm{L}_i(\mathsf{P}) \leq 0\right\} \quad \textit{with} \quad \mathrm{L}_i(\mathsf{P}) = \sum_{z\in\mathcal{Z}}\mathsf{P}[z]\log\frac{\mathsf{P}_0[z]}{\mathsf{P}_i[z]}$$

is asymptotically optimal and its advantage $\mathrm{BestAdv}_q$ *is such that*

$$1 - \mathrm{BestAdv}_q(\mathsf{H}_0, \mathsf{H}_1) \doteq \max_{1\le i\le d} 2^{-q\mathrm{C}(\mathsf{P}_0,\mathsf{P}_i)}.$$

Proof. Letting $\Pi_i = \{\mathsf{P} \in \mathcal{P} \ : \ \mathrm{L}_i(\mathsf{P}) \le 0\}$ for all $i = 1, \ldots, d$, we have that $\Pi^\star = \Pi_1 \cup \Pi_2 \cup \cdots \cup \Pi_d$. Since we know from the proof of Theorem 6.2 that the Π_i's verifies Sanov's theorem hypotheses, it is also the case for $\Pi^\star$. Therefore, from Theorem 6.1 we get

$$\Pr_{\mathsf{H}_0}[\mathsf{A}^\star_q(\mathbf{Z}^q) = 1] \doteq 2^{-q\mathrm{D}(\Pi^\star\|\mathsf{P}_0)}.$$

By definition

$$\mathrm{D}(\Pi^\star\|\mathsf{P}_0) = \min_{P\in\Pi_1\cup\cdots\cup\Pi_d} \mathrm{D}(P\|\mathsf{P}_0) = \min_{1\le i\le d}\min_{P\in\Pi_i} \mathrm{D}(P\|\mathsf{P}_0) = \min_{1\le i\le d} \mathrm{D}(\Pi_i\|\mathsf{P}_0).$$

Since we know from the proof of Theorem 6.2 that $\mathrm{D}(\Pi_i\|\mathsf{P}_0) = \mathrm{C}(\mathsf{P}_0, \mathsf{P}_i)$, we deduce

$$\Pr_{\mathsf{H}_0}[\mathsf{A}^\star_q(\mathbf{Z}^q) = 1] \doteq \max_{1\le i\le d} 2^{-q\mathrm{C}(\mathsf{P}_0,\mathsf{P}_i)}. \tag{6.8}$$

On the other hand, we have

$$\Pr_{\mathsf{H}_1}[\mathsf{A}^\star_q(\mathbf{Z}^q) = 0] = \Pr_{\mathsf{H}_1}[\mathsf{P}_{\mathbf{Z}^q} \notin \Pi_1, \ldots, \mathsf{P}_{\mathbf{Z}^q} \notin \Pi_d] \le \Pr_{\mathsf{H}_1}[\mathsf{P}_{\mathbf{Z}^q} \notin \Pi_1]$$

and we know that $\Pr_{\mathsf{H}_1}[\mathsf{P}_{\mathbf{Z}^q} \notin \Pi_1] \doteq 2^{-q\mathrm{C}(\mathsf{P}_0,\mathsf{P}_1)}$. Since this is clearly less than $\max_{1\le i\le d} 2^{-q\mathrm{C}(\mathsf{P}_0,\mathsf{P}_i)}$, we conclude from this and (6.8) that

$$1 - \mathrm{BestAdv}_q(\mathsf{H}_0, \mathsf{H}_1) \doteq \max_{1\le i\le d} 2^{-q\mathrm{C}(\mathsf{P}_0,\mathsf{P}_i)}.$$

We will now show that this advantage is asymptotically optimal.

Let A_q be an arbitrary q-limited distinguisher between H_0 and H_1 defined by an acceptance region Π, and let Adv_q denote its advantage. We have

$$\begin{aligned} 1 - \mathrm{Adv}_q(\mathsf{H}_0, \mathsf{H}_1) &= \Pr_{\mathsf{H}_0}[\mathsf{A}_q(\mathbf{Z}^q) = 1] + \Pr_{\mathsf{H}_1}[\mathsf{A}_q(\mathbf{Z}^q) = 0] \\ &= \Pr_{\mathsf{H}_0}[\mathsf{A}_q(\mathbf{Z}^q) = 1] + \sum_{i=1}^{d} \pi_i \Pr[\mathsf{A}_q(\mathbf{Z}^q) = 0|\mathsf{P} = \mathsf{P}_i] \\ &= \sum_{i=1}^{d} \pi_i (\Pr_{\mathsf{H}_0}[\mathsf{A}_q(\mathbf{Z}^q) = 1] + \Pr[\mathsf{A}_q(\mathbf{Z}^q) = 0|\mathsf{P} = \mathsf{P}_i]) \\ &\ge \sum_{i=1}^{d} \pi_i (1 - \mathrm{BestAdv}_q(\mathsf{P}_0, \mathsf{P}_i)), \end{aligned}$$

where $\mathrm{BestAdv}_q(\mathsf{P}_0, \mathsf{P}_i)$ denotes the advantage of the best distinguisher between P_0 and P_i. Since, according to Corollary 6.1,

$$1 - \mathrm{BestAdv}_q(\mathsf{P}_0, \mathsf{P}_i) \doteq 2^{-q\mathrm{C}(\mathsf{P}_0,\mathsf{P}_i)}$$

then from Lemma 5.1 we deduce that

$$\sum_{i=1}^{d} \pi_i(1 - \mathrm{BestAdv}_q(\mathsf{P}_0, \mathsf{P}_i)) \doteq 2^{-q \min_{1\leq i\leq d} \mathrm{C}(\mathsf{P}_0,\mathsf{P}_i)} = \max_{1\leq i\leq d} 2^{-q\mathrm{C}(\mathsf{P}_0,\mathsf{P}_i)},$$

which allows to conclude that $1 - \mathrm{Adv}_q(\mathsf{H}_0, \mathsf{H}_1) \geq c_q$ where $c_q \doteq 1 - \mathrm{BestAdv}_q(\mathsf{H}_0, \mathsf{H}_1)$. □

We will use Theorem 6.4 in Section 7.6 to compute the asymptotic value of a generalized version of the best linear distinguisher. In the meantime, we can deduce the following heuristic result.

Heuristic 6.3 Let P_0 be a distribution of support $\mathcal{Z}$ and $\mathcal{D} = \{\mathsf{P}_1, \ldots, \mathsf{P}_d\}$ be a finite set of distributions of support $\mathcal{Z}$. Let

$$\mathrm{C}(\mathsf{P}_0, \mathsf{P}_i) = - \inf_{0<\lambda<1} \log \sum_{z\in\mathcal{Z}} \mathsf{P}_0[z]^{1-\lambda}\mathsf{P}_i[z]^{\lambda}$$

be the Chernoff information between P_0 and P_i, for $i = 1, 2, \ldots, d$. The best q-limited distinguisher between $\mathsf{H}_0 : \mathsf{P} = \mathsf{P}_0$ and $\mathsf{H}_1 : \mathsf{P} \in \mathcal{D}$ reaches a non-negligible advantage when

$$q = \frac{1}{\min\limits_{1\leq i\leq d} \mathrm{C}(\mathsf{P}_0, \mathsf{P}_i)}.$$

When P_0 is uniform, then the best q-limited distinguisher between the two previous hypotheses reaches a non-negligible advantage when

$$q = \frac{8 \ln 2}{\min\limits_{1\leq i\leq d} \Delta(\mathsf{P}_i)}.$$

□

6.9 A General Heuristic Method to Compute the Advantage of an Arbitrary Distinguisher

In the previous sections, we focused on the best distinguisher (either asymptotic or not) and derived its advantage using, essentially, Sanov's theorem (Theorem 6.1). We can learn from the techniques we used and extract a general heuristic method that can be applied to compute the advantage of (almost) *any* q-limited distinguisher $\mathcal{A}_q$ between two simple hypotheses.

We assume that $\mathcal{A}_q$ is defined by an acceptance region

$$\Pi = \{\mathsf{P} \in \mathcal{P} \ : \ \mathrm{L}(\mathsf{P}) \leq 0\}$$

where L is some continuous function, and that both Π and its complement satisfy Sanov's theorem hypotheses. This is notably the case of any non-empty convex set. We also assume that the distributions we are considering are close to each other.

Since Π satisfies Sanov's theorem hypotheses, we know that

$$\Pr_{\mathsf{H}_0}[\mathsf{A}_q(\mathbf{Z}^q)=1] \doteq 2^{-q\mathrm{D}(\Pi\|\mathsf{P}_0)}, \tag{6.9}$$

so that all we need to do is to approximate $\mathrm{D}(\Pi\|\mathsf{P}_0)=\min_{P\in\Pi}\mathrm{D}(P\|\mathsf{P}_0)$. Similarly to what we had in the proof of Theorem 6.2 we know that (according to the method of Lagrange multipliers) the minimum can only be achieved for a distribution P such that

$$\nabla\mathrm{D}(P\|\mathsf{P}_0)=\lambda\nabla\mathrm{L}(P)+\mu\nabla\mathrm{N}(P)$$

for some λ and μ, where $\mathrm{N}(P)=\sum_{z\in\mathcal{Z}}P[z]$, under the constraint that $\mathrm{L}(P)=0$ and that $\mathrm{N}(P)=1$.

For any distribution P converging towards P_0 as $q\to\infty$, we have

$$\begin{aligned}\mathrm{D}(P\|\mathsf{P}_0) &= \sum_{z\in\mathcal{Z}}P[z]\log\frac{P[z]}{\mathsf{P}_0[z]}\\ &= \frac{1}{2\ln 2}\sum_{z\in\mathcal{Z}}\frac{(P[z]-\mathsf{P}_0[z])^2}{\mathsf{P}_0[z]}+o\left(\sum_{z\in\mathcal{Z}}\frac{(P[z]-\mathsf{P}_0[z])^2}{\mathsf{P}_0[z]}\right).\end{aligned}$$

When P is fixed but close to P_0, we can approximate $\mathrm{D}(P\|\mathsf{P}_0)$ by the first term of the right-hand side of the previous equation and deduce that

$$\frac{\partial\mathrm{D}(P\|\mathsf{P}_0)}{\partial P[a]}\approx\frac{P[a]-\mathsf{P}_0[a]}{\mathsf{P}_0[a]\ln 2}$$

for all $a\in\mathcal{Z}$. We deduce that the distribution P for which $\mathrm{D}(P\|\mathsf{P}_0)$ is minimal satisfies

$$\frac{P[a]-\mathsf{P}_0[a]}{\mathsf{P}_0[a]\ln 2}\approx\lambda\frac{\partial\mathrm{L}(P)}{\partial P[a]}+\mu$$

from which we deduce that $P[a]-\mathsf{P}_0[a]=\lambda\mathsf{P}_0[a]\ln 2\frac{\partial\mathrm{L}(P)}{\partial P[a]}+\mu\mathsf{P}_0[a]\ln 2$. Summing over $a\in\mathcal{Z}$ leads to $\lambda\sum_a\mathsf{P}_0[a]\frac{\partial\mathrm{L}(P)}{\partial P[a]}+\mu=0$, from which we deduce that the distribution P for which $\mathrm{D}(P\|\mathsf{P}_0)$ is minimal must satisfy

$$\mathrm{L}(P)=0\quad\text{and}\quad\frac{P[a]-\mathsf{P}_0[a]}{\mathsf{P}_0[a]\ln 2}\approx\lambda\sum_{z\in\mathcal{Z}}\mathsf{P}_0[z]\left(\frac{\partial\mathrm{L}}{\partial P[a]}-\frac{\partial\mathrm{L}}{\partial P[z]}\right) \tag{6.10}$$

for some constant λ. Equation (6.10) leads to a system of equations with $|\mathcal{Z}|+1$ unknowns in total. Solving this system allows to estimate $\mathrm{D}(\Pi\|\mathsf{P}_0)$ and thus the error probability $\Pr_{\mathsf{H}_0}[\mathsf{A}_q(\mathbf{Z}^q)=1]$ by approximating the left-hand side of (6.9) by its right-hand side.

The computation of $\mathrm{D}(\overline{\Pi}\|\mathsf{P}_1)$ can be performed in a similar way and leads to an approximation of the other error probability. Finally, the approximate value of the advantage is easy to deduce from the previous computations.

6.10 Case where One of the Distributions is Unknown: the *Squared* Distinguishers Family

In order to implement the best distinguisher between $\mathsf{H}_0 : \mathsf{P} = \mathsf{P}_0$ and $\mathsf{H}_1 : \mathsf{P} = \mathsf{P}_1$, the precise knowledge of *both* distributions is needed, since the likelihood ratio depends on them. This is also the case for achieving the best asymptotic advantage when testing a simple hypothesis against a composite one, as shown in Section 6.8. Yet, this information might not be always available in practice (e.g., when attacking a pseudo-random generator which specifications are unknown). In this subsection, we investigate what can be done in the situation where the distinguisher has only access to *one* of the two distributions. In this case, the *null hypothesis* is $\mathsf{H}_0 : \mathsf{P} = \mathsf{P}_0$ and the *alternate hypothesis* is $\mathsf{H}_1 : \mathsf{P} \neq \mathsf{P}_0$. We assume that the adversary has precise knowledge of P_0.

Definition 6.11 *(**χ^2 statistics**) Let P_0 be a probability distribution over a finite set $\mathcal{Z}$. Let q be a positive integer. Pearson's chi-square (χ^2) statistic [129] is the function*

$$\begin{array}{rcl} \Sigma_{\chi^2} : \mathcal{Z}^q & \longrightarrow & \mathbf{R} \\ \mathbf{z}^q & \longmapsto & \Sigma_{\chi^2}(\mathbf{z}^q) = q \sum_{a \in \mathcal{Z}} \frac{(\mathsf{P}_{\mathbf{z}^q}[a] - \mathsf{P}_0[a])^2}{\mathsf{P}_0[a]}. \end{array}$$

The logarithmic likelihood ratio (G^2) statistic is the function

$$\begin{array}{rcl} \Sigma_{G^2} : \mathcal{Z}^q & \longrightarrow & \mathbf{R} \\ \mathbf{z}^q & \longmapsto & \Sigma_{G^2}(\mathbf{z}^q) = 2q \sum_{a \in \mathcal{Z}} \mathsf{P}_{\mathbf{z}^q}[a] \ln \frac{\mathsf{P}_{\mathbf{z}^q}[a]}{\mathsf{P}_0[a]}. \end{array}$$

The Freeman-Tukey (T^2) statistic [52] is the function

$$\begin{array}{rcl} \Sigma_{T^2} : \mathcal{Z}^q & \longrightarrow & \mathbf{R} \\ \mathbf{z}^q & \longmapsto & \Sigma_{T^2}(\mathbf{z}^q) = 4q \sum_{a \in \mathcal{Z}} \left(\sqrt{\mathsf{P}_{\mathbf{z}^q}[a]} - \sqrt{\mathsf{P}_0[a]} \right)^2. \end{array}$$

The Neyman modified chi-square (NM^2) statistic is the function

$$\begin{array}{rcl} \Sigma_{NM^2} : \mathcal{Z}^q & \longrightarrow & \mathbf{R} \\ \mathbf{z}^q & \longmapsto & q\Sigma_{NM^2}(\mathbf{z}^q) = q \sum_{a \in \mathcal{Z}} \frac{(\mathsf{P}_{\mathbf{z}^q}[a] - q\mathsf{P}_0[a])^2}{\mathsf{P}_{\mathbf{z}^q}[a]}. \end{array}$$

The modified logarithmic likelihood ratio (GM^2) statistic is the function

$$\begin{array}{rcl} \Sigma_{GM^2} : \mathcal{Z}^q & \longrightarrow & \mathbf{R} \\ \mathbf{z}^q & \longmapsto & \Sigma_{G^2}(\mathbf{z}^q) = 2q \sum_{a \in \mathcal{Z}} \mathsf{P}_0[a] \ln \frac{\mathsf{P}_{\mathbf{z}^q}[a]}{\mathsf{P}_0[a]}. \end{array}$$

Let $X \in \{\chi, G, T, NM, GM\}$. A q-limited distinguisher $\mathsf{P}_q^{X^2}(T)$ defined by the sample acceptance region

$$\mathcal{A}_q^{X^2}(T) = \{\mathbf{z}^q \in \mathcal{Z}^q \ : \ \Sigma_{X^2}(\mathbf{z}^q) > qT\}$$

is a distinguisher performing a X^2 statistical test with threshold T.

Intuitively, if the null hypothesis is true, the value of each of these X^2 statistical tests Σ_{X^2} should be small since the experimental frequencies should be close to the expected ones, so that $\mathsf{P}_{\mathbf{z}^q}[a] - \mathsf{P}_0[a]$ should be close to 0 for all $a \in \mathcal{Z}$. In the simple situation where only two distributions are possible, larger values of the tests allow one to conclude that the source follows P_1.

Presumably because of its simplicity, the Pearson's chi-square (χ^2) statistic [129] is the most commonly used. According to Horn [68], the reason why the G^2 was not as used as the χ^2 in the past may lie in the difficulty of undertaking logarithmic computations. Modern computers obviously erased this drawback.

In what follows, we recall the proof of Cressie and Read [39] who show that all the X^2 tests introduced in Definition 6.11 are asymptotically equivalent. As a consequence, all the distinguishers performing a X^2 statistical test are equivalent from the point of view of the advantage when q is large enough. We will thus focus on the test which makes it possible to easily compute the advantage using Theorem 6.1, in a similar way than what we did for the best distinguisher in Theorem 6.2.

All the X^2 Tests Asymptotically Follow a χ^2 Distribution.

We first introduce Cressie and Read λ power statistic and show that it actually encompass all the X^2 tests introduced so far.

Definition 6.12 *Let P_0 be a probability distribution over a finite set $\mathcal{Z}$. Let q be a positive integer and let $\lambda \in \mathbf{R}$. For $\lambda \neq 0, -1$, the CR^λ power statistic is the function*

$$\begin{array}{rcl} \Sigma_{CR^\lambda} : \mathcal{Z}^q & \longrightarrow & \mathbf{R} \\ \mathbf{z}^q & \longmapsto & \Sigma_{CR^\lambda}(\mathbf{z}^q) = \dfrac{1}{\lambda(\lambda+1)} \displaystyle\sum_{a \in \mathcal{Z}} \mathsf{P}_{\mathbf{z}^q}[a] \left(\left(\frac{\mathsf{P}_{\mathbf{z}^q}[a]}{\mathsf{P}_0[a]} \right)^\lambda - 1 \right). \end{array}$$

For $\lambda = 0$ and $\lambda = -1$, Σ_{CR^λ} is defined by continuity:

$$\Sigma_{CR^0}(\mathbf{z}^q) = \sum_{a \in \mathcal{Z}} \mathsf{P}_{\mathbf{z}^q}[a] \ln \frac{\mathsf{P}_{\mathbf{z}^q}[a]}{\mathsf{P}_0[a]} \quad \textit{and} \quad \Sigma_{CR^{-1}}(\mathbf{z}^q) = \sum_{a \in \mathcal{Z}} \mathsf{P}_0[a] \ln \frac{\mathsf{P}_0[a]}{\mathsf{P}_{\mathbf{z}^q}[a]}.$$

Lemma 6.8 *Under the notations of definitions 6.11 and 6.12, we have:*

$$\begin{array}{rcl} \Sigma_{\chi^2}(\mathbf{z}^q) & = & 2q\Sigma_{CR^1}(\mathbf{z}^q), \\ \Sigma_{G^2}(\mathbf{z}^q) & = & 2q\Sigma_{CR^0}(\mathbf{z}^q), \\ \Sigma_{T^2}(\mathbf{z}^q) & = & 2q\Sigma_{CR^{-1/2}}(\mathbf{z}^q), \\ \Sigma_{NM^2}(\mathbf{z}^q) & = & 2q\Sigma_{CR^{-2}}(\mathbf{z}^q) \\ \Sigma_{GM^2}(\mathbf{z}^q) & = & 2q\Sigma_{CR^{-1}}(\mathbf{z}^q) \end{array}$$

We will now show that all the CR^λ tests are asymptotically equivalent, which obviously implies the same for the X^2 tests.

Lemma 6.9 *Let P_0 be a probability distribution over a finite set $\mathcal{Z}$ and $\lambda \in \mathbf{R}$. Let $z_1, z_2, \ldots, z_q \in \mathcal{Z}$ be a sequence of q elements. Let*

$$\epsilon_a = \frac{\mathsf{P}_{\mathbf{z}^q}[a] - \mathsf{P}_0[a]}{\mathsf{P}_0[a]}$$

for all $a \in \mathcal{Z}$. Under the null hypothesis, we have $\epsilon_a \to 0$ when $q \to \infty$ for all $a \in \mathcal{Z}$ and

$$\Sigma_{CR^\lambda}(\mathbf{z}^q) = \frac{1}{2}\sum_{a\in\mathcal{Z}} \mathsf{P}_0[a](\epsilon_a^2 + o(\epsilon_a^2)).$$

Proof. For $\lambda \neq 0, -1$ it is easy to show that

$$\Sigma_{CR^\lambda}(\mathbf{z}^q) = \frac{1}{\lambda(\lambda+1)}\sum_{a\in\mathcal{Z}} \mathsf{P}_0[a]\left((1+\epsilon_a)^{\lambda+1} - (1+\epsilon_a)\right).$$

Since

$$(1+\epsilon_a)^{\lambda+1} - (1+\epsilon_a) = \lambda\epsilon_a + \frac{\lambda(\lambda+1)}{2}\epsilon_a^2 + o(\epsilon_a)$$

and since $\sum_a \mathsf{P}_0[a]\epsilon_a = 0$ we obtain the announced result for $\lambda \neq 0, -1$. Similarly it can be shown that

$$\Sigma_{CR^0}(\mathbf{z}^q) = \sum_{a\in\mathcal{Z}} \mathsf{P}_0[a](1+\epsilon_a)\ln(1+\epsilon_a),$$

which leads to the announced result for $\lambda = 0$ using the fact that

$$(1+\epsilon_a)\ln(1+\epsilon_a) = \epsilon_a + \frac{1}{2}\epsilon_a^2 + o(\epsilon_a^2).$$

For $\lambda = -1$ we have

$$\Sigma_{CR^{-1}}(\mathbf{z}^q) = -\sum_{a\in\mathcal{Z}} \mathsf{P}_0[a]\ln(1+\epsilon_a)$$

and

$$-\ln(1+\epsilon_a) = -\epsilon_a + \frac{1}{2}\epsilon_a^2 + o(\epsilon_a^2)$$

which completes the proof. □

As a consequence of lemmas 6.8 and 6.9, we have that for all $\lambda \in \mathbf{R}$,

$$2q\Sigma_{CR^\lambda}(\mathbf{z}^q) \sim 2q\Sigma_{CR^1}(\mathbf{z}^q) = \Sigma_{\chi^2}(\mathbf{z}^q)$$

as $q \to \infty$. It is well known that under the null hypothesis (i.e., when the distribution followed by the source if P_0) Pearson's χ^2 statistical test converges towards a χ^2 distribution with $|\mathcal{Z}| - 1$ degrees of freedom [15, 70, 98]. We have just shown that this is consequently the case for all the tests introduced in this section. Since we don't assume anything about the alternate case, this also mean that all the distinguishers based on these respective tests are *a priori* equivalent in terms of advantage when q becomes large. In what follows, we focus on the G^2 test and compute the advantage of $\mathsf{A}_q^{G^2}(T)$.

Computing the Advantage of a G^2 Distinguisher

We first note that the G^2 statistic can be expressed in terms of the relative entropy between the type of $\mathbf{z}^q$ and the distribution P_0 as

$$\Sigma_{G^2}(\mathbf{z}^q) = 2q \sum_{a \in \mathcal{Z}} \mathsf{P}_0[a] \ln \frac{\mathsf{P}_{\mathbf{z}^q}[a]}{\mathsf{P}_0[a]} = 2\ln(2) q \mathrm{D}(\mathsf{P}_{\mathbf{z}^q} \| \mathsf{P}_0).$$

The sample acceptance region of a G^2 distinguisher is

$$\mathcal{A}_q^{G^2}(T) = \{\mathbf{z}^q \in \mathcal{Z}^q \ : \ \Sigma_{G^2}(\mathbf{z}^q) > qT\} = \{\mathbf{z}^q \in \mathcal{Z}^q \ : \ \mathrm{D}(\mathsf{P}_{\mathbf{z}^q} \| \mathsf{P}_0) > T'\},$$

where $T' = T/(2\ln 2)$. Thus, exchanging T by T' in Definition 6.11 only "shifts" the advantage of the distinguishers. For simplicity, we adopt from now on the following definition for a G^2 distinguisher.

Definition 6.13 *Let P_0 be a probability distribution over a finite set $\mathcal{Z}$ and let*

$$\Pi^{G^2}(T) = \{\mathsf{P} \in \mathcal{P} \ : \ \mathrm{D}(\mathsf{P} \| \mathsf{P}_0) > T\}.$$

The q-limited distinguisher $\mathsf{A}_q^{G^2}(T)$ defined by the type acceptance region $\Pi_q^{G^2}(T) = \Pi^{G^2}(T) \cap \mathcal{P}_q$ is a G^2 distinguisher with threshold T.

From Proposition 6.4, we note that the type acceptance region of a G^2 distinguisher corresponds to the one of the perfect distinguisher, except that the $\mathrm{D}(\mathsf{P}\|\mathsf{P}_1)$ term is now replaced by a constant T.

The following heuristic theorem describes the asymptotic behavior of a distinguisher $\mathsf{A}_q^{G^2}$ in a simplified situation where the alternate hypothesis is simple, $\mathsf{H}_1 : \mathsf{P} = \mathsf{P}_1$ (where P_1 is close to P_0), but where P_1 is *unknown* to the distinguisher.

Theorem 6.5 *Let P_0 and P_1 be two probability distributions of full support over a finite set $\mathcal{Z}$. For all $z \in \mathcal{Z}$ let*

$$\mathsf{P}_0[z] = p_z, \quad \mathsf{P}_1[z] = p_z + \delta_z \quad \textit{and} \quad \epsilon_z = \frac{\delta_z}{p_z}.$$

Let $0 < T < \max_z \log \frac{1}{\mathsf{P}_0[z]}$. Restricting Taylor series expansion in terms of the ϵ_z's to order 2, the advantage of a q-limited G^2 distinguisher with threshold T between $\mathsf{H}_0 : \mathsf{P} = \mathsf{P}_0$ and $\mathsf{H}_1 : \mathsf{P} = \mathsf{P}_1$ (P_1 being unknown to the distinguisher) is such that

$$1 - \mathrm{Adv}_q^{G^2}(\mathsf{H}_0, \mathsf{H}_1) \doteq 2^{-q \min(T, (\sqrt{T} - \sqrt{\rho})^2)}, \tag{6.11}$$

where

$$\rho = \frac{1}{2}\sum_{z\in\mathcal{Z}} p_z\epsilon_z^2.$$

Proof. Let $\mathrm{L}(\mathsf{P}) = T - \mathrm{D}(\mathsf{P}\|\mathsf{P}_0)$. According to Definition 6.13, the type acceptance region of $\mathrm{A}_q^{G^2}(T)$ is $\Pi_q^{G^2}(T) = \Pi^{G^2}(T) \cap \mathcal{P}_q$ where

$$\Pi^{G^2}(T) = \{\mathsf{P} \in \mathcal{P} \, : \, \mathrm{D}(\mathsf{P}\|\mathsf{P}_0) > T\}.$$

Clearly $\Pi^{G^2}(T)$ is not empty since the distribution P which is always zero except at the point which minimizes P_0 belongs to $\Pi^{G^2}(T)$ since $\mathrm{D}(P\|\mathsf{P}_0) = \log\frac{1}{\min_z \mathsf{P}_0[z]} = \max_z \log\frac{1}{\mathsf{P}_0[z]} > T$ by assumption. On the other hand, $\Pi^{G^2}(T)^c$ is non-empty either since P_0 belongs to it. The continuity of L ensures that the hypotheses of Sanov's theorem (Theorem 6.1) are verified, and thus we get

$$1 - \mathrm{Adv}_q^{G^2}(T) \doteq 2^{q\min(\mathrm{D}(\Pi^{G^2}(T)\|\mathsf{P}_0),\mathrm{D}(\Pi^{G^2}(T)^c\|\mathsf{P}_1))}.$$

By definition we easily obtain that

$$\mathrm{D}(\Pi^{G^2}(T)\|\mathsf{P}_0) = \inf_{P\in\Pi^{G^2}(T)} \mathrm{D}(P\|\mathsf{P}_0) = T.$$

Computing $\mathrm{D}(\Pi^{G^2}(T)\|\mathsf{P}_1)$ is more involving. For similar reasons than those we had in the proof of Theorem 6.2, this computation reduces to an optimization problem in which we must minimize $\mathsf{P} \mapsto \mathrm{D}(\mathsf{P}\|\mathsf{P}_1)$ under the constraints that $\mathrm{L}(\mathsf{P}) = 0$ and that $\mathrm{N}(\mathsf{P}) = \sum_{z\in\mathcal{Z}} \mathsf{P}[z] = 1$. According to the method of Lagrange's multipliers, a minimum can only be obtained in a point P such that

$$\nabla\mathrm{D}(\mathsf{P}\|\mathsf{P}_1) = \lambda\nabla\mathrm{L}(\mathsf{P}) + \mu\nabla\mathrm{N}(\mathsf{P})$$

for some $\lambda, \mu \in \mathbf{R}$. Solving the previous equation under the two constraints leads to a solution of the form

$$\mathsf{P}[a] = \frac{\mathsf{P}_1[a]^{\frac{1}{1+\lambda}}\mathsf{P}_0[a]^{\frac{\lambda}{1+\lambda}}}{\sum_b \mathsf{P}_1[b]^{\frac{1}{1+\lambda}}\mathsf{P}_0[b]^{\frac{\lambda}{1+\lambda}}} = \frac{\mathsf{P}_0[a]\left(\frac{\mathsf{P}_1[a]}{\mathsf{P}_0[a]}\right)^{\mu}}{\sum_b \mathsf{P}_0[b]\left(\frac{\mathsf{P}_1[b]}{\mathsf{P}_0[b]}\right)^{\mu}},$$

where $\mu = 1/(1+\lambda)$. Introducing the notations of the theorem, the previous equation becomes

$$\mathsf{P}[a] = \frac{p_a\,(1+\epsilon_a)^{\mu}}{\sum_b p_b\,(1+\epsilon_b)^{\mu}}. \tag{6.12}$$

From the expression obtained for $\mathsf{P}[a]$, we can compute $\mathrm{D}(\mathsf{P}\|\mathsf{P}_0)$:

$$\begin{aligned}\mathrm{D}(\mathsf{P}\|\mathsf{P}_0) &= \sum_a \frac{p_a\,(1+\epsilon_a)^{\mu}}{\sum_b p_b\,(1+\epsilon_b)^{\mu}}\log\left(\frac{(1+\epsilon_a)^{\mu}}{\sum_b p_b\,(1+\epsilon_b)^{\mu}}\right)\\ &= \frac{\mu\sum_a p_a\,(1+\epsilon_a)^{\mu}\log(1+\epsilon_a)}{\sum_b p_b\,(1+\epsilon_b)^{\mu}} - \log\left(\sum_a p_a\,(1+\epsilon_a)^{\mu}\right).\end{aligned}$$

Developing the last equation using Taylor series, we obtain at order 2

$$\mathsf{D}(\mathsf{P}\|\mathsf{P}_0) \approx \frac{1}{2}\mu^2 \sum_a p_a \epsilon_a^2,$$

so that the condition $\mathsf{D}(\mathsf{P}\|\mathsf{P}_0) = T$ gives $\mu \approx \sqrt{T/\rho}$ using the notations of the theorem. Similarly, we have

$$\begin{aligned}\mathsf{D}(\mathsf{P}\|\mathsf{P}_1) &= \sum_a \frac{p_a(1+\epsilon_a)^\mu}{\sum_b p_b(1+\epsilon_b)^\mu} \log \frac{(1+\epsilon_a)^{\mu-1}}{\sum_b p_b(1+\epsilon_b)^\mu} \\ &= \frac{(\mu-1)\sum_a p_a(1+\epsilon_a)^\mu \log(1+\epsilon_a)}{\sum_b p_b(1+\epsilon_b)^\mu} - \log\left(\sum_b p_b(1+\epsilon_b)^\mu\right).\end{aligned}$$

Developing the last equation in Taylor series again, we get

$$\mathsf{D}(\mathsf{P}\|\mathsf{P}_1) \approx \frac{(\mu-1)^2}{2} \sum_a p_a \epsilon_a^2.$$

Since $\mu \approx \sqrt{T/\rho}$ we finally obtain

$$\mathsf{D}(\mathsf{P}\|\mathsf{P}_1) \approx (\sqrt{T/\rho} - 1)^2 \rho = (\sqrt{T} - \sqrt{\rho})^2.$$

□

On the contrary of Theorem 6.2 which gives an rigorous expression of the advantage of the best distinguisher, Theorem 6.5 is only heuristic in the sense that we assume that the approximation of ρ obtained using Taylor series can be used in place of its exact value. Assuming that the advantage of the G^2 distinguisher can be approximated by its asymptotic value, we can compare its efficiency with respect to that of the best distinguisher. According to Heuristic 6.2 we can assume that the best distinguisher reaches a non-negligible advantage when

$$q_{\text{best}} = \frac{8 \ln 2}{\sum_{z\in\mathcal{Z}} \mathsf{P}_0[z]\epsilon_z^2}.$$

From Theorem 6.5, we can similarly assume that

$$q_{\chi^2} = \frac{1}{\min(T, (\sqrt{T} - \sqrt{\rho})^2)}$$

are sufficient to the G^2 distinguisher to distinguish H_0 from H_1. This value is minimized by choosing $T = \frac{1}{4}\rho$. In the best case from the point of view of the distinguisher, the previous equation thus reads

$$q_{\chi^2} = \frac{8}{\sum_{z\in\mathcal{Z}} \mathsf{P}_0[z]\epsilon_z^2},$$

which is of the same order of magnitude as q_{best}.

Obviously, the optimal choice of the threshold cannot always been made in practice since we assumed that the distribution P_1 is unknown. Yet, if at least the "distance" between P_0 and P_1 can be evaluated, then it not necessary to know the exact details of P_1 in order to obtain an efficient G^2 distinguisher. In that case, a G^2 distinguisher with an optimal threshold behaves just as well as the best distinguisher. This confirms previous results of Vaudenay [151], who obtained the same conclusions but under stronger assumptions on the sample distribution and considering an simpler way of measuring the efficiency of a distinguisher instead of the classical notion of advantage.

Non-asymptotic case: The best χ^2-like Test in Practice

The results presented so far are asymptotic. Whereas we considered that all χ^2-like tests are equivalent (as they all follow a χ^2 distribution with $\nu = |\mathcal{Z}| - 1$ degrees of freedom as $q \to \infty$), some of them might be more accurate than others in practice (for bounded values of q), namely, those that converge faster to the asymptotic distribution. To find out which of these tests is best, we compare the mean of the statistic to that of the asymptotic distribution, and choose the value of λ for which the convergence is the fastest possible.

Proposition 6.5 *Let $\mathcal{Z}$ be a finite set, P_0 be a probability distribution over $\mathcal{Z}$, and $Z_1, Z_2, \ldots, Z_q \sim \mathsf{P}_0$ be i.i.d. random variables. Let*

$$p_a = \mathsf{P}_0[a], \quad \delta_a = \mathsf{P}_{\mathbf{Z}^q}[a] - \mathsf{P}_0[a], \quad \text{and} \quad \epsilon_a = \frac{\delta_a}{p_a}$$

for all $a \in \mathcal{Z}$. Letting $S = \sum_a \frac{1}{p_a}$ we have

$$\mathrm{E}\left(2q\Sigma_{CR^\lambda}(\mathbf{z}^q)\right) = |\mathcal{Z}| - 1 + \frac{1}{q}\Bigg(\frac{(\lambda-1)}{3}(S - 3|\mathcal{Z}| + 2) + \frac{(\lambda-1)(\lambda-2)}{4}(S - 2|\mathcal{Z}| + 1)\Bigg) + o(q^{-3/2}).$$

Proof. Since

$$\frac{\epsilon_a}{1/q} = q\epsilon_a = \frac{q\delta_a}{p_a} = \frac{q(\mathsf{P}_{\mathbf{Z}^q}[a] - \mathsf{P}_0[a])}{p_a} = \frac{(\mathrm{N}[a|\mathbf{Z}^q] - q\mathsf{P}_0[a])}{p_a},$$

the law of large numbers guarantees that, under the null hypothesis, $\frac{\epsilon_a}{1/q} \to 0$ as $q \to \infty$ with probability 1. Therefore, $\epsilon_a = o(1/q)$ for all $a \in \mathcal{Z}$ with probability 1. With these notations, we have

$$
\begin{aligned}
&2q\Sigma_{CR^\lambda}(\mathbf{Z}^q)\\
&= \frac{2q}{\lambda(\lambda+1)}\sum_{a\in\mathcal{Z}} p_a(1+\epsilon_a)\left((1+\epsilon_a)^\lambda - 1\right)\\
&= q\sum_{a\in\mathcal{Z}} p_a\left(\lambda\epsilon_a + \epsilon_a^2 + \frac{(\lambda-1)}{3}\epsilon_a^3 + \frac{(\lambda-1)(\lambda-2)}{12}\epsilon_a^4\right) + o(1/q^3)\\
&= \sum_{a\in\mathcal{Z}}\left(qp_a\epsilon_a^2 + \frac{(\lambda-1)}{3}qp_a\epsilon_a^3 + \frac{(\lambda-1)(\lambda-2)}{12}qp_a\epsilon_a^4\right) + o(1/q^3).
\end{aligned}
$$

It can be shown that

$$
\begin{aligned}
\mathrm{E}(\mathrm{N}[a|\mathbf{Z}^q]) &= \mathrm{E}\left(\sum_i \mathbf{1}_{Z_i=a}\right) = \sum_i \mathrm{E}\left(\mathbf{1}_{Z_i=a}\right) = qp_a,\\
\mathrm{E}(\mathrm{N}[a|\mathbf{Z}^q]^2) &= \sum_{i,j}\mathrm{E}\left(\mathbf{1}_{Z_i=a}\mathbf{1}_{Z_j=a}\right) = \sum_i p_a + \sum_{i,j\neq i} p_a^2 = qp_a + q(q-1)p_a^2,\\
\mathrm{E}(\mathrm{N}[a|\mathbf{Z}^q]^3) &= \sum_{i,j,k}\mathrm{E}\left(\mathbf{1}_{Z_i=a}\mathbf{1}_{Z_j=a}\mathbf{1}_{Z_k=a}\right)\\
&= qp_a + 3q(q-1)p_a^2 + q(q-1)(q-2)p_a^3,\\
\mathrm{E}(\mathrm{N}[a|\mathbf{Z}^q]^4) &= \sum_{i,j,k,\ell}\mathrm{E}\left(\mathbf{1}_{Z_i=a}\mathbf{1}_{Z_j=a}\mathbf{1}_{Z_k=a}\mathbf{1}_{Z_\ell=a}\right)\\
&= qp_a + 7q(q-1)p_a^2 + 6q(q-1)(q-2)p_a^3 + q(q-1)(q-2)(q-3)p_a^4.
\end{aligned}
$$

$$
\begin{aligned}
q\mathrm{E}(\mathsf{P}_{\mathbf{Z}^q}[a]) &= qp_a,\\
\frac{q\mathrm{E}(\mathsf{P}_{\mathbf{Z}^q}[a]^2)}{p_a} &= 1 + (q-1)\,p_a,\\
\frac{q\mathrm{E}(\mathsf{P}_{\mathbf{Z}^q}[a]^3)}{p_a^2} &= \frac{1}{qp_a} + 3 - \frac{3}{q} + \left(q - 3 + \frac{2}{q}\right)p_a,\\
\frac{q\mathrm{E}(\mathsf{P}_{\mathbf{Z}^q}[a]^4)}{p_a^3} &= \frac{7}{qp_a} + 6 - \frac{18}{q} + \left(q - 6 + \frac{11}{q}\right)p_a + o(q^{-3/2}).
\end{aligned}
$$

From the previous equations, we compute respective the values of $qp_a\mathrm{E}(\epsilon_a^2)$, $qp_a\mathrm{E}(\epsilon_a^3)$, and $qp_a\mathrm{E}(\epsilon_a^4)$ and obtain

$$
\begin{aligned}
qp_a\mathrm{E}(\epsilon_a^2) &= 1 - p_a,\\
qp_a\mathrm{E}(\epsilon_a^3) &= \frac{1}{q}\left(\frac{1}{p_a} - 3 + 2p_a\right),\\
qp_a\mathrm{E}(\epsilon_a^4) &= \frac{3}{q}\left(\frac{1}{p_a} - 2 + p_a\right) + o(q^{-3/2}).
\end{aligned}
$$

Plugging these three values in the expression of $\mathrm{E}(2q\Sigma_{CR^\lambda}(\mathbf{Z}^q))$ we get the announced result. □

Since the mean of a χ^2 distribution with $\nu = |\mathcal{Z}| - 1$ degrees of freedom is $\nu = |\mathcal{Z}| - 1$, we see that the best χ^2-like test (in general) is the one such that

$$\frac{(\lambda-1)}{3}(S - 3\,|\mathcal{Z}| + 2) + \frac{(\lambda-1)(\lambda-2)}{4}\,(S - 2\,|\mathcal{Z}| + 1)$$

is zero, which is obviously the case for $\lambda = 1$, the second root depending on the values of S and $|\mathcal{Z}|$. Of course, in certain particular situations, another test might just behave as well. For example, in the particular case where $S - 3\,|\mathcal{Z}| + 2 = 0$, the test corresponding to $\lambda = 2$ appears to be just as good with respect to the speed at which the mean tends towards that of the asymptotic distribution. Applying the same approach than the one proposed here (and using moments of higher order), Cressie and Read concluded in [39] that the value $\lambda = \frac{2}{3}$ leads to particularly good results in practice (for certain fixed values of S and $|\mathcal{Z}|$). Using Lemma 6.8 we conclude as follows.

Proposition 6.6 *In the general case, the best χ^2-like statistical test among the family of tests included in the CR^λ power statistics is the one such that $\lambda = 1$, that is, Pearson's chi-square (χ^2) statistic.*

Chapter 7

Projection-Based Distinguishers Between two Sources

7.1 On the Need for New Distinguishers

In the general case, the memory requirement of a distinguisher between two probability distributions on a set $\mathcal{Z}$ is, roughly, the sum of a quantity proportional to $q \log |\mathcal{Z}|$ (in order to store the q samples sent by the source) and of a quantity proportional to $|\mathcal{Z}|$ (in order to store the descriptions of both P_0 and P_1). In the case of the best distinguisher, this requirements drops down to $O(|\mathcal{Z}|)$ when implementing it as described in Algorithm 7.1. In this case, one essentially only needs to store the description of both probability distributions, so that the memory requirement of the best distinguisher is proportional to $|\mathcal{Z}|$. We also note that since both $\mathrm{LR}(\mathbf{z}^q)$ and $\mathrm{LLR}(\mathbf{z}^q)$ are computed for occurring $\mathbf{z}^q$, the `NaN` case cannot occur.

The memory requirement of the χ^2 distinguisher is essentially the same, since one needs to store $|\mathcal{Z}|$ counters to compute the chi-square test of the q samples.

However, the cardinality of $\mathcal{Z}$ might be large in practice so that very often neither the best distinguisher nor the χ^2 distinguisher can be implemented. This is the case for example when considering modern block ciphers, which typically output 128 bit strings. Since we have $|\mathcal{Z}| = 2^{256}$ in this case (see Section 8.1), implementing the

```
Storage: Two probability distributions $\mathsf{P}_0$ and $\mathsf{P}_1$ over a finite set $\mathcal{Z}$, a counter
         $\mathsf{llr} \in \mathbf{R} \cup \{-\infty, +\infty\}$.
$\mathsf{llr} \leftarrow 0$
for $i = 0, \ldots, q$ do
    Receive $z_i \in \mathcal{Z}$ from the source $S$
    $\mathsf{llr} \leftarrow \mathsf{llr} + \log \mathsf{P}_0[z_i] - \log \mathsf{P}_1[z_i]$
end
if $\mathsf{llr} \leq 0$ then return $1$ else return $0$
```

Algorithm 7.1: Implementing the best q-limited distinguisher $\mathsf{A}_q^\star$ between two probability distributions P_0 and P_1 over a finite set $\mathcal{Z}$.

distinguishers presented in Chapter 6 is inconceivable.

7.2 Best Distinguisher made Practical Using Compression

We consider the same game than in Chapter 6, except that we assume now that the cardinality of the set from which the samples are drawn is large, so that none of the previous distinguishers can be implemented directly. From now on, we denote by $\mathcal{L}$ this set of large cardinality. In the simple hypothesis case, we denote by $\widetilde{\mathsf{P}}_0$ and $\widetilde{\mathsf{P}}_1$ the two possible distributions that the source can follow. To deal with this situation, a possible solution is to reduce the samples' size by means of a projection [7, 151, 163]

$$h : \mathcal{L} \longrightarrow \mathcal{Z},$$

where $\mathcal{Z}$ is a finite set of "reasonable" cardinality. If $L \in \mathcal{L}$ is a random variable of distribution $\widetilde{\mathsf{P}}_i$, this projection defines a random variable $Z = h(L)$ of distribution P_i, with $i \in \{0, 1\}$.

It is usually convenient to restrict to *balanced* projection as a uniform distribution on $\mathcal{L}$ leads to a uniform distribution on $\mathcal{Z}$ in this case.

Definition 7.1 *Let $\mathcal{L}$ and $\mathcal{Z}$ be two finite sets such that $|\mathcal{Z}|$ divides $|\mathcal{L}|$. A function $h : \mathcal{L} \longrightarrow \mathcal{Z}$ is said to be* balanced *if, for all $z \in \mathcal{Z}$, the subset $h^{-1}(z) \subset \mathcal{L}$ of all preimages of z by h is such that*

$$\left|h^{-1}(z)\right| = \frac{|\mathcal{L}|}{|\mathcal{Z}|}.$$

We call *projection-based distinguishers* the class of distinguishers (see Definition 6.1) that reduce the sample space using a projection before trying to distinguish an hypothesis from another.

Definition 7.2 *Let $\mathbf{L}^q = L_1, L_2, \ldots, L_q$ be q i.i.d. random variables sampled in a finite set $\mathcal{L}$ according to a distribution $\widetilde{\mathsf{P}}$. Let $\mathcal{Z}$ be finite set such that $|\mathcal{Z}| \leq |\mathcal{L}|$, let $h : \mathcal{L} \rightarrow \mathcal{Z}$ and let $Z_i = h(L_i)$ for $i = 1, \ldots, q$. Let H_0 and H_1 be two incompatible hypotheses on $\widetilde{\mathsf{P}}$ such that one is true. A q-limited* projection-based *distinguisher SA_q between H_0 and H_1 is a q-limited distinguisher between H_0 and H_1 which takes $\mathbf{Z}^q = Z_1, Z_2, \ldots, Z_q$ as an input instead of $\mathbf{L}^q = L_1, L_2, \ldots, L_q$.*

From this sole definition, both the perfect and the χ^2 distinguishers of Chapter 6 can be seen as particular projection-based distinguisher where the projection h would simply be the identity. In what follows we consider the case where reducing the sample space by a large factor is necessary to implement the best distinguisher, i.e., that $|\mathcal{L}| \gg |\mathcal{Z}|$. Algorithm 7.2 describes the game played by a generic projection-based distinguisher (in the simple hypothesis case) that reduces the samples using a

```
$b \xleftarrow{\mathsf{u}} \{0,1\}$                    /* Random choice between $\widetilde{\mathsf{P}}_0$ and $\widetilde{\mathsf{P}}_1$ */
$\mathsf{view} \leftarrow \{\widetilde{\mathsf{P}}_0, \widetilde{\mathsf{P}}_1\}$
$\mathcal{A} \leftarrow \mathsf{SA}_q(\mathsf{view})$ such that $\mathcal{A} \subset \mathcal{Z}^q$
for $i = 1, \ldots, q$ do
    $L_i \xleftarrow{\widetilde{\mathsf{P}}_b} \mathcal{L}$
    $Z_i = h(L_i)$
    $\mathsf{view} \leftarrow \mathsf{view} \cup \{Z_i\}$
end
$\widehat{b} \leftarrow \mathsf{SA}_q(\mathsf{view})$          /* $\widehat{b} = 1$ when $(Z_1, \ldots, Z_q) \in \mathcal{A}$ and 0 otherwise */
if $\widehat{b} = b$ then return 1 else return 0
```

Algorithm 7.2: Game played by a q-limited projection-based distinguisher SA_q, using a balanced projection $h : \mathcal{L} \to \mathcal{Z}$, between two probability distributions $\widetilde{\mathsf{P}}_0$ and $\widetilde{\mathsf{P}}_1$ over a finite set $\mathcal{L}$.

projection h. Without anticipating too much, we can already give examples of typical projection-based distinguishers:

- Linear Distinguisher: In the case where $\mathcal{L} = \{0,1\}^N$ (for some large N), one can choose some nonzero *mask* $a \in \mathcal{L}$ and let $h(L) = a \bullet L$, where $\bullet$ denotes the bit-wise exclusive-or operation. In this case $\mathcal{Z} = \{0,1\}$.
- Extended Linear Distinguisher: Letting $\mathcal{L} = \{0,1\}^N$ again, a natural way to extend linear distinguishers is to consider a projections $h : \mathcal{L} \to \mathcal{Z}$ where $\mathcal{Z} = \{0,1\}^n$ for some small n, and such that h is GF(2) linear.
- Multiple Linear Distinguisher: A simple particular case of extended linear distinguisher arises when considering several linear projections $h^{(i)} : \{0,1\}^N \to \{0,1\}$ for $i = 1, \ldots, n$, and letting $h(L) = (h_1(L), \ldots, h_n(L)) \in \{0,1\}^n$.

We study these examples in more details in the following sections. In all cases, we assume that the sample space is reduced enough so that the best distinguisher can be implemented on the Z_i's. Intuitively, one can only *loose* information by considering less data and thus, a projection-based distinguisher cannot always perform as well as a well chosen (standard) distinguisher. The following lemma shows that certain projection-based distinguishers *cannot* behave as well as well-chosen standard distinguishers, since projections reduce the Squared Euclidean Imbalance (SEI, see Definition 6.10), which was shown in Section 6.7 to be a fundamental measure, inverse-proportional to the data complexity of a distinguisher.

Lemma 7.1 (Projections reduce the SEI). *Let $\mathcal{L}$ and $\mathcal{Z}$ be two finite sets such that $|\mathcal{Z}|$ divides $|\mathcal{L}|$. Let $h : \mathcal{L} \to \mathcal{Z}$ be a balanced function. Let $\widetilde{\mathsf{P}}$ be a probability distribution of support $\mathcal{L}$ and let $L \in \mathcal{L}$ be a random variable following $\widetilde{\mathsf{P}}$. Let P denote* *the distribution of* $h(L) \in \mathcal{Z}$. *We have*

$$\Delta(\mathsf{P}) \leq \Delta(\widetilde{\mathsf{P}}).$$

Proof. By definition

$$\Delta(\mathsf{P}) = |\mathcal{Z}| \sum_{z \in \mathcal{Z}} \left(\Pr_{\widetilde{\mathsf{P}}}[h(L) = z] - \frac{1}{|\mathcal{Z}|} \right)^2 .$$

Using the fact that h is balanced

$$\Pr_{\widetilde{\mathsf{P}}}[h(L) = z] - \frac{1}{|\mathcal{Z}|} \quad = \quad \sum_{\ell \in \mathcal{L}} \mathbf{1}_{h(\ell)=z} \left(\Pr_{\widetilde{\mathsf{P}}}[\ell] - \frac{1}{|\mathcal{L}|} \right),$$

so that using Cauchy's inequality (and a simple trick, which consists in distributing the $\mathbf{1}_{h(\ell)=z}$ term over both sums)

$$\left(\Pr_{\widetilde{\mathsf{P}}}[h(L) = z] - \frac{1}{|\mathcal{Z}|} \right)^2 \leq \frac{|\mathcal{L}|}{|\mathcal{Z}|} \sum_{\ell \in \mathcal{L}} \mathbf{1}_{h(\ell)=z} \left(\Pr_{\widetilde{\mathsf{P}}}[\ell] - \frac{1}{|\mathcal{L}|} \right)^2 .$$

We conclude by summing over $z \in \mathcal{Z}$. □

7.3 Linear Distinguishers for Binary Sources

We assume in this subsection that the source is binary, i.e., that it generates samples in $\mathcal{L} = \{0,1\}^N$ for some large positive integer N (e.g. $N = 128$). A linear distinguisher is a projection-based distinguisher that applies to each sample $L \in \{0,1\}^N$ a projection

$$\begin{array}{rcl} h_a : \mathcal{L} & \longrightarrow & \mathcal{Z} = \{0,1\} \\ L & \longmapsto & h(L) = a \bullet L, \end{array}$$

where $a \in \mathcal{L} \setminus \{0\}$ is called a *mask* and where $\bullet$ denotes the bit-wise exclusive-or operation. Clearly, h_a is balanced. For simplicity, we restrict to the simple hypothesis problem where one of the two distributions is uniform: we let $\widetilde{\mathsf{P}}_0$ and $\widetilde{\mathsf{P}}_1$ be two probability distributions over $\mathcal{L}$, such that $\widetilde{\mathsf{P}}_0$ is uniform, let $\widetilde{\mathsf{P}}$ be the sample distribution, and consider the two hypotheses $\mathsf{H}_0 : \widetilde{\mathsf{P}} = \widetilde{\mathsf{P}}_0$ and $\mathsf{H}_1 : \widetilde{\mathsf{P}} = \widetilde{\mathsf{P}}_1$. When $L \sim \widetilde{\mathsf{P}}_b$, we denote by P_b the distribution of the bit $Z = h_a(L)$, for $b \in \{0,1\}$. Since h_a is balanced, P_0 is uniform over $\mathcal{Z} = \{0,1\}$. It is well known that the data complexity of a linear distinguisher is roughly inverse-proportional to the *linear probability* of Z (as noted in [32], using the notations of [112]), a fact that we now show to be a direct consequence from previous results.

Definition 7.3 *Let* $B \in \{0,1\}$ *be a random bit. The linear probability of* B *is denoted* $\mathrm{LP}(B)$ *and is defined by*

$$\mathrm{LP}(B) = (2\Pr[B=0] - 1)^2 = (\Pr[B=0] - \Pr[B=1])^2 = \left(\mathrm{E}\left((-1)^B\right)\right)^2 .$$

Let N be a positive integer and let L be a random binary variable in the set $\mathcal{L} = \{0,1\}^N$. Let $a \in \mathcal{L} \setminus \{0\}$. The linear probability of L with respect to the mask a is the linear probability of $a \bullet L = a_1 L_1 \oplus a_2 L_2 \oplus \cdots \oplus c_N L_N \in \{0,1\}$, i.e.,

$$\mathrm{LP}_a(L) = \mathrm{LP}(a \bullet L).$$

Let $\widetilde{\mathsf{P}}$ be a probability distribution over $\mathcal{L}$. The linear probability of $\widetilde{\mathsf{P}}$ with respect to the mask a is the linear probability (with respect to the same mask) of a random variable following this distribution, i.e., if $L \sim \widetilde{\mathsf{P}}$ then

$$\mathrm{LP}_a(\widetilde{\mathsf{P}}) = \mathrm{LP}_a(L).$$

We will now derive an expression of the advantage of a linear distinguisher and then show that, when $\widetilde{\mathsf{P}}_1$ is close to the uniform distribution $\widetilde{\mathsf{P}}_0$, an approximation to the first order allows to deduce that the data complexity is inverse-proportional to the linear probability of the biased distribution $\widetilde{\mathsf{P}}_1$. The following result is a direct consequence of Corollary 6.1.

Corollary 7.1 *Let N be a positive integer. Let $\widetilde{\mathsf{P}}_0$ and $\widetilde{\mathsf{P}}_1$ be two probability distributions over the binary set $\mathcal{L} = \{0,1\}^N$, where $\widetilde{\mathsf{P}}_0$ is uniform. Let $\mathcal{Z} = \{0,1\}$, $a \in \mathcal{L} \setminus \{0\}$, and let $h_a : \mathcal{L} \to \mathcal{Z}$ be the projection defined by $h_a(L) = a \bullet L$. For $b \in \{0,1\}$, we denote by P_b the distribution of $h(L)$ where $L \sim \widetilde{\mathsf{P}}_b$. The advantage is such that*

$$1 - \mathrm{Adv}_{\mathsf{LA}_q}(\widetilde{\mathsf{P}}_0, \widetilde{\mathsf{P}}_1) \doteq 2^{-q\mathrm{C}(\mathsf{P}_0,\mathsf{P}_1)}. \tag{7.1}$$

Proof. The result in the case where $\mathsf{P}_0 = \mathsf{P}_1$ is trivial. We assume now that $\mathsf{P}_0 \neq \mathsf{P}_1$. We denote by $\widetilde{\mathsf{P}}$ the distribution of the L_i's in $\mathcal{L}$ and by P the distribution of Z_i's in $\mathcal{Z}$, where $Z_i = h(L_i)$, for $i = 1, \ldots, q$. Since $a \neq 0$, h_a is balanced so that P_0 is uniform. Since $\mathsf{P}_0 \neq \mathsf{P}_1$, we have $\widetilde{\mathsf{P}} = \widetilde{\mathsf{P}}_0 \Leftrightarrow \mathsf{P} = \mathsf{P}_0$ and $\widetilde{\mathsf{P}} = \widetilde{\mathsf{P}}_1 \Leftrightarrow \mathsf{P} = \mathsf{P}_1$. Consequently,

$$\begin{aligned} 1 - \mathrm{Adv}_{\mathsf{LA}_q}(\widetilde{\mathsf{P}}_0, \widetilde{\mathsf{P}}_1) &= \Pr_{\widetilde{\mathsf{P}}=\widetilde{\mathsf{P}}_0}[\mathsf{LA}_q(\mathbf{Z}^q) = 1] + \Pr_{\widetilde{\mathsf{P}}=\widetilde{\mathsf{P}}_1}[\mathsf{LA}_q(\mathbf{Z}^q) = 0] \\ &= \Pr_{\mathsf{P}=\mathsf{P}_0}[\mathsf{LA}_q(\mathbf{Z}^q) = 1] + \Pr_{\mathsf{P}=\mathsf{P}_1}[\mathsf{LA}_q(\mathbf{Z}^q) = 0] \\ &= 1 - \mathrm{BestAdv}_q(\mathsf{P}_0, \mathsf{P}_1). \end{aligned}$$

The result follows from Corollary 6.1. □

Lemma 7.2 *Let $\mathcal{Z} = \{0,1\}$. Under the assumptions of Lemma 6.5 and assuming that P_0 is uniform, we have*

$$\mathrm{C}(\mathsf{P}_0, \mathsf{P}_1) = \frac{\mathrm{LP}(\mathsf{P}_1)}{8 \ln 2} + o\left(\mathrm{LP}(\mathsf{P}_1)\right).$$

Proof. According to Lemma 6.5, $\mathrm{C}(\mathsf{P}_0, \mathsf{P}_1) = (\epsilon_0^2 + \epsilon_1^2)/(16 \ln 2) + o(\epsilon_0^2 + \epsilon_1^2)$, where $\epsilon_b = 2\mathsf{P}_1[b] - 1$. Obviously, $\mathrm{LP}(\mathsf{P}_1) = \epsilon_0^2 = \epsilon_1^2$, which leads to the announced result. □

From Corollary 7.1 and Lemma 7.2 we can easily deduce the well accepted fact that the data complexity of a linear distinguisher is inverse-proportional to the linear probability of the biased distribution. For this, we approximate the left-hand side of (7.1) by its right-hand side and we further replace the Chernoff information between P_0 and P_1 by its first order approximation. This leads to the following heuristic.

Heuristic 7.1 Let N be a positive integer. Let $\widetilde{\mathsf{P}}_0$ and $\widetilde{\mathsf{P}}_1$ be two distributions of support $\mathcal{L} = \{0,1\}^N$, such that $\widetilde{\mathsf{P}}_0$ is uniform. Let $a \in \mathcal{L} \setminus \{0\}$. Assuming that $\mathrm{LP}_a(\widetilde{\mathsf{P}}_1) \ll 1$, the q-limited linear distinguisher LA_q between $\widetilde{\mathsf{P}}_0$ and $\widetilde{\mathsf{P}}_1$ based on the mask a reaches a non-negligible advantage when

$$q = \frac{8 \ln 2}{\mathrm{LP}_a(\widetilde{\mathsf{P}}_1)}.$$

□

7.4 Links between Best, Projection-Based, and Linear Distinguishers for Binary Sources

In what follows, we introduce some tools that will facilitate the study of the relations between the various types of distinguishers we have considered so far.

Definition 7.4 *Let N be a positive integer. Let $\widetilde{\mathsf{P}}$ be an arbitrary distribution over the set $\mathcal{L} = \{0,1\}^N$ and let $\epsilon_\ell = \widetilde{\mathsf{P}}[\ell] - \frac{1}{|\mathcal{L}|}$. The Fourier transform of $\widetilde{\mathsf{P}}$ at point $u \in \mathcal{L}$ is defined as*

$$\widehat{\epsilon}_u = \sum_{\ell \in \mathcal{L}} (-1)^{u \bullet \ell} \epsilon_\ell. \tag{7.2}$$

Lemma 7.3 *Under the notations of Definition 7.4 we have*

$$\epsilon_\ell = \frac{1}{2^N} \sum_{u \in \mathcal{L}} (-1)^{u \bullet \ell} \widehat{\epsilon}_u. \tag{7.3}$$

Proof. Starting from the right hand side of (7.3) and plugging (7.2) in,

$$\begin{aligned}
\frac{1}{2^N}\sum_{u\in\mathcal{L}}(-1)^{u\bullet\ell}\widehat{\epsilon}_u &= \frac{1}{2^N}\sum_{u\in\mathcal{L}}(-1)^{u\bullet\ell}\sum_{\ell'\in\mathcal{L}}(-1)^{u\bullet\ell'}\epsilon_{\ell'} \\
&= \frac{1}{2^N}\sum_{\ell'\in\mathcal{L}}\epsilon_{\ell'}\sum_{u\in\mathcal{L}}(-1)^{u\bullet(\ell\oplus\ell')} \\
&= \sum_{\ell'\in\mathcal{L}}\epsilon_{\ell'}\mathbf{1}_{\ell=\ell'} \\
&= \epsilon_\ell.
\end{aligned}$$

□

The next proposition can be compared to Parseval's Theorem and relates the squared Euclidean Imbalance of a distribution (SEI, see Definition 6.10) to its Fourier coefficients.

Proposition 7.1 *Let $\widetilde{\mathsf{P}}$ be a probability distribution over a finite set $\mathcal{L} = \{0,1\}^N$. The* SEI *of $\widetilde{\mathsf{P}}$ is related to its Fourier coefficient by*

$$\Delta(\widetilde{\mathsf{P}}) = \sum_{u\in\mathcal{L}}\widehat{\epsilon}_u^2.$$

Proof. By definition we have

$$\begin{aligned}
\sum_{u\in\mathcal{L}}\widehat{\epsilon}_u^2 &= \sum_{u\in\mathcal{L}}\left(\sum_{\ell\in\mathcal{L}}(-1)^{u\bullet\ell}\epsilon_\ell\right)\left(\sum_{\ell'\in\mathcal{L}}(-1)^{u\bullet\ell'}\epsilon_{\ell'}\right) \\
&= \sum_{\ell,\ell'\in\mathcal{L}}\epsilon_\ell\epsilon_\ell'\sum_{u\in\mathcal{L}}(-1)^{u\bullet(\ell\oplus\ell')} \\
&= 2^N\sum_{\ell,\ell'\in\mathcal{L}}\epsilon_\ell\epsilon_\ell'\mathbf{1}_{\ell=\ell'} \\
&= 2^N\sum_{\ell\in\mathcal{L}}\epsilon_\ell^2 \\
&= \Delta(\widetilde{\mathsf{P}}).
\end{aligned}$$

□

The following proposition relates the squared Euclidean imbalance of a distribution to its linear probability.

Proposition 7.2 *Let N be a positive integer. Let $\widetilde{\mathsf{P}}$ be a probability distribution over the set $\mathcal{L} = \{0,1\}^N$. The squared Euclidean imbalance (SEI) of $\widetilde{\mathsf{P}}$ is related to its linear* *probabilities by:*

$$\Delta(\widetilde{\mathsf{P}}) = \sum_{a \in \mathcal{L}\setminus\{0\}} \mathrm{LP}_a(\widetilde{\mathsf{P}}).$$

Proof. From (7.2) we have

$$\widehat{\epsilon}_a = \sum_{\ell \in \mathcal{L}} (-1)^{a \bullet \ell} \epsilon_\ell = \sum_{\ell \in \mathcal{L}} (-1)^{a \bullet \ell} \left(\widetilde{\mathsf{P}}[\ell] - \frac{1}{|\mathcal{L}|} \right) = \mathrm{E}\left((-1)^{a \bullet L}\right) - \mathbf{1}_{a=0},$$

where $L \sim \widetilde{\mathsf{P}}$. From Proposition 7.1 we obtain

$$\Delta(\widetilde{\mathsf{P}}) = \sum_{a \in \mathcal{L}} \widetilde{\epsilon}_a^2 = \sum_{a \in \mathcal{L}} \left(\mathrm{E}\left((-1)^{a \bullet L}\right) - \mathbf{1}_{u=0} \right)^2 = \sum_{a \in \mathcal{L}\setminus\{0\}} \left(\mathrm{E}\left((-1)^{a \bullet L}\right) \right)^2$$

from which we easily conclude by considering Definition 7.3. □

Corollary 7.2 *Let N be a positive integer. Let $\widetilde{\mathsf{P}}$ be a probability distribution over the set $\mathcal{L} = \{0,1\}^N$. Letting*

$$\mathrm{LP}_{\max}(\widetilde{\mathsf{P}}) = \max_{a \in \mathcal{L}\setminus\{0\}} \mathrm{LP}_a(\widetilde{\mathsf{P}})$$

we have

$$\Delta(\widetilde{\mathsf{P}}) \le (2^N - 1)\mathrm{LP}_{\max}(\widetilde{\mathsf{P}}).$$

Based on Heuristic 7.1, we know that the smallest data complexity achievable by a linear distinguisher (with a non-negligible advantage) between the uniform distribution $\widetilde{\mathsf{P}}_0$ and a biased distribution $\widetilde{\mathsf{P}}_1$ over $\mathcal{L} = \{0,1\}^N$ is of the order of magnitude of

$$q_{\mathrm{lin}} = \frac{8 \ln 2}{\mathrm{LP}_{\max}(\widetilde{\mathsf{P}}_1)}.$$

From Heuristic 6.2 and Definition 6.10, this is also the case for the best distinguisher limited to

$$q_{\mathrm{best}} = \frac{8 \ln 2}{\Delta(\widetilde{\mathsf{P}}_1)}.$$

Consequently, Corollary 7.2 shows that the data complexity of the best distinguisher between two distributions of random bit strings can decrease with a factor up to 2^N when compared to the best linear distinguisher, i.e.,

$$q_{\mathrm{best}} \ge \frac{q_{\mathrm{lin}}}{2^N - 1}.$$

It is interesting to note that this bound is actually tight as the following example shows.

Example 7.1 Let $0 < \gamma < 1$. We consider the distribution $\widetilde{\mathsf{P}}_1$ defined over $\mathcal{L} = \{0,1\}^N$ (for some large positive integer N) by

$$\widetilde{\mathsf{P}}_1[\ell] = \begin{cases} \frac{1}{2^N} + \left(1 - \frac{1}{2^N}\right)\gamma & \text{if } \ell = 0, \\ \frac{1}{2^N} - \frac{1}{2^N}\gamma & \text{otherwise.} \end{cases}$$

For all $a \in \mathcal{L} \setminus \{0\}$ and $L \sim \widetilde{\mathsf{P}}_1$ we have

$$\begin{aligned} \mathrm{LP}_a(\widetilde{\mathsf{P}}_1) &= \left(\mathrm{E}\left((-1)^{a \bullet L}\right)\right)^2 \\ &= \left(\sum_{\ell \in \mathcal{L}} (-1)^{a \bullet \ell}\, \widetilde{\mathsf{P}}_1[\ell]\right)^2 \\ &= \left(\frac{1}{2^N} + \left(1 - \frac{1}{2^N}\right) + \left(\frac{1}{2^N} - \frac{1}{2^N}\gamma\right) \sum_{\ell \in \mathcal{L}\setminus\{0\}} (-1)^{a \bullet \ell}\right)^2 \\ &= \gamma^2. \end{aligned}$$

On the other hand,

$$\Delta(\widetilde{\mathsf{P}}_1) = \frac{1}{2^N}\left((2^N - 1)^2\gamma^2 + (2^N - 1)\gamma^2\right) = \left(2^N - 1\right)\gamma^2.$$

We see that in this example $\Delta(\widetilde{\mathsf{P}}_1) = (2^N - 1)\mathrm{LP}_a(\widetilde{\mathsf{P}}_1)$. □

A natural extension of linear distinguishers would be to consider a specific class of projection-based distinguishers which reduce the sample space using a projection h which is GF(2)-linear. We call these distinguishers *extended linear distinguishers* and wonder about the complexity gap between classical linear distinguishers and their extended versions. The following theorem proves that if a biased distribution cannot be distinguished from a uniform one by means of a classical linear distinguisher, then (to some extent) an extended linear distinguisher won't succeed either.

Theorem 7.1 *Let N and n be two positive integers such that $N > n$. Let $\widetilde{\mathsf{P}}$ be a probability distribution over the set $\mathcal{L} = \{0,1\}^N$. Let $h : \{0,1\}^N \to \{0,1\}^n$ be* GF(2)*-linear projection. Let P denote the distribution of $h(L)$ where $L \sim \widetilde{\mathsf{P}}$. We have*

$$\Delta(\mathsf{P}) \leq (2^n - 1)\mathrm{LP}_{\max}(\widetilde{\mathsf{P}}).$$

Proof. Let $L \in \mathcal{L}$ be a random variable sampled according to the distribution $\widetilde{\mathsf{P}}$, so that $h(L) \sim \mathsf{P}$. From Proposition 7.2,

$$\Delta(\mathsf{P}) = \sum_{a \in \{0,1\}^n \setminus \{0\}} \mathrm{LP}_a(\mathsf{P}) = \sum_{a \in \{0,1\}^n \setminus \{0\}} \left(\mathrm{E}\left((-1)^{a \bullet h(L)}\right)\right)^2.$$

Since h is GF(2)-linear we have $a \bullet h(L) = {}^t h(a) \bullet L$ for all $a \in \mathcal{L}$, where ${}^t h$ denotes the transpose of h. Consequently,

$$\Delta(\mathsf{P}) = \sum_{a \in \{0,1\}^n \setminus \{0\}} \mathrm{LP}_{{}^t h(a)}(\widetilde{\mathsf{P}}) \leq (2^n - 1)\mathrm{LP}_{\max}(\widetilde{\mathsf{P}}).$$

□

The previous theorem is meaningful for practical cases, where N large (e.g., $N = 128$) so that the best distinguisher cannot be implemented, and where n is small enough so that the extended linear distinguisher can be easily implemented (e.g. $n < 30$). When it is the case, the factor $2^n - 1$ can be assumed to be small. Consequently, if a biased distribution $\widetilde{\mathsf{P}}$ on $\mathcal{L} = \{0,1\}^N$ cannot be distinguished from the uniform distribution by a linear distinguisher (which happens iff $\mathrm{LP}_{\max}(\widetilde{\mathsf{P}})$ is negligible), then the previous theorem shows that an extended linear distinguisher (which roughly needs $1/\Delta(\mathsf{P})$ samples to reach a non-negligible advantage) cannot be much more efficient.

Example 7.2 **(Multiple linear characteristics)** As an example of extended linear distinguisher, we consider the concatenation of several linear projections. More precisely, we let $h^{(1)}, h^{(2)}, \ldots, h^{(n)} : \{0,1\}^N \rightarrow \{0,1\}$ be n linear projections and consider $h : \{0,1\}^N \rightarrow \{0,1\}^n$ defined by $h = (h^{(1)}, h^{(2)}, \ldots, h^{(n)})$. Letting $L \in \mathcal{L}$ be a random variable sampled according to a biased distribution $\widetilde{\mathsf{P}}$, we denote by $\mathsf{P}^{(i)}$ the distribution of $h^{(i)}(L)$ for $i = 1, 2, \ldots, n$ and by $\mathsf{P} = \mathsf{P}^{(1)} \times \mathsf{P}^{(2)} \times \cdots \times \mathsf{P}^{(n)}$ the distribution of $h(L)$. Theorem 7.1 implies that

$$\Delta(\mathsf{P}) \leq (2^n - 1)\mathrm{LP}_{\max}(\widetilde{\mathsf{P}}).$$

This has a notable implication on multiple-linear cryptanalysis, where several characteristics are concatenated. It shows that one cannot expect to need less than q/n samples to distinguish $\widetilde{\mathsf{P}}$ from the uniform distribution when n distinct characteristics are used and if q samples would be needed by a linear distinguisher. This result is correct regardless of the dependency between the n characteristics. □

We now consider a more general case than that considered in Example 7.2: we consider the concatenation of several (not necessarily Boolean nor linear) projections, but in the case where the $h^{(i)}(L)$'s are mutually independent random variables.

Proposition 7.3 *Let N, n, d be three positive integers such that $N \geq d \cdot n$. Let L be a random variable sampled in a finite set $\{0,1\}^N$. Let $h^{(i)} : \{0,1\}^N \rightarrow \{0,1\}^n$ for $i = 1, 2, \ldots, d$ and $h = (h^{(1)}, h^{(n)}, \ldots, h^{(d)})$, such that the $h^{(i)}(L)$'s are mutually independent random variables. Denoting P_i the distribution of $h^{(i)}(L)$ for all $i = 1, 2, \ldots, d$ and letting $\mathsf{P} = \mathsf{P}_1 \times \mathsf{P}_2 \times \cdots \times \mathsf{P}_d$ we have*

$$\Delta(\mathsf{P}) + 1 = \prod_{i=1}^{d} (\Delta(\mathsf{P}_i) + 1).$$

Proof. We let $Z_i = h^{(i)}(L)$ for $i = 1, 2, \ldots, d$, so that $Z_i \sim \mathsf{P}_i$. Starting from Proposition 7.2,

$$\begin{aligned}\Delta(\mathsf{P}) &= \sum_{(a_1,\ldots,a_d)\in\mathcal{Z}^d\setminus\{\mathbf{0}\}} \left(\mathrm{E}\left((-1)^{a_1\bullet Z_1\oplus\cdots\oplus a_d\bullet Z_d}\right)\right)^2 \\ &= \sum_{(a_1,\ldots,a_d)\in\mathcal{Z}^d\setminus\{\mathbf{0}\}} \prod_{i=1}^{d} \left(\mathrm{E}\left((-1)^{a_i\bullet Z_i}\right)\right)^2\end{aligned}$$

where we used the mutual independence of the Z_i's. The announced result easily follows by applying Proposition 7.2 again. □

This result shows that merging n independent biases should only be considered when their respective amplitudes are within the same order of magnitude.

Let us summarize the results we obtained about linear distinguishers in this subsection. Given a probability distribution $\widetilde{\mathsf{P}}$ on a "large" set $\mathcal{L} = \{0,1\}^N$, we have compared the best distinguisher between $\widetilde{\mathsf{P}}$ and the uniform distribution to the best linear distinguisher (Corollary 7.2) and showed that the ratio between their respective data complexities is bounded by 2^N, which is large by assumption. This result is more of theoretical interest since the best distinguisher cannot be implemented anyway, due to the assumption made on the cardinality of $\mathcal{L}$. On the practical side, we considered a wide class of projection-based distinguishers, namely extended linear distinguishers (which include multiple linear distinguishers) and showed that, to a certain extent, if $\widetilde{\mathsf{P}}$ cannot be distinguished from a uniform distribution by means of a linear distinguisher, then an extended linear distinguisher won't succeed either. In the light of this discussion, one may wonder if resistance to linear distinguishers always implies a certain resistance to any projection-based distinguishers (that would reduce the sample size enough, so that the best distinguisher can be implemented). The following example shows that this is *not* the case, as it is possible to find a biased distribution P which cannot be distinguished from the uniform distribution by a linear distinguisher (the value of $\mathrm{LP}_{\max}(\widetilde{\mathsf{P}})$ is negligible) but which can be in the absolute using a (non-linear) distinguisher.

Example 7.3 We consider the ring $\mathbf{Z}_4$ of integers modulo 4 and use their binary representation (i.e., 0 is `00`, 1 is `01`, and so forth). For a positive integer n such that $n+1$ is divisible by 4, let $\mathcal{L} = \mathbf{Z}_4^{n+1}$. An element of $\mathcal{L}$ can be represented as a N-bit string where $N = 2n+2$. We let $\widetilde{\mathsf{P}}_0$ be the uniform distribution over $\mathcal{L}$ and $\widetilde{\mathsf{P}}_1$ be such that when sampled according to this distribution, $(X_1, X_2, \ldots, X_{n+1}) \in \mathcal{L}$ is such that $(X_1, X_2, \ldots, X_n) \in \mathbf{Z}_4^n$ is uniformly distributed and $X_{n+1} = (Y + \sum_{i=1}^{n} X_i) \bmod 4$, where $Y \in \{0,1\}$ is uniformly distributed. Let $\widetilde{\mathsf{P}}$ be the sample distribution. It is easy to construct a (Boolean) projection-based distinguisher that easily distinguishes $\widetilde{\mathsf{P}} = \widetilde{\mathsf{P}}_1$

from $\widetilde{\mathsf{P}} = \widetilde{\mathsf{P}}_0$. We let $h : \mathcal{L} \to \{0,1\}$ be such that

$$h(x_1, x_2, \ldots, x_{n+1}) = \mathrm{msb}\left(\left(x_{n+1} - \sum_{i=1}^{n} x_n\right) \bmod 4\right).$$

Clearly, when the X_i's are uniformly distributed, $h(X_1, X_2, \ldots, X_{n+1})$ is uniformly distributed. When $(X_1, X_2, \ldots, X_{n+1}) \sim \widetilde{\mathsf{P}}_1$ (i.e., when $\widetilde{\mathsf{P}} = \widetilde{\mathsf{P}}_1$), then we always have $h(X_1, X_2, \ldots, X_{n+1}) = 0$. As a consequence, if we denote by P_1 the SEI of $h(X_1, X_2, \ldots, X_{n+1})$ when the $(X_1, X_2, \ldots, X_{n+1}) \sim \widetilde{\mathsf{P}}_1$, we have $\Delta(\mathsf{P}_1) = 1$. This implies (according to the discussion of Section 6.7) that the projection-based distinguisher based on h easily distinguishes $\widetilde{\mathsf{P}} = \widetilde{\mathsf{P}}_1$ from $\widetilde{\mathsf{P}} = \widetilde{\mathsf{P}}_0$ with a few samples. On the other hand, we will show that $\mathrm{LP}_{\max}(\widetilde{\mathsf{P}}_1) = 2^{-(n+1)}$, which is small by assumption.

Since each x_i lies in $\mathbf{Z}_4$, it can be described by two bits x_i^{H} and x_i^{L}, such that $x_i = 2x_i^{\mathrm{H}} + x_i^{\mathrm{L}} = x_i^{\mathrm{H}} \| x_i^{\mathrm{L}}$. Any linear distinguisher can be defined by a projection h_{lin} such that

$$h_{\mathrm{lin}}(x_1, x_2, \ldots, x_{n+1}) = \left(\bigoplus_{j=1}^{n+1} a_j x_j^{\mathrm{L}}\right) \oplus \left(\bigoplus_{j=1}^{n+1} b_j x_j^{\mathrm{H}}\right),$$

where $a_1, \ldots, a_{n+1}, b_1, \ldots, b_{n+1} \in \{0,1\}$ with at least one non-zero value. In the case where $(X_1, X_2, \ldots, X_{n+1}) \sim \widetilde{\mathsf{P}}_1$, it is easy so show that

$$\begin{aligned} X_{n+1}^{\mathrm{L}} \oplus Y &= \bigoplus_{j=1}^{n} X_j^{\mathrm{L}}, \text{ and} \\ X_{n+1}^{\mathrm{H}} &= \left(\bigoplus_{j=1}^{n} X_j^{\mathrm{H}}\right) \oplus \left(\bigoplus_{0 \le j < k \le n} X_j^{\mathrm{L}} X_k^{\mathrm{L}}\right) \oplus \left(\bigoplus_{j=1}^{n} X_j^{\mathrm{L}} Y\right). \end{aligned}$$

Thus denoting B the value of the bit $h_{\mathrm{lin}}(X_1, X_2, \ldots, X_{n+1})$ in this case, we have

$$\begin{aligned} B = &\left(\bigoplus_{j=1}^{n} (a_j \oplus a_{n+1}) X_j^{\mathrm{L}}\right) \oplus \left(\bigoplus_{j=1}^{n} (b_j \oplus b_{n+1}) X_j^{\mathrm{H}}\right) \oplus a_{n+1} Y \\ &\oplus \left(b_{n+1} \bigoplus_{1 \le j < k \le n} X_j^{\mathrm{L}} X_k^{\mathrm{L}}\right) \oplus \left(b_{n+1} \bigoplus_{j=1}^{n} X_j^{\mathrm{L}} Y\right). \end{aligned}$$

If $b_{n+1} = 0$ we can see that $\Pr_{\widetilde{\mathsf{P}} = \widetilde{\mathsf{P}}_1}[B = 0] = \frac{1}{2}$ (as at least one of the $a_1, \ldots, a_{n+1}$, $b_1, \ldots, b_n$ is strictly positive), hence $\mathrm{LP}(\widetilde{\mathsf{P}}_1) = \mathrm{LP}(B) = 0$ in this case. If $b_{n+1} = 1$, we have

$$\begin{aligned} B = &\left(\bigoplus_{j=1}^{n} (a_j \oplus a_{n+1}) X_j^{\mathrm{L}}\right) \oplus \left(\bigoplus_{j=1}^{n} \overline{b_j} X_j^{\mathrm{H}}\right) \oplus a_{n+1} Y \\ &\oplus \left(\bigoplus_{1 \le j < k \le n} X_j^{\mathrm{L}} X_k^{\mathrm{L}}\right) \oplus \left(\bigoplus_{j=1}^{n} X_j^{\mathrm{L}} Y\right). \end{aligned}$$

If one of the $\overline{b_j}$'s is non-zero, then B is uniformly distributed and the linear probability is zero again. We now assume that $b_j = 1$ for all $j = 1, \ldots, n$ and get

$$B = \left(\bigoplus_{j=1}^{n} (a_j \oplus a_{n+1}) X_j^{\mathrm{L}} \right) \oplus a_{n+1} Y \oplus \left(\bigoplus_{1 \le j < k \le n} X_j^{\mathrm{L}} X_k^{\mathrm{L}} \right) \oplus \left(\bigoplus_{j=1}^{n} X_j^{\mathrm{L}} Y \right).$$

For $j = 1, 2, \ldots, n$ we let $\alpha_j = a_j \oplus a_{n+1}$, $U_j = X_j^{\mathrm{L}}$, $\alpha_{n+1} = a_{n+1}$, and $U_{n+1} = Y$. Note that the U_i's are mutually independent random bits and that there is no constraint on the α_i's. The previous equation reduces to

$$B = \left(\bigoplus_{j=1}^{n+1} \alpha_j U_j \right) \oplus \left(\bigoplus_{1 \le j < k \le n+1} U_j U_k \right). \tag{7.4}$$

We are looking for the α_i's that maximize $\mathrm{LP}(B)$. We will now show that, for any choice of $\alpha_1, \ldots, \alpha_n$, the choice of α_{n+1} makes no difference on $\mathrm{LP}(B)$. By symmetry, this implies that for any choice of $\alpha_1, \ldots, \alpha_{\ell-1}, \alpha_{\ell+1}, \ldots, \alpha_{n+1}$, where $\ell = 1, \ldots, n+1$, the choice of α_ℓ has no influence on $\mathrm{LP}(B)$. As a consequence, all the possible choices of $\alpha_1, \ldots, \alpha_{n+1}$ are equivalent with respect to the resulting value of $\mathrm{LP}(B)$, so that we will choose $\alpha_i = 0$ for all $i = 1, \ldots, n+1$. We now show that for any choice of $\alpha_1, \ldots, \alpha_n$, the choice of α_{n+1} makes no difference on $\mathrm{LP}(B)$.

- If we set $\alpha_{n+1} = 0$, equation (7.4) can be written as

$$B = (1 \oplus U_{n+1}) \left(\bigoplus_{j=1}^{n} \alpha_j U_j \right) \oplus \left(\bigoplus_{1 \le j < k \le n} U_j U_k \right) = (1 \oplus U_{n+1}) B_1 \oplus B_2,$$

  where $B_1 = \bigoplus_{j=1}^{n} \alpha_j U_j$ and $B_2 = \left(\bigoplus_{1 \le j < k \le n} U_j U_k \right)$ are random bits, independent from U_{n+1}. Consequently,

$$\begin{aligned} \Pr[B = 0] &= \frac{1}{2} \Pr[B = 0 | U_{n+1} = 0] + \frac{1}{2} \Pr[B = 0 | U_{n+1} = 1] \\ &= \frac{1}{2} \Pr[B_1 \oplus B_2 = 0] + \frac{1}{2} \Pr[B_2 = 0]. \end{aligned} \tag{7.5}$$

- Similarly, if we set $\alpha_{n+1} = 1$, equation (7.4) can be written as

$$B = (1 \oplus U_{n+1}) B_1 \oplus B_2 \oplus U_{n+1},$$

  using the same notations. Consequently,

$$\begin{aligned} \Pr[B = 0] &= \frac{1}{2} \Pr[B = 0 | U_{n+1} = 0] + \frac{1}{2} \Pr[B = 0 | U_{n+1} = 1] \\ &= \frac{1}{2} \Pr[B_1 \oplus B_2 = 0] + \frac{1}{2} \Pr[B_2 = 1]. \end{aligned} \tag{7.6}$$

Using Lemma 3.1 in Appendix C, we note that

$$B_2 = \sum_{1\leq 1<k\leq n} U_jU_k \bmod 2 = \frac{W(W-1)}{2} \bmod 2,$$

where W denotes the Hamming weight of the binary string $U_1\|\cdots\|U_n$. Consequently,

$$\begin{aligned}
\Pr[B_2 = 0] &= \Pr[W \bmod 4 = 0 \text{ or } 1] \\
&= \Pr[W \bmod 4 = 0] + \Pr[W \bmod 4 = 1], \quad \text{and} \\
\Pr[B_2 = 1] &= \Pr[W \bmod 4 = 2 \text{ or } 3] \\
&= \Pr[W \bmod 4 = 2] + \Pr[W \bmod 4 = 3].
\end{aligned}$$

It is easy to see that, since we assumed that $n+1$ is divisible by 4,

$$\Pr[W \bmod 4 = 0] = \sum_{k=0}^{\frac{n-3}{4}} \Pr[W = 4k] = 2^{-n}\sum_{k=0}^{\frac{n-3}{4}} \binom{n}{4k}. \tag{7.7}$$

Similarly:

$$\begin{aligned}
\Pr[W \bmod 4 = 1] &= 2^{-n}\sum_{k=0}^{\frac{n-3}{4}} \binom{n}{4k+1} \\
\Pr[W \bmod 4 = 2] &= 2^{-n}\sum_{k=0}^{\frac{n-3}{4}} \binom{n}{4k+2} \\
\Pr[W \bmod 4 = 3] &= 2^{-n}\sum_{k=0}^{\frac{n-3}{4}} \binom{n}{4k+3}
\end{aligned}$$

Letting $\ell = \frac{n-3}{4} - k$ in (7.7) and using Pascal's rule, it is easy to see that

$$\Pr[W \bmod 4 = 0] = 2^{-n}\sum_{\ell=0}^{\frac{n-3}{4}} \binom{n}{n-(\ell+3)} = \Pr[W \bmod 4 = 3].$$

Similarly, $\Pr[W \bmod 4 = 1] = \Pr[W \bmod 4 = 2]$ and we conclude that

$$\Pr[B_2 = 0] = \Pr[B_2 = 1] = \frac{1}{2}.$$

Plugging this result in (7.5) and in (7.6) we see that, whatever the choice of $\alpha_{n+1} \in \{0,1\}$, the value of $\Pr[B = 0]$ (and thus of $\mathrm{LP}(B)$) remains the same. Any choice being equivalent, we choose to set $\alpha_{n+1} = 0$. Proceeding in the same way for all α_i's we see that (7.4) reduces to

$$B = \bigoplus_{1\leq j<k\leq n+1} U_jU_k. \tag{7.8}$$

Using Lemma 3.1 in Appendix C again, we deduce from the previous equation that

$$B = \frac{W'(W'-1)}{2} \mod 2,$$

where W' denotes the Hamming weight of the bit string $U_1\|\cdots\|U_{n+1}$. We thus have

$$\Pr[B=0] = \Pr[W' \bmod 4 = 0] + \Pr[W' \bmod 4 = 1].$$

Using the law of total probability (based on the condition $U_{n+1} = 0$ or 1) we easily obtain that

$$\begin{aligned}
\Pr[B=0] &= \frac{1}{2}(2\Pr[W \bmod 4 = 0] + \Pr[W \bmod 4 = 1] + \Pr[W \bmod 4 = 3]) \\
&= \frac{1}{2} + \frac{1}{2}(\Pr[W \bmod 4 = 0] - \Pr[W \bmod 4 = 2]) \\
&= \frac{1}{2} + 2^{-(n+1)}\left(\sum_{k=0}^{\frac{n-3}{4}}\binom{n}{4k} - \sum_{k=0}^{\frac{n-3}{4}}\binom{n}{4k+2}\right) \\
&= \frac{1}{2} + 2^{-(n+3)}((1+i)^{n+1} + (1-i)^{n+1}) \qquad (7.9)
\end{aligned}$$

using Lemma 3.2 in Appendix C. Noticing that

$$\begin{aligned}
(1+i)^{n+1} + (1-i)^{n+1} &= 2^{n+1}(e^{i(n+1)\frac{\pi}{4}} + e^{-i(n+1)\frac{\pi}{4}}) \\
&= 2\cdot 2^{\frac{n+1}{2}}\cos((n+1)\frac{\pi}{4}) \\
&= 2\cdot 2^{\frac{n+1}{2}}(-1)^{\frac{n+1}{4}},
\end{aligned}$$

we easily conclude from this and from (7.9) that

$$\mathrm{LP}_{\max}(\widetilde{\mathsf{P}}_1) = \mathrm{LP}(B) = 2^{-(n+1)}.$$

□

7.5 Extending the Notion of Linear Probability to Arbitrary Sets

In the digital age, information is mostly seen as a sequence of bits and, naturally, most block ciphers and cryptanalytic tools assume that the sample space is made of binary strings. This restriction is quite questionable though, as it is easy to think of specific settings in which it could be desirable to adapt the block size to the data being encrypted. For example, when considering credit card numbers, social security numbers, payment orders, schedules, telegrams, calendars, or string of alphabetical characters, it seems that there is no reason what so ever to restrict to binary strings. Whereas an apparently straightforward solution would be to encode the data prior encryption, the loss in terms of simplicity (inevitably affecting the security analysis) and of efficiency would be unfortunate.

Although most modern block ciphers (e.g., [1,3,6,41,76,101,145]) are defined on a binary set, practical and efficient examples of block ciphers defined on a set of arbitrary size exist (see for example Schroeppel's "omnicipher" Hasty Pudding [138]). Some others, although still defined on binary sets, suggest to use a mixture of group laws over the same set. For example, IDEA [96] combines three group structures: exclusive bit or, addition modulo 2^{16} and a tweaked multiplication modulo $2^{16}+1$. Designing a block cipher with an arbitrary block space can be particularly challenging since the state of the art concerning alternate group structures is very limited. Although differential cryptanalysis [21], through the theory of Markov ciphers [97], can be specified over an arbitrary group, linear cryptanalysis [110, 111] is based on a measurement (the linear probability) that sticks to bit strings. Applying this attack against a non-binary block cipher would at least require to generalize this notion.

In the following sections, we re-visit linear distinguishers but without assuming that the underlying set is made of bit strings. Consequently, the only structure we can consider on these sets is that of finite *Abelian Groups*. We first recall essential results on characters which will play a central role in the generalization of the linear probability.

Characters over Finite Abelian Groups

Let G be a finite group of order n. We let $L^2(G)$ denote the n-dimensional vector space of complex-valued functions f on G. The conjugate $\overline{f}$ of f is defined by $\overline{f}(a) = \overline{f(a)}$ for all $a \in \mathsf{G}$. We define an *inner product* on $L^2(G)$ by

$$(f_1, f_2) = \sum_{a \in \mathsf{G}} f_1(a)\overline{f_2}(a).$$

The Euclidean norm of $f \in L^2(G)$ is simply

$$\|f\|_2 = (f,f)^{1/2} = \left(\sum_a |f(a)|^2\right)^{1/2}.$$

Consequently, $L^2(G)$ is actually a Hilbert Space.

Definition 7.5 *A* character *of an Abelian group* G *is a homomorphism* $\chi : \mathsf{G} \to \mathbf{C}^\times$, *where* $\mathbf{C}^\times$ *is the multiplicative group of nonzero complex numbers.*

If $\chi : \mathsf{G} \to \mathbf{C}^\times$ is a character, then $\chi(1) = 1$ and $\chi(a_1a_2) = \chi(a_1)\chi(a_2)$ for all $a_1, a_2 \in \mathsf{G}$. Clearly, $\chi(a)$ is a nth root of unity, hence $\overline{\chi}(a) = \chi(a)^{-1}$. The *product* of two characters χ_1 and χ_2 is defined as

$$\chi_1\chi_2(a) = \chi_1(a)\chi_2(a)$$

for all $a \in \mathsf{G}$. The character χ_0 defined by $\chi_0(a) = 1$ for all $a \in \mathsf{G}$ is the neutral element for this operation. Clearly, $\chi^{-1} = \overline{\chi}$. The set of all characters of G is a group, called the *dual group* of G, and denoted $\widehat{\mathsf{G}}$. We know that G is isomorphic to $\widehat{\mathsf{G}}$ [121].

Lemma 7.4 *[Theorems 4.6 and 4.7 in [121]] Let* G *be a finite Abelian group of order* n*, and let* $\widehat{\mathsf{G}}$ *be its dual group. If* $\chi \in \widehat{\mathsf{G}}$ *(resp.* $a \in \mathsf{G}$*) then*

$$\sum_{a\in\mathsf{G}} \chi(a) = \begin{cases} n & \textit{if } \chi = \chi_0, \\ 0 & \textit{otherwise,} \end{cases} \qquad \textit{resp.} \qquad \sum_{\chi\in\widehat{\mathsf{G}}} \chi(a) = \begin{cases} n & \textit{if } a = 1, \\ 0 & \textit{otherwise.} \end{cases}$$

If $\chi_1, \chi_2 \in \widehat{\mathsf{G}}$ *(resp.* $a, b \in \mathsf{G}$*) then*

$$\sum_{a\in\mathsf{G}} \chi_1(a)\overline{\chi_2}(a) = \begin{cases} n & \textit{if } \chi_1 = \chi_2, \\ 0 & \textit{otherwise,} \end{cases} \qquad \textit{resp.} \qquad \sum_{\chi\in\widehat{\mathsf{G}}} \chi(a)\overline{\chi}(b) = \begin{cases} n & \textit{if } a = b, \\ 0 & \textit{otherwise.} \end{cases}$$

If χ_1, χ_2 are characters of G, we deduce $(\chi_1, \chi_2) = n$ if $\chi_1 = \chi_2$ and 0 otherwise. Therefore, the n characters of the dual group $\widehat{\mathsf{G}}$ is an orthogonal basis of the vector space $L^2(G)$.

Definition 7.6 *[Fourier transform] The Fourier transform of* $f \in L^2(\mathsf{G})$ *is the function* $\widehat{f} \in L^2(\widehat{\mathsf{G}})$ *such that*

$$\widehat{f}(\chi) = (f, \chi) = \sum_{a\in\mathsf{G}} f(a)\overline{\chi}(a) \quad \textit{for all } \chi \in \widehat{\mathsf{G}}.$$

If $\widehat{f} \in L^2(\widehat{\mathsf{G}})$ is the Fourier transform of $f \in L^2(G)$, then the Fourier inversion is

$$f = \frac{1}{n} \sum_{\chi\in\widehat{\mathsf{G}}} \widehat{f}(\chi)\chi.$$

Theorem 7.2 *[Plancherel's formula] If* $\widehat{f} \in L^2(\widehat{\mathsf{G}})$ *is the Fourier transform of* $f \in L^2(G)$*, then*

$$\|\widehat{f}\|_2 = \sqrt{n}\|f\|_2.$$

Extending the Notion of Linear Probability

Consider the particular case where $\mathsf{G} = \{0,1\}^k$, $\chi_u(a) = (-1)^{u\bullet a}$ for all $u, a \in \mathsf{G}$, and where $\bullet$ denotes the inner dot product in G. The mapping $u \mapsto \chi_u$ is an isomorphism between G and $\widehat{\mathsf{G}}$. Consequently, when $\mathsf{G} = \{0,1\}^k$ any character χ of G

can be expressed as $\chi(a) = (-1)^{u \bullet a}$ for some $u \in \mathsf{G}$. It is easy to make the parallel with linear cryptanalysis, where u is a mask, so that there is a one-to-one mapping between masks and characters in this case. So, it seems reasonable to generalize linear cryptanalysis on any finite Abelian group by using characters instead of masks.

Definition 7.7 *Let H be a finite subgroup of $\mathbf{C}^\times$ of order d. Let $H \in \mathsf{H}$ be a random variable. The linear probability of H is denoted $\mathrm{LP}(H)$ and is defined by*

$$\mathrm{LP}(H) = |\mathrm{E}(H)|^2 = \left| \sum_{h \in \mathsf{H}} h \Pr[H = h] \right|^2 .$$

Let G be an Abelian group and let $\chi : \mathsf{G} \to \mathbf{C}^\times$ be a character of order d. The linear probability of random variable $G \in \mathsf{G}$ with respect to the character χ is the linear probability of $\chi(G)$, i.e.,

$$\mathrm{LP}_\chi(G) = \mathrm{LP}(\chi(G)).$$

Let $\widetilde{\mathsf{P}}$ be a probability distribution over G. The linear probability of $\widetilde{\mathsf{P}}$ with respect to the character χ is the linear probability (with respect to the same character) of a random variable following this distribution, i.e., if $G \sim \widetilde{\mathsf{P}}$ then

$$\mathrm{LP}_\chi(\widetilde{\mathsf{P}}) = \mathrm{LP}_\chi(G).$$

Note that $\mathrm{LP}_\chi(\widetilde{\mathsf{P}}) = |\widehat{\widetilde{\mathsf{P}}}[\chi]|^2$, so that the linear probability of χ is simply the square of *magnitude* of the discrete Fourier transform of the probability distribution. In the particular case where $\mathsf{G} = \{0,1\}^\ell$, we can see that for any u we have $\mathrm{LP}_u(\widetilde{\mathsf{P}}) = \mathrm{LP}_{\chi_u}(\widetilde{\mathsf{P}})$, so that Definition 7.7 indeed generalizes the earlier notion of linear probability (see Definition 7.3).

7.6 Linear Distinguishers for Sources over Arbitrary Sets

From the study of the classical setting described in Subsection 7.3 we see that, essentially, a linear distinguisher tries to distinguish a uniform distribution P_0 on $\mathcal{Z} = \{0,1\}$ from a biased distribution P_1 which is completely described by its *bias* $\epsilon \in \mathbf{R}$ with respect to P_0, i.e., $\mathsf{P}_1 = (\frac{1+\epsilon}{2}, \frac{1-\epsilon}{2}) = (\frac{1-\epsilon}{2} + \epsilon, \frac{1-\epsilon}{2})$. In that case, the linear probability of $B \sim \mathsf{P}_1$ is $\mathrm{LP}(B) = \epsilon^2$. When extending linear cryptanalysis to arbitrary sets, we will assume the exact same setting when the character used to reduce the sample space is of order 2. For characters of higher order d, considering two simple hypotheses appears too restrictive. Instead, we will assume that the alternate hypothesis is composite but only *once the sample space is reduced.* More precisely, let G be a large group, $\chi : \mathsf{G} \to \mathbf{C}$ of order d, and $\mathsf{H} = \chi(\mathsf{G})$ which is a subgroup of $\mathbf{C}^\times$ of order d. We consider two distributions $\widetilde{\mathsf{P}}_0, \widetilde{\mathsf{P}}_1$ over G, where $\widetilde{\mathsf{P}}_0$ is uniform. For $G \in \mathsf{G}$ we let $H = \chi(G)$. We assume that when $G \sim \widetilde{\mathsf{P}}_0$ then $H \sim \mathsf{P}_0$ where P_0 is the uniform distribution over H (in particular, this implies that d divides the order of G and that χ is balanced). When $G \sim \widetilde{\mathsf{P}}_1$ then $H \sim \mathsf{P}_u$, where $u \in \mathsf{H}$ is unknown, and where P_u is the distribution over H defined by

$$\mathsf{P}_u[h] = \begin{cases} \frac{1-\epsilon}{d} + \epsilon & \text{when } h = u \\ \frac{1-\epsilon}{d} & \text{otherwise,} \end{cases} \tag{7.10}$$

where $0 < \epsilon < 1$. Letting $\widetilde{\mathsf{P}}$ be the distribution of $G \in \mathsf{G}$ and P the distribution of $H = \chi(G)$, we can write the hypothesis testing problem

$$\mathsf{H}_0 : \widetilde{\mathsf{P}} = \widetilde{\mathsf{P}}_0 \quad \text{vs.} \quad \mathsf{H}_1 : \widetilde{\mathsf{P}} = \widetilde{\mathsf{P}}_1$$

as

$$\mathsf{H}_0 : \mathsf{P} = \mathsf{P}_0 \quad \text{vs.} \quad \mathsf{H}_1 : \mathsf{P} \in \{\mathsf{P}_u : u \in \mathsf{H}\}.$$

Lemma 7.5 *Let P_0 be the uniform distribution on a finite subgroup H of $\mathbf{C}^\times$ of order d. Let $\mathcal{D} = \{\mathsf{P}_u : u \in \mathsf{H}\}$ be a set of d distributions on H defined by (7.10). The q-limited distinguisher between the null hypothesis $\mathsf{H}_0 : \mathsf{P} = \mathsf{P}_0$ and the alternate hypothesis $\mathsf{H}_1 : \mathsf{P} \in \mathcal{D}$ defined by the distribution acceptance region $\Pi_q^\star = \Pi^\star \cap \mathcal{P}_q$, where*

$$\Pi^\star = \left\{ \mathsf{P} \in \mathcal{P} \ : \ \|\mathsf{P}\|_\infty \geq \frac{\log(1-\epsilon)}{\log(1-\epsilon) - \log(1+(d-1)\epsilon)} \right\}, \tag{7.11}$$

is asymptotically optimal and its advantage $\mathrm{BestAdv}_q$ is such that

$$1 - \mathrm{BestAdv}_q(\mathsf{H}_0, \mathsf{H}_1) \doteq 2^{q \inf_{0<\lambda<1} \log \frac{1}{d}\left((1+(d-1)\epsilon)^\lambda + (d-1)(1-\epsilon)^\lambda\right)}.$$

Proof. According to Theorem 6.4, the best distinguisher is defined by the acceptance region

$$\Pi^\star = \{\mathsf{P} \in \mathcal{P} \ : \ \min_{u \in \mathsf{H}} \mathsf{L}_u(\mathsf{P}) \leq 0\} \quad \text{with} \quad \mathsf{L}_u(\mathsf{P}) = \sum_{h \in \mathsf{H}} \mathsf{P}[h] \log \frac{\mathsf{P}_0[h]}{\mathsf{P}_u[h]}.$$

Since

$$\mathsf{L}_u(\mathsf{P}) = \mathsf{P}[u] \log \frac{1-\epsilon}{1+(d-1)\epsilon} - \log(1-\epsilon),$$

the minimum is obtained for the $u \in \mathsf{H}$ which maximizes P (recall that $\epsilon > 0$). From this we easily deduce (7.11). In that case, Theorem 6.4 also states that

$$1 - \mathrm{BestAdv}_q(\mathsf{H}_0, \mathsf{H}_1) \doteq \max_{u \in \mathsf{H}} 2^{-q\mathrm{C}(\mathsf{P}_0, \mathsf{P}_u)}.$$

It is easy to see that $\mathrm{C}(\mathsf{P}_0, \mathsf{P}_u) = \mathrm{C}(\mathsf{P}_0, \mathsf{P}_{u'})$ for $u \neq u'$, so that

$$1 - \mathrm{BestAdv}_q(\mathsf{H}_0, \mathsf{H}_1) \doteq 2^{-q\mathrm{C}(\mathsf{P}_0, \mathsf{P}_u)}$$

for any $u \in \mathsf{H}$. The definition of the Chernoff information allows to conclude. □

It should be noted that for all $u \in \mathsf{H}$, if $H \sim \mathsf{P}_u$ we have

$$\mathrm{LP}(H) = \left| u\left(\frac{1-\epsilon}{d} + \epsilon\right) + \sum_{h \in \mathsf{H}\setminus\{u\}} h\frac{1-\epsilon}{d} \right|^2 = \left| u\epsilon + \frac{1-\epsilon}{d}\sum_{h\in\mathsf{H}} h \right|^2 = |\epsilon|^2,$$

since $\sum_{h\in\mathsf{H}} h = 0$ (as the h's are the d roots of unity) and since $|u| = 1$. Consequently, if $G \sim \widetilde{\mathsf{P}}_1$ then $\mathrm{LP}_\chi(G) = \epsilon^2$, regardless of which distribution among the P_u's is actually followed by $\chi(G)$. It makes thus sense so write

$$\mathrm{LP}_\chi(\widetilde{\mathsf{P}}_1) = \epsilon^2.$$

Moreover, for close distributions we have

$$\inf_{0<\lambda<1} \log\left(\frac{1}{d}\left((1+(d-1)\epsilon)^\lambda + (d-1)(1-\epsilon)^\lambda\right)\right) \approx \frac{d-1}{8\ln 2}\epsilon^2$$

and

$$\frac{\log(1-\epsilon)}{\log(1-\epsilon) - \log(1+(d-1)\epsilon)} \approx \frac{1}{d} + \frac{1}{2}\left(1 - \frac{1}{d}\right)\epsilon,$$

so that we can deduce the following heuristic from Lemma 7.5.

Heuristic 7.2 Let $\widetilde{\mathsf{P}}_0$ and $\widetilde{\mathsf{P}}_1$ be two distributions of full support over a finite Abelian group G, such that $\widetilde{\mathsf{P}}_0$ is uniform. Let $\chi : \mathsf{G} \to \mathbf{C}^\times$ be a character of order d. Assuming that $\mathrm{LP}_\chi(\widetilde{\mathsf{P}}_1) \ll 1$, the q-limited linear distinguisher LA_q between $\widetilde{\mathsf{P}}_0$ and $\widetilde{\mathsf{P}}_1$ based on the character χ reaches a non-negligible advantage when

$$q = \frac{8\ln 2}{(d-1)\mathrm{LP}_\chi(\widetilde{\mathsf{P}}_1)}.$$

It is optimal among all possible linear distinguishers reducing the sample space by means of characters and outputs 1 when

$$\left(\|\mathsf{P}_{\mathbf{H}^q}\|_\infty - \frac{1}{d}\right)^2 \geq \frac{1}{4}\left(1-\frac{1}{d}\right)^2 \mathrm{LP}_\chi(\widetilde{\mathsf{P}}_1),$$

where $\mathbf{H}^q = H_1, H_2, \ldots, H_q$ are the q samples such that $H_i = \chi(G_i)$, where the G_i's are the original source samples. □

Case Study: $\mathbf{Z}_m^r$-based Linear Cryptanalysis

We illustrate the theory with a concrete example, that is, linear cryptanalysis over the additive group $\mathbf{Z}_m^r$. For any positive integer d such that $d|m$, we define φ_a^d for $a = (a_1, \ldots, a_r)$ where $a_\ell \in \{0, 1, \ldots, d-1\}$ for $\ell = 1, \ldots, r$ by

$$\begin{array}{rccl} \varphi_a^d : & \mathbf{Z}_m^r & \longrightarrow & \mathbf{C}^\times \\ & x & \longmapsto & \varphi_a^d(x) = e^{\frac{2\pi i}{d}\sum_{\ell=1}^r a_\ell x_\ell}. \end{array}$$

The m^r characters of the additive group $\mathbf{Z}_m^r$ are called *additive characters modulo* m [121] and are the φ_a^m's. Note that a character φ_a^m of order d can be expressed as $\varphi_{ma/d}^m$.

We revisit Example 7.3 on page 77 where a source generating a random variable $X = (X_1, \ldots, X_{n+1}) \in \mathbf{Z}_4^{n+1}$ is considered (where $n+1$ is divisible by 4). When the source follows the distribution $\widetilde{\mathsf{P}}_0$, X is uniformly distributed. When the source follows distribution $\widetilde{\mathsf{P}}_1$, $X_1, \ldots, X_n$ are uniformly distributed mutually independent random variables in $\mathbf{Z}_4$ and $X_{n+1} = Y + \sum_{i=1}^n X_i$, where Y is either 0 or 1 with equal probability and where the addition is performed modulo 4. Considering X as a bit string of length $2n+2$. We showed in Example 7.3 that $\max_\alpha \mathrm{LP}_{\varphi_\alpha^2}(\widetilde{\mathsf{P}}_1) = 2^{-(n+1)}$ (the max being taken over classical linear masks), which means that the source cannot be distinguished from a perfectly random one using a classical linear distinguisher.

We will now show that $\widetilde{\mathsf{P}}_1$ can easily be distinguished from $\widetilde{\mathsf{P}}_0$ by a generalized linear distinguisher, and more precisely, by a linear distinguisher of order 4. Let $a = (-1, \ldots, -1, 1) \in \mathbf{Z}_4^{n+1}$ and consider the character φ_a^4 over $\mathbf{Z}_4^{n+1}$. In this case we have

$$\mathrm{LP}_{\varphi_a^4}(\widetilde{\mathsf{P}}_1) = \left|\mathrm{E}\left(e^{\frac{\pi i}{2}(X_{n+1} - \sum_{\ell=1}^n X_\ell)}\right)\right|^2 = \left|\mathrm{E}\left(e^{\frac{\pi i}{2}Y}\right)\right|^2 = \frac{1}{2}.$$

According to Heuristic 7.2, only a few samples are then needed to a linear distinguisher based on φ_a^4 in order to distinguish (with a non-negligible advantage) $\widetilde{\mathsf{P}}_1$ from the uniform distribution. Through this example, we notice that there can be a huge gap between linear distinguishers of order 2 and linear distinguishers of order 4.

7.7 A Fundamental Link Between Projection-Based and Linear Distinguishers

In Proposition 7.2 (on page 73), we showed that the squared Euclidean imbalance (SEI) of a distribution $\widetilde{\mathsf{P}}$ over $\mathcal{L} = \{0,1\}^N$ is linked to its (classical) linear probabilities by

$$\Delta(\widetilde{\mathsf{P}}) = \sum_{a \in \mathcal{L}\setminus\{0\}} \mathrm{LP}_a(\widetilde{\mathsf{P}}).$$

This result can be easily adapted to the generalized linear probabilities defined earlier.

We first note that the distribution $\widetilde{\mathsf{P}}$ over the group G of order N is completely defined by the mapping

$$\begin{array}{rcrcl} \mathsf{f}_{\widetilde{\mathsf{P}}} & : & \mathsf{G} & \longrightarrow & \mathbf{R} \\ & & a & \longmapsto & \mathsf{f}_{\widetilde{\mathsf{P}}}(a) = \epsilon_a = \widetilde{\mathsf{P}}[a] - \frac{1}{N}. \end{array} \tag{7.12}$$

Using this notation and the elementary Fourier analysis introduced in Subsection 7.5, we obtain the following expression of the squared Euclidean imbalance.

Lemma 7.6 *Let* G *be a finite Abelian group of order* N *and let* $\widetilde{\mathsf{P}}$ *be a probability distribution over* G*. We have*

$$\Delta(\widetilde{\mathsf{P}}) = N\|\mathsf{f}_{\widetilde{\mathsf{P}}}\|_2^2 = \|\widehat{\mathsf{f}_{\widetilde{\mathsf{P}}}}\|_2^2,$$

where $\mathsf{f}_{\widetilde{\mathsf{P}}}$ *is defined as in* (7.12).

Proof. From Definition 6.10, and using the notations of this section concerning the ϵ_a's, we have

$$\Delta(\widetilde{\mathsf{P}}) = N \sum_{a\in\mathsf{G}} \epsilon_a^2 = N \sum_{a\in\mathsf{G}} \mathsf{f}_{\widetilde{\mathsf{P}}}(a)^2 = N\|\mathsf{f}_{\widetilde{\mathsf{P}}}\|_2^2.$$

Plancherel's formula (Theorem 7.2) allows to conclude. □

Lemma 7.7 *Let* G *be a finite Abelian group and let* $\widetilde{\mathsf{P}}$ *be a probability distribution over* G. *Let* $A \in \mathsf{G}$ *be a random variable sampled according to* $\widetilde{\mathsf{P}}$. *For all characters* $\chi : \mathsf{G} \to \mathbf{C}^\times$ *we have*

$$\widehat{\mathsf{f}_{\widetilde{\mathsf{P}}}}(\chi) = \begin{cases} \mathrm{E}(\overline{\chi}(A)) & \textit{when } \chi \neq \chi_0 \\ 0 & \textit{otherwise,} \end{cases}$$

where $\mathsf{f}_{\widetilde{\mathsf{P}}}$ *is defined as in* (7.12).

Proof. Let N denote the order of G. By definition, for all $\chi \in \widehat{\mathsf{G}}$ we have

$$\widehat{\mathsf{f}_{\widetilde{\mathsf{P}}}}(\chi) = \sum_{a\in\mathsf{G}} \mathsf{f}_{\widetilde{\mathsf{P}}}(a)\overline{\chi}(a) = \sum_{a\in\mathsf{G}} \left(\widetilde{\mathsf{P}}[a] - \frac{1}{N}\right)\overline{\chi}(a) = \sum_{a\in\mathsf{G}} \widetilde{\mathsf{P}}[a]\overline{\chi}(a) - \mathbf{1}_{\chi=\chi_0},$$

where the last equality relies on Lemma 7.4. □

Based on lemmas 7.6 and 7.7, we can now easily generalize Proposition 7.2.

Proposition 7.4 *[Generalization of Proposition 7.2] Let* $\widetilde{\mathsf{P}}$ *be a probability distribution over the finite Abelian group* G. *The squared Euclidean imbalance (SEI) of* $\widetilde{\mathsf{P}}$ *is related to its linear probabilities by:*

$$\Delta(\widetilde{\mathsf{P}}) = \sum_{\chi\in\widehat{\mathsf{G}}\setminus\{\chi_0\}} \mathrm{LP}_\chi(\widetilde{\mathsf{P}}). \tag{7.13}$$

Proof. The result easily follows by successively using lemmas 7.6 and 7.7. □

Equation 7.13 can be pretty insightful: in situations where one particular character χ is such that $\mathrm{LP}_\chi(\widetilde{\mathsf{P}})$ overwhelms all other linear probabilities, then this single character can be used to approximate the linear hull (that is, the cumulative effect of all characteristics). In that case, there exists a linear distinguisher which is nearly optimal in terms of the number of samples.

Proposition 7.4 also allows to study what happens when combining *independent* sources. We consider two examples where we respectively *add* and *concatenate* independent samples.

Lemma 7.8 *(Addition of Sources) Let* G *be a finite Abelian group. Let* $A_1, A_2 \in \mathsf{G}$ *be two independent random variables of respective distributions* $\widetilde{\mathsf{P}}_1$ *and* $\widetilde{\mathsf{P}}_2$. *Let* $\widetilde{\mathsf{P}}$ *be the* *distribution of the random variable $A_1 + A_2$. We have*

$$\Delta(\widetilde{\mathsf{P}}) \leq \Delta(\widetilde{\mathsf{P}}_1)\Delta(\widetilde{\mathsf{P}}_2).$$

Proof. Successively using the fact that characters are homomorphisms and that the random variables A_1 and A_2 are independent we obtain that for all $\chi \in \widehat{\mathsf{G}}$

$$\mathrm{LP}_\chi(\widetilde{\mathsf{P}}) = |\mathrm{E}(\overline{\chi}(A_1))\mathrm{E}(\overline{\chi}(A_2))|^2 = \mathrm{LP}_\chi(\widetilde{\mathsf{P}}_1)\mathrm{LP}_\chi(\widetilde{\mathsf{P}}_2).$$

From this and from Proposition 7.4 we deduce

$$\Delta(\widetilde{\mathsf{P}}) \leq \left(\sum_{\chi\in\widehat{\mathsf{G}}\setminus\{\chi_0\}} \mathrm{LP}_\chi(\widetilde{\mathsf{P}}_1)\right)\left(\sum_{\chi\in\widehat{\mathsf{G}}\setminus\{\chi_0\}} \mathrm{LP}_\chi(\widetilde{\mathsf{P}}_2)\right) = \Delta(\widetilde{\mathsf{P}}_1)\Delta(\widetilde{\mathsf{P}}_2).$$

□

Note that the previous bound is tight whenever there exists $\chi \in \widehat{\mathsf{G}}$ such that $\Delta(\mathsf{P}_1) \approx \mathrm{LP}_\chi(\widetilde{\mathsf{P}}_1)$ and $\Delta(\mathsf{P}_2) \approx \mathrm{LP}_\chi(\widetilde{\mathsf{P}}_2)$.

Lemma 7.9 *(Concatenation of Sources) Let G_1 and G_2 be two finite Abelian groups, and let $\mathsf{G} = \mathsf{G}_1 \times \mathsf{G}_2$. Let $A_1 \in \mathsf{G}_1$ and $A_2 \in \mathsf{G}_2$ be two independent random variables of respective distributions $\widetilde{\mathsf{P}}_1$ and $\widetilde{\mathsf{P}}_2$. Let $\widetilde{\mathsf{P}}$ be the distribution of the random variable $(A_1, A_2) \in \mathsf{G}$. We have*

$$\Delta(\widetilde{\mathsf{P}}) = (\Delta(\widetilde{\mathsf{P}}_1) + 1)(\Delta(\widetilde{\mathsf{P}}_2) + 1) - 1.$$

Proof. From Proposition 7.4 we know that $\Delta(\widetilde{\mathsf{P}}) = \sum_{\chi\in\widehat{\mathsf{G}}} \mathrm{LP}_\chi(\widetilde{\mathsf{P}})$. Since $\widehat{\mathsf{G}} \cong \mathsf{G} = \mathsf{G}_1 \times \mathsf{G}_2 \cong \widehat{\mathsf{G}}_1 \times \widehat{\mathsf{G}}_2$, this gives

$$\Delta(\widetilde{\mathsf{P}}) = \sum_{(\mu,\kappa)\in\widehat{\mathsf{G}}_1\times\widehat{\mathsf{G}}_2} \mathrm{LP}_\mu(\widetilde{\mathsf{P}}_1)\mathrm{LP}_\kappa(\widetilde{\mathsf{P}}_2) = (\Delta(\widetilde{\mathsf{P}}_1) + 1)(\Delta(\widetilde{\mathsf{P}}_2) + 1) - 1.$$

□

We note that when both $\Delta(\widetilde{\mathsf{P}}_1)$ and $\Delta(\widetilde{\mathsf{P}}_2)$ are small, then the previous lemma shows that $\Delta(\widetilde{\mathsf{P}}) \approx \Delta(\widetilde{\mathsf{P}}_1) + \Delta(\widetilde{\mathsf{P}}_2)$.

In the rest of this subsection, we will reconsider Example 7.3 and show that the fact that a (well chosen) generalized linear distinguisher succeeds where a classical linear distinguisher eventually fails is not exceptional. More precisely, we will show that if a given biased distribution can be distinguished from the uniform distribution with a non negligible advantage by some distinguisher, then there exists a generalized linear distinguisher which can also distinguish it with a non negligible advantage, i.e., there exists a group structure on the underlying set and a powerful linear distinguisher defined with respect to this structure.

Definition 7.8 *Let $\widetilde{\mathsf{P}}$ be a probability distribution over a finite Abelian group G. We denote by $\mathrm{LP}^m_{\max}(\widetilde{\mathsf{P}})$ the maximum value of $\mathrm{LP}_\chi(\widetilde{\mathsf{P}})$ over $\chi \in \widehat{\mathsf{G}} \setminus \{\chi_0\}$ where the order of χ divides m, i.e.,*

$$\mathrm{LP}^m_{\max}(\widetilde{\mathsf{P}}) = \max_{\substack{\chi \in \widehat{\mathsf{G}} \setminus \{\chi_0\} \\ \chi^m = \chi_0}} \mathrm{LP}_\chi(\widetilde{\mathsf{P}}).$$

We denote by $\mathrm{LP}^m_{\mathrm{MAX}}(\widetilde{\mathsf{P}})$ the maximum value of $\mathrm{LP}^m_{\max}(\widetilde{\mathsf{P}})$ over all group laws that can be defined on the finite set G, i.e., if $\diamondsuit$ denotes an arbitrary group law on the finite set G we let

$$\mathrm{LP}^m_{\mathrm{MAX}}(\widetilde{\mathsf{P}}) = \max_{\diamondsuit} \mathrm{LP}^m_{\max}(\widetilde{\mathsf{P}}).$$

In the previous definition, we note that $\mathrm{LP}^m_{\max}$ is a measure that depends of the underlying group structure whereas $\mathrm{LP}^m_{\mathrm{MAX}}$ is not. Using these notations, we can deduce the following lemma from Proposition 7.4.

Lemma 7.10 *Let $\widetilde{\mathsf{P}}$ be a probability distribution over a finite Abelian group G of order N. Let m be the exponent of G. Then,*

$$\Delta(\widetilde{\mathsf{P}}) \leq (N-1)\mathrm{LP}^m_{\max}(\widetilde{\mathsf{P}}) \quad \textit{and} \quad \Delta(\widetilde{\mathsf{P}}) \leq (N-1)\mathrm{LP}^m_{\mathrm{MAX}}(\widetilde{\mathsf{P}}).$$

Proof. Since m is the exponent of G and since $\mathsf{G} \cong \widehat{\mathsf{G}}$, then $\mathrm{LP}_\chi(\widetilde{\mathsf{P}}) \leq \mathrm{LP}^m_{\max}(\widetilde{\mathsf{P}})$ for all $\chi \in \widehat{\mathsf{G}}$. Proposition 7.4 allows to conclude. □

This result shows that the best distinguisher between a biased distribution $\widetilde{\mathsf{P}}$ and the uniform distribution has a data complexity at least $n-1$ times smaller than the one of the best distinguisher between $\widetilde{\mathsf{P}}$ and the uniform distribution. This result is not really of practical interest since one usually considers linear (or more generally, projection-based) distinguishers when the best distinguisher cannot be implemented. The following theorem (which is a generalization of Theorem 7.1) links the data complexity of the best distinguisher on the reduced sample space and the best generalized linear distinguisher. It shows that in the particular case where the sample space is reduced by a homomorphic projection, bounding the linear probability of the source is sufficient to bound the advantage of the best distinguisher on the reduced sample space.

Theorem 7.3 *Let G and H be two finite Abelian groups of order N and n respectively, such that $n|N$. Let $h : \mathsf{G} \to \mathsf{H}$ be a surjective group homomorphism. Let $\widetilde{\mathsf{P}}$ be a probability distribution of support G and let $G \in \mathsf{G}$ be a random variable sampled according to $\widetilde{\mathsf{P}}$. Let P be the distribution of $h(G) \in \mathsf{H}$. Then:*

$$\Delta(\mathsf{P}) \leq (n-1)\mathrm{LP}^n_{\max}(\widetilde{\mathsf{P}}).$$

Proof. From Proposition 7.4 we have

$$\Delta(\mathsf{P}) = \sum_{\chi \in \widehat{\mathsf{H}} \setminus \{\chi_0\}} \mathrm{LP}_{\chi}(\mathsf{P}) = \sum_{\chi \in \widehat{\mathsf{H}} \setminus \{\chi_0\}} \mathrm{LP}_{\chi \circ h}(\widetilde{\mathsf{P}}) \leq (n-1)\mathrm{LP}^{n}_{\max}(\widetilde{\mathsf{P}}).$$

Concerning the inequality, we note that $\kappa = \chi \circ h$ is a character of G such that $\kappa^n = \kappa_0$. Consequently,

$$\max_{\chi \in \widehat{\mathsf{H}} \setminus \{\chi_0\}} \mathrm{LP}_{\chi \circ h}(\widetilde{\mathsf{P}}) \leq \max_{\substack{\kappa \in \widehat{\mathsf{G}} \setminus \{\kappa_0\} \\ \text{s.t. } \kappa^n = \kappa_0}} \mathrm{LP}_{\kappa}(\mathsf{P}) = \mathrm{LP}^{n}_{\max}(\mathsf{P}).$$

□

We stress that the previous theorem only applies when the sample space is reduced through a group homomorphism, i.e., in a *linear* way. Indeed, there exists practical examples of random sources with a small $\mathrm{LP}^{n}_{\max}(\widetilde{\mathsf{P}})$ that are significantly broken when the source space is reduced by a (well chosen) non-homomorphic projection (see the case study on page 86). Consequently, the previous result tells us nothing about the advantage of an adversary using an arbitrary projection. In what follows we show a security criterion which is *sufficient* to obtain provable security against *any* projection-based distinguisher based on a balanced projection.

Theorem 7.4 *Let* G *and* H *be two finite Abelian groups of order* N *and* n *respectively, such that* $n|N$. *Let* $h : \mathsf{G} \to \mathsf{H}$ *be a balanced projection. Let* $\widetilde{\mathsf{P}}$ *be a probability distribution of support* G *and let* $G \in \mathsf{G}$ *be a random variable sampled according to* $\widetilde{\mathsf{P}}$. *Let* P *be the distribution of* $h(G) \in \mathsf{H}$. *Then:*

$$\Delta(\mathsf{P}) \leq (n-1)\mathrm{LP}^{n}_{\mathrm{MAX}}(\widetilde{\mathsf{P}}).$$

Proof. We first define a group structure on H such that h is a homomorphism. Let $\mathsf{H} = \{h_1, h_2, \ldots, h_n\}$ (where h_1 is the neutral element) and let $\mathsf{G}_i = h^{-1}(h_i) \subset \mathsf{G}$ for $i = 1, 2, \ldots, n$. Since h is balanced, the G_i's form a partition of G and are such that $|\mathsf{G}_i| = \frac{N}{n}$ for $i = 1, 2, \ldots, n$. Based on the group law on H, we can define the product $\mathsf{G}_i\mathsf{G}_j$ by

$$\mathsf{G}_k = \mathsf{G}_i\mathsf{G}_j \Leftrightarrow h_k = h_i h_j,$$

for all $i, j, k = 1, 2, \ldots, n$. This directly defines a group law on the G_i's, the identity element being G_1, the inverse of G_i being $h^{-1}(h_i^{-1})$, the associativity following directly from the one of the law defined on H.

Consider an arbitrary group law on G_1 and let $\tau_i : \mathsf{G}_1 \to \mathsf{G}_i$ define a bijection between G_1 and G_i for $i = 1, 2, \ldots, n$ (where τ_1 is the identity). Let $x, y \in \mathsf{G}$ be two arbitrary elements and let i,j,k be such that $x \in \mathsf{G}_i$, $y \in \mathsf{G}_j$ and $\mathsf{G}_k = \mathsf{G}_i\mathsf{G}_j$. We defined the product xy on G by

$$xy = \tau_k(\tau_i^{-1}(x)\tau_j^{-1}(y)).$$

It is easy to see that this product is a group law on G. The neutral element is actually the neutral element of the group law defined on G_1, the inverse of $x \in \mathsf{G}_i$ is $\tau_\ell(\tau_i^{-1}(x)^{-1})$ (where ℓ is such that $\mathsf{G}_i^{-1} = \mathsf{G}_\ell$). It is moreover easy to see that the projection $h : \mathsf{G} \to \mathsf{H}$ is a group homomorphism with respect to the group law we have just defined on G. Indeed, using the previous notations we have $h(x)h(y) = g_ig_j = g_k = h(xy)$.

Given the group law we just defined on G, we can apply Theorem 7.3 (since h is a homomorphism with respect to this group law) and get

$$\Delta(\mathsf{P}) \leq (n-1)\mathrm{LP}^n_{\max}(\widetilde{\mathsf{P}}).$$

Since $\mathrm{LP}^n_{\max}(\widetilde{\mathsf{P}}) \leq \mathrm{LP}^n_{\mathrm{MAX}}(\widetilde{\mathsf{P}})$ and since $\mathrm{LP}^n_{\mathrm{MAX}}(\widetilde{\mathsf{P}})$ does not depend on the group law on G, the result follows. □

Consequently, assuming there exists an "efficient" projection-based distinguisher on $\widetilde{\mathsf{P}}$ based on a balanced projection h on a "small" set H, $\Delta(\mathsf{P})$ must be large, n must be small and thus, according to the previous theorem, $\mathrm{LP}^n_{\mathrm{MAX}}(\widetilde{\mathsf{P}})$ must be large. This means there exists a group structure on G and a character on this group of small order that define an effective linear cryptanalysis: *if we can efficiently distinguish by compressing the samples, we can also do it linearly.*

To the best of our knowledge, all widespread block ciphers provably secure against linear cryptanalysis consider in their security proof a *specific* group or field structure on the text space. Usually, the most convenient is the one used to actually define the block cipher. Obviously, a potential adversary is not limited to the description considered by the designers. The previous theorem shows that, provided that a known plaintext attack on the block cipher exists, then some *change* to the group structure of the text space is sufficient to perform a successful linear cryptanalysis of the cipher (note that finding the correct group structure might be a non-trivial task). In other words, although the cipher is stated to be provably secure against linear cryptanalysis, it might not be the case when generalizing linear cryptanalysis to other group structures. This is mainly due to the fact that the SEI does not depend on the group structure given to the text space (only the distance of P from the uniform distribution is relevant) whereas the linear probability is a measure that *depends* on the group structure. Consequently, when proving the resistance to linear cryptanalysis, one should ideally bound the value of $\mathrm{LP}^n_{\mathrm{MAX}}(\widetilde{\mathsf{P}})$ and not of $\mathrm{LP}^n_{\max}(\widetilde{\mathsf{P}})$ (as it is currently the case for most block ciphers).

7.8 Links with Differential Cryptanalysis

Differential cryptanalysis [21,22,24] is a chosen plaintext attack where pairs of texts are chosen with a fixed difference. In the case of block ciphers, the adversary looks for a high correlation between a specific input difference and a specific output difference. In the case of random sources we consider a natural way of expressing the main quantity differential cryptanalysis is based on, namely, the *differential probability* [124].

Definition 7.9 *Let G be a finite Abelian group, let $\widetilde{\mathsf{P}}$ be a probability distribution over G, and let $A, B \in \mathsf{G}$ be two independent random variables sampled according to $\widetilde{\mathsf{P}}$. The differential probability of the distribution $\widetilde{\mathsf{P}}$ over G with respect to the mask $u \in \mathsf{G}$ is*

$$\mathrm{DP}_u(\widetilde{\mathsf{P}}) = \Pr[A^{-1} \cdot B = u] = \Pr[B = A \cdot u].$$

It is known that, in the binary case, linear and differential cryptanalysis are linked, and in particular that the linear probability is equal to the Fourier transform of the differential probability [32]. This duality extends to our generalization of the linear probability as the following lemma shows.

Lemma 7.11 *Let G be a finite Abelian group of order N and let $\widetilde{\mathsf{P}}$ be a probability distribution over G. Let $\chi \in \widehat{\mathsf{G}}$ be a character of G and $u \in \mathsf{G}$. The inverse Fourier transform of $\mathrm{LP}_\chi(\widetilde{\mathsf{P}})$ at the point u is*

$$\widehat{\mathrm{LP}}_u(\widetilde{\mathsf{P}}) = \mathrm{DP}_u(\widetilde{\mathsf{P}}).$$

and the Fourier transform of $\mathrm{DP}_u(\widetilde{\mathsf{P}})$ at the point χ is

$$\widehat{\mathrm{DP}}_\chi(\widetilde{\mathsf{P}}) = \mathrm{LP}_\chi(\widetilde{\mathsf{P}}).$$

Proof. By definition, $\mathrm{LP}_\chi(\widetilde{\mathsf{P}}) = \mathrm{E}(\chi(A))\mathrm{E}(\overline{\chi}(B))$ where $A, B \in \mathsf{G}$ are two independent random variables sampled according to $\widetilde{\mathsf{P}}$. Successively using the inverse of the Fourier transform, the fact that A and B are independent, that the mean is linear, and that χ is a homomorphism, we have for all $u \in \mathsf{G}$:

$$\widehat{\mathrm{LP}}_u(\widetilde{\mathsf{P}}) = \frac{1}{N} \sum_{\chi \in \widehat{\mathsf{G}}} \mathrm{E}\left(\chi(A)\overline{\chi}(B)\right)\chi(u) = \frac{1}{N}\mathrm{E}\left(\sum_{\chi \in \widehat{\mathsf{G}}} \chi(A \cdot u)\overline{\chi}(B)\right),$$

which is an expression that we can simplify using Lemma 7.4 in order to obtain $\widehat{\mathrm{LP}}_u(\widetilde{\mathsf{P}}) = \mathrm{E}\left(\mathbf{1}_{A \cdot u = B}\right) = \Pr[A \cdot u = B] = \mathrm{DP}_u(\widetilde{\mathsf{P}})$, which proves the first equality. Conversely, substituting DP by $\widehat{\mathrm{LP}}$ in the Fourier transform of the differential probability and expanding the expression leads to

$$\widehat{\mathrm{DP}}_\chi(\widetilde{\mathsf{P}}) = \sum_{u \in \mathsf{G}} \left(\frac{1}{N} \sum_{\rho \in \widehat{\mathsf{G}}} \mathrm{LP}_\rho(\widetilde{\mathsf{P}})\rho(u)\right)\overline{\chi}(u) = \frac{1}{N} \sum_{\rho \in \widehat{\mathsf{G}}} \mathrm{LP}_\rho(\widetilde{\mathsf{P}}) \sum_{u \in \mathsf{G}} \rho(u)\overline{\chi}(u)$$

which is an expression that can be simplified using the orthogonality relations given in Lemma 7.4. We obtain

$$\widehat{\mathrm{DP}}_\chi(\widetilde{\mathsf{P}}) = \sum_{\rho \in \widehat{\mathsf{G}}} \mathrm{LP}_\rho(\widetilde{\mathsf{P}}) \mathbf{1}_{\rho=\chi} = \mathrm{LP}_\chi(\widetilde{\mathsf{P}}).$$

□

Chapter 8

Projection-Based Distinguishers Between two Oracles

So far we discussed how to distinguish random values. Now we investigate applications for distinguishing random functions, such as block ciphers, and in particular, how to transform this into the previous problem. This distinction may look completely useless from a mathematical point of view since we can consider both cases with the abstract notion of random variable. The crucial difference comes from implementation reasons as they live in spaces of quite different sizes. For instance, a random value in $\{0,1\}^{128}$ is represented by 128 bits whereas a block cipher with 128-bit blocks requires $\log_2(2^{128}!) \approx 2^{135}$ bits. In practice, the distinguisher doesn't have access to the full description of the block cipher but rather to a few input/output pairs.

8.1 From Random Sources to Random Oracles

A block cipher on a finite set is a family of permutations on that set, indexed by a parameter called the key. More formally, let $\mathcal{T}$ and $\mathcal{K}$ be two finite sets, respectively called the *text space* and the *key space*. A block cipher C on the text space $\mathcal{T}$ and key space $\mathcal{K}$ is a set of $|\mathcal{K}|$ permutations on $\mathcal{T}$, i.e.,

$$\mathsf{C} = \{\mathsf{C}_k : \mathcal{T} \to \mathcal{T} : k \in \mathcal{K}\},$$

where each C_k is a permutation. For simplicity, we assume from now on that $\mathsf{C}_k \neq \mathsf{C}_{k'}$ when $k \neq k'$. In practice this might not be always the case, like for example with the DES which is known to have weak keys [43]. Yet, removing these weak keys from the key space suffices to solve this issue. When C corresponds to the set of *all* possible permutations on $\mathcal{T}$ (in which case $|\mathcal{K}| = |\mathcal{T}|!$) it is called the *perfect cipher* and is denoted $\mathsf{C}^\star$.

We are mainly interested in distinguishing attacks as they often easily lead to key recovery attacks. These can be formalized as an hypothesis problem. Let $\mathcal{T}$ and $\mathcal{K}$ be two finite sets and let C be a block cipher defined on the text space $\mathcal{T}$ and the key space $\mathcal{K}$. Let $\mathsf{C}^\star$ denote the perfect cipher on $\mathcal{T}$. We consider a random oracle $\mathcal{O}$ which is either sampled uniformly at random among all possible permutations (hypothesis

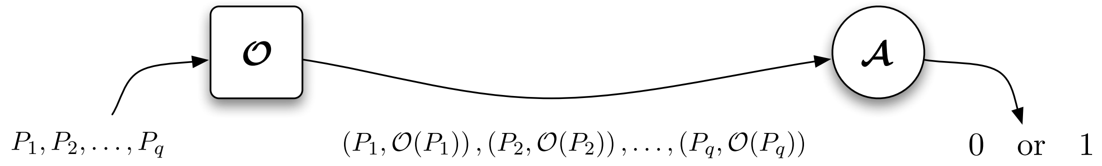

$\mathcal{O}$
$\mathcal{A}$
$P_1, P_2, \ldots, P_q$
$(P_1, \mathcal{O}(P_1)), (P_2, \mathcal{O}(P_2)), \ldots, (P_q, \mathcal{O}(P_q))$
0 or 1

Figure 8.1: Distinguishing attack in a known plaintext setting

H_0) or among all the permutations defined by the block cipher C (hypothesis H_1). We denote these hypotheses $\mathsf{H}_0 : \mathcal{O} \leftarrow \mathsf{C}^\star$ and $\mathsf{H}_1 : \mathcal{O} \leftarrow \mathsf{C}$.

We restrict to known plaintext attacks (like linear cryptanalysis). In this setting, the q plaintexts $P_1, P_2, \ldots, P_q \in \mathcal{T}$ are assumed to be mutually independent and uniformly distributed. The random oracle is evaluated in each of these q points, outputting $C_i = \mathcal{O}(P_i)$ for $i = 1, \ldots, q$. We denote $\mathcal{L} = \mathcal{T} \times \mathcal{T}$ and $L_i = (P_i, \mathcal{O}(P_i)) = (P_i, C_i) \in \mathcal{L}$ for $i = 1, 2, \ldots, q$ the resulting samples that are finally submitted to the distinguisher. This situation is represented on Figure 8.1.

Under hypothesis H_0, we note that for all $\ell = (p, c) \in \mathcal{L}$ and $L = (P, \mathcal{O}(P))$ where $P \in \mathcal{T}$ is uniformly distributed,

$$\Pr[L = \ell] = \Pr[P = p, \mathcal{O}(p) = c] = \Pr[P = p]\Pr[\mathcal{O}(p) = c] = \frac{1}{|\mathcal{T}|^2} = \frac{1}{|\mathcal{L}|},$$

where the probabilities hold over the random oracle and the random plaintexts. We see that the L_i's are uniformly distributed under hypothesis H_0. The distinguishing problem between $\mathsf{H}_0 : \mathcal{O} \leftarrow \mathsf{C}^\star$ and $\mathsf{H}_1 : \mathcal{O} \leftarrow \mathsf{C}$ can now be turned into a new equivalent one, in which the two hypotheses are $\mathsf{H}_0 : \widetilde{\mathsf{P}} = \widetilde{\mathsf{U}}$ and $\mathsf{H}_1 : \widetilde{\mathsf{P}} \in \widetilde{\mathcal{D}}$, where

- $\widetilde{\mathsf{P}}$ is the distribution of $(P, \mathcal{O}(P))$,
- $\widetilde{\mathsf{U}}$ is the uniform distribution over $\mathcal{L}$, and
- $\widetilde{\mathcal{D}} = \{\widetilde{\mathsf{P}}_1, \widetilde{\mathsf{P}}_2, \ldots, \widetilde{\mathsf{P}}_{|\mathcal{K}|}\}$ is a set of $|\mathcal{K}|$ distributions over $\mathcal{L}$, where $\widetilde{\mathsf{P}}_k$ is the distribution of $L = (P, \mathsf{C}_k(P))$ when P is uniformly distributed.

This situation exactly corresponds to the composite hypothesis testing problem studied in Section 6.8.

For practical reasons we will exclusively focus on *projection-based* distinguishers which reduce the sample space by restricting the information kept about each plaintext and each ciphertext. Similarly to what was introduced in Section 7.2 in the case of distinguishers between random sources, projection-based distinguishers between random oracles reduce the sample space by means of projections. Let $\mathcal{X}$ and $\mathcal{Y}$ be two finite sets (the cardinalities of which being typically much smaller than $|\mathcal{T}|$) and let $\mathcal{Z} = \mathcal{X} \times \mathcal{Y}$. Note that the cardinalities of $\mathcal{X}$ and $\mathcal{Y}$ may differ. We consider two *balanced* projections

$$\rho : \mathcal{T} \longrightarrow \mathcal{X} \quad \text{and} \quad \mu : \mathcal{T} \longrightarrow \mathcal{Y}$$

and let $h : \mathcal{L} \to \mathcal{Z}$ be such that $h = (\rho, \mu)$. We let P be the distribution of $h(L)$ when $L \sim \widetilde{\mathsf{P}}$, and respectively denote U and P_k the possible values of this distribution when $\widetilde{\mathsf{P}} = \widetilde{\mathsf{U}}$ and $\widetilde{\mathsf{P}} = \widetilde{\mathsf{P}}_k$ respectively (note that since h is balanced and since $\widetilde{\mathsf{U}}$ is uniform, then U is uniform too). As the adversary does not know the key k, we assume that the projections are the same for all possible values of k. For the same reason, we also assume that the decision rule is the same in all cases. The original distinguishing problem between $\mathsf{H}_0 : \mathcal{O} \leftarrow \mathsf{C}^\star$ and $\mathsf{H}_1 : \mathcal{O} \leftarrow \mathsf{C}$ now reads

$$\mathsf{H}_0 : \mathsf{P} = \mathsf{U} \quad \text{against} \quad \mathsf{H}_1 : \mathsf{P} \in \mathcal{D},$$

where

- P is the distribution of $h(L) = (\rho(P), \mu(\mathcal{O}(P)))$,
- U is the uniform distribution over $\mathcal{Z} = \mathcal{X} \times \mathcal{Y}$,
- $\mathcal{D} = \{\mathsf{P}_1, \mathsf{P}_2, \ldots, \mathsf{P}_{|\mathcal{K}|}\}$ is a set of $|\mathcal{K}|$ distributions over $\mathcal{Z}$, where P_k is the distribution of $(\rho(P), \mu(\mathsf{C}_k(P)))$ when P is uniformly distributed.

According to Theorem 6.4, the best (asymptotic) q-limited distinguisher is in that case defined by the acceptance region $\Pi_q^\star = \Pi^\star \cap \mathcal{P}_q$ where

$$\Pi^\star = \{\mathsf{P} \in \mathcal{P} \;:\; \min_{1 \le k \le |\mathcal{K}|} \mathsf{L}_k(\mathsf{P}) \le 0\} \quad \text{with} \quad \mathsf{L}_k(\mathsf{P}) = \sum_{z \in \mathcal{Z}} \mathsf{P}[z] \log \frac{1}{|\mathcal{Z}|\, \mathsf{P}_k[z]}.$$

8.2 Cryptanalysis Complexity by means of Transition and Bias Matrices

Definition 8.1 *Let $\mathcal{T}$ be a finite set and $\mathcal{O} : \mathcal{T} \to \mathcal{T}$ be an oracle on that set. Let $\mathcal{X}$ and $\mathcal{Y}$ be two finite sets, such that $|\mathcal{X}|, |\mathcal{Y}| < |\mathcal{T}|$, and let $\rho : \mathcal{T} \longrightarrow \mathcal{X}$ and $\mu : \mathcal{T} \longrightarrow \mathcal{Y}$ be two balanced projections. The transition matrix of $\mathcal{O}$ with respect to the projections ρ, μ is the $\mathcal{X} \times \mathcal{Y}$ matrix $\mathbf{T}^{\rho,\mu}$ defined by*

$$[\mathbf{T}^{\rho,\mu}]_{x,y} = \Pr[\mu(\mathcal{O}(P)) = y | P \leftarrow \rho^{-1}(x)]$$

for all $(x, y) \in \mathcal{X} \times \mathcal{Y}$.

Considering the hypothesis testing problem described in the previous section, we note that the distribution of each $Z_i = (X_i, Y_i) = (\rho(P_i), \mu(\mathcal{O}(P_i)))$ can be expressed by means of the transition matrix $\mathbf{T}^{\rho,\mu}$ of the random oracle under both hypotheses H_0 and H_1. Indeed, for $\mathsf{P} \in \{\mathsf{U}, \mathsf{P}_1, \ldots, \mathsf{P}_{|\mathcal{K}|}\}$ and $z = (x, y) \in \mathcal{Z}$ we have

$$\mathsf{P}[z] = \Pr[\mu(\mathcal{O}(P)) = y | \rho(P) = x] \Pr[\rho(P) = x] = \frac{[\mathbf{T}^{\rho,\mu}]_{x,y}}{|\mathcal{X}|}. \tag{8.1}$$

Since U is the uniform distribution, the transition matrix under hypothesis H_0 is a uniform matrix that we denote by $\mathbf{U}$, where for all $(x, y) \in \mathcal{X} \times \mathcal{Y}$ we have

$$[\mathbf{U}]_{x,y} = \frac{1}{|\mathcal{Y}|}. \tag{8.2}$$

We denote by $\mathbf{T}_k^{\rho,\mu}$ the transition matrix corresponding to P_k.

Definition 8.2 *Let $\mathcal{T}$ be a finite set and $\mathcal{O} : \mathcal{T} \to \mathcal{T}$ be an oracle on that set. Let $\mathcal{X}$ and $\mathcal{Y}$ be two finite sets, such that $|\mathcal{X}|, |\mathcal{Y}| < |\mathcal{T}|$, and let $\rho : \mathcal{T} \longrightarrow \mathcal{X}$ and $\mu : \mathcal{T} \longrightarrow \mathcal{Y}$ be two balanced projections. The bias matrix of $\mathcal{O}$ with respect to the projections ρ, μ is the $\mathcal{X} \times \mathcal{Y}$ matrix $\mathbf{B}^{\rho,\mu}$ defined by*

$$\mathbf{B}^{\rho,\mu} = \mathbf{T}^{\rho,\mu} - \mathbf{U},$$

where $\mathbf{T}^{\rho,\mu}$ is the transition matrix of $\mathcal{O}$ with respect to the projections ρ μ and where $\mathbf{U}$ is the $\mathcal{X} \times \mathcal{Y}$ matrix such that $[\mathbf{U}]_{x,y} = \frac{1}{|\mathcal{Y}|}$ for all $(x, y) \in \mathcal{X} \times \mathcal{Y}$.

The following lemma shows how the bias matrix relates to the squared Euclidean imbalance

Lemma 8.1 *Let $\mathcal{T}$ and $\mathcal{K}$ be two finite sets and C be a block cipher over the text space $\mathcal{T}$ and the key space $\mathcal{K}$. Let $k \in \mathcal{K}$. Let $\mathcal{X}$ and $\mathcal{Y}$ be two finite sets, such that $|\mathcal{X}|, |\mathcal{Y}| < |\mathcal{L}|$, and let $\rho : \mathcal{T} \longrightarrow \mathcal{X}$ and $\mu : \mathcal{T} \longrightarrow \mathcal{Y}$ be two balanced projections. Let $\mathbf{B}_k^{\rho,\mu}$ be the bias matrix of C_k with respect to ρ and μ and let P_k be the distribution of $(\rho(P), \mu(\mathsf{C}_k(P))) \in \mathcal{X} \times \mathcal{Y}$, where $P \in \mathcal{T}$ is uniformly distributed. Then*

$$\Delta(\mathsf{P}_k) = \frac{|\mathcal{Y}|}{|\mathcal{X}|} \|\mathbf{B}_k^{\rho,\mu}\|_2^2.$$

Proof. Denoting by $\mathbf{T}_k^{\rho,\mu}$ the transition matrix of C_k with respect to ρ and μ, we have according to Definition 6.10

$$\Delta(\mathsf{P}_k) = |\mathcal{X}|\,|\mathcal{Y}| \sum_{x,y} \left(\mathsf{P}_k[x, y] - \frac{1}{|\mathcal{X}|\,|\mathcal{Y}|} \right)^2 = \frac{|\mathcal{Y}|}{|\mathcal{X}|} \sum_{x,y} \left([\mathbf{T}_k^{\rho,\mu}]_{x,y} - \frac{1}{|\mathcal{Y}|} \right)^2$$

since $\mathsf{P}_k[x, y] = \frac{[\mathbf{T}_k^{\rho,\mu}]_{x,y}}{|\mathcal{X}|}$ as noted in (8.1). □

The following result is a direct implication of the previous lemma and of Heuristic 6.3. It leads to the conclusion that $\frac{8 \ln 2}{\min_i \Delta(\mathsf{P}_i)}$ samples are sufficient to distinguish P from the uniform distribution (see page 56).

Heuristic 8.1 Let $\mathcal{T}$ and $\mathcal{K}$ be two finite sets and C be a block cipher over the text space $\mathcal{T}$ and the key space $\mathcal{K}$. Let $\mathcal{X}$ and $\mathcal{Y}$ be two finite sets, such that $|\mathcal{X}|, |\mathcal{Y}| < |\mathcal{T}|$, and let $\rho : \mathcal{T} \longrightarrow \mathcal{X}$ and $\mu : \mathcal{T} \longrightarrow \mathcal{Y}$ be two balanced projections. For all $k \in \mathcal{K}$, let

$\mathbf{B}_k^{\rho,\mu}$ be the bias matrix of C_k with respect to ρ and μ. Assuming that $\|\mathbf{B}_k^{\rho,\mu}\|_2 \ll 1$, the q-limited projection-based distinguisher between $\mathsf{H}_0 : \mathcal{O} \leftarrow \mathsf{C}^\star$ and $\mathsf{H}_1 : \mathcal{O} \leftarrow \mathsf{C}$ based on the projections ρ and μ reaches a non-negligible advantage when

$$q = \frac{|\mathcal{X}|}{|\mathcal{Y}|} \frac{8 \ln 2}{\min\limits_{1 \le k \le |\mathcal{K}|} \|\mathbf{B}_k^{\rho,\mu}\|_2^2}. \tag{8.3}$$

□

This heuristic tells us what is the approximate number of queries required to reach a non-negligible advantage. Since the key space can be assumed to be rather large, it is impossible in general to evaluate the right-hand side of (8.3). In certain cases, it might be possible to reduce the min for all keys to one running over equivalence classes of bias matrices (where two such matrices would be equivalent whenever their respective Euclidean norm are equal). The most radical assumption arises when one assumes that there is only one equivalence class.

Definition 8.3 ***(Hypothesis of stochastic equivalence)*** *Under the notations of Definition 8.2, the hypothesis of stochastic equivalence states that any pair of keys $k, k' \in \mathcal{K}$ we can write*

$$\|\mathbf{B}_k^{\rho,\mu}\|_2 \approx \|\mathbf{B}_{k'}^{\rho,\mu}\|_2.$$

This assumption was initially formalized by Lai in the scope of differential cryptanalysis (see [94, 96]). Under this assumption, we see that (8.3) can be approximated by

$$q = \frac{|\mathcal{X}|}{|\mathcal{Y}|} \frac{8 \ln 2}{\|\mathbf{B}_k^{\rho,\mu}\|_2^2}$$

for any $k \in \mathcal{K}$. Similarly to what we did in Definition 7.4 for probability distributions, we can define the Fourier transform $\widehat{\mathbf{B}}$ of a bias matrix $\mathbf{B}$.

Definition 8.4 *Let n and m be two positive integers, let $\mathcal{X} = \{0,1\}^n$ and $\mathcal{Y} = \{0,1\}^m$. Let $\mathbf{B}$ be a $2^n \times 2^m$ bias matrix indexed over $\mathcal{X} \times \mathcal{Y}$. The Fourier transform of $\mathbf{B}$ is the $2^n \times 2^m$ matrix $\widehat{\mathbf{B}}$ defined by*

$$[\widehat{\mathbf{B}}]_{u,v} = \sum_{(x,y) \in \mathcal{X} \times \mathcal{Y}} (-1)^{u \bullet x \oplus v \bullet y} [\mathbf{B}]_{x,y} \tag{8.4}$$

for all $(u,v) \in \mathcal{X} \times \mathcal{Y}$.

Lemma 8.2 *Under the notations of Definition 8.4 we have*

$$[\mathbf{B}]_{x,y} = \frac{1}{|\mathcal{X}|\,|\mathcal{Y}|} \sum_{(u,v) \in \mathcal{X} \times \mathcal{Y}} (-1)^{u \bullet x \oplus v \bullet y} [\widehat{\mathbf{B}}]_{u,v}. \tag{8.5}$$

Proof. Starting from the right-hand side of (8.5) and plugging (8.4) in,

$$\begin{aligned}
&\frac{1}{|\mathcal{X}|\,|\mathcal{Y}|}\sum_{(u,v)\in\mathcal{X}\times\mathcal{Y}}(-1)^{u\bullet x\oplus v\bullet y}[\widehat{\mathbf{B}}]_{u,v}\\
&\quad=\frac{1}{|\mathcal{X}|\,|\mathcal{Y}|}\sum_{(u,v)\in\mathcal{X}\times\mathcal{Y}}(-1)^{u\bullet x\oplus v\bullet y}\sum_{(x',y')\in\mathcal{X}\times\mathcal{Y}}(-1)^{u\bullet x'\oplus v\bullet y'}[\mathbf{B}]_{x',y'}\\
&\quad=\frac{1}{|\mathcal{X}|\,|\mathcal{Y}|}\sum_{(x',y')\in\mathcal{X}\times\mathcal{Y}}[\mathbf{B}]_{x',y'}\sum_{(u,v)\in\mathcal{X}\times\mathcal{Y}}(-1)^{u\bullet(x\oplus x')\oplus v\bullet(y\oplus y')}\\
&\quad=\sum_{(x',y')\in\mathcal{X}\times\mathcal{Y}}[\mathbf{B}]_{x',y'}\mathbf{1}_{x=x'}\mathbf{1}_{y=y'}\\
&\quad=[\mathbf{B}]_{x,y}.
\end{aligned}$$

□

The next proposition can be compared to Parseval's Theorem and relates the Euclidean norm of a bias matrix to that of its Fourier transform.

Proposition 8.1 *Let n and m be two positive integers, let $\mathcal{X}=\{0,1\}^n$ and $\mathcal{Y}=\{0,1\}^m$. Let $\mathbf{B}$ be a $2^n\times 2^m$ bias matrix indexed over $\mathcal{X}\times\mathcal{Y}$. The Euclidean norms of $\mathbf{B}$ and that of its Fourier transform $\widehat{\mathbf{B}}$ are related by*

$$\|\mathbf{B}\|_2^2=\frac{1}{|\mathcal{X}|\,|\mathcal{Y}|}\|\widehat{\mathbf{B}}\|_2^2.$$

Proof. By definition we have

$$\begin{aligned}
\|\widehat{\mathbf{B}}\|_2^2 &= \sum_{u,v}\left(\sum_{x,y}(-1)^{u\bullet x\oplus v\bullet y}[\mathbf{B}]_{x,y}\right)\left(\sum_{x',y'}(-1)^{u\bullet x'\oplus v\bullet y'}[\mathbf{B}]_{x',y'}\right)\\
&= \sum_{x,y,x',y'}[\mathbf{B}]_{x,y}[\mathbf{B}]_{x',y'}\sum_{u,v}(-1)^{u\bullet(x\oplus x')\oplus v\bullet(y\oplus y')}\\
&= |\mathcal{X}|\,|\mathcal{Y}|\sum_{x,y,x',y'}[\mathbf{B}]_{x,y}[\mathbf{B}]_{x',y'}\mathbf{1}_{x=x'}\mathbf{1}_{y=y'}\\
&= |\mathcal{X}|\,|\mathcal{Y}|\sum_{x,y}[\mathbf{B}]_{x,y}^2=|\mathcal{X}|\,|\mathcal{Y}|\,\|\mathbf{B}\|_2^2
\end{aligned}$$

□

Proposition 8.1 together with Lemma 8.1 show that

$$\Delta(\mathsf{P})=\frac{1}{|\mathcal{X}|^2}\|\widehat{\mathbf{B}}^{\rho,\mu}\|_2^2,$$

using the notations of the lemma.

Example 8.1 We consider a block cipher C defined over the text space $\mathcal{T} = \{0,1\}^N$ (for some positive integer N) and key space $\mathcal{K}$, and two linear balanced Boolean projections

$$\rho, \mu : \{0,1\}^N \to \{0,1\}.$$

A projection-based distinguisher based on these projections exactly corresponds to what is known as a (classical) linear distinguisher on the block cipher. It is easy to see that the bias matrix of C_k (for some $k \in \mathcal{K}$) can be written as

$$\mathbf{B}_k^{\rho,\mu} = \begin{pmatrix} \epsilon & -\epsilon \\ -\epsilon & \epsilon \end{pmatrix}$$

where ϵ is a real value (which exactly corresponds to the bias of Matsui's linear expressions). We have $\|\mathbf{B}_k^{\rho,\mu}\|_2^2 = 4\epsilon^2$. Under the hypothesis of stochastic equivalence (see Definition 8.3), it is easy to see that the bias matrix of $\mathsf{C}_{k'}$ (for $k' \neq k$) is either equal to that of C_k or such that

$$\mathbf{B}_{k'}^{\rho,\mu} = \begin{pmatrix} -\epsilon & \epsilon \\ \epsilon & -\epsilon \end{pmatrix}.$$

According to Heuristic 8.1, we see that a linear distinguisher reaches a non-negligible advantage when

$$q = \frac{2\ln 2}{\epsilon^2}$$

which (up to a constant) is a well accepted result in linear cryptanalysis [110,111]. □

8.3 Piling-up Transition Matrices

A distinguishing attack on an iterated cipher is practical on the condition that the cryptanalyst knows a transition matrix spanning several rounds. In practice, she derives a transition matrix on each round and, provided that the projections were chosen carefully, pile them in order to obtain a transition matrix on several rounds of the cipher.

We consider the scenario where a block cipher C is made of two rounds, which we assume to have the same structure to simplify the notations. In other words $\mathsf{C} = \{\mathsf{c}' \circ \mathsf{c}'' : \mathsf{c}', \mathsf{c}'' \in \mathsf{R}\}$, where R is a set of permutations on the text space $\mathcal{T}$ and key space $\mathcal{K}$. With our notations, the key space of the block cipher C is $\mathcal{K}^2$. We consider three balanced projections $\rho : \mathcal{T} \to \mathcal{X}$, $\mu : \mathcal{T} \to \mathcal{W}$, and $\phi : \mathcal{T} \to \mathcal{Y}$, respectively applied to the input of the first round R_{k_1}, on the input of the second round R_{k_2}, and on the output of the block cipher C_k, where $k = (k_1, k_2)$. We respectively denote by P_1, P_2, and P_3 the input of R_{k_1}, the input of R_{k_2}, and the output of C_k. Finally the random variables

$$\begin{tikzcd}
P_1 \arrow[r, "\rho"] \arrow[d, "\mathsf{R}_{k_1}"'] & X \arrow[d, "\mathbf{T}_1^{\rho,\mu}"] \\
P_2 \arrow[r, "\mu"] \arrow[d, "\mathsf{R}_{k_2}"'] & W \arrow[d, "\mathbf{T}_2^{\mu,\phi}"] \\
P_3 \arrow[r, "\phi"] & Y
\end{tikzcd}$$

Figure 8.2: A commutative diagram illustrating how to pile the transition matrices on a two rounds iterated block cipher

X, W, and Y respectively denote $\rho(P_1)$, $\mu(P_2)$, and $\phi(P_3)$. With these notations, the respective transition matrices of R_{k_1}, R_{k_2}, and C_k are defined by

$$\left[\mathbf{T}_{k_1}^{\rho,\mu}\right]_{x,w} = \Pr[W = w|X = x], \quad [\mathbf{T}_{k_2}^{\mu,\phi}]_{w,y} = \Pr[Y = y|W = w],$$
$$\text{and } [\mathbf{T}_{k}^{\rho,\phi}]_{x,y} = \Pr[Y = y|X = x].$$

This situation is represented on Figure 8.2. Note that we use a representation which is very similar to Wagner's unified view of block cipher cryptanalysis commutative diagrams [163].

Definition 8.5 *A sequence $X_1, X_2, X_3 \ldots$ of random variables taking values in some finite set $\mathcal{X}$ has the Markov property if for all $x_1, x_2, \ldots$ in $\mathcal{X}$ and all positive integer n we have*

$$\Pr[X_{n+1} = x_{n+1}|X_n = x_n, \ldots, X_1 = x_1] = \Pr[X_{n+1} = x_{n+1}|X_n = x_n].$$

In that case, the sequence $X_1, X_2, X_3 \ldots$ is a Markov chain and is denoted $X_1 \leftrightarrow X_2 \leftrightarrow X_3 \leftrightarrow \ldots$

Proposition 8.2 *Let $\mathcal{X}$, $\mathcal{W}$, and $\mathcal{Y}$ be three finite sets and $X \in \mathcal{X}$, $W \in \mathcal{W}$, and $Y \in \mathcal{Y}$ be three uniformly distributed random variables defined on these sets such that*

$$X \leftrightarrow W \leftrightarrow Y$$

is a Markov chain. Let $\mathbf{T}_1$ be the $|\mathcal{X}| \times |\mathcal{W}|$ transition matrix defined by $[\mathbf{T}_1]_{x,w} = \Pr[W = w|X = x]$ and $\mathbf{T}_2$ be the $|\mathcal{W}|\times|\mathcal{Y}|$ transition matrix defined by $[\mathbf{T}_2]_{w,y} = \Pr[Y = y|W = w]$. The $|\mathcal{X}| \times |\mathcal{Y}|$ transition matrix $\mathbf{T}$ defined by $[\mathbf{T}]_{x,y} = \Pr[Y = y|X = x]$ verifies

$$\mathbf{T} = \mathbf{T}_1 \times \mathbf{T}_2.$$

Proof. Successively using the law of total probability, the fact that the random variables are uniformly distributed, and the Markovian property of $X \leftrightarrow W \leftrightarrow Y$, we have

$$\begin{aligned}[\mathbf{T}]_{x,y} &= \sum_{w\in\mathcal{W}} \Pr[Y=y|X=x,W=w]\Pr[W=w|X=x] \\ &= \sum_{w\in\mathcal{W}} \Pr[Y=y|W=w]\Pr[W=w|X=x] \\ &= \sum_{w\in\mathcal{W}} [\mathbf{T}_1]_{x,w}[\mathbf{T}_2]_{w,y}.\end{aligned}$$

□

In what follows, we consider the idealistic situation where the random variables X, Y, and W representing the reduced samples form a Markov chain as in Proposition 8.2. Clearly, when the keys k_1 and k_2 are fixed in the situation described on Figure 8.2, then $\mathsf{P}_1 \leftrightarrow \mathsf{P}_2 \leftrightarrow \mathsf{P}_3$ is a Markov chain. Consequently, if ρ, μ, and ϕ are the identity (and thus, do not reduce the sample space at all), then this is also the case for $X \leftrightarrow W \leftrightarrow Y$. Yet, the Markovian property of X, Y, and Y is not guaranteed as soon as the projections reduce the sample space.

Lemma 8.3 *Under the notations and assumptions of Proposition 8.2, the bias matrices* $\mathbf{B}_1$*,* $\mathbf{B}_2$*, and* $\mathbf{B}$ *respectively corresponding to the transition matrices* $\mathbf{T}_1$*,* $\mathbf{T}_2$*, and* $\mathbf{T}$ *verify*

$$\mathbf{B} = \mathbf{B}_1 \times \mathbf{B}_2 \quad \textit{and} \quad \widehat{\mathbf{B}} = \frac{1}{|\mathcal{W}|}\widehat{\mathbf{B}_1} \times \widehat{\mathbf{B}_2}.$$

Therefore

$$\|\mathbf{B}\|_2 \le \|\mathbf{B}_1\|_2\|\mathbf{B}_2\|_2 \tag{8.6}$$

with equality if, and only if we can write $[\mathbf{B}_1]_{x,w} = \alpha_x\gamma_w$ *and* $[\mathbf{B}_2]_{w,y} = \gamma_w\beta_y$ *for some* $\alpha \in \mathbf{R}^{|\mathcal{X}|}$*,* $\beta \in \mathbf{R}^{|\mathcal{Y}|}$*, and* $\gamma \in \mathbf{R}^{|\mathcal{W}|}$*.*

Proof. As $\mathbf{T} = \mathbf{T}_1 \times \mathbf{T}_2$, we have

$$[\mathbf{B}]_{x,y} = [\mathbf{T}]_{x,y} - \frac{1}{|\mathcal{Y}|} = \sum_{w\in\mathcal{W}} \left([\mathbf{B}_1]_{x,w} + \frac{1}{|\mathcal{W}|}\right)\left([\mathbf{B}_2]_{w,y} + \frac{1}{|\mathcal{Y}|}\right) - \frac{1}{|\mathcal{Y}|}.$$

As $\sum_w [\mathbf{B}_1]_{x,w} = 0$, we obtain $[\mathbf{B}]_{x,y} = [\mathbf{B}_1 \times \mathbf{B}_2]_{x,y} + \frac{1}{|\mathcal{W}|}\sum_w [\mathbf{B}_2]_{w,y}$. Since W and Y are uniformly distributed $\sum_{w\in\mathcal{W}} [\mathbf{B}_2]_{w,y} = 0$, which proves that $\mathbf{B} = \mathbf{B}_1 \times \mathbf{B}_2$. We also have

$$\begin{aligned}
\left[\widehat{\mathbf{B}_1} \times \widehat{\mathbf{B}_2}\right]_{u,v} &= \sum_{a\in\mathcal{W}} \left[\widehat{\mathbf{B}_1}\right]_{u,a}\left[\widehat{\mathbf{B}_2}\right]_{a,v} \\
&= \sum_{\substack{(x,w)\in\mathcal{X}\times\mathcal{W}\\(w',y)\in\mathcal{W}\times\mathcal{Y}}} (-1)^{u\bullet x\oplus v\bullet y} \left[\mathbf{B}_1\right]_{x,w}\left[\mathbf{B}_2\right]_{w',y} \sum_{a\in\mathcal{W}} (-1)^{a\bullet(w\oplus w')} \\
&= |\mathcal{W}| \sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} (-1)^{u\bullet x\oplus v\bullet y} \sum_{w\in\mathcal{W}} \left[\mathbf{B}_1\right]_{x,w}\left[\mathbf{B}_2\right]_{w,y} \\
&= |\mathcal{W}| \sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} (-1)^{u\bullet x\oplus v\bullet y} \left[\mathbf{B}\right]_{x,y} \\
&= |\mathcal{W}| \left[\widehat{\mathbf{B}}\right]_{u,v},
\end{aligned}$$

which proves that $\widehat{\mathbf{B}} = \frac{1}{|\mathcal{W}|}\widehat{\mathbf{B}_1} \times \widehat{\mathbf{B}_2}$. Finally, from the Cauchy-Schwarz inequality:

$$\begin{aligned}
\|\mathbf{B}_1 \times \mathbf{B}_2\|_2^2 &= \sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} \left(\sum_{w\in\mathcal{W}} \left[\mathbf{B}_1\right]_{x,w}\left[\mathbf{B}_2\right]_{w,y}\right)^2 \\
&\leq \sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} \left(\sum_{w\in\mathcal{W}} \left[\mathbf{B}_1\right]_{x,w}^2\right)\left(\sum_{w'\in\mathcal{W}} \left[\mathbf{B}_2\right]_{w',y}^2\right) \\
&= \|\mathbf{B}_1\|_2^2\|\mathbf{B}_2\|_2^2,
\end{aligned}$$

with equality if, and only if, for all $(x,y) \in \mathcal{X} \times \mathcal{Y}$ there exists some $\lambda_{x,y}$ such that $\left[\mathbf{B}_1\right]_{x,w} = \lambda_{x,y}\left[\mathbf{B}_2\right]_{w,y}$, so that $\left[\mathbf{B}_1\right]_{x,w} = \lambda_{x,0}\left[\mathbf{B}_2\right]_{w,0} = \alpha_x\gamma_w$. Taking β_y equal to $\alpha_0/\lambda_{0,y}$ when $\lambda_{0,y} \neq 0$ and to zero otherwise leads to the announced result. □

From Lemma 8.3 we can deduce the original version of the piling-up lemma.

Corollary 8.1 *Under the notations and assumptions of Lemma 8.3 and assuming that $\mathcal{X} = \mathcal{Y} = \mathcal{W} = \{0,1\}$, there exists $\epsilon_1, \epsilon_2 \in \mathbf{R}$ such that*

$$\mathbf{B}_1 = \begin{bmatrix} \epsilon_1 & -\epsilon_1 \\ -\epsilon_1 & \epsilon_1 \end{bmatrix} \quad \textit{and} \quad \mathbf{B}_2 = \begin{bmatrix} \epsilon_2 & -\epsilon_2 \\ -\epsilon_2 & \epsilon_2 \end{bmatrix}$$

so that (8.6) *is actually an equality:*

$$\|\mathbf{B}\|_2 = \|\mathbf{B}_1\|_2\|\mathbf{B}_2\|_2.$$

How to find the projections ρ, μ, and ϕ on larger spaces exhibiting such a Markovian property for a given block cipher in the most general case remains however an open question to us. Yet, there are practical cases where this property holds, namely when considering Markov ciphers (see Definition 8.8) and well chosen homomorphic projections. This is in particular the case of the characters (over Abelian groups) that we use in Section 8.4 to generalize linear cryptanalysis. If the projections satisfy the Markovian property, we obtain under the natural notations implied by Figure 8.2 that

$$\|\mathbf{B}_k^{\rho,\phi}\|_2^2 \leq \|\mathbf{B}_{k_1}^{\rho,\mu}\|_2^2 \|\mathbf{B}_{k_2}^{\mu,\phi}\|_2^2$$

which shows that one needs at least

$$q = \frac{8 \ln 2}{\|\mathbf{B}_{k_1}^{\rho,\mu}\|_2^2 \|\mathbf{B}_{k_2}^{\mu,\phi}\|_2^2}$$

samples to distinguish C from the perfect cipher. In the case of our generalization of linear cryptanalysis, we will obtain a more precise approximate of q.

8.4 Generalized Linear Cryptanalysis of Block Ciphers

We will now consider a particular type of projection-based distinguishers, namely, generalized linear distinguishers. These distinguishers reduce the sample space by means of *linear* projections, where the linearity relates to the group law defined on the text space. In a classical linear cryptanalysis of a block cipher C, the text space is $\mathcal{T} = \{0,1\}^n$ (for some positive integer n) and the group law is the exclusive-or operation. In that case, the adversary typically runs over q plaintext/ciphertext pairs $(P_i, \mathcal{O}(P_i))$, for $i = 1, 2, \ldots, q$, and add the value of $a \bullet P_i \oplus b \bullet \mathcal{O}(P_i)$ to a counter, where a and b are input/output masks defined on the text space $\mathcal{T}$. The adversary eventually guesses whether the generator is implementing an instance of the block cipher or not by measuring the bias of the counter with respect to $q/2$ (which is the expected value of the counter in the case where $\mathcal{O}$ is the perfect cipher). By choosing the masks with care, the bias may be large when $\mathcal{O} = \mathsf{C}_k$ for some key k. In this situation, the linear probability

$$\mathrm{LP}_{a,b}(\mathsf{C}_k) = (2 \cdot \Pr_P(a \bullet P \oplus b \bullet \mathsf{C}_k(P) = 0) - 1)^2 = \left| \mathrm{E}((-1)^{a \bullet P \oplus b \bullet \mathsf{C}_k(P)}) \right|^2$$

estimates the efficiency of the attack against C_k. We will now see how to generalize this metric to arbitrary group structures.

As in Section 8.1, we consider the plaintext/ciphertext pairs as random variables in $\mathcal{T}^2$. Namely, we let $L = (P, \mathcal{O}(P))$ where P is uniformly distributed and denote by $\widetilde{\mathsf{P}}$ its distribution. This turns the distinguish problem between random oracles in a distinguishing problem between random sources. We let G_1 and G_2 denote two group structures on $\mathcal{T}$, let $\mathsf{G} = \mathsf{G}_1 \times \mathsf{G}_2$ be the group product, and consider the characters $\rho \in \widehat{\mathsf{G}}_1$ and $\mu \in \widehat{\mathsf{G}}_2$. Finally, we let $\chi \in \widehat{\mathsf{G}}$ be the character defined by $\chi(a,b) = \overline{\rho}(a)\mu(b)$. Following Definition 7.7 and the discussion of Section 7.6, we know that our generalized version of linear cryptanalysis on random sources is based on

$$\mathrm{LP}_\chi(\widetilde{\mathsf{P}}) = \mathrm{LP}(\chi(P, \mathcal{O}(P))) = \mathrm{LP}\left(\overline{\rho}(P)\mu(\mathcal{O}(P))\right) = \left|\mathrm{E}\left(\overline{\rho}(P)\mu(\mathcal{O}(P))\right)\right|^2,$$

where $P \in \mathcal{T}$ is uniformly distributed. It seems therefore natural to extend the classical linear probability for block ciphers as follows.

Definition 8.6 *Let $\mathcal{T}$ and $\mathcal{K}$ be two finite sets and C be a block cipher over the text space $\mathcal{T}$ and the key space $\mathcal{K}$. Let G_1 and G_2 be two group structures over the same set $\mathcal{T}$. For all group characters $\rho \in \widehat{\mathsf{G}}_1$ and $\mu \in \widehat{\mathsf{G}}_2$, the linear probability of C over $\mathcal{T}$ with respect to ρ and μ is defined by*

$$\mathrm{LP}_{\rho,\mu}(\mathsf{C}_k) = \left|\mathrm{E}_{P \in_{\mathrm{U}} \mathcal{T}}\Big(\overline{\rho}(P)\mu(\mathsf{C}_k(P))\Big)\right|^2 .$$

We denote the expected *linear probability over all keys by*

$$\mathrm{ELP}_{\rho,\mu}(\mathsf{C}) = \mathrm{E}_K(\mathrm{LP}_{\rho,\mu}(\mathsf{C}_K)) = \frac{1}{|\mathcal{K}|}\sum_{k\in\mathcal{K}} \mathrm{LP}_{\rho,\mu}(\mathsf{C}_k).$$

We note that in the particular case where $\mathcal{T} = \{0,1\}^n$ and the group law considered is the exclusive-or operation, then there always exists a and b in $\mathcal{T}$ such that $\rho(x) = (-1)^{a\bullet x}$ and $\mu(x) = (-1)^{b\bullet x}$. In that case, the previous definition rewrites

$$\mathrm{LP}_{\rho,\mu}(\mathsf{C}_k) = \left|\mathrm{E}_{P\in_{\mathrm{U}}\mathcal{T}}\Big((-1)^{a\bullet P}(-1)^{b\bullet \mathsf{C}_k(P)}\Big)\right|^2 = \mathrm{LP}_{a,b}(\mathsf{C}_k).$$

If $\widetilde{\mathsf{P}}_k$ denotes the distribution of $(P, \mathsf{C}_k(P))$ for a uniformly distributed P, we have by construction

$$\mathrm{LP}_{\chi}(\widetilde{\mathsf{P}}_k) = \left|\mathrm{E}\left(\overline{\rho}(P)\mu(\mathsf{C}_k(P))\right)\right|^2 = \mathrm{LP}_{\rho,\mu}(\mathsf{C}_k).$$

Therefore, we can deduce from Lemma 7.5 that an optimal linear distinguisher between $\mathsf{H}_0 : \mathcal{O} \leftarrow \mathsf{C}^\star$ and $\mathsf{H}_1 : \mathcal{O} \leftarrow \mathsf{C}$ based on the characters ρ and μ should accept H_1 whenever

$$\max_h \mathrm{N}[h|\mathbf{H}^q] \geq \frac{q\log(1-\epsilon)}{\log(1-\epsilon)-\log(1+(d-1)\epsilon)}, \tag{8.7}$$

where $H_i = \overline{\rho}(P_i)\mu(\mathcal{O}(P_i))$ and $\epsilon^2 = \mathrm{LP}_{\rho,\mu}(\mathsf{C}_k)$. The optimal linear distinguisher should thus accept H_1 when the maximum value of the relative frequency of the plaintext/ciphertext pairs exceeds a threshold which is completely determined by the linear probability of C_k with respect to chosen characters. Since the key k is unknown, a distinguisher will in practice choose ϵ such that

$$\epsilon^2 = \mathrm{ELP}_{\rho,\mu}(\mathsf{C}).$$

Note that assuming the hypothesis of stochastic equivalence (see Definition 8.3) also lead to this strategy since in that case one assumes that

$$\mathrm{LP}_{\rho,\mu}(\mathsf{C}_k) \approx \mathrm{LP}_{\rho,\mu}(\mathsf{C}_{k'})$$

for all $k \neq k'$, and thus $\mathrm{LP}_{\rho,\mu}(\mathsf{C}_k) \approx \mathrm{ELP}_{\rho,\mu}(\mathsf{C})$ for all $k \in \mathcal{K}$.

We now deduce from Heuristic 7.2 that the distinguisher based on (8.7) reaches a non-negligible advantage when

$$q = \frac{8 \ln 2}{(d-1)\mathrm{ELP}_{\rho,\mu}(\mathsf{C})}$$

where d is the order of the character $\chi = (\rho, \mu)$, which is the least common multiple of the order of ρ and of the order of μ. We deduce the following heuristic on which we will rely to compute the attack complexities against TOY100 and against SAFER.

Heuristic 8.2 Let $\mathcal{T}$ and $\mathcal{K}$ be two finite sets and C be a block cipher over the text space $\mathcal{T}$ and the key space $\mathcal{K}$. Let $\mathsf{C}^\star$ be the perfect cipher over $\mathcal{T}$. Let G_1 and G_2 be two group structures over the same set $\mathcal{T}$. Let $\rho \in \widehat{\mathsf{G}}_1$ and $\mu \in \widehat{\mathsf{G}}_2$ be two characters of order d_1 and d_2 respectively. Letting $\epsilon^2 = \mathrm{ELP}_{\rho,\mu}(\mathsf{C})$, the best q-limited distinguisher between $\mathsf{H}_0 : \mathcal{O} \leftarrow \mathsf{C}^\star$ and $\mathsf{H}_1 : \mathcal{O} \leftarrow \mathsf{C}$ based on the characters ρ and μ should accept H_1 if, and only if

$$\max_h \mathrm{N}[h|\mathbf{H}^q] \geq \frac{q \log(1-\epsilon)}{\log(1-\epsilon) - \log(1+(d-1)\epsilon)}$$

is verified, where $H_i = \overline{\rho}(P_i)\mu(\mathcal{O}(P_i))$ for each sample P_i, with $i = 1, \ldots, q$, and where d is the least common multiple of the order of ρ and of the order of μ. Assuming that $\mathrm{ELP}_{\rho,\mu}(\mathsf{C}) \ll 1$, the distinguisher reaches a non-negligible advantage when

$$q = \frac{8 \ln 2}{(d-1)\mathrm{ELP}_{\rho,\mu}(\mathsf{C})}.$$

□

Fourier Transforms and Links with Differential Cryptanalysis

Differential cryptanalysis over arbitrary groups was formalized by Lai, Massey, and Murphy in [97]. The complexity of this attack relates to the differential probability.

Definition 8.7 *Let $\mathcal{T}$ and $\mathcal{K}$ be two finite sets and C be a block cipher over the text space $\mathcal{T}$ and the key space $\mathcal{K}$. Let $\boxplus$ and $\otimes$ be two group laws on $\mathcal{T}$ and let $a, b \in \mathcal{T}$ and $k \in \mathcal{K}$. The differential probability of C_k with respect to the input difference a and output difference b is*

$$\mathrm{DP}_{a,b}(\mathsf{C}_k) = \Pr[\mathsf{C}_k(P \boxplus a) = \mathsf{C}_k(P) \otimes b],$$

where the probability holds over the uniform distribution of the plaintext $P \in \mathcal{T}$. The expected differential probability of C with respect to the same masks is

$$\mathrm{EDP}_{a,b}(\mathsf{C}) = \mathrm{E}_K\left(\mathrm{DP}_{a,b}(\mathsf{C}_K)\right) = \frac{1}{|\mathcal{K}|}\sum_{k \in \mathcal{K}} \mathrm{DP}_{a,b}(\mathsf{C}_k).$$

By construction, we know that the linear probability verifies

$$\mathrm{LP}_{\rho,\mu}(\mathsf{C}_k) = \mathrm{LP}_{\chi}(\widetilde{\mathsf{P}}_k), \tag{8.8}$$

where $\widetilde{\mathsf{P}}_k$ denotes the distribution of $(P, \mathsf{C}_k(P))$ and $\chi(a,b) = \overline{\rho}(a)\mu(b)$. Consequently, the inverse Fourier transform of $\mathrm{LP}_{\rho,\mu}(\mathsf{C}_k)$ at point $(a,b) \in \mathsf{G}_1 \times \mathsf{G}_2$ is equal to $\widehat{\mathrm{LP}}_{(a,b)}(\widetilde{\mathsf{P}}_k)$, which leads to

$$\widehat{\mathrm{LP}}_{a,b}(\mathsf{C}_k) = \frac{1}{|\mathcal{T}|^2} \sum_{(\rho,\mu)\in\widehat{\mathsf{G}}_1\times\widehat{\mathsf{G}}_2} \mathrm{LP}_{\rho,\mu}(\mathsf{C}_k)\overline{\rho}(a)\mu(b) \tag{8.9}$$

We have a similar property for the differential probability since, with the same notations, we have

$$\begin{aligned}
\mathrm{DP}_{(a,b)}(\widetilde{\mathsf{P}}_k) &= \Pr_{P,P'}[(P', \mathsf{C}_k(P')) = (P, \mathsf{C}_k(P)) \cdot (a,b)] \\
&= \Pr_{P,P'}[P' = P \boxplus a, \mathsf{C}_k(P') = \mathsf{C}_k(P) \otimes b] \\
&= \Pr_{P,P'}[P' = P \boxplus a, \mathsf{C}_k(P \boxplus a) = \mathsf{C}_k(P) \otimes b] \\
&= \frac{1}{|\mathcal{T}|}\Pr_P[\mathsf{C}_k(P \boxplus a) = \mathsf{C}_k(P) \otimes b]
\end{aligned}$$

and thus

$$\mathrm{DP}_{(a,b)}(\widetilde{\mathsf{P}}_k) = \frac{1}{|\mathcal{T}|}\mathrm{DP}_{a,b}(\mathsf{C}_k). \tag{8.10}$$

The Fourier transform of $\mathrm{DP}_{a,b}(\mathsf{C}_k)$ at point $\chi \in \widehat{\mathsf{G}}$ (where $\mathsf{G} = \mathsf{G}_1 \times \mathsf{G}_2$) is equal to $|\mathcal{T}|\,\widehat{\mathrm{DP}}_{\chi}(\widetilde{\mathsf{P}}_k)$ which leads to

$$\widehat{\mathrm{DP}}_{\rho,\mu}(\mathsf{C}_k) = \sum_{(a,b)\in\mathsf{G}_1\times\mathsf{G}_2} \mathrm{DP}_{a,b}(\mathsf{C}_k)\rho(a)\overline{\mu}(b). \tag{8.11}$$

Based on Lemma 7.11 which shows the link between linear and differential probabilities for random sources, we can reformulates this link between the corresponding metrics for linear and differential distinguishers on block ciphers (as in [32]).

Lemma 8.4 *Let $\mathcal{T}$ and $\mathcal{K}$ be two finite sets and C be a block cipher over the text space $\mathcal{T}$ and key space $\mathcal{K}$. Let G_1 and G_2 be two group structures over $\mathcal{T}$ and $\chi \in \widehat{\mathsf{G}}_1$ and $\rho \in \widehat{\mathsf{G}}_2$. Let $(\rho,\mu) \in \widehat{\mathsf{G}}_1 \times \widehat{\mathsf{G}}_2$ and $(a,b) \in \mathsf{G}_1 \times \mathsf{G}_2$. For all $k \in \mathcal{K}$, the inverse Fourier transform of $\mathrm{LP}_{\rho,\mu}(\mathsf{C}_k)$ at point (a,b) is*

$$\widehat{\mathrm{LP}}_{a,b}(\mathsf{C}_k) = \frac{1}{|\mathcal{T}|}\mathrm{DP}_{a,b}(\mathsf{C}_k) \quad \textit{and thus} \quad \widehat{\mathrm{ELP}}_{a,b}(\mathsf{C}) = \frac{1}{|\mathcal{T}|}\mathrm{EDP}_{a,b}(\mathsf{C}). \tag{8.12}$$

Conversely, the Fourier transform of $\mathrm{DP}_{a,b}(\mathsf{C}_k)$ at the point (ρ,μ) is

$$\widehat{\mathrm{DP}}_{\rho,\mu}(\mathsf{C}_k) = |\mathcal{T}|\,\mathrm{LP}_{\rho,\mu}(\mathsf{C}_k) \quad \textit{and thus} \quad \widehat{\mathrm{EDP}}_{\rho,\mu}(\mathsf{C}) = |\mathcal{T}|\,\mathrm{ELP}_{\rho,\mu}(\mathsf{C}). \tag{8.13}$$

Proof. Denoting $\widetilde{\mathsf{P}}_k$ the distribution of $(P, \mathsf{C}_k(P))$ where $P \in \mathcal{T}$ is uniformly distributed, it follows from Lemma 7.11 and (8.10) that

$$\widehat{\mathrm{DP}}_{\rho,\mu}(\mathsf{C}_k) = |\mathcal{T}|\, \widehat{\mathrm{DP}}_{\chi}(\widetilde{\mathsf{P}}) = |\mathcal{T}|\, \mathrm{LP}_{\chi}(\widetilde{\mathsf{P}}) = |\mathcal{T}|\, \mathrm{LP}_{\rho,\mu}(\mathsf{C}_k).$$

The equality for the mean follows from the fact that

$$\begin{aligned}
\mathrm{E}_K\left(\widehat{\mathrm{DP}}_{\rho,\mu}(\mathsf{C}_K)\right) &= \frac{1}{|\mathcal{K}|}\sum_{k\in\mathcal{K}} \widehat{\mathrm{DP}}_{\rho,\mu}(\mathsf{C}_k)\\
&= \frac{1}{|\mathcal{K}|}\sum_{k\in\mathcal{K}}\sum_{a,b} \mathrm{DP}_{a,b}(\mathsf{C}_k)\rho(a)\overline{\mu}(b)\\
&= \sum_{a,b} \mathrm{EDP}_{a,b}(\mathsf{C})\rho(a)\overline{\mu}(b)\\
&= \widehat{\mathrm{EDP}}_{\rho,\mu}(\mathsf{C}).
\end{aligned}$$

Substituting $\widehat{\mathrm{DP}}_{\rho,\mu}(\mathsf{C}_K)$ by $|\mathcal{T}|\, \mathrm{LP}_{\rho,\mu}(\mathsf{C}_K)$ in the left-hand side of the previous equation leads to (8.13).

Conversely, based on (8.8) and on the same lemma, we similarly obtain that

$$\widehat{\mathrm{LP}}_{a,b}(\mathsf{C}_k) = \widehat{\mathrm{LP}}_{a,b}(\widetilde{\mathsf{P}}) = \mathrm{DP}_{(a,b)}(\widetilde{\mathsf{P}}) = \frac{1}{|\mathcal{T}|}\mathrm{DP}_{a,b}(\mathsf{C}_k).$$

Proving the rest of (8.12) can be done in a similar way than what we did for the mean of the differential probability. □

In the case the block cipher considered is a Markov cipher [97] (which is the case of almost any iterated block cipher when the round keys are mutually independent), then it is easy to relate the expected differential probability on the block cipher to the expected differential probability over the individual rounds [97].

Definition 8.8 *Let $\mathcal{T}$ and $\mathcal{K}$ be two finite sets. Let $\boxplus$ and $\otimes$ be two group laws on $\mathcal{T}$. A block cipher C over the text space $\mathcal{T}$ and the key space $\mathcal{K}$ is a Markov cipher if for any $a, b, p \in \mathcal{T}$ we have*

$$\mathrm{EDP}_{a,b}(\mathsf{C}) = \mathrm{Pr}_K[\mathsf{C}_K(p \boxplus a) = \mathsf{C}_K(p) \otimes b].$$

where the probability holds over the uniformly distributed key $K \in \mathcal{K}$.

Theorem 8.1 *Let $\mathcal{T}$ and $\mathcal{K}$ be two finite sets and $\mathsf{C}^{(1)}, \mathsf{C}^{(2)}, \ldots, \mathsf{C}^{(r)}$ be $r > 1$ mutually independent Markov ciphers on the text space $\mathcal{T}$ and the key space $\mathcal{K}$. For any input input/output differences $w_0, w_r \in \mathcal{T}$ we have*

$$\mathrm{EDP}_{w_0,w_r}\left(\mathsf{C}^{(r)} \circ \cdots \circ \mathsf{C}^{(2)} \circ \mathsf{C}^{(1)}\right) = \sum_{w_1,\ldots,w_{r-1}} \prod_{i=1}^{r} \mathrm{EDP}_{w_{i-1},w_i}\left(\mathsf{C}^{(i)}\right).$$

The link between the linear and differential probabilities given in Lemma 8.4 allows to deduce an iterative formula for linear probabilities over iterated cipher similar to the well known expression given by Theorem 8.1 for the differential probabilities. This corresponds to what is known as Nyberg's linear hull effect [125].

Theorem 8.2 *Let $\mathcal{T}$ and $\mathcal{K}$ be two finite sets. Let $\mathsf{G}_0 \dots, \mathsf{G}_r$ be $r+1$ Abelian group structures over $\mathcal{T}$ and let $\mathsf{C} = \mathsf{C}^{(r)} \circ \cdots \circ \mathsf{C}^{(1)}$ be a product cipher of r independent Markov ciphers $\mathsf{C}^{(i)} : \mathsf{G}_{i-1} \longrightarrow \mathsf{G}_i$. For any $\chi_0 \in \widehat{\mathsf{G}}_0$ and $\chi_r \in \widehat{\mathsf{G}}_r$ we have*

$$\mathrm{ELP}_{\chi_0,\chi_r}(\mathsf{C}) = \sum_{\chi_1 \in \widehat{\mathsf{G}}_1} \cdots \sum_{\chi_{r-1} \in \widehat{\mathsf{G}}_{r-1}} \prod_{i=1}^{r} \mathrm{ELP}_{\chi_{i-1},\chi_i}\left(\mathsf{C}^{(i)}\right).$$

Proof. We first prove the theorem when $r=2$. From Lemma 8.4 we have

$$\mathrm{ELP}_{\chi_0,\chi_2}(\mathsf{C}) = \frac{1}{|\mathcal{T}|}\widehat{\mathrm{EDP}}_{\chi_0,\chi_2}(\mathsf{C}).$$

From (8.11), the right-hand side of the previous equation is equal to

$$\frac{1}{|\mathcal{T}|}\sum_{w_0,w_2} \mathrm{EDP}_{w_0,w_2}(\mathsf{C})\chi_0(w_0)\overline{\chi_2}(w_2).$$

Since we consider independent Markov ciphers, Theorem 8.1 shows that the last equation is equal to

$$\frac{1}{|\mathcal{T}|}\sum_{w_0,w_1.w_2} \mathrm{EDP}_{w_0,w_1}(\mathsf{C}^{(1)})\mathrm{EDP}_{w_1,w_2}(\mathsf{C}^{(2)})\chi_0(w_0)\overline{\chi_2}(w_2).$$

Applying Lemma 8.4 shows that this expression is equal to

$$|\mathcal{T}|\sum_{w_0,w_1.w_2} \widehat{\mathrm{ELP}}_{w_0,w_1}(\mathsf{C}^{(1)})\widehat{\mathrm{ELP}}_{w_1,w_2}(\mathsf{C}^{(2)})\chi_0(w_0)\overline{\chi_2}(w_2). \tag{8.14}$$

According to (8.9) we have by definition:

$$\begin{aligned}
\widehat{\mathrm{ELP}}_{w_0,w_1}(\mathsf{C}^{(1)}) &= \frac{1}{|\mathcal{T}|^2}\sum_{(\rho_0,\chi_1)\in\widehat{\mathsf{G}}_0\times\widehat{\mathsf{G}}_1} \mathrm{ELP}_{\rho_0,\chi_1}(\mathsf{C}^{(1)})\overline{\rho_0}(w_0)\chi_1(w_1)\\
\widehat{\mathrm{ELP}}_{w_1,w_2}(\mathsf{C}^{(2)}) &= \frac{1}{|\mathcal{T}|^2}\sum_{(\rho_1,\mu_2)\in\widehat{\mathsf{G}}_1\times\widehat{\mathsf{G}}_2} \mathrm{ELP}_{\rho_1,\mu_2}(\mathsf{C}^{(2)})\overline{\rho_1}(w_1)\mu_2(w_2)
\end{aligned}$$

Plugging these values in (8.14) and reducing the result using the orthogonal relations given in Lemma 7.4 leads to

$$\mathrm{ELP}_{\chi_0,\chi_2}(\mathsf{C}) = \sum_{\chi_1\in\widehat{\mathsf{G}}_1} \mathrm{ELP}_{\chi_0,\chi_1}(\mathsf{C}^{(1)})\mathrm{ELP}_{\chi_1,\chi_2}(\mathsf{C}^{(2)}),$$

which exactly corresponds to the linear hull formula for $r = 2$. To obtain the result for any $r > 2$, it suffices to apply the previous formula $r - 1$ times in a recursive way on $\mathsf{C} = \mathsf{C}^{(r)} \circ \cdots \circ \mathsf{C}^{(2)} \circ \mathsf{C}^{(1)}$. □

As direct computation of the expected linear probability on a realistic instance of a block cipher is not practical, the cryptanalyst typically follows a bottom-up approach, in which she first computes the linear probability of small building blocks of the cipher and then extends the result to the whole construction. In the remaining of this section, we study several typical building blocks on which block ciphers are often based.

A Toolbox for the Generalized Linear Cryptanalysis

We can look at a block cipher as a circuit made of building blocks and in which every edge is attached to a specific group. From this point of view, a linear characteristic is a family mapping every edge to one character of the attached group. The building blocks we will consider are represented on Figure 8.3. If χ_1 and χ_2 are characters on G_1 and G_2 respectively, we denote by $\chi_1 \| \chi_2 : \mathrm{G}_1 \times \mathrm{G}_2 \to \mathbf{C}^{\times}$ the character mapping $(a, b) \in \mathrm{G}_1 \times \mathrm{G}_2$ on $\chi_1(a)\chi_2(b)$. We assume that the cryptanalyst constructs a linear characteristic in a reversed way [16] (i.e., starting from the end of the block cipher towards the beginning), her objective being to carefully choose the characters in order to maximize the linear probability on each individual building block.

Building Block (a): We consider a *duplicate gate* such that $a, b, c \in \mathrm{G}$ and $a = b = c$. Let χ_1, χ_2 be two characters defined over G so that $\chi_1 \| \chi_2$ is a character on the output of the gate, we have (by definition) $\chi_1(b)\chi_2(c) = \chi_1(a)\chi_2(a) = \chi_1\chi_2(a)$. Simply denoting (a) the duplicate gate, we have

$$\mathrm{LP}_{\chi_1\chi_2, \chi_1\|\chi_2}((\mathrm{a})) = 1,$$

so that $\chi_1\chi_2$ is an appropriate character on the input of the gate.

Building Block (b): We consider a layer that applies a *group homomorphism* from $\mathsf{G} = \mathsf{G}_1 \times \cdots \times \mathsf{G}_m$ to $\mathsf{H} = \mathsf{H}_1 \times \cdots \times \mathsf{H}_n$. We denote the homomorphism by hom, the m inputs as $a_1, a_2, \ldots, a_m$ and the n outputs $b_1, b_2, \ldots, b_n$, so that $\mathrm{hom}(a_1, a_2, \ldots, a_m) = (b_1, b_2, \ldots, b_n)$. Given n characters χ_i on H_i, $i = 1, \ldots, n$, we have $\chi(b_1, \ldots, b_n) =$

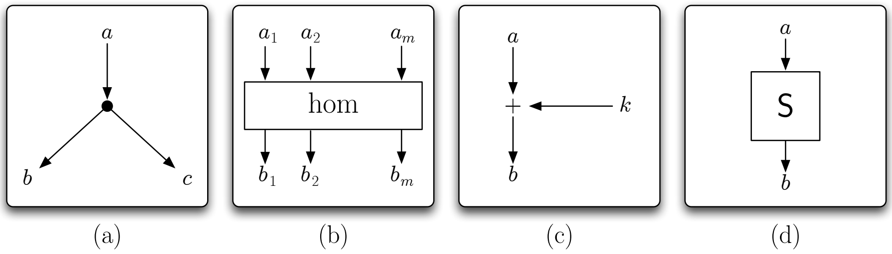


Figure 8.3: Typical Building Blocks of Block Ciphers

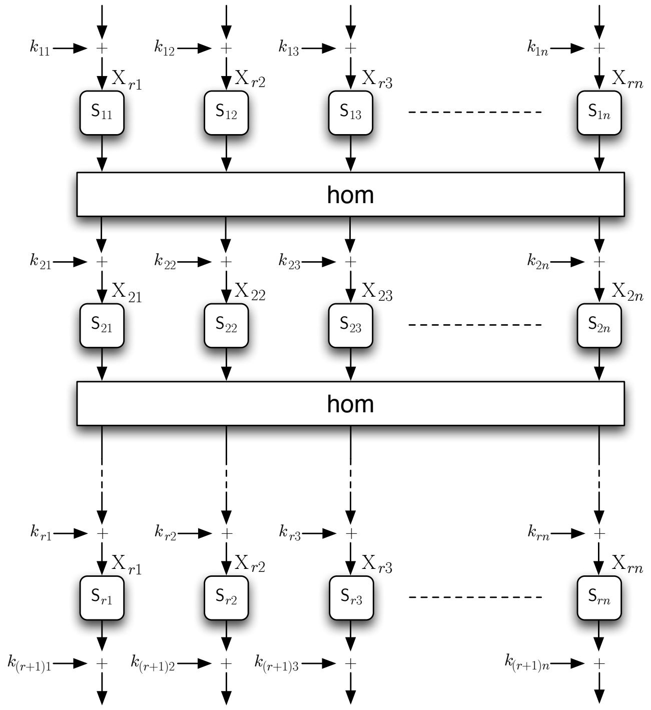


Figure 8.4: A Typical Substitution-Permutation Network with r Rounds

$(\chi \circ \mathrm{hom})(a_1, \ldots, a_m)$ for $\chi = \chi_1 \| \cdots \| \chi_n$. As $\chi \circ \mathrm{hom}$ is still a character on G we obtain $\mathrm{LP}_{\chi\circ\mathrm{hom},\chi}((\mathrm{b})) = 1$. Note that we do have $\chi \circ \mathrm{hom} = \chi'_1 \| \cdots \| \chi'_m$ for some $(\chi'_1, \ldots, \chi'_m) \in \widehat{\mathsf{G}}_1 \times \cdots \times \widehat{\mathsf{G}}_m$, so that χ'_i is an appropriate character for a_i.

Building Block (c): Given $\mathrm{hom}(a) = a + k$ on a given group G (with additive notation), we have $\chi(b) = \chi(a)\chi(k)$. Since k is constant, $\mathrm{LP}_{\chi,\chi}((\mathrm{c})) = 1$, so that χ is an appropriate character on the input.

Building Block (d): When considering a (non-homomorphic) permutation S, the linear probability $\mathrm{LP}_{\rho,\mu}(\mathsf{S})$ should be computed by considering the substitution table of S.

Consider for example a typical substitution-permutation network C as shown on Figure 8.4. By piling all relations up on a typical substitution-permutation network C, we obtain a relation of the form

$$\overline{\chi}(P)\rho(\mathsf{C}(P)) = \left(\prod_{i,j} \overline{\chi_{i,j}}(X_{i,j})\rho_{i,j}(\mathsf{S}_{i,j}(X_{i,j}))\right) \times \left(\prod_{i,j} \chi_{i,j}(k_{i,j})\right)$$

where the first product runs over all building blocks of type (d) and the second over building blocks of type (c). Hence, by making the heuristic approximation of independence of all X_i's (which is commonly done in classical linear cryptanalysis), we obtain that

$$\mathrm{ELP}_{\chi,\rho}(\mathsf{C}) \approx \prod_{i,j} \mathrm{LP}_{\chi_{i,j},\rho_{i,j}}(\mathsf{S}_{i,j}).$$

This is the classical single-path linear characteristic. Provided that we can lower bound (e.g., by exploiting the properties of the homeomorphism) the number of active substitution boxes (i.e., boxes with non-trivial input/output masks) by b and that we have

$$\max_{i,j} \mathrm{LP}_{\max}(\mathsf{S}_{i,j}) \leq \lambda$$

we obtain that $\mathrm{ELP}_{\max}(\mathsf{C})$ is heuristically bounded by λ^b for single-path characteristics. We can obtain an exact formula for $\mathrm{ELP}_{\chi,\rho}(\mathsf{C})$ by using linear hulls and assuming that all the subkeys are mutually independent, since we are indeed considering a Markov cipher here, so that Theorem 8.2 applies.

In the next section, we show how to use the tools presented so far to build and study the security of a block cipher that encrypts blocks of decimal digits.

8.5 The Block Cipher DEAN: a Toy Example for our Generalization of Linear Cryptanalysis

We introduce DEAN18 (as for Digital Encryption Algorithm for Numbers), a toy cipher that encrypts blocks of 18 decimal digits (which approximately corresponds to a block size of 60 bits), which could be used to encrypt a credit-card number for example. The structure of this toy cipher is inspired from that of the AES [41]. We consider an R-round substitution-permutation network, each round being based on the same structure. Blocks are represented as 3×3 arrays of elements of the additive group $\mathbf{Z}_{10} \times \mathbf{Z}_{10}$. Each round successively applies to the plaintext the following operations:

- `AddRoundKeys`, that performs a digit-wise addition of a round key to the input (the addition being taken modulo 10),
- `SubBytes`, that applies a fixed bijective substitution box S (defined in Table 8.1, where an element $(a, b) \in \mathbf{Z}_{10}^2$ is represented as an integer $10 \cdot a + b \in \{0, 1, \ldots, 99\}$) on each 2-digit element of the array,
- `ShiftRows`, that shifts to the left each row of the input over a given offset (equal to the row number, starting from 0 at the top),
- `MixColumns`, that multiplies each column of the input by the matrix

$$\mathsf{M} = \begin{pmatrix} \alpha & 1 & 1 \\ 1 & \alpha & 1 \\ 1 & 1 & \alpha \end{pmatrix} \tag{8.15}$$

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
27	48	46	31	63	30	91	56	47	26	10	34	8	23	78	77	80	65	71	43
20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39
36	72	29	79	83	7	58	95	69	74	67	35	32	59	82	14	75	99	24	87
40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59
16	90	76	51	28	93	50	38	25	3	13	97	55	60	49	86	57	89	62	45
60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79
18	37	1	6	98	68	39	17	19	20	64	44	33	40	96	2	12	41	52	85
80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99
81	5	0	15	54	88	92	21	84	22	53	11	4	94	42	66	70	9	61	73

Table 8.1: The fixed substitution box on $\mathbf{Z}_{10}^2$ of DEAN18.

where the multiplication of an arbitrary element $(a, b) \in \mathbf{Z}_{10}^2$ by α (resp. 1) is defined[1] by $\alpha \cdot (a, b) = (a+b, -a)$ (resp. $1 \cdot (a, b) = (a, b)$). One can easily see that this defines a structure on $\mathbf{Z}_{10}^2$ that is isomorphic to GF(4) × GF(25), on which the matrix is an MDS matrix [77, 150].

The branch number of the matrix multiplication is equal to 4, i.e., the total number of non-zero elements of the input and output columns is either 0, 4 or more. Consequently, given a non-trivial character $\rho = (\rho_1, \rho_2, \rho_3)$ on the output of the transformation, we obtain (given that we are considering a building block of type (b)) that the appropriate character $\chi = (\chi_1, \chi_2, \chi_3)$ on the input is non-trivial and that among the six characters $\chi_1, \ldots, \rho_3$, at least four are non-trivial. When at least one of the six characters is non-trivial, we say that the column is active.

Extending this result to the whole `MixColumns` transformation and applying similar arguments than those used on the AES [41], one can obtain that any two rounds characteristic (i.e., succession of three characters on the text space) has a weight lower bounded by $4Q$, where the weight is simply the number of non-trivial characters on $\mathbf{Z}_{10}^2$ among the 27 components of the three round characters and Q is the number of active columns at the output of the first round. Similar arguments also lead to the fact that the sum of the number of active columns at the output of the first and of the third round of a 4-round characteristic is at least 4. Consequently, the weight of a 4-round characteristic is at least 16.

Denoting by $\mathrm{LP}_{\max}(\mathsf{S})$ the maximum value of $\mathrm{LP}_{\chi,\rho}(\mathsf{S})$ over pairs of non-trivial characters, we conclude (under standard heuristic assumptions on the independence of the output of the characters at each round) that the linear probability of a $4r$-rounds characteristic is upper-bounded by $(\mathrm{LP}_{\max}(\mathsf{S}))^{16r}$. Assuming that one characteristic among the linear hull [125] is overwhelming, we conclude from Heuristic 8.2 that in the best case (from an adversary point of view), a distinguishing attack against a $4r$-round version of our toy cipher needs $q \approx ((d-1)\mathrm{LP}_{\max}(\mathsf{S}))^{-16r}$ samples, where d is the order of the linear cryptanalysis considered (i.e., d is the least common multiple of the orders of the input and of the output characters).

For the substitution box of our toy cipher, we obtain $\mathrm{LP}_{\max}(\mathsf{S}) \approx 0.069$. If we consider a generalized linear cryptanalysis of order 2, the number of samples that is nec-

[1] Considering the elements of $\mathbf{Z}_{10}^2$ as elements of $\mathbf{Z}_{10}[\alpha]/(\alpha^2 - \alpha + 1)$ naturally leads to this definition.

essary to attack four rounds is close to $3.8 \times 10^{18} \approx 2^{61}$. We conclude that $R = 8$ rounds are enough for DEAN18 to keep a high security margin (as far as linear cryptanalysis of order 2 is concerned). If we consider instead a generalized linear cryptanalysis of order 5, the number of samples to attack 8 rounds is close to $7.78 \times 10^{17} \approx 2^{59}$ so that 8 rounds are still sufficient to resist generalized linear cryptanalysis of order 5. Finally, if we consider order 10, the number of samples to attack 8 rounds drops to $4.18 \times 10^{6} \approx 2^{22}$, so that more rounds should be considered in that case. With 20 rounds for example, the number of samples needed is $3.57 \times 10^{16} \approx 2^{55}$.

It is possible to extend the previous construction to larger blocks. As an example, we introduce DEAN27, a toy cipher similar to DEAN18, that encrypts blocks of 27 decimal digits (which approximately corresponds to a block size of 90 bits). The structure of DEAN27 is the same than that of DEAN18 with some exceptions. The blocks of DEAN27 are represented as 3×3 arrays of element of the additive group $\mathbf{Z}_{10}^{3}$. Each rounds successively applies to the plaintext the following operations:

- `AddRoundKeys`, that performs a digit-wise addition of a round key,
- `SubBytes`, that applies a fixed bijective substitution box (defined in tables D.1 and D.2 in Appendix D, where an element $(a, b, c) \in \mathbf{Z}_{10}^{2}$ is represented as an integer $100 \cdot a + 10 \cdot b + c \in \{0, 1, \ldots, 999\}$) on each 3-digit element of the array,
- `ShiftRows`, which is the same than in DEAN18,
- `MixColumns`, that multiplies the input by the matrix M in (8.15), where the multiplication of an arbitrary element $(a, b, c) \in \mathbf{Z}_{10}^{3}$ by α (resp. 1) is defined[2] by $\alpha \cdot (a, b, c) = (a + b, c, a)$ (resp. $1 \cdot (a, b, c) = (a, b, c)$). This defines a structure on $\mathbf{Z}_{10}^{3}$ that is isomorphic to $\mathrm{GF}(8) \times \mathrm{GF}(125)$, on which the matrix is MDS.

The security analysis of DEAN27 is the same than that of DEAN18. For the substitution box S of DEAN27, we obtain $\mathrm{LP}_{\max}(\mathsf{S}) \approx 0.01$ (we obtained the table by drawing 150 tables at random, keeping the one having the smallest $\mathrm{LP}_{\max}$). If we consider a generalized cryptanalysis of order 2 on DEAN27, the number of samples necessary to attack four rounds is close to 10^{32} (which is to be compared to the size of the code book, which is 10^{27}). If we consider a generalized linear cryptanalysis of order 5, the number of samples required is approximately 2.32×10^{22} for four rounds and 5.42×10^{44} for height rounds. Finally, a successful attack against height rounds with a linear cryptanalysis of order 10 approximately requires 2.9×10^{33} samples.

8.6 A $\mathbf{Z}_{100}^{16}$ Generalized Linear Cryptanalysis of TOY100

In [58], Granboulan et al. introduce TOY100, a block cipher that encrypts blocks of 32 decimal digits. The structure of TOY100 is similar to that of the AES. An r rounds version of TOY100 is made of $r - 1$ identical rounds followed by a slightly

[2] Considering the elements of $\mathbf{Z}_{10}^{3}$ as elements of $\mathbf{Z}_{10}[\alpha]/(\alpha^3 - \alpha^2 - 1)$ naturally leads to this definition.

different final round. Each block is represented as a 4×4 matrix $A = (a_{i,j})_{i,j\in\{0,\ldots,3\}}$, the $a_{i,j}$'s being called subblocks. Round i (for $i = 1, \ldots, r-1$) first adds modulo 100 a round key to the subblocks (we do not describe the key schedule here as we assume that the round keys are mutually independent), then applies a fixed substitution box to each resulting round key, and finally mixes the subblocks together by applying a linear transformation. The last round replaces the diffusion layer by a modulo 100 round key addition. The round key addition, confusion, and diffusion layers are respectively denoted $\sigma[K]$, γ, and θ. The diffusion layer applied to a block A can be represented as a matrix product $M \times A \times M$ where

$$M = \begin{pmatrix} 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \end{pmatrix}$$

and where all computations are performed modulo 100. The best attack against $\mathsf{TOY100}$ presented so far is based on the generalization of linear cryptanalysis suggested in [58]. It breaks $\mathsf{TOY100}$ reduced to 7 rounds with a data/time complexity of $0.66 \cdot 10^{31}$. We propose here a linear cryptanalysis that breaks up to 8 rounds. We first observe that any block

$$A(\delta) = \begin{pmatrix} \delta & 0 & 100-\delta & 0 \\ 0 & 0 & 0 & 0 \\ 100-\delta & 0 & \delta & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

where $\delta \in \{1, \ldots, 99\}$ is such that $M \times A(\delta) \times M = A(\delta)$, i.e., is not changed by the diffusion layer. We let $\mathcal{I} = \{A(\delta), \delta = 1, \ldots, 99\}$ be the set of these 99 blocks. Our attack against $\mathsf{TOY100}$ reduced to r rounds first guesses 4 subblocks of the first round key and 4 subblocks of the last (the positions of which exactly correspond to the non-zero subblocks of $A(\delta)$). This allows to peel-off the first and last layers of substitution boxes, so that we now consider the transformation $(\theta \circ \gamma \circ \sigma[K])^{r-2} \circ \theta$ (where it is understood that the round keys are mutually independent). For any 4×4 input/output masks (i.e., blocks) $\alpha = (\alpha_{i,j})_{i,j\in\{1,\ldots,4\}}$ and $\beta = (\beta_{i,j})_{i,j\in\{1,\ldots,4\}}$ we let, for any transformation C on $\mathbf{Z}_{100}^{16}$,

$$\mathrm{ELP}_{\alpha,\beta}(\mathsf{C}) = \left|\mathrm{E}_M\left(\overline{\varphi_\alpha}(M)\varphi_\beta(\mathsf{C}(M))\right)\right|^2$$

r	Lower bound on $\max_{\alpha_0,\alpha_{r-2}} \mathrm{ELP}_{\alpha_0,\alpha_{r-2}}(\theta \circ \gamma \circ \sigma[K])^{r-2} \circ \theta$	d	Data/Time Complexity of the attack against r rounds
4	$0.37 \cdot 10^{-9}$	50	$0.55 \cdot 10^{8}$
5	$0.47 \cdot 10^{-14}$	50	$0.42 \cdot 10^{13}$
6	$0.66 \cdot 10^{-19}$	50	$0.31 \cdot 10^{18}$
7	$0.10 \cdot 10^{-23}$	50	$0.20 \cdot 10^{23}$
8	$0.18 \cdot 10^{-28}$	50	$0.11 \cdot 10^{28}$
9	$0.34 \cdot 10^{-33}$	50	$0.61 \cdot 10^{32}$

Table 8.2: Complexities of the best linear cryptanalysis we obtained on reduced round versions of $\mathsf{TOY100}$.

where

$$\varphi_\alpha(M) = e^{\frac{2\pi i}{100}\sum_{i,j=1}^{4}\alpha_{i,j}m_{i,j}}.$$

Applying the linear hull [125] formula of Theorem 8.2 and the observation on the diffusion layer of TOY100 we obtain that the linear probability on $(\theta \circ \gamma)^{r-2} \circ \theta$ with input (resp. output) masks $\alpha_0 \in \mathcal{I}$ (resp. $\alpha_{r-2} \in \mathcal{I}$) is such that

$$\begin{aligned}
&\mathrm{ELP}_{\alpha_0,\alpha_{r-2}}((\theta \circ \gamma{\circ}\sigma[K])^{r-2} \circ \theta) \\
&\qquad = \mathrm{ELP}_{\alpha_0,\alpha_{r-2}}((\theta \circ \gamma \circ \sigma[K])^{r-2}) \\
&\qquad = \sum_{\alpha_1 \in \mathbf{Z}_{100}^4} \cdots \sum_{\alpha_{r-3} \in \mathbf{Z}_{100}^4} \prod_{i=1}^{r-2} \mathrm{ELP}_{\alpha_{i-1},\alpha_i}(\theta \circ \gamma \circ \sigma[K]) \\
&\qquad \geq \sum_{\alpha_1 \in \mathcal{I}} \cdots \sum_{\alpha_{r-3} \in \mathcal{I}} \prod_{i=1}^{r-2} \mathrm{ELP}_{\alpha_{i-1},\alpha_i}(\theta \circ \gamma \circ \sigma[K]) \\
&\qquad = \sum_{\alpha_1 \in \mathcal{I}} \cdots \sum_{\alpha_{r-3} \in \mathcal{I}} \prod_{i=1}^{r-2} \mathrm{LP}_{\alpha_{i-1},\alpha_i}(\gamma).
\end{aligned}$$

Practical computations of the previous equations are given in Table 8.2 where d denotes the least common multiple of the orders of the input and of the output characters which maximize the linear probability. Using an 8-round linear hull and guessing the necessary keys on an extra round, we can thus break 9 rounds of TOY100 with data complexity $0.11 \cdot 10^{28}$ (and possibly 10 rounds with a data complexity of $0.61 \cdot 10^{32}$, which represents more than half the code book).

Chapter 9

A Generalized Linear Cryptanalysis of SAFER K/SK

SAFER is a family of block ciphers. The first member of this family was introduced by Massey and is called SAFER K-64 [107]. It encrypts 64-bit blocks under 64-bit keys. It is an iterated block cipher, meaning that it is made of a succession of rounds all identical in their structure. Each round is parameterized by two 64-bit round keys which are derived from the main 64-bit secret key using a key-schedule algorithm. SAFER K-64 is "byte-oriented" in the sense that the elementary operations operate on chunks of 8 bits. Moreover, it has the particularity to use two distinct group operations to mix key bits with text bits, namely, the exclusive-or and the addition modulo 256. Soon after the original publication, Massey announced SAFER K-128, a block cipher identical to SAFER K-64 (except for the recommended number of rounds) but with a different key schedule [108] that allowed to use 128-bit keys. The first real security issue on SAFER K/SK was pointed out by Knudsen [87] and lead Massey to update the key schedules, announcing SAFER SK-64 and SAFER SK-128 [109].

In a joint work with Khachatrian and Kuregian, Massey proposed SAFER+ [85] as an AES candidate. Based on the same design principle of the earlier versions of SAFER, SAFER+ encrypts 128-bit blocks. Although SAFER+ was not among the AES finalists, it is still widespread. For example, the E1 algorithm used during authentication in Bluetooth is based on SAFER+. Finally, the same authors submitted SAFER + + to the NESSIE project [86]. The main improvements of SAFER + + against SAFER+ concerns the diffusion layer, which improvements allowed to reduce the total number of round, increasing the encryption speed.

In this section, we focus on the two first members of the SAFER family, namely SAFER K-64 and SAFER SK-64. To the best of our knowledge, the best chosen plaintext attacks (CPA) against these block ciphers are Wu et al. truncated differentials [164] (which improve on previous work by Knudsen and Berson [90, 91]) which break up two 6 rounds of both versions of SAFER. The best known plaintext attacks are due to Nakahara et al. who manage to find a 3.75-round non-homomorphic linear relation [65] with bias $\epsilon = 2^{-29}$ for certain classes of weak keys, concluding that linear cryptanalysis [110, 111] does not seem to be a serious threat against SAFER K.

In this section, we will apply the generalization of linear cryptanalysis intro-

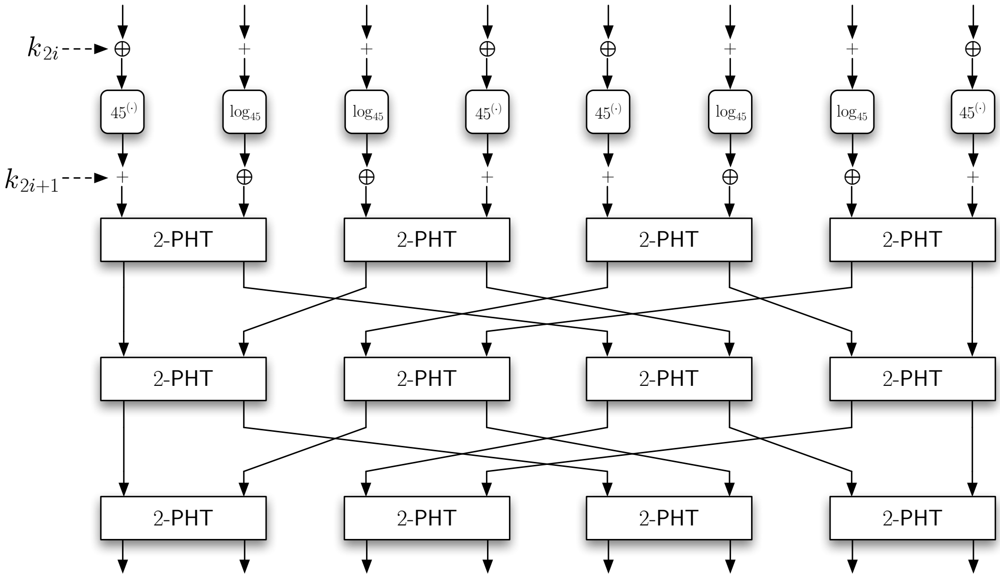


Figure 9.1: The ith encryption round function of SAFER

duced in Section 8.4, which breaks up to five rounds of SAFER K/SK. Because our definitions of linear probabilities, linear relations, etc. differ from those used in the classical version of linear cryptanalysis, it may seem that our results do not contradict the statement that SAFER is secure against Matsui's linear cryptanalysis (see [65] or [119] for example). In fact, they do, as we will see for the best attacks we could find on four rounds of SAFER K/SK.

In Section 9.1 we describe the encryption procedure of SAFER K/SK, give an overview of the properties of the key schedules that we exploit in our attacks, and give more details about previous cryptanalytic results. In Section 9.2 we study the main building blocks of SAFER K/SK with respect to our generalization of linear cryptanalysis. We introduce the notion of *reduced hull*, which simply corresponds to restrict the full linear hull [125] (see Theorem 8.2) to some (carefully chosen) characteristics, and explain how to build reduced hull of low weight (i.e., activating a small number of substitution boxes). In Section 9.3, we make use of the previous concepts to attack up to five rounds of SAFER.

9.1 The SAFER Family

A Short Description of the Encryption Procedure

The encryption procedures of SAFER K-64, SAFER K-128, SAFER SK-64, and SAFER SK-128 are almost identical. They all iterate the exact same round function, the only difference being that the recommended number of iteration of this round function is 6 for SAFER K-64 [107], 8 for SAFER SK-64 [109], and 10 for both 128-bit versions of

SAFER [107,109]. The round function is represented on Figure 9.1. An r-round version of SAFER encrypts 8 bytes of text by applying the round function r times followed by a final mixed key addition (whose structure is identical to the first mixed key addition layer of the round function). Each round is parameterized by two 8-byte round keys so that a $2r+1$ round keys must be derived from the secret key.

The round function first applies a byte-wise key addition, mixing exclusive-or's and additions modulo 256. Then, each byte goes through a substitution box. Two kinds of boxes are used on SAFER: $x \mapsto (45^x \bmod 257) \bmod 256$ and its inverse. The output of the substitution box layer goes through another byte-wise key addition before being processed by a diffusion layer made of boxes called 2-PHT and defined by $\text{2-PHT}(a,b) = (2a+b, a+b)$, the addition being performed modulo 256. Denoting $x \in \mathbf{Z}_{256}^8$ the input of the linear layer, the output $y \in \mathbf{Z}_{256}^8$ can be written as $y = \mathbf{M} \times x$ where

$$\mathbf{M} = \begin{pmatrix} 8 & 4 & 4 & 2 & 4 & 2 & 2 & 1 \\ 4 & 2 & 2 & 1 & 4 & 2 & 2 & 1 \\ 4 & 4 & 2 & 2 & 2 & 2 & 1 & 1 \\ 2 & 2 & 1 & 1 & 2 & 2 & 1 & 1 \\ 4 & 2 & 4 & 2 & 2 & 1 & 2 & 1 \\ 2 & 1 & 2 & 1 & 2 & 1 & 2 & 1 \\ 2 & 2 & 2 & 2 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}. \tag{9.1}$$

Finally, we will adopt a special notation to denote *reduced-round* versions of SAFER. We will consider each of the four round layers as one fourth of a complete round. Consequently, a 2.5 reduced-round version of SAFER will correspond to two full rounds followed by the first mixed key addition and substitution layer of the third round. With these notations, the encryption procedure of SAFER K-64 is actually made of 6.25 rounds. To be consistent with the notations of the original publications, when we refer to a r-round version of SAFER, we actually mean a $r + 0.25$ reduced-round version of SAFER.

A Very Short Description of the Key Schedules

For the sake of simplicity, we restrict to give the dependencies of each round key byte with respect to the main secret key instead of describing the key schedules of the various versions of SAFER.

- SAFER K-64: The jth round key byte $(1 \leq j \leq 8)$ only depends on the jth main secret key byte. For example, guessing the third byte of the main secret key allows to derive the third byte of each round key.
- SAFER SK-64: The jth byte $(1 \leq j \leq 8)$ of round key number i $(1 \leq i \leq 2r+1)$, depends on the ℓth byte of the secret key, where $\ell = (i+j-2) \bmod 9 + 1$ and where the 9th byte of the secret key is simply the exclusive-or of its previous 8 bytes.

Type	# rounds	Type of the Attack	Time	Plaintexts	Reference
SAFER K-64	3	KPA/Weak keys	2^{25}	2^{15}	[119]
SAFER K-64	4	KPA/Weak keys	2^{41}	2^{31}	[119]
SAFER K-64	5	KPA/Weak keys	2^{71}	2^{61}	[119]
SAFER K-64/128	5	CPA	2^{61}	2^{39}	[90,91]
SAFER K-64	5	CPA	2^{49}	2^{44}	[90,91]
SAFER K/SK-64	5	CPA	2^{46}	2^{38}	[164]
SAFER K/SK-64	6	CPA	2^{61}	2^{53}	[164]

Table 9.1: Previous cryptanalytic results on SAFER. The time complexity unit is a SAFER encryption.

In our analysis we assume that the key is a full vector of subkeys. When studying the average complexity of our attack, we further assume that they are randomly picked with uniform distribution.

Previous Cryptanalysis Results

We summarize known attacks against SAFER in Table 9.1

Differential Cryptanalysis and Friends. The resistance of SAFER to differential cryptanalysis [21] was extensively studied by Massey in [108], where it is argued that 5 rounds are sufficient to resist to this attack. It is shown by Knudsen and Berson [90,91] that 5 rounds can actually be broken using truncated differentials [88], a result which is extended to 6 rounds by Wu et al. in [164].

Linear Cryptanalysis and Friends. In [65], Harpes et al. apply a generalization of linear cryptanalysis [110, 111] to SAFER K-64 but do not manage to find an effective homomorphic threefold sum (which generalize the notion of linear characteristics) for 1.5 rounds or more. Nakahara et al. showed in [119] that for certain weak key classes, one can find a 3.75-round non-homomorphic linear relation with bias $\epsilon = 2^{-29}$ (which leads to a plaintext complexity of $8/\epsilon^2 = 2^{61}$ known plaintexts on five rounds and a time complexity of 2^{71} since the probability that a random key belongs to the correct weak key class is 2^{-10}). One of the conclusions of the authors of the latter article is that linear cryptanalysis does not seem to be a serious threat to SAFER K-64.

Concerns about the Diffusion Layer. The diffusion properties of the linear layer of SAFER have also been widely studied and, compared to the confusion layer, seem to be its major weakness. In [116], Murphy proposes an algebraic analysis of the 2-PHT layer, showing in particular that by considering the message space as a $\mathbf{Z}$-module, one can find a particular submodule which is an *invariant* of the 2-PHT transformation. In [150], Vaudenay shows that by replacing the original substitution boxes in a 4 round version of SAFER by random permutations, one obtains in 6.1% of the cases a construction that can be broken by linear cryptanalysis. This also lead Brincat and Meijer to explore potential alternatives of the 2-PHT layer [31].

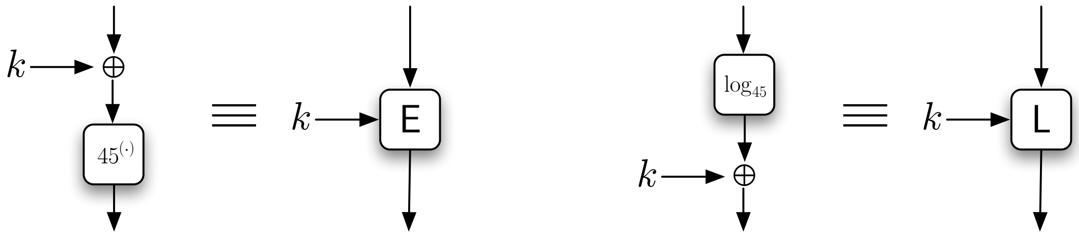


Figure 9.2: Replacing key exclusive-or and fixed substitution boxes by equivalent keyed substitution boxes

Concerns about the Original Key Schedule. The other major weakness of SAFER K is indubitably its key schedule. The analysis proposed in [90, 116] lead Massey to choose the one proposed by Knudsen in [90] for SAFER SK.

9.2 Linear Cryptanalysis of SAFER: from $\mathbf{Z}_2^8$ to $\mathbf{Z}_{2^8}$

A possible reason why linear cryptanalysis does not seem to be a threat for SAFER is that Matsui's linear characteristics (that fits so well the operations made in DES) are in fact *not* linear when it comes to the diffusion layer of SAFER except when they only focus on the least significant bit of the bytes. Yet, those bits are not biased through the substitution boxes [150]. Indeed, whereas a classical linear cryptanalysis combines text and key bits by performing exclusive-or's (i.e., additions in $\mathbf{Z}_2$), SAFER mostly relies on additions in $\mathbf{Z}_{2^8}$. In other words, the group structure that is classically assumed in linear cryptanalysis does not fit when it comes to study SAFER.

Our attack is focused on the additive group $(\mathbf{Z}_m^r, +)$. The m^r characters of this group are called *additive character modulo m* [121] and are the $\chi_{\mathbf{a}}$'s for $\mathbf{a} = (a_1, \ldots, a_r) \in \{0, 1, \ldots, m-1\}^r$ defined by

$$\begin{array}{cccc} \chi_{\mathbf{a}}: & \mathbf{Z}_m^r & \longrightarrow & \mathbf{C}^\times \\ & \mathbf{x} = (x_1, \ldots, x_r) & \longmapsto & \chi_{\mathbf{a}}(\mathbf{x}) = e^{\frac{2\pi i}{m} \sum_{\ell=1}^r a_\ell x_\ell}. \end{array} \tag{9.2}$$

The attack on SAFER that we describe in Section 9.3 only involves additive characters modulo 256. To simplify the notations (and to somehow stick to the vocabulary we are used to in classical linear cryptanalysis), we denote the linear probability of the permutation C over $\mathbf{Z}_{256}^r$ with respect to $\chi_{\mathbf{a}}$ and $\chi_{\mathbf{b}}$ (where $\mathbf{a}, \mathbf{b} \in \{0, 1, \ldots, 255\}^r$) by

$$\mathrm{LP}_{\mathbf{a},\mathbf{b}}(\mathsf{C}) = \left|\mathrm{E}_P(\overline{\chi_{\mathbf{a}}}(P)\chi_{\mathbf{b}}(\mathsf{C}(P)))\right|^2, \tag{9.3}$$

where P is a uniformly distributed random variable, and call it the linear probability of C with input mask $\mathbf{a}$ and output mask $\mathbf{b}$. Note that a mask byte equal to 128 means that we focus on the least significant bit of the text byte x as $128 \times x \bmod 256$ only depends on it.

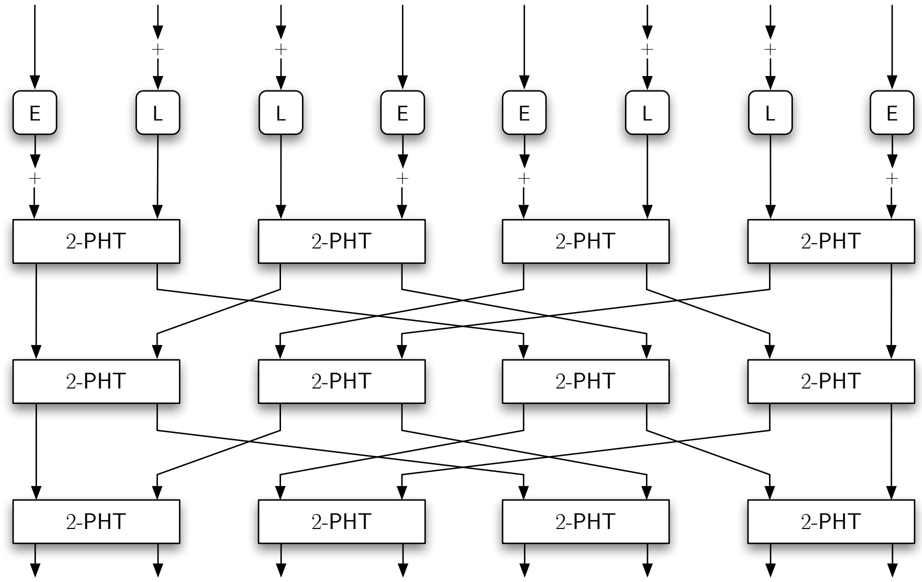


Figure 9.3: Another view of SAFER

Hiding the $\mathbf{Z}_2^8$ Group

Because the encryption procedure uses additions modulo 256 together with bit-wise exclusive-or, we have to deal with two types of characters. Nevertheless, one can notice that the mixture of group operations only occurs within the *confusion* layer. To simplify the analysis we can think of the succession of a round key exclusive-or and a fixed substitution box as a *keyed substitution box* (see Figure 9.2). Using this point of view, we represent on round of SAFER on Figure 9.3.

Studying SAFER's Building Blocks

We consider several building blocks that SAFER is made of, and study their behavior with respect to the linear probability (as given by (9.3)). In what follows, $a, b \in \{0, 1, \ldots, 255\}$ and $P \in \mathbf{Z}_{256}$ is a uniformly distributed random variable. Using the notations used in this section, we can reformulate some of the results of the toolbox introduced in Section 8.4.2 and represented on Figure 8.3.

Building Block (b): We consider the 2-PHT transformation and denote by $\mathbf{a} \in \mathbf{Z}_{256}^2$ and $\mathbf{b} \in \mathbf{Z}_{256}^2$ the input and output masks on this transformation. According to the result obtained in Section 8.4.2, we know that we have

$$\mathrm{LP}^{\text{2-PHT}}(\mathbf{a}, \mathbf{b}) = 1 \quad \Leftrightarrow \quad \mathbf{a} = \text{2-PHT}(\mathbf{b}).$$

This comes from the fact that the 2-PHT transformation is a symmetric linear operator (in the sense that $\text{2-PHT}^T = \text{2-PHT}$).

Building Block (c): We consider a key addition in $\mathbf{Z}_{256}$. For all $k \in \mathbf{Z}_{256}$ we have $\chi_a(P+k) = \chi_a(P)\chi_a(k)$ (as χ_a is a group homomorphism), so that

$$\mathrm{LP}_{a,a}(\cdot + k) = |\mathrm{E}_P(\overline{\chi_a}(P)\chi_a(P+k))|^2 = |\mathrm{E}_P(\overline{\chi_a}(P)\chi_a(P)\chi_a(k))|^2 = 1.$$

Note that if key K is random, the previous equation implies that

$$\mathrm{E}_K\left(\mathrm{LP}_{(a,a)}(\cdot + K)\right) = 1.$$

Building Block (d'): We consider the parallel computation through two fixed substitution boxes S_1 and S_2 over $\mathbf{Z}_{256}$ and denote by $\mathbf{a} = (a_1, a_2) \in \mathbf{Z}_{256}^2$ and $\mathbf{b} = (b_1, b_2) \in \mathbf{Z}_{256}^2$ the input and output masks on these boxes. We assume that the plaintext $P = (P_1, P_2) \in \mathbf{Z}_{256}^2$ is such that P_1 and P_2 are independent. Letting $\Theta = e^{\frac{2\pi i}{256}}$ we have

$$\begin{aligned}\mathrm{LP}_{\mathbf{a},\mathbf{b}}(\mathsf{S}_1\|\mathsf{S}_2) &= \left|\mathrm{E}_P(\Theta^{-(a_1P_1+a_2P_2)}\Theta^{(b_1\mathsf{S}_1(P_1)+b_2\mathsf{S}_2(P_2))})\right|^2 \\ &= \left|\mathrm{E}_{P_1}(\Theta^{-a_1P_1}\Theta^{b_1\mathsf{S}_1(P_1)})\mathrm{E}_{P_2}(\Theta^{-a_2P_2}\Theta^{b_2\mathsf{S}_2(P_2)})\right|^2 \\ &= \mathrm{LP}_{a_1,b_1}(\mathsf{S}_1)\cdot\mathrm{LP}_{a_2,b_2}(\mathsf{S}_2).\end{aligned}$$

When the boxes are random and independent, this leads to

$$\mathrm{E}_{\mathsf{S}_1,\mathsf{S}_2}\left(\mathrm{LP}_{\mathbf{a},\mathbf{b}}(\mathsf{S}_1\|\mathsf{S}_2)\right) = \mathrm{E}_{\mathsf{S}_1}(\mathrm{LP}_{a_1,b_1}(\mathsf{S}_1))\cdot\mathrm{E}_{\mathsf{S}_2}(\mathrm{LP}_{a_2,b_2}(\mathsf{S}_2)).$$

Assuming that the key bits are mutually independent, the previous building blocks make it possible to compute the linear probability of *one* full round of SAFER. Indeed if an input/output pair of masks $\mathbf{a}, \mathbf{b}$ are given, and letting $\mathbf{b}' = \mathbf{M}^T \times \mathbf{b}$ (where $\mathbf{M}$ is the matrix given in (9.1)), then the linear probability on one full round, simply denoted R, is given by

$$\mathrm{ELP}_{\mathbf{a},\mathbf{b}}(\mathsf{R}) = \prod_{i=1}^{8}\mathrm{ELP}_{a_i,b'_i}(\mathsf{S}_i)$$

where S_i corresponds to a keyed E box for $i = 1,4,5,8$ and to a keyed L box for $i = 2,3,6,7$.

Considering Several Rounds of SAFER: the Reduced Hull Effect

When several rounds are considered, Nyberg's linear hull effect [125] applies just as for classical linear cryptanalysis of Markov ciphers (see Theorem 8.2). Considering a succession of $r > 1$ rounds with independent round keys, and denoting $\mathbf{a}_0$ and $\mathbf{a}_r$ the input and the output masks respectively, this means that

$$\mathrm{ELP}_{\mathbf{a}_0,\mathbf{a}_r}(\mathsf{R}_r\circ\cdots\circ\mathsf{R}_1) = \sum_{\mathbf{a}_1,\ldots,\mathbf{a}_{r-1}}\prod_{i=1}^{r}\mathrm{ELP}_{\mathbf{a}_{i-1},\mathbf{a}_i}(\mathsf{R}_i).$$

We stress that this equation is a *real* equality (namely, *not* a heuristic approximation) under the hypothesis that the round keys are independent.

When cryptanalysing a block cipher, it is often considered that one specific characteristic (i.e., a succession of $r+1$ masks $\mathbf{a}_0, \mathbf{a}_1, \ldots, \mathbf{a}_r$) is overwhelming (i.e., approximates the hull) so that

$$\mathrm{ELP}_{\mathbf{a}_0,\mathbf{a}_r}(\mathsf{R}_r \circ \cdots \circ \mathsf{R}_1) \approx \prod_{i=1}^{r} \mathrm{ELP}_{\mathbf{a}_{i-1},\mathbf{a}_i}(\mathsf{R}_i).$$

This approach was taken by Matsui when cryptanalysing DES. In that particular case, the correctness of this approximation could be experimentally verified [111]. We will not consider the full linear hull effect nor restrict ourselves to one specific characteristic. Instead, we consider the characteristics among the hull following a specific *pattern*.

Definition 9.1 *Let* $\mathbf{a} \in \mathbf{Z}_{256}^8$ *be an arbitrary mask. The* pattern *corresponding to the mask* $\mathbf{a}$ *is the binary vector of length eight, with zeroes at the zero position of* $\mathbf{a}$ *and* $*$ *at the non-zero positions of* $\mathbf{a}$*. The weight* $w(\mathsf{p})$ *of a pattern* p *is the number of* $*$ *in this pattern. We denote the fact that a mask* $\mathbf{a}$ *corresponds to pattern* p *by* $\mathbf{a} \in \mathsf{p}$*. We denote by* $\mathtt{and}$ *the byte-wise masking operation, i.e., given an element* $m \in \mathbf{Z}_{256}^8$ *and a pattern* p*,* $m' = m \,\mathtt{and}\, \mathsf{p}$ *is such that* $m'_i = 0$ *if* $\mathsf{p}_i = 0$ *and* $m'_i = m_i$ *otherwise, for* $i = 1, \ldots, 8$*. We denote by* $\mathtt{int}_\mathsf{p}(m)$ *the integer representation of the concatenation of the bytes of* $m \,\mathtt{and}\, \mathsf{p}$ *corresponding to the non-zero positions of* p*, and by* $\mathcal{I}(\mathsf{p}) = \{\mathtt{int}_\mathsf{p}(m) : m \in \mathbf{Z}_{256}^8\}$*. Finally, for an arbitrary integer* $i \in \mathcal{I}(\mathsf{p})$*, we denote* $\mathtt{int}_\mathsf{p}^{-1}(i)$ *the element* $m \in \mathsf{p}$ *such that* $\mathtt{int}_\mathsf{p}(m) = i$*.*

For example, the pattern corresponding to the mask

$$\mathbf{a} = \mathtt{[0,128,0,0,0,255,7,1]}$$

is $\mathsf{p} = \mathtt{[0*000***]}$ (which is of weight 4). If

$$m = \mathtt{[3,128,128,255,0,255,7,1]},$$

then $m \,\mathtt{and}\, \mathsf{p} = \mathbf{a}$, and

$$\mathtt{int}_\mathsf{p}(m) = \mathsf{10000000\,11111111\,00000111\,00000001}_2 = 2164197121.$$

Note that for an arbitrary element $m \in \mathbf{Z}_{256}^8$ and any pattern p,

$$\mathtt{int}_\mathsf{p}^{-1}(\mathtt{int}_\mathsf{p}(m)) = m \,\mathtt{and}\, \mathsf{p}.$$

The fact that we only consider, among the hull, the characteristics following a given sequence of patterns $\mathsf{p}_0, \mathsf{p}_1, \ldots, \mathsf{p}_r$ can be written as

$$\mathrm{ELP}_{\mathbf{a}_0,\mathbf{a}_r}(\mathsf{R}_r \circ \cdots \circ \mathsf{R}_1) \approx \sum_{\substack{\mathbf{a}_1 \in \mathsf{p}_1 \\ \cdots \\ \mathbf{a}_{r-1} \in \mathsf{p}_{r-1}}} \prod_{i=1}^{r} \mathrm{ELP}_{\mathbf{a}_{i-1},\mathbf{a}_i}(\mathsf{R}_i). \tag{9.4}$$

where $a_0 \in \mathsf{p}_0$ and $a_r \in \mathsf{p}_r$. We call this approximation the *reduced hull effect*. Note that in any case, (9.4) actually underestimates the true linear hull.

Building Reduced Hulls on Two Rounds

In order to build such reduced hulls on SAFER, we start by enumerating the *possible* sequences of patterns on the linear diffusion layer (see tables E.1 through E.5 in Appendix E). In the tables, we denote the existence of an input mask $\mathbf{a}_1$ of pattern p_1 corresponding to an output mask $\mathbf{a}_2$ of pattern p_2 (i.e., $\mathbf{a}_1 = \mathbf{M}^T \times \mathbf{a}_2$, so that $\mathrm{LP}_{\mathbf{a}_1,\mathbf{a}_2}(\mathbf{M}) = 1$) by $\mathsf{p}_1 \rightarrow \mathsf{p}_2$. Moreover, we denote the fact that n *distinct* pairs of input/output masks following the pattern $\mathsf{p}_1/\mathsf{p}_2$ can be found by $\mathsf{p}_1 \xrightarrow{n} \mathsf{p}_2$. For example, the output mask corresponding to the input mask $\mathbf{a} =$ [0,0,0,0,0,0,0,128] on the linear layer is $\mathbf{b} =$ [128,0,0,0,0,0,0,0], and there is no other possible mask with the same input/output patterns. This fact is denoted by

$$\texttt{[0000000*]} \xrightarrow{1} \texttt{[*0000000]}.$$

If we consider the input pattern [0000000*] and the output pattern [***0*000] for example, two distinct pairs of masks on the linear layer following these patterns can be found (namely, [0,0,0,0,0,0,0,64] corresponds to [192,128,128,0,128,0,0,0] and [0,0,0,0,0,0,0,192] to [64,128,128,0,128,0,0,0]). This is denoted by

$$\texttt{[0000000*]} \xrightarrow{2} \texttt{[***0*000]}.$$

In the tables, these patterns are ordered by input/output weights, where $w_1 \rightarrow w_2$ ($1 \leq w_1, w_2 \leq 8$) denotes the list of all possible input/output patterns $\mathsf{p}_1, \mathsf{p}_2$ on the linear layer such that p_1 is of weight w_1 and p_2 is of weight w_2. To reduce the size of the list, we restrict it to patterns of weight sum less than 7.

Next, we need to build characteristics on several rounds based on the lists of possible succession of patterns on the linear layer. We proceed step-by-step, starting with characteristics on two rounds. Two characteristics *on full rounds* can only be concatenated if, and only if, the output mask of the first one is equal to the input mask of the second one. This translates for patterns as follows: two successions of patterns on the linear layer can only be concatenated if the output pattern of the first succession is equal to the input pattern of the second succession.

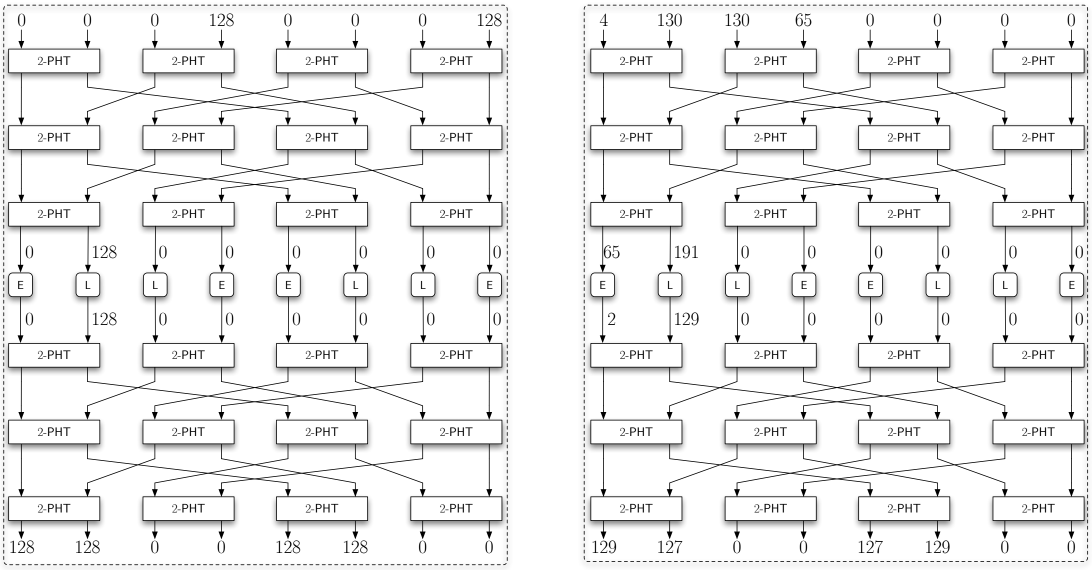
Example 9.1 We can concatenate

$$\texttt{[000*000*]} \xrightarrow{1} \texttt{[0*000000]} \quad \text{and} \quad \texttt{[0*000000]} \xrightarrow{1} \texttt{[**00**00]}.$$

We denote this by $\texttt{[000*000*]} \xrightarrow{1} \texttt{[0*000000]} \xrightarrow{1} \texttt{[**00**00]}$. This means that succession of patterns of weights $2 \rightarrow 1 \rightarrow 4$ on two rounds exist. In this particular example, there is only one characteristic corresponding to this succession of masks, which is represented on Figure 9.4(a). □

Example 9.2 Similarly, one can obtain the succession

$$\texttt{[****0000]} \xrightarrow{252} \texttt{[**000000]} \xrightarrow{254} \texttt{[**00**00]}$$



(a) (b)

Figure 9.4: Examples of characteristics on two successive linear layers

which is a succession of pattern of weights 4 → 2 → 4 on two rounds. In this case, $252 \times 254 = 64008$ distinct characteristics correspond to this succession (one of which is represented on Figure 9.4(b)). □

Finally, it should be noted that the characteristic of Example 9.1 actually leads to an ELP equal to 0. This is due to the fact that both input and output masks on the substitution box are equal to 128, which is equivalent to compute the traditional linear probability by only considering the least significant bit. In the second example, computing the reduced hull leads to a non-zero linear probability. On Table E.6 in Appendix E we list all possible sequences of three weights less than 6. A ✓ indicates that a non-zero reduced hull with the corresponding weight patterns exists, a 0 indicates that a reduced hull exists but always lead to a ELP equal to 0 (like in Example 9.1), a $\emptyset$ means that no characteristic corresponds to the succession of weights. If nothing is specified, it means that we do not need the corresponding patterns for our attacks.

9.3 Attacks on Reduced-Round Versions of SAFER

From Distinguishing Attacks to Key Recovery

In this section, a *reduced hull on r diffusion layers of* SAFER corresponds to a succession patterns on r successive linear layers separated by confusion layers. The *weight* of a reduced hull is the number of active substitution boxes (i.e., the number of boxes with non-zero input/output masks) for any characteristic of the hull. For

```
Input: A reduced hull on $r$ rounds with input mask $\mathbf{a_0} \in \mathsf{p}_0$ and output mask $\mathbf{a_r} \in \mathsf{p}_r$.
Output: A set of counters $\mathrm{lh}_{\kappa_1,\kappa_2,\kappa_{2r+1}}$ with $\kappa_1, \kappa_2 = 0, \ldots, 2^{8w(\mathsf{p}_0)}-1$ and $\kappa_{2r+1} = 0, \ldots, 2^{8w(\mathsf{p}_r)}-1$.
Memory: A set of counters $N_{i,j}$ initialized to 0, with $i = 0, \ldots, 2^{8 \cdot w(\mathsf{p}_0)}-1$ and $j = 0, \ldots, 2^{8 \cdot w(\mathsf{p}_r)}-1$.
foreach of the $d$ plaintext/ciphertext pair $(m, c)$ do
    $i \leftarrow \texttt{int}_{\mathsf{p}_0}(m)$ and $j \leftarrow \texttt{int}_{\mathsf{p}_r}(c)$
    $N_{i,j} \leftarrow N_{i,j} + 1$
done
foreach $(\kappa_1, \kappa_2, \kappa_{2r+1}) \in \mathcal{I}(\mathsf{p}_0) \times \mathcal{I}(\mathsf{p}_0) \times \mathcal{I}(\mathsf{p}_r)$ do
    $k_1 \leftarrow \texttt{int}_{\mathsf{p}_0}^{-1}(\kappa_1)$, $k_2 \leftarrow \texttt{int}_{\mathsf{p}_0}^{-1}(\kappa_2)$, and $k_{2r+1} \leftarrow \texttt{int}_{\mathsf{p}_r}^{-1}(\kappa_{2r+1})$
    /* compute the likelihood lh corresponding to the round keys guess */
    $\mathrm{counter}_h \leftarrow 0$ for all $h$ in the subgroup of $\mathbf{C}^\times$ induced by $\chi_{\mathbf{a}_0}$ and $\chi_{\mathbf{a}_r}$
    foreach $(i, j) \in \{0, 1, \ldots, 2^{8w(\mathsf{p}_0)} - 1\} \times \{0, 1, \ldots, 2^{8w(\mathsf{p}_r)} - 1\}$ such that $N_{i,j} > 0$ do
        $m \leftarrow \texttt{int}_{\mathsf{p}_0}^{-1}(i)$ and $c \leftarrow \texttt{int}_{\mathsf{p}_r}^{-1}(j)$
        Add/xor $k_1$ to $m$, apply the substitution box layer, add/xor $k_2$, call the result $m'$.
        Subtract $k_{2r+1}$ to $c$, call the result $c'$
        $h \leftarrow \overline{\chi_{\mathbf{a_0}}}(m')\chi_{\mathbf{a}_r}(c')$ and $\mathrm{counter}_h \leftarrow \mathrm{counter}_h + N_{i,j}$
    done
    $\mathrm{lh}_{\kappa_1,\kappa_2,\kappa_{2r+1}} \leftarrow \max_h(\mathrm{counter}_h)$
done
```

Table 9.2: Key Recovery Attack against a r reduced-round version of SAFER.

example, the succession

$$[\texttt{****0000}] \xrightarrow{252} [\texttt{**000000}] \xrightarrow{254} [\texttt{**00**00}]$$

(of Example 9.2) is a reduced hull of weight 2 on two diffusion layers. A reduced hull easily leads to a distinguishing attack on a reduced-round version of SAFER that would start and end by a diffusion layer.

Table 9.2 describes a key recovery attack on a SAFER reduced to r rounds by means of a reduced hull on r diffusion layers. Each of the counters obtained with this algorithm measures the likelihood of the corresponding round key bits (for round keys 1, 2, and $2r+1$) of being the correct ones. The way these counters are computed relies on the decision rule of Heuristic 8.2. Based on this heuristic, we expect the correct guess to be near the top of a list sorted according to these counters when the number of plaintexts/ciphertext pairs is close to

$$q = \frac{8 \ln 2}{(d-1)\mathrm{ELP}^{\mathsf{C}}(\mathbf{a}_0, \mathbf{a}_r)},$$

where d is the least common multiple of the respective orders of the input and of the output characters on the r rounds.

In the worst case, line 4 loops $2^{8\cdot(2w(\mathsf{p}_0)+w(\mathsf{p}_r))}$ times. In practice, the complexity is much lower (by considering key dependence due to the key schedule) and depends on the number of bits n_k that we need to guess in our attacks. When considering SAFER K-64 for example, a guess for the meaningful bytes of k_1 uniquely determines the bytes of k_2 (for the reasons given in Section 9.1.2). Similarly, the meaningful bytes of k_{2r+1} that are at the same positions than those of k_1 are also uniquely determined. When considering SAFER SK-64, similar techniques may apply, depending on the specific shapes of the input/output masks and the number of rounds. In all cases, if the meaningful bytes of k_2 and k_{2r+1} are actually added modulo 256, then they don't need to be guessed (as for Building Block (a), they don't alter the linear probability). If we only consider SAFER SK, this observation also applies to k_2. Finally, line 8 loops 2^{n_p} times where $n_p = \min(8\cdot(w(\mathsf{p}_0)+w(\mathsf{p}_r)), \log_2 q)$ (as $\sum_{i,j} N_{i,j} = q$). Consequently, given any input/output masks $\mathbf{a}_0 \in \mathsf{p}_0$ and $\mathbf{a}_r \in \mathsf{p}_r$, the time complexity of the attack is given by

$$T = \frac{8\ln 2}{(d-1)\mathrm{ELP}^{\mathsf{C}}(\mathbf{a}_0,\mathbf{a}_r)} + 2^{n_k+n_p}. \tag{9.5}$$

An attack on 2 Rounds

The best attacks we could find on two rounds are based on reduced hull of weight 2 and are listed in Table 9.3. The best attack on SAFER K exploits the reduced hull represented on Figure 9.5. To perform the attack, one needs to guess 8 bits of K_1, no bits of K_2 (as those that could be meaningful are added modulo 256 and thus do not influence the linear probability), and 8 bits of K_5 (as those in position 4 are uniquely determined by the guess made on K_1). We thus obtain $n_k = 16$. The algorithm then loops through the

$$q = \frac{8\ln 2}{\max_{\mathbf{a}_0,\mathbf{a_2}}\left((d-1)\mathrm{ELP}^{\mathsf{H}^{(2)}}(\mathbf{a}_0,\mathbf{a_2})\right)}$$

pairs, where $\mathsf{H}^{(2)}$ here denotes the reduced hull and where $\mathbf{a}_0$ (resp. $\mathbf{a}_2$) denote the input (resp. output) mask on $\mathsf{H}^{(2)}$. The final complexity is computed according to (9.5) and is approximately equal to $2^{23.62}$. Table 9.3 gives other complexities for various characteristics.

For SAFER SK, the previous reduced hull leads to a higher complexity than

Reduced hull	$\min_{\mathbf{a}_0,\mathbf{a}_2} \frac{8\ln 2}{(d-1)\mathrm{ELP}^{\mathsf{H}^{(2)}}(\mathbf{a}_0,\mathbf{a}_2)}$	2^{n_p}	2^{n_k}	Complexity
$[000*0000] \xrightarrow{1} [**000000] \xrightarrow{254} [**00**00]$	$2^{7.35}$	$2^{7.35}$	$2^{24}/2^{24}$	$2^{31.35}/\mathbf{2^{31.35}}$
$[000*0000] \xrightarrow{1} [**000000] \xrightarrow{255} [00**00**]$	$2^{7.62}$	$2^{7.62}$	$2^{16}/2^{24}$	$\mathbf{2^{23.62}}/2^{31.62}$
$[00000*00] \xrightarrow{1} [*000*000] \xrightarrow{254} [*0*0*0*0]$	$2^{6.61}$	$2^{6.61}$	$2^{24}/2^{32}$	$2^{30.61}/2^{38.61}$
$[00000*00] \xrightarrow{1} [*000*000] \xrightarrow{255} [0*0*0*0*]$	$2^{6.87}$	$2^{6.87}$	$2^{24}/2^{32}$	$2^{30.87}/2^{38.87}$
$[000000*0] \xrightarrow{1} [*0*00000] \xrightarrow{254} [****0000]$	$2^{7.35}$	$2^{7.35}$	$2^{24}/2^{24}$	$2^{31.35}/2^{31.35}$
$[000000*0] \xrightarrow{1} [*0*00000] \xrightarrow{255} [0000****]$	$2^{7.62}$	$2^{7.62}$	$2^{24}/2^{32}$	$2^{31.62}/2^{39.62}$

Table 9.3: Reduced hull on two diffusion layers and attack complexities against two rounds of SAFER K/SK.

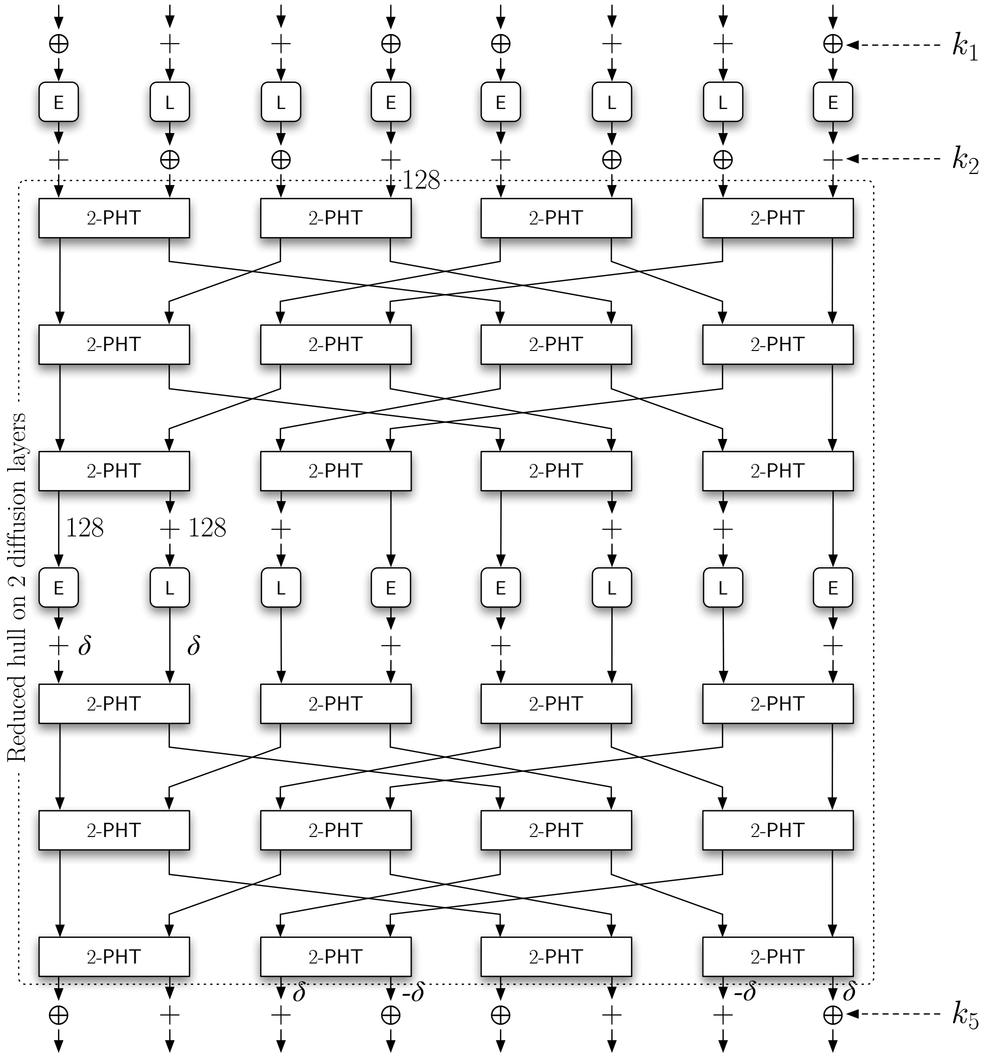


Figure 9.5: The reduced hull on 2 diffusion layers used to attack 2 rounds of SAFER K

for SAFER K as 8 more bits of K_5 must be guessed. It appears that the best attack on two rounds of SAFER SK makes use of the first characteristics given in Table 9.3 and has a complexity approximately equal to $2^{31.35}$.

Attacks on 3, 4, and 5 Rounds

To attack three rounds of SAFER K/SK, we make use of reduced hulls on two diffusion layers of weight 6. We list all such possibles reduced hulls in Table E.7 in Appendix E, restricting to input/output patterns of weight 1 to limit the number of key bits guess. Using similar techniques than for the two rounds case, we manage to mount an attack against both versions of SAFER reduced to three rounds within a complexity close to $2^{38.75}$.

To attack four rounds, we use the reduced hulls on four diffusion layers listed on Table E.8 in Appendix E. The first reduced hull in the table shows that both versions of SAFER can be attacked within a complexity close to 2^{49}. Some of our attacks actually *exactly* correspond to the original version of linear cryptanalysis. It is the case here as the non-zero byte of both input/output masks maximizing the expected linear probability is equal to 128. This means that both input/output masks only focus on one single bit. It is not clear to us whether this correlation can easily be found by other means than ours.

Finally, the first reduced hull of Table E.9 in Appendix E shows that 5 rounds of SAFER K can be broken within a complexity of 2^{56}. Finally, we note that among the output masks that maximize the expected linear probability, several end by an even byte. For example the best reduced hull is obtained when the last output masks ends by a 2. The same remarks applies to the fourth byte of the output mask. Consequently, strictly less than 16 key bits need to be guessed in the last round key, so that the same reduced hull can also be used break 5 rounds of SAFER SK.

9.4 Implementation of the Attack on 2 Rounds

We implemented the best attack on two rounds of SAFER K. As show on Table 9.3, the best choice of the reduced hull is in this case:

$$\texttt{[000*0000]} \xrightarrow{1} \texttt{[**000000]} \xrightarrow{255} \texttt{[00**00**]}.$$

More precisely, the input/output masks we choose are

$$\mathbf{a}_0 = \texttt{[0 0 0 128 0 0 0 0]} \quad \text{and} \quad \mathbf{a}_2 = \texttt{[0 0 183 73 0 0 73 183]}.$$

We tweaked the attacked of Table 9.2 to our particular two round attack. For example, since the total number of samples that we expect to require is less than $2^{8\cdot w(\mathsf{p}_0)} \times 2^{8\cdot w(\mathsf{p}_2)} = 2^{40}$, we did not implemented the $N_{i,j}$ counters but simply used tables to store the plaintext/ciphertext pairs. The pseudo-code of our key ranking experiments is given in Table 9.4. The parameter KEYS corresponds to the number of keys used

```
 1:  for $k = 1, 2, \ldots, \mathtt{KEYS}$
 2:    $k \leftarrow \{0,1\}^{64}$ and $(k_1, k_2, \ldots, k_5) \leftarrow \mathsf{SAFER\_K\_KEY\_SCHEDULE}(k)$
 3:    for $t = 1, 2, \ldots, \mathtt{PAIRS}$
 4:      $m[t] \leftarrow \{0,1\}^{64}$ and $c[t] \leftarrow \mathsf{SAFER\_K\_TWO\_ROUNDS\_ENCRYPT}_{k_1,k_2\ldots,k_5}(m[t])$
 5:    done
 6:    $\mathrm{lh}_{\mathtt{correct\_key}} \leftarrow \mathtt{compute\_lh}(c, m, k_1, \ldots, k_5)$ and $\mathsf{rank}[k] \leftarrow 1$
 7:    foreach $(k'_1, k'_2, \ldots, k'_5) \neq (k_1, k_2, \ldots, k_5)$ do
 8:      $\mathrm{lh}_{\mathtt{incorrect\_key}} \leftarrow \mathtt{compute\_lh}(c, m, k'_1, \ldots, k'_5)$
 9:      if $\mathrm{lh}_{\mathtt{incorrect\_key}} > \mathrm{lh}_{\mathtt{correct\_key}}$ then $\mathsf{rank}[k] \leftarrow \mathsf{rank} + 1$ end if
10:      if $\mathrm{lh}_{\mathtt{incorrect\_key}} = \mathrm{lh}_{\mathtt{correct\_key}}$ then $\mathsf{rank}[k] \leftarrow \mathsf{rank} + 0.5$ end if
11:    done
12:  done
13:  output $\frac{1}{\mathsf{KEYS}} \sum_k \mathsf{rank}[k]$

14:  procedure $\mathtt{compute\_lh}(c, m, k_1, \ldots, k_5)$
15:    $\mathsf{counter}[0, 1, 2, \ldots, 255] \leftarrow 0$
16:    for $t = 1, 2, \ldots, \mathtt{PAIRS}$
17:      $m'_3 \leftarrow \mathsf{E}[m_3[t] \oplus k_{1,3}] + k_{2,3}$
18:      $c'_2 \leftarrow c_2[t] - k_{5,2}$, $c'_3 \leftarrow c_3[t] \oplus k_{5,3}$, $c'_6 \leftarrow c_6[t] - k_{5,6}$, $c'_7 \leftarrow c_7[t] \oplus k_{5,7}$
19:      $\ell \leftarrow -128 \cdot m'_3 + 183 \cdot c'_2 + 73 \cdot c'_3 + 73 \cdot c'_6 + 183 \cdot c'_7$ and $\mathtt{counter}[\ell] \leftarrow \mathtt{counter}[\ell] + 1$
20:    done
21:    return $\max_e \mathtt{counter}[e]$
```

Table 9.4: Key ranking experiments based on the best two rounds attack on SAFER K.

to compute the mean position of the correct round key guess among all the possible key candidates. The parameter PAIRS is the number of samples used to rank each key. In this algorithm, the SAFER_K_KEY_SCHEDULE at line 2 actually corresponds to a reduced version of the original key schedule of SAFER K where we only output the first five round keys. The SAFER_K_TWO_ROUNDS_ENCRYPT (line 4) encrypts a 64-bit message by applying two full rounds followed by the mixed addition with the fifth key. In line 6 we compute the likelihood of the correct round key. This likelihood is based on the decision rule of Heuristic 8.2. Here we count the number of occurrences of each $\ell \in \{0, 1, 2, \ldots, 255\}$ instead of counting the number of occurrences of each $e^{\frac{2i\pi}{256}\ell}$, which is completely equivalent. Between lines 7 and 12, we run through all the possible wrong round keys and increment the position of the correct round key guess each time the likelihood of a wrong round key is higher (or equal) to that of the correct round key. On line 13, the algorithm outputs the average value of the ranks.

The results of our experiments are shown on figures 9.6 and 9.7. Figure 9.6 illustrates how the correct round key rank depends on the number of plaintext/ciphertext

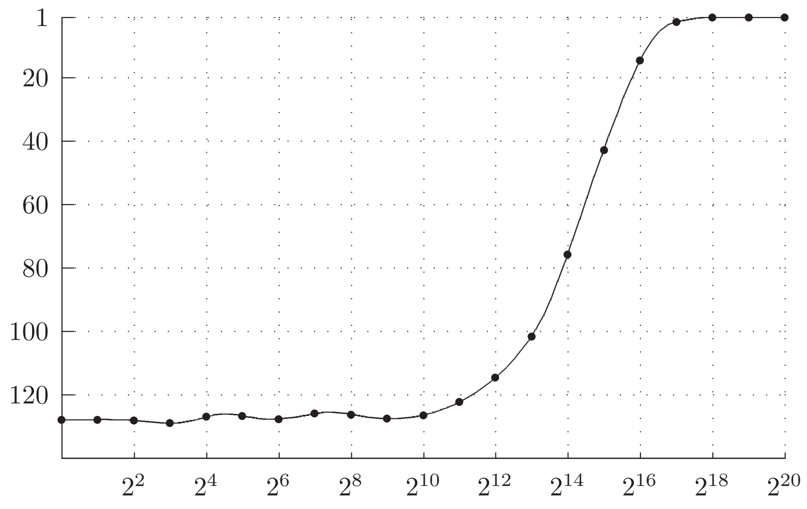


Figure 9.6: Rank of the correct key when the third, seventh, and eighth bytes of k_5 are *correct* for all the wrong keys

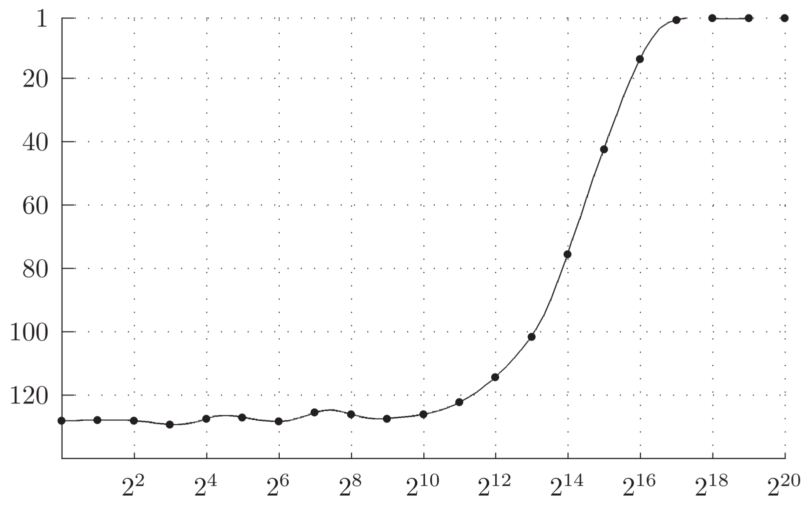


Figure 9.7: Rank of the correct key when the third, seventh, and eighth bytes of k_5 are *incorrect* for all the wrong keys

r	Reduced hull	Attack Complexity
2	$\texttt{[000*0000]} \xrightarrow{1} \texttt{[**000000]} \xrightarrow{254} \texttt{[**00**00]}$	$2^{31.5}/2^{31.35}$
2	$\texttt{[000*0000]} \xrightarrow{1} \texttt{[**000000]} \xrightarrow{254} \texttt{[00**00**]}$	$2^{23.62}/2^{31.62}$
3	$\texttt{[0*000000]} \xrightarrow{1} \texttt{[**00**00]} \xrightarrow{255} \texttt{[000*000*]} \xrightarrow{1} \texttt{[0*000000]}$	$2^{38.75}/2^{38.75}$
4	$\texttt{[000*0000]} \xrightarrow{1} \texttt{[**000000]} \xrightarrow{254} \texttt{[**00**00]} \xrightarrow{255} \texttt{[000*000*]} \xrightarrow{1} \texttt{[0*000000]}$	$2^{49.22}/2^{49.22}$
4	$\texttt{[0*0*0000]} \xrightarrow{1} \texttt{[0000**00]} \xrightarrow{254} \texttt{[**00**00]} \xrightarrow{255} \texttt{[000*000*]} \xrightarrow{1} \texttt{[0*000000]}$	$2^{49.18}/2^{56}$
5	$\texttt{[000*0000]} \xrightarrow{1} \texttt{[**000000]} \xrightarrow{254} \texttt{[**00**00]} \xrightarrow{254} \texttt{[0*000*00]} \xrightarrow{1} \texttt{[0*000*00]} \xrightarrow{254} \texttt{[0*0*0*0*]}$	$2^{56}/2^{56}$

Table 9.5: Selected reduced hulls on r diffusion layers and attack complexities against r rounds of SAFER K/SK.

pairs. Here we do *not* rank the correct guess among all possible guesses, but only consider (on line 7 of the algorithm given in Table 9.4) the wrong round keys such that

$$k'_{5,2} = k_{5,2},\ k'_{5,6} = k_{5,6},\ \text{and } k'_{5,7} = k_{5,7}.$$

Note that a guess for $k'_{1,3}$ uniquely determines $k'_{2,3}$ and $k'_{5,3}$. Consequently, the correct round key is ranked among 256 possible keys. From Figure 9.6 we can see that the correct round key can be distinguished from wrong keys with non-negligible probability when the number of samples q is roughly greater than 2^{10}. For $q > 2^{16}$ the correct guess is systematically ranked in the top ten keys. Figure 9.7 shows similar results except that the wrong keys are such that

$$k'_{5,2} \neq k_{5,2},\ k'_{5,6} \neq k_{5,6},\ \text{and } k'_{5,7} \neq k_{5,7}.$$

In this case also, the correct key can be distinguished from the wrong ones with non-negligible probability when $q > 2^{10}$ and is almost always guessed correctly when more than 2^{16} samples are available.

These results show that Heuristic 8.2 might underestimate a little bit the number of samples needed to rank the right key among the most likely candidates. Future work could consider small versions of SAFER (like for example the one suggested in [119]) in order to perform attacks on more rounds.

9.5 Conclusion

We have presented a generalized linear cryptanalysis of SAFER K/SK, the complexity of which are summarized in Table 9.5. Our attacks do not break the full versions of these ciphers but manage to attack up to 5 rounds. This improves on previous results from Nakahara et al. We showed that in certain cases (for example, in the attack against four rounds), our attack actually correspond to a classical linear cryptanalysis. This seems to contradicts the belief that SAFER is particularly strong against linear cryptanalysis.

It is not clear to us whether our techniques can be used to attack the 128-bit block versions of the SAFER family. Because of the block length, the search of useful reduced hull is much more complex than for SAFER K/SK. Moreover, in the case of SAFER + +, the good diffusion properties may lead to high complexities within a small number of rounds.

Part III

Block Cipher Designs and Security Proofs

Chapter 10

Provable Security and the Decorrelation Theory

Most modern block ciphers are designed to resist a wide range of cryptanalytic techniques. Among them, one may cite linear cryptanalysis [110, 111, 147], differential cryptanalysis [21, 22], as well as several variants such as impossible differentials [18], the boomerang attack [162] or the rectangle attack [19]. Proving resistance to all these attacks is often tedious and does not give any guarantee that a subtle new variant would not break the construction. Rather than considering all known attacks individually, it would obviously be preferable to give a *unique* proof, valid for a family of attacks.

In [155], Vaudenay shows that the Decorrelation Theory provides tools to prove security results in the Luby-Rackoff model [102], i.e., against adversaries only limited by the number of plaintext/ciphertext pairs they can access. When these pairs are randomly chosen and mutually independent, these adversaries exactly correspond to the q-limited distinguisher studied in the previous sections and are referred to as q-limited *non-adaptive* adversaries. When the adversary is allowed to choose each query depending on the results of the previous ones, it is referred to as a q-limited *adaptive* adversary. In Section 10.1 we detail the Luby-Rackoff model and in Section 10.2 we recall the notion of *q-wise distribution matrix* of a block cipher and how it relates to the advantage of the best (non-)adaptive q-limited adversary.

In practice, one is more interested by the security against "practical" attacks (such as linear and differential cryptanalysis) rather than provable security against abstract adversaries. Vaudenay shows in [154] that it is actually possible to relate the security against q-limited adversaries to both linear and differential cryptanalysis, but also to a wider class of attacks known as *iterated attacks.* This class of attacks was initially inspired by linear and differential cryptanalysis and actually formalizes most of the possible statistical attacks against block ciphers. In particular, linear cryptanalysis is an iterated attack of order 1, differential cryptanalysis is of order 2, and higher order differential cryptanalysis [88, 95] of order i is an iterated attack of order 2^i. We recall some of these results in Section 10.3.

As an example, we recall in Section 10.4 how the famous result from Luby and Rackoff about the security of the Feistel scheme [50, 102] translates in the Decorrelation Theory. This security result will play an important role in the security proof of KFC [5], one of the two provable secure constructions that we will introduce in chapters 11 and 12.

Computing the exact adversaries' advantage against a *practical* block cipher can prove to be a hard task in general, even by means of the Decorrelation Theory. As a possible solution, Vaudenay suggests to use so-called decorrelation modules. These can be though as building blocks, with perfect decorrelation up to a given (small) order, that can be assembled to construct a block cipher which security easily follows from that of the modules. In [152, 155], Vaudenay proposes practical constructions based on these modules. In particular, COCONUT98 is one of the first efficient block cipher provably secure against 2-limited adversaries. Yet, since decorrelation results do not prove anything more than what they claim, security against 2-limited adversaries does not give any kind of guaranty against q-limited adversaries for $q > 2$. This is illustrated by Wagner's boomerang attack [162] that breaks COCONUT98 within a complexity close to 2^{38}, which is less than the 2^{64} level of security that one would expect. Although it was made clear by Wagner that this attack "*is not to suggest that the decorrelation approach is fundamentally flawed [...], but rather that the theoretical results must be interpreted with caution*" [162, p.159], it lead to a certain confusion in the academic world. For example, according to Knudsen and Rijmen, "*although the decorrelation theory may be a valuable contribution to cryptographic research, it does not guarantee resistance against state-of-the-art differential attacks*" [92, p.94].

Since a model is essentially a simplified abstraction of reality, the argument about the significance of the results obtained within the one introduced by Luby and Rackoff is perfectly admissible. Yet we stress the fact that this debate would have nothing to do with the Decorrelation Theory which, in its basic form, is essentially a mean to compute or bound the advantage of a q-limited distinguisher in this model. One could possibly discuss the practicality of the tools introduced by this theory, but that would probably essentially be a matter of taste. Of course, we do not contest the validity of the attacks against ciphers which security is based on decorrelation results. Our argument is that it would be preferable to take advantage of the best of both worlds, namely, of the provable security aspects of the block ciphers based on decorrelation techniques and of the practical security aspects of ad-hoc constructions.

To do so, we suggest to avoid *algebraic* decorrelation modules, which surely provide perfect level of decorrelation up to a given order, but which security collapse after that. Instead, we propose to use typical building blocks which widespread ciphers are usually made of, and try to see what kind of results can be proved afterwards. Although this approach seems to correspond to the classical one, we will see that by bringing more randomness within these building blocks, one can point out symmetries within the distribution matrices that make it possible to actually *prove* several security results about the whole construction. We introduce these building blocks in Section 10.5.

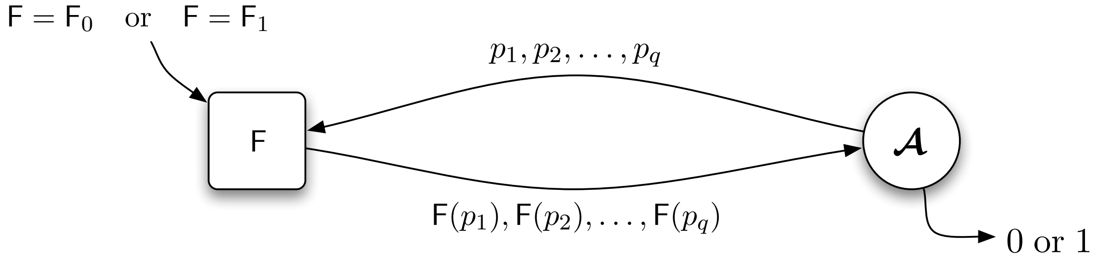

$\mathsf{F} = \mathsf{F}_0$ or $\mathsf{F} = \mathsf{F}_1$
$p_1, p_2, \ldots, p_q$
F
$\mathcal{A}$
$\mathsf{F}(p_1), \mathsf{F}(p_2), \ldots, \mathsf{F}(p_q)$
0 or 1

Figure 10.1: A q-limited non-adaptive distinguisher between $\mathsf{H}_0 : \mathsf{F} = \mathsf{F}_0$ and $\mathsf{H}_1 : \mathsf{F} = \mathsf{F}_1$

10.1 The Luby-Rackoff Model

In their seminal work, Luby and Rackoff showed how to construct a secure block cipher from a secure pseudo-random bit generator [102]. Their definition of *security* for a block cipher is the one we consider. Essentially, they assume that a block cipher is secure when no algorithm can distinguish between a black box implementing a random instance of the block cipher and a black box containing a random instance of the perfect cipher[1] by submitting (a limited number of) input strings and by looking at the outputs. In what follows, we give a more formal definition of this security notion and generalize it to random functions. We consider a game in which an adversary is given a black box access to either of these two functions, its objective being to guess whether it has access to F_0 or to F_1. This can be modelized as an hypothesis testing problem, where the two hypotheses are $\mathsf{H}_0 : \mathsf{F} = \mathsf{F}_0$ and $\mathsf{H}_1 : \mathsf{F} = \mathsf{F}_1$, in which the adversary is allowed to learn the value of the random function F in q points. The algorithm is assumed to be *computationally unbounded* (and therefore, we can assume it is deterministic) and only limited by the number of queries to the black box. When the q queries are made at once, the adversary is *non-adaptive* (see Figure 10.1). When the distinguisher is allowed to adaptively choose a query depending on the outcomes of the previous ones, it is *adaptive* (see Figure 10.2). Denoting p_i the ith query of the adversary and letting $Z_i = (p_i, \mathsf{F}(p_i))$ for $i = 1, 2, \ldots, q$, the advantage of an adversary A between H_0 and H_1 is

$$\mathrm{Adv}_{\mathsf{A}}(\mathsf{H}_0, \mathsf{H}_1) = \left|\Pr_{\mathsf{H}_0}[\mathsf{A}(\mathbf{Z}^q) = 1] - \Pr_{\mathsf{H}_1}[\mathsf{A}(\mathbf{Z}^q) = 1]\right|,$$

as in Definition 6.2, the probabilities holding over the random function F. In what follows we either write $\mathrm{Adv}_{\mathsf{A}}(\mathsf{H}_0, \mathsf{H}_1)$ or $\mathrm{Adv}_{\mathsf{A}}(\mathsf{F}_0, \mathsf{F}_1)$.

To evaluate the randomness of a pseudo-random function F, we assume in the previous game that F_0 is actually a uniformly distributed random function drawn among the $|\mathcal{Y}|^{|\mathcal{X}|}$ possible functions on the given sets and that F_1 is equal to F. In that setting, if the advantage of any *adaptive* distinguisher between both hypotheses is negligible, the function F is said to be *pseudorandom*.

[1] That is, a permutation drawn uniformly at random among all possible permutations on the given set.

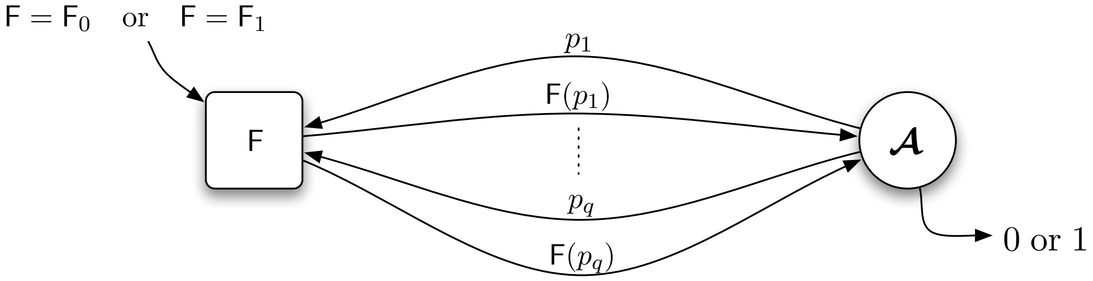


Figure 10.2: A q-limited adaptive distinguisher between $\mathsf{H}_0 : \mathsf{F} = \mathsf{F}_0$ and $\mathsf{H}_1 : \mathsf{F} = \mathsf{F}_1$

The security notions introduced for random functions are easy to adapt to random permutations (or block ciphers). Let $\mathcal{T}$ be a finite set and let $\mathsf{C}_0, \mathsf{C}_1 : \mathcal{T} \to \mathcal{T}$ be two random permutations on that set. We consider a game in which an adversary is given a black box access to either of these two permutations, its objective being to guess whether it has access to C_0 or to C_1. The corresponding two hypotheses are $\mathsf{H}_0 : \mathsf{C} = \mathsf{C}_0$ and $\mathsf{H}_1 : \mathsf{C} = \mathsf{C}_1$, the adversary being allowed to learn the values of the random permutation C in q points. In this case also, the adversary is assumed to be computationally unbounded. The notions of adaptive and non-adaptive distinguishers naturally apply to this setting and the definition of the advantage is unchanged.

We can apply the previous security notion to evaluate the security of a block cipher. Let

$$\mathsf{C} = \{\mathsf{C}_k : \mathcal{T} \to \mathcal{T} : k \in \mathcal{K}\}$$

be a block cipher on the text space $\mathcal{T}$ and the finite key space $\mathcal{K}$. Let $\mathsf{C}^\star$ be the perfect cipher on $\mathcal{T}$, that is, a permutation drawn uniformly at random among all $|\mathcal{T}|!$ permutations on $\mathcal{T}$. We say that C is secure against non-adaptive attacks when all non-adaptive distinguishers between C and $\mathsf{C}^\star$ have a negligible advantage. The block cipher is secure against adaptive attacks when this is also the case of the advantage of any adaptive adversary.

10.2 Computing the Advantage by means of Distribution Matrices

The distribution matrix of a random function or a random permutation is a fundamental notion of the Decorrelation Theory.

Definition 10.1 *Let q be a positive integer. Let $\mathcal{X}$ and $\mathcal{Y}$ be two finite sets and let $\mathsf{F} : \mathcal{X} \to \mathcal{Y}$ be a random function. The q-wise distribution matrix of F is the $|\mathcal{X}|^q \times |\mathcal{Y}|^q$ matrix $[\mathsf{F}]^q$ defined by*

$$[\mathsf{F}]^q_{(x_1,\ldots,x_q),(y_1,\ldots,y_q)} = \Pr_{\mathsf{F}}[\mathsf{F}(x_1) = y_1, \ldots, \mathsf{F}(x_q) = y_q] = \Pr_{\mathsf{F}}[(x_1,\ldots,x_q) \xrightarrow{\mathsf{F}} (y_1,\ldots,y_q)]$$

where $x_1, \ldots, x_q \in \mathcal{X}$ *and* $y_1, \ldots, y_q \in \mathcal{Y}$.

As an example, the 2-wise distribution matrix of the uniformly distributed random function $\mathsf{F}^\star : \{0,1\} \to \{0,1\}$ is the $2^2 \times 2^2$ matrix

$$[\mathsf{F}^\star]^2 = \begin{bmatrix} 1/2 & 0 & 0 & 1/2 \\ 1/4 & 1/4 & 1/4 & 1/4 \\ 1/4 & 1/4 & 1/4 & 1/4 \\ 1/2 & 0 & 0 & 1/2 \end{bmatrix}.$$

The 2-wise distribution matrix of the perfect cipher $\mathsf{C}^\star : \{0,1\} \to \{0,1\}$ is the $2^2 \times 2^2$ matrix

$$[\mathsf{C}^\star]^2 = \begin{bmatrix} 1/2 & 0 & 0 & 1/2 \\ 0 & 1/2 & 1/2 & 0 \\ 0 & 1/2 & 1/2 & 0 \\ 1/2 & 0 & 0 & 1/2 \end{bmatrix}.$$

Intuitively, the role of the distribution matrix of an arbitrary function F (resp. permutation C) is to evaluate to what extend the function (resp. permutation) behaves like its ideal counterpart $\mathsf{F}^\star$ (resp. $\mathsf{C}^\star$). In other words, if $[\mathsf{F}]^q$ looks just like $[\mathsf{F}^\star]^q$, then q queries won't be enough to distinguish F from a uniformly distributed random function (this is more formally stated later). Clearly, the information contained in the q-wise distribution matrix of a function F is also included in its $(q+1)$-wise distribution matrix, since

$$[\mathsf{F}]^q_{(x_1,\ldots,x_q),(y_1,\ldots,y_q)} = \sum_{x_{q+1}\in\mathcal{X}} [\mathsf{F}]^{q+1}_{(x_1,\ldots,x_{q+1}),(y_1,\ldots,y_{q+1})} = \sum_{y_{q+1}\in\mathcal{Y}} [\mathsf{F}]^{q+1}_{(x_1,\ldots,x_{q+1}),(y_1,\ldots,y_{q+1})}.$$

Therefore, it is clear that there might be huge gap between the best $(q+1)$-limited distinguisher between F and $\mathsf{F}^\star$ and the best q-limited distinguisher.

Example 10.1 Let $\mathsf{F} : \{0,1\} \to \{0,1\}$ be a random function which is either $f_0 : \{0,1\} \to \{0,1\}$ or $f_1 : \{0,1\} \to \{0,1\}$ with equal probability, where $f_b(x) = b$ for all x. The 2-wise distribution matrix of a random function F is

$$[\mathsf{F}]^2 = \begin{bmatrix} 1/2 & 0 & 0 & 1/2 \\ 1/2 & 0 & 0 & 1/2 \\ 1/2 & 0 & 0 & 1/2 \\ 1/2 & 0 & 0 & 1/2 \end{bmatrix}.$$

It is easy to distinguish it from $\mathsf{F}^\star$ (with a high advantage) by simply asking two distinct queries and checking whether both answers are distinct. If they are, the black box cannot be implementing F. The advantage of the distinguisher following this strategy is $\frac{1}{2}$. Clearly, the 1-wise distribution matrix of F is

$$[\mathsf{F}]^1 = \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix},$$

which allows to clearly see that F cannot be distinguished from $\mathsf{F}^\star$ with one query only, i.e., the advantage of any distinguisher limited to one query is zero. □

This suggests that the *distance* between the respective distribution matrices of two random functions (or permutations) might be a good measure of how distinct these two functions (or permutations) are. This is formally stated in the following theorem (which corresponds to theorems 10 and 11 in [155]). Note that although the theorem is stated for random functions here, it also applies to random permutations.

Definition 10.2 *Let q be a positive integer. Let $\mathcal{X}$ and $\mathcal{Y}$ be two finite sets and let A be a $|\mathcal{X}|^q \times |\mathcal{Y}|^q$ matrix indexed by q-tuples $(x,y) = ((x_1,\ldots,x_q),(y_1,\ldots,y_q)) \in \mathcal{X}^q \times \mathcal{Y}^q$. We define*

$$|||A|||_\infty = \max_x \sum_y |A_{x,y}|$$

and

$$\|A\|_{\mathrm{a}} = \max_{x_1} \sum_{y_1} \cdots \max_{x_q} \sum_{y_q} |A_{x,y}|\,.$$

Theorem 10.1 *Let q be a positive integer. Let $\mathcal{X}$ and $\mathcal{Y}$ be two finite sets and $\mathsf{F}_0,\mathsf{F}_1 : \mathcal{X} \to \mathcal{Y}$ be two random functions. The advantage of the best q-limited non-adaptive distinguisher A_{na} between $\mathsf{H}_0 : \mathsf{F} = \mathsf{F}_0$ and $\mathsf{H}_1 : \mathsf{F} = \mathsf{F}_1$ is such that*

$$\mathrm{Adv}_{\mathsf{A}_{\mathrm{na}}}(\mathsf{H}_0,\mathsf{H}_1) = \frac{1}{2}|||[\mathsf{F}_1]^q - [\mathsf{F}_0]^q|||_\infty.$$

The advantage of the best q-limited adaptive distinguisher A_{a} between $\mathsf{H}_0 : \mathsf{F} = \mathsf{F}_0$ and $\mathsf{H}_1 : \mathsf{F} = \mathsf{F}_1$ is such that

$$\mathrm{Adv}_{\mathsf{A}_{\mathrm{a}}}(\mathsf{H}_0,\mathsf{H}_1) = \frac{1}{2}\|[\mathsf{F}_1]^q - [\mathsf{F}_0]^q\|_{\mathrm{a}}.$$

Example 10.2 If we re-consider the random function of Example 10.1, we can see that

$$[\mathsf{F}]^2 - [\mathsf{F}^\star]^2 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1/4 & -1/4 & -1/4 & 1/4 \\ 1/4 & -1/4 & -1/4 & 1/4 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

and thus

$$\frac{1}{2}|||[\mathsf{F}]^2 - [\mathsf{F}^\star]^2|||_\infty = \frac{1}{2}\|[\mathsf{F}]^2 - [\mathsf{F}^\star]^2\|_{\mathrm{a}} = \frac{1}{2}.$$

□

Example 10.3 As an application of Theorem 10.1, one can compute the RF/RP-advantage [140], which is the advantage of the best distinguisher between $\mathsf{F}^\star : \mathcal{T} \to \mathcal{T}$

and $\mathsf{C}^\star : \mathcal{T} \to \mathcal{T}$. Applying decorrelation techniques to prove this lemma was originally suggested by Junod [74]. We let $N = |\mathcal{T}|$. Obviously, we can assume that the q queries $x_1, x_2, \ldots, x_q$ made by the best distinguisher are distinct, since asking the same query twice cannot increase its advantage. In that case

$$[\mathsf{F}^\star]^q_{(x_1,\ldots,x_q),(y_1,\ldots,y_q)} = \Pr_{\mathsf{F}^\star}[\cap_{i=1}^q \mathsf{F}^\star(x_i) = y_i] = \prod_{i=1}^q \Pr_{\mathsf{F}^\star}[\mathsf{F}^\star(x_i) = y_i] = \frac{1}{N^q}.$$

and

$$[\mathsf{C}^\star]^q_{(x_1,\ldots,x_q),(y_1,\ldots,y_q)} = \Pr_{\mathsf{C}^\star}[\cap_{i=1}^q \mathsf{C}^\star(x_i) = y_i] = \begin{cases} 0 & \text{if } y_i = y_j \text{ for some } i \neq j, \\ \frac{(N-q)!}{N!} & \text{otherwise.} \end{cases}$$

From the two previous equations, one can see that the difference $[\mathsf{F}^\star]^q_{(x_1,\ldots,x_q),(y_1,\ldots,y_q)} - [\mathsf{C}^\star]^q_{(x_1,\ldots,x_q),(y_1,\ldots,y_q)}$ does *not* depend on the particular choice of the q inputs (as long as they are distinct). This immediately allows to conclude that the best strategy is simply to choose q distinct queries, which can be done at once, so that the best adaptive adversary does not present any advantage compared to the best non-adaptive one. Simply denoting A the best distinguisher, letting $x_1, \ldots, x_q \in \mathcal{T}$ be q distinct elements, and denoting

$$D = N(N-1)\cdots(N-q+1)$$

the number of strings of q distinct elements of $\mathcal{T}$, we have

$$\begin{aligned} \mathrm{Adv}_{\mathsf{A}}(\mathsf{F}^\star, \mathsf{C}^\star) &= \sum_{y_1,\ldots,y_q} \left| [\mathsf{F}^\star]^q_{(x_1,\ldots,x_q),(y_1,\ldots,y_q)} - [\mathsf{C}^\star]^q_{(x_1,\ldots,x_q),(y_1,\ldots,y_q)} \right| \\ &= D \left| \frac{1}{N^q} - \frac{(N-q)!}{N!} \right| + (N^q - D)\frac{1}{N^q} \\ &= 1 - \frac{1}{N^q}\frac{N!}{(N-q)!}. \end{aligned}$$

We derived the exact advantage of the best distinguisher between $\mathsf{F}^\star$ and $\mathsf{C}^\star$. Letting $q = \theta\sqrt{N}$, the previous equation leads to (see [157, p.71] for example)

$$\mathrm{Adv}_{\mathsf{A}}(\mathsf{F}^\star, \mathsf{C}^\star) = 1 - \frac{N!}{N^{\theta\sqrt{N}}(N - \theta\sqrt{N})!} \xrightarrow{N\to\infty} 1 - e^{-\theta^2/2}.$$

This shows that one can distinguish $\mathsf{F}^\star$ from $\mathsf{C}^\star$ with a high advantage when the number of queries q is of the order of magnitude of $\sqrt{N}$. □

In the scope of provable security of block cipher, one is essentially interested in computing the advantage of the best (non-)adaptive distinguisher between the block cipher C considered and the perfect cipher $\mathsf{C}^\star$. Theorem 10.1 provides a neat way to do so, provided that the q-wise distribution matrix of C can be computed. In the rest of this section, we recall several essential properties about distribution matrices and about their norms, which we will extensively use in the following sections to prove security results on real block cipher constructions. These results are either trivial or proved in [155].

Lemma 10.1 *Let q be a positive integer and let $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$ be three finite sets. Let $\mathsf{F}_0 : \mathcal{X} \to \mathcal{Y}$ and $\mathsf{F}_1 : \mathcal{Y} \to \mathcal{Z}$ be two independent random functions. Then*

$$[\mathsf{F}_1 \circ \mathsf{F}_0]^q = [\mathsf{F}_0]^q \times [\mathsf{F}_1]^q.$$

Lemma 10.1 will be particularly helpful for computing the distribution matrix of a product cipher, based on the distribution matrices of each individual rounds (assuming that the round keys are mutually independent).

Lemma 10.2 *Let q be a positive integer and let $\mathcal{T}$ be a finite set. Let $\mathsf{C} : \mathcal{T} \to \mathcal{T}$ be a random permutation. We have*

$$[\mathsf{C} \circ \mathsf{C}^\star]^q = [\mathsf{C}^\star \circ \mathsf{C}]^q = [\mathsf{C}^\star]^q.$$

Note that the previous lemma is not true when considering random *functions*. Indeed, consider for example the function $f : \{0,1\} \to \{0,1\}$ such that $f(x) = 0$ for all x, then $(f \circ \mathsf{F}^\star)(x) = 0$ for all x whereas $(\mathsf{F}^\star \circ f)(x)$ is a random value.

Based on Lemma 10.2 and on the fact that both $|||\cdot|||_\infty$ and $\|\cdot\|_\mathrm{a}$ are matrix norms (i.e., such that $\|A \times B\| \leq \|A\| \times \|B\|$), it is easy to prove the following theorem (see Theorem 4 in [155]).

Theorem 10.2 *Let q be a positive integer, let $\mathcal{T}$ be a finite set and $\mathsf{C}_1, \ldots, \mathsf{C}_r$ be r mutually independent random permutations on $\mathcal{T}$. Let $\mathsf{C} = \mathsf{C}_r \circ \cdots \circ \mathsf{C}_1$ and $\mathsf{C}^\star$ be the perfect cipher over $\mathcal{T}$. Letting $\|\cdot\|$ be either $|||\cdot|||_\infty$ or $\|\cdot\|_\mathrm{a}$, we have*

$$\|[\mathsf{C}]^q - [\mathsf{C}^\star]^q\| \leq \prod_{i=1}^{r} \|[\mathsf{C}_i]^q - [\mathsf{C}^\star]^q\|.$$

Theorem 10.2 is essential when studying product ciphers. For example, given a cipher which iterates the same round r times (with mutually independent round keys), it is usually sufficient to derive the distribution matrix of one round, compute the norm of the difference between this matrix and that of the perfect cipher, and raise the result to the power r to get (using Theorem 10.1) a practical upper bound on the advantage of the best (non-)adaptive distinguisher on the whole construction.

```
$c \leftarrow 0$
for $t = 1, 2, \ldots, q$ do
  $P \leftarrow \{0,1\}^n$ and $C \leftarrow \mathsf{c}(P)$
  if $a \bullet P = b \bullet C$ then $c \leftarrow c + 1$ endif
done
if $\left|c - \frac{q}{2}\right| > T$ then output 1 else output 0
```

Algorithm 10.1: A q-limited linear distinguisher with oracle access to a permutation c on $\{0,1\}^n$, based on the input/output masks $a, b \in \{0,1\}^n \setminus \{0\}$ and the threshold T

10.3 From Linear Cryptanalysis and Differential Cryptanalysis to other Iterated Attacks

Linear cryptanalysis is an attack proposed by Matsui [110, 111] based on previous ideas from Tardy-Corfdir and Gilbert [147]. It is a projection based attack (see Chapter 8) that applies to block ciphers defined on bit strings, in which the adversary linearly derives one bit of information from each plaintext/ciphertext pair available. Algorithm 10.1 gives a general description of a q-limited linear distinguisher.

Based on Heuristic 8.2, we know that the data complexity of an effective (classical) linear distinguisher between the block cipher C and the perfect cipher $\mathsf{C}^\star$ should be at least close to

$$\frac{8 \ln 2}{\mathrm{ELP}_{a,b}(\mathsf{C})}$$

where $a, b \in \{0,1\}^n \setminus \{0\}$ are the input/output masks used to derive the bit from each plaintext/ciphertext pair, and where (see also Definition 8.6)

$$\mathrm{ELP}_{a,b}(\mathsf{C}) = \mathrm{E}_{\mathsf{C}}\left(\mathrm{LP}_{a,b}(\mathsf{C})\right) \quad \text{with} \quad \mathrm{LP}_{a,b}(\mathsf{c}) = \left(2\mathrm{Pr}_P[a \bullet P = b \bullet \mathsf{c}(P)] - 1\right)^2,$$

the random variable $P \in \{0,1\}^n$ being uniformly distributed. In [155], Vaudenay shows that there is a link between the expected linear probability of a block cipher C and its 2-wise distribution matrix, namely,

$$\left|\mathrm{ELP}_{a,b}(\mathsf{C}) - \frac{1}{2^n - 1}\right| = \left|\mathrm{ELP}_{a,b}(\mathsf{C}) - \mathrm{ELP}_{a,b}(\mathsf{C}^\star)\right| \leq |||[\mathsf{C}]^2 - [\mathsf{C}^\star]^2|||_\infty. \tag{10.1}$$

Through Theorem 10.1, this shows that upper-bounding the advantage of the best 2-limited non-adaptive adversary against C by a negligible value allows to prove that C is immune against linear cryptanalysis.

Differential cryptanalysis [21, 23, 24] is a chosen plaintext attack introduced by Biham and Shamir. It works by randomly selecting *pairs* of plaintexts having a chosen fixed difference, asking the corresponding ciphertexts and checking whether their

```
for $t = 1, 2, \ldots, q$ do
    $P \leftarrow \{0,1\}^n$
    if $\mathsf{c}(P \oplus a) = \mathsf{c}(P) \oplus b$ then output 1 endif
done
output 0
```

Algorithm 10.2: A q-limited differential distinguisher with oracle access to a permutation c on $\{0,1\}^n$, based on the input/output differences $a, b \in \{0,1\}^n \setminus \{0\}$.

difference is equal to particular chosen value. Algorithm 10.2 gives a general description of a q-limited differential distinguisher.

It is a well accepted fact that the data complexity of an effective differential distinguisher between C and $\mathsf{C}^\star$ should at least be close to

$$\frac{1}{\mathrm{EDP}_{a,b}(\mathsf{C})},$$

where $a, b \in \{0,1\}^n \setminus \{0\}$ are the input/output differences and where

$$\mathrm{EDP}_{a,b}(\mathsf{C}) = \mathrm{E}_{\mathsf{C}}\left(\mathrm{DP}_{a,b}(\mathsf{C})\right) \quad \text{with} \quad \mathrm{DP}_{a,b}(\mathsf{c}) = \Pr_P[\mathsf{c}(P \oplus a) = \mathsf{c}(P) \oplus b],$$

the random variable $P \in \{0,1\}^n$ being uniformly distributed [124]. Similarly than what we have in (10.1), Vaudenay shows in [155] that

$$\left|\mathrm{EDP}_{\mathsf{C}}(a,b) - \frac{1}{2^n - 1}\right| = |\mathrm{EDP}_{\mathsf{C}}(a,b) - \mathrm{EDP}_{\mathsf{C}^\star}(a,b)| \leq \frac{1}{2}|||[\mathsf{C}]^2 - [\mathsf{C}^\star]^2|||_\infty. \qquad (10.2)$$

Theorem 10.1 allows to conclude that upper-bounding the advantage of any 2-limited non-adaptive adversary against C by some negligible value allows to conclude that C is secure against differential cryptanalysis.

In certain circumstances, bounding the advantage of the best 2-limited non-adaptive adversary also allows to prove some resistance to *iterated attacks* [154, 155], which formalize a large class of attacks against block ciphers, including linear and differential cryptanalysis. In an iterated attack of order d, the adversary is given a sample $(P, C) = ((P_1, P_2, \ldots, P_d), (C_1, C_2, \ldots, C_d))$, where $C_i = \mathsf{C}(P_i)$ for $i = 1, 2, \ldots, d$, and keeps one bit of information denoted $\mathcal{F}(P, C)$. After q iterations, the distinguisher decides which hypothesis is most likely, based on the q bits. The q samples are assumed to be mutually independent and identically distributed. Algorithm 10.3 describes a iterated distinguisher of order d.

It is easy to see that linear cryptanalysis is an iterated attack of order 1: let $d = 1$, $\mathcal{T} = \{0,1\}^n$, assume that the distribution of P_1 is uniform and define the increment rule as $\mathcal{D} = \{(p, c) \in \{0,1\}^n \times \{0,1\}^n : a \bullet p = b \bullet c\}$. Letting $\mathcal{A} = \{c \in \{0, 1, \ldots, q\} : \left|c - \frac{q}{2}\right| > T\}$ leads to a linear distinguisher based on the input/output masks $a, b \in \{0,1\}^n$ and threshold T.

```
$c \leftarrow 0$
for $t = 1, 2, \ldots, q$ do
    $P = (P_1, P_2, \ldots, P_d) \leftarrow \mathcal{T}^d$ and $C = (C_1, C_2, \ldots, C_d) \leftarrow (\mathsf{c}(P_1), \mathsf{c}(P_2), \ldots, \mathsf{c}(P_d))$
    if $(P, C) \in \mathcal{D}$ then $c \leftarrow c + 1$ endif
done
if $c \in \mathcal{A}$ then output 1 else output 0 endif
```

Algorithm 10.3: A q-limited iterated distinguisher of order d with oracle access to a permutation c on the finite set $\mathcal{T}$, with an increment rule $\mathcal{D} \subset \mathcal{T}^d \times \mathcal{T}^d$ and an acceptance region $\mathcal{A} \subset \{0, 1, \ldots, q\}$.

Similarly, differential cryptanalysis is an iterated attack of order 2: let $d = 2$, $\mathcal{T} = \{0,1\}^n$ (for simplicity we stick to binary ciphers, although this is not mandatory here). Let the increment rule be

$$\mathcal{D} = \{(p_1, p_2, c_1, c_2) \in \{0,1\}^{4n} : p_1 \oplus p_2 = a \text{ and } c_1 \oplus c_2 = b\}$$

for some $a, b \in \{0,1\}^n \setminus \{0\}$ and assume that the distribution of P_1 is uniform and that $P_2 = P_1 \oplus a$. Letting finally $\mathcal{A} = \{1, 2. \ldots, q\}$ (i.e., the distinguisher outputs 1 whenever the counter has been incremented, that is, whenever $\mathsf{c}(P_1 \oplus a) = \mathsf{c}(P_1) \oplus b$ for one of the q random values of P_1) we obtain a differential distinguisher based on the input/output differences a, b.

Unlike linear and differential cryptanalysis, bounding the advantage of the best d-limited non-adaptive distinguisher is not sufficient in general to provide security against iterated attacks of order d, Vaudenay provides a counter example in [155]. However he shows that bounding the advantage of the best $2d$-limited non-adaptive adversary can be sufficient.

Theorem 10.3 *(Theorem 18 in [155]) Let $\mathsf{C} : \mathcal{T} \to \mathcal{T}$ be a block cipher such that*

$$|||[\mathsf{C}]^{2d} - [\mathsf{C}^\star]^{2d}|||_\infty \le \epsilon$$

for some $d \le \frac{|\mathcal{T}|}{2}$ and $\epsilon > 0$, where $\mathsf{C}^\star$ is the perfect cipher on $\mathcal{T}$. Let q be a positive integer. The advantage Adv *of the best q-limited iterated distinguisher of order d between C and $\mathsf{C}^\star$ is such that*

$$\mathrm{Adv} \le 5\sqrt[3]{\left(2\delta + \frac{5d^2}{2\,|\mathcal{T}|} + \frac{3\epsilon}{2}\right) q^2} + q\epsilon,$$

where δ is the probability that any two different iterations send at least one query in common.

Note that the bound given by the previous theorem is meaningful only if δ is not too large, which can only occur if $\mathcal{T}$ is large enough. In what follows, we assume that whenever ϵ is negligible, then the block cipher is immune against iterated attacks of order d. For the particular case of iterated attacks of order 1, we easily obtain the following corollary.

Corollary 10.1 *Let $\mathsf{C} : \mathcal{T} \to \mathcal{T}$ be a block cipher such that the advantage of the best 2-limited non-adaptive distinguisher between C and $\mathsf{C}^\star$ is upper-bounded by $\epsilon > 0$, where $\mathsf{C}^\star$ is the perfect cipher on $\mathcal{T}$. Let q be a positive integer. The advantage* Adv *of the best q-limited iterated distinguisher of order* 1 *between C and $\mathsf{C}^\star$ is such that*

$$\mathrm{Adv} \leq 9\sqrt[3]{\left(\frac{1}{|\mathcal{T}|} + \epsilon\right) q^2} + 2q\epsilon,$$

Proof. Noting that $\delta = \frac{1}{|\mathcal{T}|}$ in this case easily leads to the announced result. □

The previous lemma shows that if $\epsilon \approx \frac{1}{|\mathcal{T}|}$, then no q-limited iterated distinguisher of order 1 can efficiently distinguish the block cipher from the perfect cipher when q is negligible compared to $\sqrt{|\mathcal{T}|}$.

10.4 Decorrelation of Feistel Ciphers

A r-rounds Feistel scheme [50] is a construction that turns r random functions

$$\mathsf{F}_1, \mathsf{F}_2, \ldots, \mathsf{F}_r : \mathcal{T} \to \mathcal{T}$$

(where $\mathcal{T}$ is some finite set) into a random permutation

$$\Psi(\mathsf{F}_1, \mathsf{F}_2, \ldots, \mathsf{F}_r) : \mathcal{T}^2 \to \mathcal{T}^2$$

as shown on Figure 10.3. It is easy to see that this defines a permutation since

$$\Psi^{-1}(\mathsf{F}_1, \mathsf{F}_2, \ldots, \mathsf{F}_r) = \Psi(\mathsf{F}_r, \mathsf{F}_{r-1}, \ldots, \mathsf{F}_1).$$

A Feistel cipher is a block cipher based on a Feistel scheme, the best known example being probably the DES [122] which is based on a 16-rounds Feistel scheme. We recall here a famous result by Luby and Rackoff about 3-rounds Feistel schemes.

Theorem 10.4 ***(Theorem 1 in [102])*** *Let n be a positive integer and let $\mathsf{F}_1^\star, \mathsf{F}_2^\star, \mathsf{F}_3^\star : \{0,1\}^n \to \{0,1\}^n$ be three independent and uniformly distributed random functions. Let q be a positive integer. The advantage of the best q-limited adaptive distinguisher between $\Psi(\mathsf{F}_1^\star, \mathsf{F}_2^\star, \mathsf{F}_3^\star)$ and the perfect cipher $\mathsf{C}^\star$ on $\{0,1\}^{2n}$ is upper-bounded by $q^2 \cdot 2^{-n}$.*

This result generalizes to r-rounds Feistel schemes as follows.

Theorem 10.5 ***(Theorem 21 in [155])*** *Let n and q be two positive integers. Let $\mathsf{F}^\star$ be a uniformly distributed random function on $\{0,1\}^n$. Let $\mathsf{F}_1, \mathsf{F}_2, \ldots, \mathsf{F}_r$ be r independent*

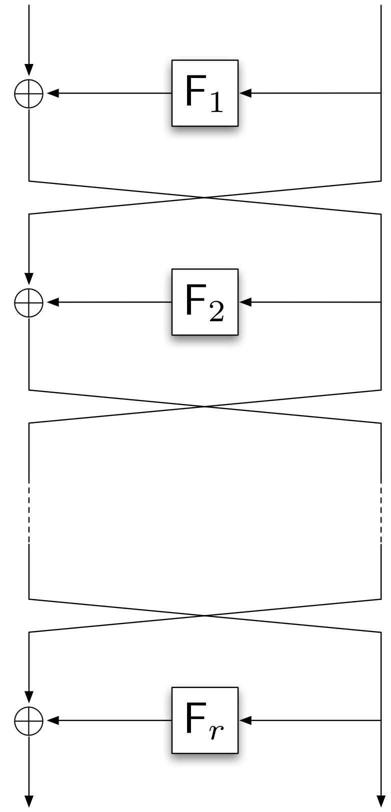


Figure 10.3: An r rounds Feistel scheme $\Psi(\mathsf{F}_1, \mathsf{F}_2, \dots, \mathsf{F}_r)$

random functions on $\{0,1\}^n$ *such that*

$$\mathrm{Adv}_{\mathsf{A}_q}(\mathsf{F}_i, \mathsf{F}^\star) \leq \epsilon$$

for $i = 1, 2, \dots, r$ *and for any* q*-limited adversary* A_q*. Let* $\mathsf{C} = \Psi(\mathsf{F}_1, \mathsf{F}_2, \dots, \mathsf{F}_r)$ *be an* r*-rounds Feistel cipher on* $\{0,1\}^{2n}$ *and let* $\mathsf{C}^\star$ *denote the perfect cipher on the same set. For any* q*-limited adversary* A_q *and for any integer* $k \geq 3$ *we have*

$$\mathrm{Adv}_{\mathsf{A}_q}(\mathsf{C}, \mathsf{C}^\star) \leq \frac{1}{2}\left(2k\epsilon + \frac{2q^2}{2^n}\right)^{\lfloor r/k \rfloor}.$$

If we let $\mathsf{F}_i = \mathsf{F}^\star$ for all $i = 1, 2, \dots, r$ in the previous theorem, we obtain $\epsilon = 0$. Setting $r = k = 3$ allows to fall back on the Luby and Rackoff result.

This theorem shows that if we can instantiate *independent* random functions secure against all q-limited distinguishers, we can obtain a block cipher provably secure against any q-limited distinguisher.

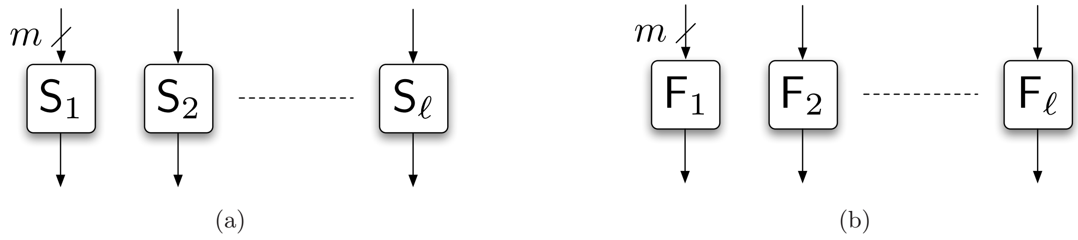


Figure 10.4: Layers of ℓ bijective random S-boxes and of ℓ random F-boxes on m-bit strings

10.5 Decorrelation Modules: Avoiding Algebraic Constructions

In this section we will study the main building blocks that we will use in the secure constructions that we will introduce in chapters 11 and 12. As we will see, these building blocks are particularly well suited to prove results against 2-limited adversaries. Throughout this section, we will be using the following definitions and notations about arrays of bit strings.

Definition 10.3 *Let $a = (a_1, a_2, \ldots, a_\ell)$ be an array of m-bit strings. The* support *of a is the array of $\{0,1\}^\ell$ with 0's at the positions where the entry of a is equal to zero and with 1's where the entry of a is non-zero. We denote the support of a by $\textsc{supp}(a)$. The support of a is said to be included in the support of b when $a_i \neq 0 \Rightarrow b_i \neq 0$ for all $i = 1, 2, \ldots, \ell$. This is denoted $\textsc{supp}(a) \subseteq \textsc{supp}(b)$. The Hamming weight of a (or of $\textsc{supp}(a)$) is the number of 1's of $\textsc{supp}(a)$. We denote this weight by $w(a)$. When $w(a) = \ell$, it means that all the entries of a are non-zero, in which case we say that a is of* full support.

Layer of S-Boxes

We consider a layer made of ℓ mutually independent and uniformly distributed permutations on $\{0,1\}^m$, arranged side by side. This situation is represented on Figure 10.4(a). These permutations are called *substitution boxes* and we denote them by S_1 through S_ℓ. The complete layer is denoted S and is a random permutation defined on $\{0,1\}^{m\ell}$. Let $M = 2^m$. For any of the ℓ uniformly distributed random substitution boxes S_i and any $a_i, a'_i, b_i, b'_i \in \{0,1\}^m$, we have

$$\Pr[\mathsf{S}_i(a_i) = b_i, \mathsf{S}_i(a'_i) = b'_i] = \begin{cases} \frac{1}{M} & \text{if } a_i = a'_i \text{ and } b_i = b'_i, \\ \frac{1}{M(M-1)} & \text{if } a_i \neq a'_i \text{ and } b_i \neq b'_i, \\ 0 & \text{otherwise.} \end{cases}$$

The last equation can be also written in a more compact form, namely,

$$\Pr[\mathsf{S}_i(a_i) = b_i, \mathsf{S}_i(a'_i) = b'_i] = \mathbf{1}_{\text{SUPP}(a_i \oplus a'_i) = \text{SUPP}(b_i \oplus b'_i)} M^{-1}(M-1)^{-w(a_i \oplus a'_i)}. \qquad (10.3)$$

Letting $a = (a_i)_i$, $a' = (a'_i)_i$, $b = (b_i)_i$, and $b' = (b'_i)_i$ with $i = 1, 2, \ldots, \ell$ and $a_i, a'_i, b_i, b'_i \in \{0,1\}^m$ for all i, we have

$$[\mathsf{S}]^2_{(a,a'),(b,b')} = \Pr\left[\cap_{i=1}^{\ell} (a_i, a'_i) \xrightarrow{\mathsf{S}_i} (b_i, b'_i)\right] = \prod_{i=1}^{\ell} \Pr[\mathsf{S}_i(a_i) = b_i, \mathsf{S}_i(a'_i) = b'_i], \qquad (10.4)$$

as the ℓ substitution boxes are assumed to be independent. Equations (10.3) and (10.4) lead to

$$[\mathsf{S}]^2_{(a,a'),(b,b')} = \mathbf{1}_{\text{SUPP}(a \oplus a') = \text{SUPP}(b \oplus b')} M^{-\ell}(M-1)^{-w(a \oplus a')}. \qquad (10.5)$$

Layer of F-Boxes

We consider a layer made of ℓ mutually independent and uniformly distributed functions on $\{0,1\}^m$, arranged side by side. This situation is represented on Figure 10.4(b). These functions are called F-Boxes and we denote them by F_1 through F_ℓ. The complete layer is denoted F and is a random function defined on $\{0,1\}^{m\ell}$. Let $M = 2^m$. For any of the ℓ uniformly distributed random substitution box S_i and any $a_i, a'_i, b_i, b'_i \in \{0,1\}^m$, we have

$$\Pr[\mathsf{F}_i(a_i) = b_i, \mathsf{F}_i(a'_i) = b'_i] = \begin{cases} \frac{1}{M^2} & \text{if } a_i \neq a'_i, \\ \frac{1}{M} & \text{if } a_i = a'_i \text{ and } b_i = b'_i, \\ 0 & \text{otherwise.} \end{cases}$$

In a more compact form, this reads

$$\Pr[\mathsf{F}_i(a_i) = b_i, \mathsf{F}_i(a'_i) = b'_i] = \mathbf{1}_{\text{SUPP}(b_i \oplus b'_i) \subseteq \text{SUPP}(a_i \oplus a'_i)} M^{-1-w(a_i \oplus a'_i)},$$

using the notations of Definition 10.3. From the last equation and the fact that the F-boxes are assumed to be independent, we conclude that

$$[\mathsf{F}]^2_{(a,a'),(b,b')} = \mathbf{1}_{\text{SUPP}(b \oplus b') \subseteq \text{SUPP}(a \oplus a')} M^{-\ell-w(a \oplus a')}, \qquad (10.6)$$

where $a = (a_i)_i$, $a' = (a'_i)_i$, $b = (b_i)_i$, and $b' = (b'_i)_i$ with $i = 1, 2, \ldots, \ell$ and $a_i, a'_i, b_i, b'_i \in \{0,1\}^m$ for all i.

Transition Matrices: Pair of Texts ↔ Support of Pair

From (10.5) and (10.6), one can see that the 2-wise distribution matrices of both S and F-box layers only depend on the supports of the exclusive-or of the inputs and of the support of the exclusive-or of the outputs. To take advantage of this fact in futures computations, we will introduces two *transition matrices*, that we denote PS and SP, which respectively map pair of texts to support of pair and the converse, in a uniform way. When considering arrays of ℓ strings in $\{0,1\}^m$ (as in both previous examples), we let PS be the $2^{2m\ell} \times 2^\ell$ matrix defined by

$$\mathsf{PS}_{(a,a'),\gamma} = \mathbf{1}_{\gamma=\textsc{supp}(a\oplus a')} \tag{10.7}$$

for all $a = (a_i)_i$ (resp. $a' = (a'_i)_i$) with $i = 1, 2, \ldots, \ell$ and $a_i, a'_i \in \{0,1\}^m$ for all i, and all $\gamma \in \{0,1\}^\ell$. Similarly, we let SP be the $2^\ell \times 2^{2m\ell}$ matrix defined by

$$\mathsf{SP}_{\gamma,(a,a')} = \mathbf{1}_{\gamma=\textsc{supp}(a\oplus a')} M^{-\ell}(M-1)^{-w(\gamma)}, \tag{10.8}$$

where $M = 2^m$.

Lemma 10.3 *The transition matrices* SP *and* PS *are such that*

$$\mathsf{SP} \times \mathsf{PS} = \mathsf{Id} \quad \textit{and} \quad \mathsf{PS} \times \mathsf{SP} = [\mathsf{S}]^2,$$

where $[\mathsf{S}]^2$ *is the* 2*-wise distribution matrix of a layer of* S*-boxes as in* (10.5)*.*

Proof. We note that for all $\gamma, \gamma' \in \{0,1\}^\ell$ we have

$$\begin{aligned}
(\mathsf{SP} \times \mathsf{PS})_{\gamma,\gamma'} &= M^{-\ell}(M-1)^{-w(\gamma)} \sum_{a,a'} \mathbf{1}_{\gamma=\textsc{supp}(a\oplus a')} \mathbf{1}_{\gamma'=\textsc{supp}(a\oplus a')} \\
&= \mathbf{1}_{\gamma=\gamma'} M^{-\ell}(M-1)^{-w(\gamma)} \sum_{a,a'} \mathbf{1}_{\gamma=\textsc{supp}(a\oplus a')} \\
&= \mathbf{1}_{\gamma=\gamma'} (M-1)^{-w(\gamma)} \sum_{a} \mathbf{1}_{\gamma=\textsc{supp}(a)} \\
&= \mathbf{1}_{\gamma=\gamma'},
\end{aligned}$$

which proves the first equality. For all $a, a', b, b' \in \{0,1\}^{m\ell}$ we have

$$\begin{aligned}
(\mathsf{PS} \times \mathsf{SP})_{(a,a'),(b,b')} &= M^{-\ell} \sum_{\gamma} \mathbf{1}_{\gamma=\textsc{supp}(a\oplus a')} \mathbf{1}_{\gamma=\textsc{supp}(b\oplus b')} (M-1)^{-w(\gamma)} \\
&= \mathbf{1}_{\textsc{supp}(a\oplus a')=\textsc{supp}(b\oplus b')} M^{-\ell}(M-1)^{-w(a\oplus a')},
\end{aligned}$$

which exactly corresponds to the expression of $[\mathsf{S}]^2_{(a,a'),(b,b')}$ obtained in (10.5). □

The following lemma shows how the transition matrices SP and PS apply when considering a layer of F-boxes.

Lemma 10.4 *Using the notations of this section, letting* $\overline{\mathsf{F}}$ *be the* $2^\ell \times 2^\ell$ *matrix indexed by supports, defined by*

$$\overline{\mathsf{F}}_{\gamma,\gamma'} = \mathbf{1}_{\gamma' \subseteq \gamma} M^{-w(\gamma)}(M-1)^{w(\gamma')},$$

we obtain

$$[\mathsf{F}]^2 = \mathsf{PS} \times \overline{\mathsf{F}} \times \mathsf{SP}.$$

Proof. Starting from the expression of $[\mathsf{F}]^2$ in (10.6), we have

$$\begin{aligned}
[\mathsf{F}]^2_{(a,a'),(b,b')} &= \mathbf{1}_{\text{SUPP}(b\oplus b')\subseteq\text{SUPP}(a\oplus a')}M^{-\ell-w(a\oplus a')}\sum_{\gamma,\gamma'}\mathbf{1}_{\gamma=\text{SUPP}(a\oplus a')}\mathbf{1}_{\gamma'=\text{SUPP}(b\oplus b')}\\
&= \sum_{\gamma,\gamma'}\mathbf{1}_{\gamma=\text{SUPP}(a\oplus a')}\mathbf{1}_{\gamma'\subseteq\gamma}M^{-\ell-w(\gamma)}\mathbf{1}_{\gamma'=\text{SUPP}(b\oplus b')}\\
&= \sum_{\gamma,\gamma'}\mathsf{PS}_{(a,a'),\gamma}\mathbf{1}_{\gamma'\subseteq\gamma}M^{-w(\gamma)}(M-1)^{w(\gamma')}\mathsf{SP}_{\gamma',(b,b')}\\
&= \sum_{\gamma,\gamma'}\mathsf{PS}_{(a,a'),\gamma}\overline{\mathsf{F}}_{\gamma,\gamma'}\mathsf{SP}_{\gamma',(b,b')},
\end{aligned}$$

which allows to conclude. □

We conclude the list of fundamental properties of the transition matrices SP and PS by the following lemma, which shows how these matrices will allow us to drastically reduce the complexity of computations in the constructions' security proofs that we will present in later chapters.

Lemma 10.5 *Let* M *be a* $2^{2m\ell}\times 2^{2m\ell}$ *matrix indexed by pairs of* ℓ*-tuples of* m*-bit strings, such that there exists a* $2^\ell\times 2^\ell$ *matrix* $\overline{\mathsf{M}}$ *indexed by* ℓ*-bit strings verifying*

$$\mathsf{M} = \mathsf{PS}\times\overline{\mathsf{M}}\times\mathsf{SP}.$$

Then

$$\|\mathsf{M}\|_{\mathrm{a}} = |||\mathsf{M}|||_\infty = |||\overline{\mathsf{M}}|||_\infty.$$

Proof. Using the definition of the $\|\cdot\|_a$ given in Definition 10.2 we have

$$\begin{aligned}
\|\mathsf{M}\|_{\mathrm{a}} &= \max_a\sum_b\max_{a'}\sum_{b'}\left|\left(\mathsf{PS}\times\overline{\mathsf{M}}\times\mathsf{SP}\right)_{(a,a'),(b,b')}\right|\\
&= \max_a\sum_b\max_{a'}\sum_{b'}\left|\sum_{\gamma,\gamma'}\mathsf{PS}_{(a,a'),\gamma}\overline{\mathsf{M}}_{\gamma,\gamma'}\mathsf{SP}_{\gamma',(b,b')}\right|\\
&= \max_a\sum_b\max_{a'}\sum_{b'}\left|\overline{\mathsf{M}}_{\text{SUPP}(a\oplus a'),\text{SUPP}(b\oplus b')}M^{-\ell}(M-1)^{-w(b\oplus b')}\right|\\
&= M^{-\ell}\max_a\sum_b\max_{a'}\sum_{b'}\left|\overline{\mathsf{M}}_{\text{SUPP}(a\oplus a'),\text{SUPP}(b\oplus b')}\right|(M-1)^{-w(b\oplus b')}\\
&= M^{-\ell}\max_a\sum_b\max_{a'}\sum_{\gamma'}\left|\overline{\mathsf{M}}_{\text{SUPP}(a\oplus a'),\gamma'}\right|(M-1)^{-w(\gamma')}\sum_{b'}\mathbf{1}_{\gamma'=\text{SUPP}(b\oplus b')}.
\end{aligned}$$

Since for all b we have

$$(M-1)^{-w(\gamma')}\sum_{b'}\mathbf{1}_{\gamma'=\textsc{supp}(b\oplus b')}=1,$$

the sum over b cancels the $M^{-\ell}$ term, so that

$$\|\mathsf{M}\|_{\mathrm{a}}=\max_{a,a'}\sum_{\gamma'}\left|\overline{\mathsf{M}}_{\textsc{supp}(a\oplus a'),\gamma'}\right|=\max_{\gamma}\sum_{\gamma'}\left|\overline{\mathsf{M}}_{\gamma,\gamma'}\right|.$$

Similar computations clearly lead to the same expression for $|||\mathsf{M}|||_\infty$. □

Lemma 10.5 shows that if the 2-wise distribution matrices of two random functions (resp. permutations) only depend on the support of the exclusive-or of their inputs and on the support of the exclusive-or of their outputs, then the best 2-limited adaptive adversary between these two functions is not more powerful than the best 2-limited non-adaptive adversary. This situation occurs for example in the computation of the RF/RP advantage in Example 10.3.

Conclusion

The two building blocks presented here will build the core of the two block cipher constructions of the following two chapters. Essentially, the constructions we will introduce will alternate these S-box and F-box layers with well chosen linear layers that will make it possible to reduce even further the complexity of the computations that are necessary to compute the respective advantages of the best 2-limited adaptive and non-adaptive distinguishers.

Chapter 11

Dial $\mathbf{C}$ for Cipher: Provable Security against Common Attacks

The block cipher C is the first of the two provably secure block cipher constructions that we propose. At a very high level, C is based on the same substitution-permutation network than that of the Advanced Encryption Standard (AES [41]), except that the layers made of fixed substitution boxes are replaced by perfectly random and independent S-boxes. The key-schedule of C is based on the Blum-Blum-Shub pseudo-random generator [29] and, as a consequence, is probably the slowest (but the more secure) key schedule ever suggested for a concrete construction. We provide a detailed description of C and of its key schedule in Section 11.1. Ensues a review of all security results on C, starting with those which are proved in sections 11.2 through 11.7, and going on with some results in Section 11.8 which, though not proved, seem quite reasonable. We then present a way of considerably speeding up the key schedule while preserving all security results and finish with implementation considerations. We conclude this chapter with practical considerations in section 11.9 and 11.10.

Throughout this chapter, a *perfectly random permutation* denotes a random permutation uniformly distributed among all possible permutations on the appropriate set. Consequently, when referring to a *random permutation*, nothing is assumed about its distribution.

11.1 A Description of the Block Cipher $\mathbf{C}$

High Overview

The block cipher $\mathsf{C} : \{0,1\}^{128} \to \{0,1\}^{128}$ is an iterated block cipher. It is made of a succession of *rounds*, all identical in their structure. Each round is parameterized by a *round-key* which is derived from the main 128-bit secret key using a so-called *key schedule algorithm.* The structure of each round is made of a (non-linear) substitution layer followed by a (linear) permutation layer. The non-linear part of the round mixes the key bits with the text bits in order to bring *confusion* (in the sense of [139]). The

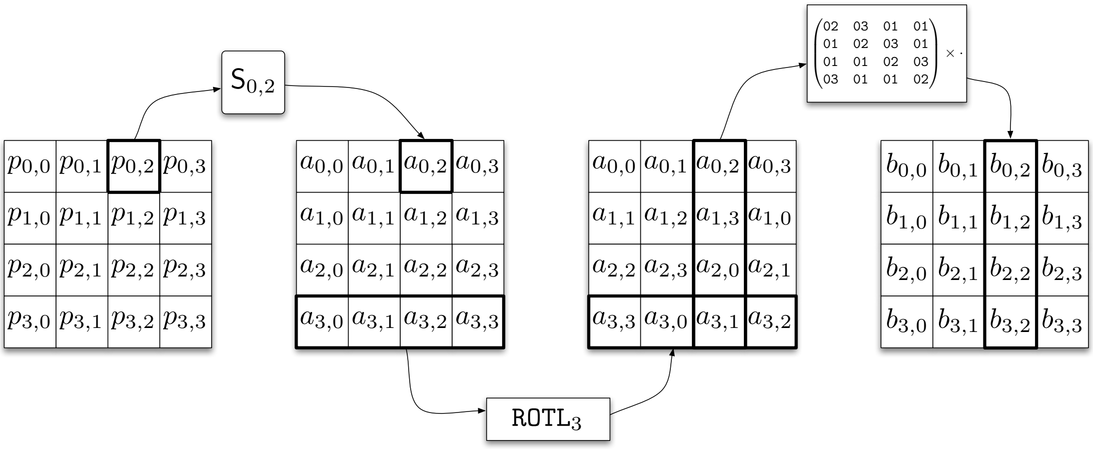

$\mathsf{S}_{0,2}$
$\begin{pmatrix}02&03&01&01\\01&02&03&01\\01&01&02&03\\03&01&01&02\end{pmatrix}\times\cdot$
$p_{0,0}$ $p_{0,1}$ $p_{0,2}$ $p_{0,3}$
$p_{1,0}$ $p_{1,1}$ $p_{1,2}$ $p_{1,3}$
$p_{2,0}$ $p_{2,1}$ $p_{2,2}$ $p_{2,3}$
$p_{3,0}$ $p_{3,1}$ $p_{3,2}$ $p_{3,3}$
$a_{0,0}$ $a_{0,1}$ $a_{0,2}$ $a_{0,3}$
$a_{1,0}$ $a_{1,1}$ $a_{1,2}$ $a_{1,3}$
$a_{2,0}$ $a_{2,1}$ $a_{2,2}$ $a_{2,3}$
$a_{3,0}$ $a_{3,1}$ $a_{3,2}$ $a_{3,3}$
$a_{0,0}$ $a_{0,1}$ $a_{0,2}$ $a_{0,3}$
$a_{1,1}$ $a_{1,2}$ $a_{1,3}$ $a_{1,0}$
$a_{2,2}$ $a_{2,3}$ $a_{2,0}$ $a_{2,1}$
$a_{3,3}$ $a_{3,0}$ $a_{3,1}$ $a_{3,2}$
$b_{0,0}$ $b_{0,1}$ $b_{0,2}$ $b_{0,3}$
$b_{1,0}$ $b_{1,1}$ $b_{1,2}$ $b_{1,3}$
$b_{2,0}$ $b_{2,1}$ $b_{2,2}$ $b_{2,3}$
$b_{3,0}$ $b_{3,1}$ $b_{3,2}$ $b_{3,3}$
$\mathtt{ROTL}_3$

Figure 11.1: One full round of C

linear part dissipates the eventual redundancy, bringing *diffusion*. Such an iterated block cipher is often referred to as a *substitution-permutation network* (SPN). Several modern block ciphers (such as the AES [41] or SAFER [107]) follow this structure. In what follows, we successively detail the SPN of C and its key schedule algorithm.

The Substitution-Permutation Network

In a nutshell, C follows the same SPN as the AES [41], except that there is no round key addition, that the fixed substitution box is replaced by independent perfectly random permutations, and that the last round of C only includes the non-linear transformation.

C is made of $r = 10$ *independent* rounds $\mathsf{R}^{(1)}, \ldots, \mathsf{R}^{(r)} : \{0,1\}^{128} \to \{0,1\}^{128}$, so that $\mathsf{C} = \mathsf{R}^{(r)} \circ \cdots \circ \mathsf{R}^{(1)}$. A r round version of C will either be denoted by $\mathsf{C}[r]$ or simply by C when the number of rounds is clear from the context. A full round of C is shown on Figure 11.1. Each round considers the 128-bit text input as a four by four array of bytes seen as elements of the finite field $\mathrm{GF}(s)$ where $s = 2^8$. Consequently, if $a \in \{0,1\}^{128}$ denotes some input of the round transformation, we will denote a_ℓ (resp. $a_{i,j}$) the ℓ-th (resp. the $(i+4j)$-th) byte of a for $0 \le \ell \le 15$ (resp. $0 \le i,j \le 3$) and call such an input a *state*. Following Definition 10.3, the *support* $\textsc{supp}(a)$ of a state a is a four by four array with 0's where the corresponding entry of a is zero and 1's everywhere else. The Hamming weight of a (or of $\textsc{supp}(a)$) is the number of 1's of $\textsc{supp}(a)$. We denote this weight by $w(a)$. When $w(a) = 16$, it means that all the entries of the state a are non-zero, in which case we say that a is of *full support*.

Except for the last one, each round $\mathsf{R}^{(i)}$ successively applies a non-linear transformation $\mathsf{S}^{(i)}$ followed by a linear transformation L so that $\mathsf{R}^{(i)} = \mathsf{L} \circ \mathsf{S}^{(i)}$ for $i = 1, \ldots, r-1$. The last round $\mathsf{R}^{(r)}$ excludes the linear transformation, i.e., $\mathsf{R}^{(r)} = \mathsf{S}^{(r)}$.

The non-linear transformation $\mathsf{S}^{(i)}$ is a set of 16 *independent* and *perfectly* *random* permutations[1] of $\mathrm{GF}(s)$. Denoting $\mathsf{S}^{(i)} = \{\mathsf{S}_0^{(i)}, \ldots, \mathsf{S}_{15}^{(i)}\}$ the 16 permutations of round i and $a, b \in \{0,1\}^{128}$ the input and the output of $\mathsf{S}^{(i)}$ respectively, we have $b = \mathsf{S}^{(i)}(a) \;\Leftrightarrow\; b_\ell = \mathsf{S}_\ell^{(i)}(a_\ell)$ for $0 \leq \ell \leq 15$. Depending on the level of security/performance one wants to achieve, the round permutations can be *de-randomized* (see Section 11.9).

The linear transformation L does not depend on the round number. It first applies a rotation to the left on each row of the input state (considered as a four by four array), over four different offsets. A linear transformation is then applied to each column of the resulting state. More precisely, if a, b denote the input and the output of L respectively, we have (considering indices modulo 4):

$$\begin{pmatrix} b_{0,j} \\ b_{1,j} \\ b_{2,j} \\ b_{3,j} \end{pmatrix} = \begin{pmatrix} \mathtt{02} & \mathtt{03} & \mathtt{01} & \mathtt{01} \\ \mathtt{01} & \mathtt{02} & \mathtt{03} & \mathtt{01} \\ \mathtt{01} & \mathtt{01} & \mathtt{02} & \mathtt{03} \\ \mathtt{03} & \mathtt{01} & \mathtt{01} & \mathtt{02} \end{pmatrix} \times \begin{pmatrix} a_{0,j} \\ a_{1,j+1} \\ a_{2,j+2} \\ a_{3,j+3} \end{pmatrix}$$

The linear transformation exactly corresponds to the succession of the transformations `ShiftRows` and `MixColumns` defined for the AES.

The Key-Schedule Algorithm

Generating a perfectly random permutation of $\{0,1\}^8$. As there are $2^8!$ possible permutations of $\{0,1\}^8$, it is possible to define a one to one mapping between $\{0,1,\ldots,2^8!-1\}$ and the set of permutations of $\{0,1\}^8$. The mapping we choose is described in Algorithm 11.1. We simply need to derive pseudo-random integers in $\{0,1,\ldots,2^8!-1\}$ from the 128-bit secret key. As each of the ten rounds involves 16 permutations, we need 160 such integers, representing a total of $160 \cdot \lceil \log_2(2^8!) \rceil = 269\,440$ pseudo-random bits.

Deriving an extended key from the secret key.

Definition 11.1 *An extended key of* $\mathsf{C}[r]$ *is a set of* $16 \cdot r$ *integers in* $\{0,1,\ldots,2^8!-1\}$.

In order to derive an extended key from the 128 secret key, we need to generate $16 \cdot r$ pseudo-random integers of $\{0,1,\ldots,2^8!-1\}$. We propose to use the Blum-Blum-Shub pseudo-random number generator [30].

Definition 11.2 *A prime* p *is a strong-prime if* $(p-1)/2$ *is prime. A prime* p *is a*

[1] Note that a random 8-bit permutation is usually more biased than the substitution box of the AES [126,167]. However this bias is key-dependent and thus does not represent a threat. Biases on the AES box are independent of the key and thus can help to distinguish (reduced rounds of) the AES from the perfect cipher when the key is unknown. Exploiting the strong bias of the substitution boxes of C requires to know the location of this bias, which is impossible without the knowledge of the permutation that was used (i.e., of the key). For instance the maximum ELP of the transformation made of a random key addition followed by the AES substitution box is 2^{-6} whereas the perfectly random substitution boxes we use have a maximum ELP of $1/(s-1) \approx 2^{-8}$. Intuitively, a cipher cannot become weaker when replacing an (arbitrary) random permutation by a perfectly random permutation.

```
Input: An integer $0 \leq \kappa < 2^8!$
Output: A table $\pi$ of size 256 such that $\pi[0], \dots, \pi[255] \in \{0, \dots, 255\}$ is a
        permutation of $\{0,1\}^8$ uniquely defined by $\kappa$
EucDiv(a,b): returns the quotient and remainder of the Euclidean division of $a$
             by $b$.
$q \leftarrow \kappa$, $\pi[0] \leftarrow 0$, $\pi[1] \leftarrow 1$ , ..., $\pi[255] \leftarrow 255$
for $m = 256, \dots, 1$ do
    $(q, r) \leftarrow$ EucDiv$(q, m)$
    Swap the values of $\pi$ at positions $r$ and $m - 1$
end
```

Algorithm 11.1: Defining a one to one mapping from integers between 0 and $2^8!$ onto the set of permutations of $\{0,1\}^8$.

strong-strong-prime if both p and $(p-1)/2$ are strong-primes.

Let p and q be two (fixed) 1024-bit strong-strong-prime numbers[2], and let $n = p \cdot q$. Considering the secret key k as a 128-bit integer, let $\{x_i \in \mathbf{Z}_n^* \ : \ i = -1, 0, 1, 2, \dots\}$ be the sequence defined by

$$\begin{cases} x_{-1} = k \cdot 2^{894} + 2^{1023} & \text{and} \\ x_i \quad = x_{i-1}^2 \bmod n & \text{for } i \geq 0. \end{cases}$$

Let BBS $= a_1 b_1 a_2 b_2 \dots$ be the pseudo-random bit string where $a_i, b_i \in \{0,1\}$ respectively denote the least and most significant[3] bits of x_i. We will use BBS to generate the 160 integers we need.

Dividing the BBS sequence into $\lceil \log_2(2^8!) \rceil$-bit substrings, we obtain pseudo-random integers in $\{0, 1, \dots, 2^{\lceil \log_2(2^8!) \rceil} - 1\}$, thus sometimes larger than $2^8!$. A naive approach to deal with those too large integers is to discard the substrings leading to such integers, thus having to generate $\lceil \log_2(2^8!) \rceil$ more bits each time this happens. This strategy requires the generation of $160 \cdot 2^{\lceil \log_2(2^8!) \rceil} / 2^8! \approx 270\,134$ pseudo-random bits in average. More efficient approaches exits (e.g., discarding only a few bits instead of a whole block), but the improvement in terms of efficiency is not worth the loss in terms of clarity.

[2] Note that strong-strong-primes are always congruent to 3 modulo 4, i.e., are Blum integers. We use strong-strong primes to ensure that the generator will have a long period. See Section 11.7 for more details.

[3] the most significant bit corresponds to being larger or smaller than $(n-1)/2$.

11.2 Exact Security against $\mathbf{2}$-limited Adversaries

Adaptive vs. Non-Adaptive Adversaries

We will now apply the Decorrelation Theory results from Chapter 10 to compute the *exact* advantage of the best 2-limited adaptive and non-adaptive distinguishers against an r round version of C, *assuming that the extended key of* $\mathsf{C}[r]$ *is uniformly distributed.* This assumption comes down to assume that the random substitution boxes are mutually independent and uniformly distributed.

We have $\mathsf{C} = \mathsf{R}^{(r)} \circ \cdots \circ \mathsf{R}^{(1)}$ where each round $\mathsf{R}^{(i)}$ is equal to $\mathsf{L} \circ \mathsf{S}^{(i)}$, except for the last round $\mathsf{R}^{(r)}$ which is equal to $\mathsf{S}^{(r)}$. Consequently we have

$$\mathsf{C} = \mathsf{S}^{(r)} \circ \mathsf{L} \circ \mathsf{S}^{(r-1)} \circ \cdots \circ \mathsf{S}^{(i)} \circ \mathsf{L} \circ \mathsf{S}^{(i-1)} \circ \cdots \circ \mathsf{S}^{(2)} \circ \mathsf{L} \circ \mathsf{S}^{(1)}.$$

Since we assume here that the substitution boxes are independent, then so are the $\mathsf{S}^{(i)}$'s. From Lemma 10.1, the previous equation leads to

$$[\mathsf{C}]^2 = [\mathsf{S}^{(1)}]^2 \times [\mathsf{L}]^2 \times [\mathsf{S}^{(2)}]^2 \times \cdots \times [\mathsf{S}^{(i-1)}]^2 \times [\mathsf{L}]^2 \times [\mathsf{S}^{(i)}]^2 \times \cdots \times [\mathsf{S}^{(r-1)}]^2 \times [\mathsf{L}]^2 \times [\mathsf{S}^{(r)}]^2.$$

Since each non-linear substitution layer has the same 2-wise distribution matrix, we simply denote it $[\mathsf{S}]^2$ and obtain

$$[\mathsf{C}]^2 = ([\mathsf{S}]^2 \times [\mathsf{L}]^2)^{r-1} \times [\mathsf{S}]^2. \tag{11.1}$$

Since we assume in this section that the random substitution boxes are independent and uniformly distributed, we note that the S layer exactly corresponds to the S-box layer studied in Section 10.5. Using the two matrices PS and SP respectively defined in (10.7) and (10.8), and which verify (according to Lemma 10.3)

$$\mathsf{SP} \times \mathsf{PS} = \mathsf{Id} \quad \text{and} \quad \mathsf{PS} \times \mathsf{SP} = [\mathsf{S}]^2,$$

we note that (11.1) can be simplified to

$$[\mathsf{C}]^2 = (\mathsf{PS} \times \mathsf{SP} \times [\mathsf{L}]^2)^{r-1} \times \mathsf{PS} \times \mathsf{SP} = \mathsf{PS} \times (\overline{\mathsf{L}})^{r-1} \times \mathsf{SP} \tag{11.2}$$

where

$$\overline{\mathsf{L}} = \mathsf{SP} \times [\mathsf{L}]^2 \times \mathsf{PS}. \tag{11.3}$$

Note that $\overline{\mathsf{L}}$ is a $2^{16} \times 2^{16}$ matrix indexed by supports. On the other hand, we have from Lemma 10.2 that

$$[\mathsf{C}^\star]^2 = [\mathsf{S}]^2 \times [\mathsf{C}^\star]^2 \times [\mathsf{S}]^2 = \mathsf{PS} \times \overline{\mathsf{C}^\star} \times \mathsf{SP}, \tag{11.4}$$

where $\overline{\mathsf{C}^\star}$ is the $2^{16} \times 2^{16}$ matrix indexed by supports and defined by

$$\overline{\mathsf{C}^\star} = \mathsf{SP} \times [\mathsf{C}^\star] \times \mathsf{PS}. \tag{11.5}$$

From (11.2) and (11.4) we obtain

$$[\mathsf{C}]^2 - [\mathsf{C}^\star]^2 = \mathsf{PS} \times \left((\overline{\mathsf{L}})^{r-1} - \overline{\mathsf{C}^\star}\right) \times \mathsf{SP}$$

and thus, according to Lemma 10.5 we deduce the following result.

Lemma 11.1 *The best 2-limited adaptive adversary A_{a} against an $r > 1$ round version of C is not more powerful than the best 2-limited non-adaptive adversary A_{na}. Moreover,*

$$\mathrm{Adv}_{\mathsf{A}_{\mathrm{a}}}(\mathsf{C}[r], \mathsf{C}^\star) = \mathrm{Adv}_{\mathsf{A}_{\mathrm{na}}}(\mathsf{C}[r], \mathsf{C}^\star) = \frac{1}{2}|||(\overline{\mathsf{L}})^{r-1} - \overline{\mathsf{C}^\star}|||_\infty, \tag{11.6}$$

where $\overline{\mathsf{L}}$ and $\overline{\mathsf{C}^\star}$ are two $2^{16} \times 2^{16}$ matrices respectively defined in (11.3) *and* (11.5).

Computation of the Advantage of the Best 2-limited Adversary

Although Lemma 11.1 shows that the computation of the advantage of the best 2-limited (non-)adaptive adversary by means of distribution matrices comes down to a computation on $2^{16} \times 2^{16}$ matrices, we see that since a matrix multiplication roughly takes $(2^{16})^3$ field operations[4] and, using a square and multiply technique, $\log r$ such multiplications are needed, the overall number of operations needed to compute $(\overline{\mathsf{L}})^{r-1}$ is roughly equal to 2^{50} (for 8 rounds) by using 2×2^{32} multiple precision rational number registers. This is still pretty hard to implement using ordinary hardware. Nevertheless, from one computation of $(\overline{\mathsf{L}})^{r-1}$ we could deduce all expected linear probabilities over all possible input/output masks almost for free.

We will now show how to make the computation of the advantage less complex. Starting from (11.3) we have for all supports γ, γ' that

$$\begin{aligned}
\overline{\mathsf{L}}_{\gamma,\gamma'} &= \sum_{a,a',b,b'} \mathsf{SP}_{\gamma,(a,a')}[\mathsf{L}]^2_{(a,a'),(b,b')}\mathsf{PS}_{(b,b'),\gamma'} \\
&= s^{-16}(s-1)^{-w(\gamma)} \sum_{a,a'} \mathbf{1}_{\gamma=\textsc{supp}(a\oplus a')}\mathbf{1}_{\gamma'=\textsc{supp}(\mathsf{L}\times(a\oplus a'))} \\
&= (s-1)^{-w(\gamma)} \sum_{a} \mathbf{1}_{\gamma=\textsc{supp}(a)}\mathbf{1}_{\gamma'=\textsc{supp}(\mathsf{L}\times(a))}.
\end{aligned}$$

We summarize this result in the following lemma.

Lemma 11.2 *For all supports γ and γ', let $\mathrm{N}[\gamma,\gamma']$ be the number of ways of connecting a support γ to a support γ' through L, i.e.,*

$$\mathrm{N}[\gamma,\gamma'] = \left|\{\textit{supports } a \textit{ such that } \textsc{supp}(a) = \gamma \textit{ and } \textsc{supp}(\mathsf{L}\times a) = \gamma'\}\right|.$$

For all supports γ and γ' we have

$$\overline{\mathsf{L}}_{\gamma,\gamma'} = (s-1)^{-w(\gamma)}\mathrm{N}[\gamma,\gamma'].$$

[4] Using Strassen's algorithm, the complexity drops to $(2^{16})^{\log 7}$ field operations [161].

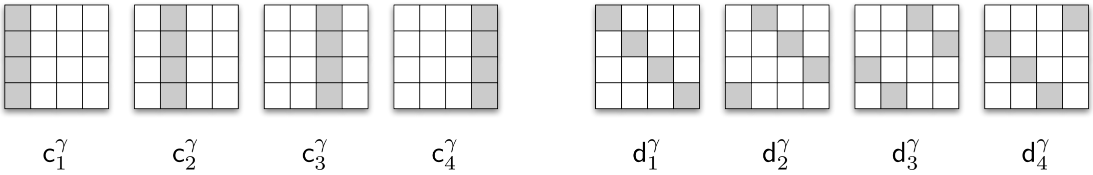


Figure 11.2: The four column's and diagonal weights of a support γ

We will see that, thanks to the properties of the linear transformation L, $\mathrm{N}[\gamma,\gamma']$ only depends on the weights of the diagonals of γ and of those of the columns of γ'. We introduce notations to deal with Hamming weights of columns and diagonals. We denote by

$$\mathsf{c}^\gamma = (\mathsf{c}_1^\gamma, \mathsf{c}_2^\gamma, \mathsf{c}_3^\gamma, \mathsf{c}_4^\gamma), \tag{11.7}$$

the vector of the four weights of γ's columns. Similarly, we denote by

$$\mathsf{d}^\gamma = (\mathsf{d}_1^\gamma, \mathsf{d}_2^\gamma, \mathsf{d}_3^\gamma, \mathsf{d}_4^\gamma), \tag{11.8}$$

the vector of the four weights of γ's diagonals. What we mean by columns and diagonals should be clear from Figure 11.2.

The `MixColumns` operation is a linear multipermutation [150], as the set of all codewords $(a, \texttt{MixColumns}(a))$ is a $[8,4,5]$ MDS code. For this reason, $\mathrm{N}[\gamma,\gamma']$ can be computed by means of a fundamental result from El-Khamy and McEliece [48] (Theorem 11.2 is actually a direct consequence of Theorem 3 in [48]).

Theorem 11.1 ***(Theorem 6 in [105])*** *Let $\mathcal{C}$ be a (n,k,d)-MDS code on $\mathrm{GF}(s)$, so that $d = n-k+1$. The weight enumerator $\mathcal{W}_\mathcal{C}$ of $\mathcal{C}$ is the weight repartition of the codewords of $\mathcal{C}$, i.e.,*

$$\mathcal{W}_\mathcal{C}(w) = |\{c \in \mathcal{C} : w(c) = w\}|$$

and verifies $\mathcal{W}_\mathcal{C}(0) = 1$, $\mathcal{W}_\mathcal{C}(w) = 0$ for all $1 \le w < d$, and

$$\mathcal{W}_\mathcal{C}(w) = \binom{n}{w} \sum_{j=d}^{w} \binom{w}{j} (-1)^{w-j} (s^{j-d+1} - 1)$$

for $d \le w \le n$.

Theorem 11.2 *Let $\mathcal{C}$ be a $(2\ell,k,d)$-MDS code on $\mathrm{GF}(s)$, so that $d = 2\ell - k + 1$. For any codeword $c = (c_1, c_2) \in \mathcal{C}$, where c_1 (resp. c_2) denotes the first (resp. last) ℓ coordinates of c, let*

$$\mathcal{W}_\mathcal{C}^2(w_1,w_2) = |\{c = (c_1,c_2) \in \mathcal{C} : w(c_1) = w_1 \text{ and } w(c_2) = w_2\}|\,.$$

We have

$$\mathcal{W}_{\mathcal{C}}^2(w_1,w_2) = \mathcal{W}_{\mathcal{C}}(w_1+w_2)\frac{\binom{\ell}{w_1}\binom{\ell}{w_2}}{\binom{2\ell}{w_1+w_2}}.$$

As a consequence, the value of $\mathrm{N}[\gamma,\gamma']$ is uniquely determined by the weights $\mathsf{d}^\gamma = (\mathsf{d}_1^\gamma,\mathsf{d}_2^\gamma,\mathsf{d}_3^\gamma,\mathsf{d}_4^\gamma)$ of the four diagonal of γ and by the weights $\mathsf{c}^{\gamma'} = (\mathsf{c}_1^{\gamma'},\mathsf{c}_2^{\gamma'},\mathsf{c}_3^{\gamma'},\mathsf{c}_4^{\gamma'})$ of the four columns of γ'. More precisely, denoting $\mathcal{C}$ the MDS code defined by the `MixColumns` operation and using Theorem 11.2, we have

$$\mathrm{N}[\gamma,\gamma'] = \prod_{s=1}^{4}\frac{\mathcal{W}_{\mathcal{C}}^2(\mathsf{d}_s^\gamma,\mathsf{c}_s^{\gamma'})}{\binom{4}{\mathsf{d}_s^\gamma}\binom{4}{\mathsf{c}_s^{\gamma'}}} = \prod_{s=1}^{4}\frac{\mathcal{W}_{\mathcal{C}}(\mathsf{d}_s^\gamma+\mathsf{c}_s^{\gamma'})}{\binom{8}{\mathsf{d}_s^\gamma+\mathsf{c}_s^{\gamma'}}}, \tag{11.9}$$

where $\mathcal{W}_{\mathcal{C}}(\cdot)$ is given by Theorem 11.1. From this last equation and from Lemma 11.2 we deduce the following lemma.

Lemma 11.3 *For all supports γ and γ' we have*

$$\overline{\mathsf{L}}_{\gamma,\gamma'} = \prod_{s=1}^{4}\frac{\mathcal{W}_{\mathcal{C}}(\mathsf{d}_s^\gamma+\mathsf{c}_s^{\gamma'})}{\binom{8}{\mathsf{d}_s^\gamma+\mathsf{c}_s^{\gamma'}}}(s-1)^{-\mathsf{d}_s^\gamma},$$

where $\mathsf{d}^\gamma = (\mathsf{d}_1^\gamma,\mathsf{d}_2^\gamma,\mathsf{d}_3^\gamma,\mathsf{d}_4^\gamma)$ are the respective weights of the four diagonals of γ and where $\mathsf{c}^{\gamma'} = (\mathsf{c}_1^{\gamma'},\mathsf{c}_2^{\gamma'},\mathsf{c}_3^{\gamma'},\mathsf{c}_4^{\gamma'})$ are the respective weights of the four columns of γ'.

The previous lemma shows that $\overline{\mathsf{L}}_{\gamma,\gamma'}$ actually only depends on d^γ and on $\mathsf{c}^{\gamma'}$. Introducing two new transition matrices will allow us to exploit this dependency in order to reduce the size of the matrices needed to compute the final advantage. We let SW be the $2^{16}\times 5^4$ matrix which rows and columns are respectively indexed by supports and 4-tuple of weights in $\{0,1,\ldots,4\}$, and defined by

$$\mathsf{SW}_{\gamma,w} = \mathsf{SW}_{\gamma,(w_1,w_2,w_3,w_4)} = \mathbf{1}_{\mathsf{d}^\gamma=w} = \prod_{s=1}^{4}\mathbf{1}_{\mathsf{d}_s^\gamma=w_s} \tag{11.10}$$

using the notation defined in (11.8). Similarly, we let WS be the $5^4\times 2^{16}$ matrix which rows and columns are respectively indexed by 4-tuple of weights in $\{0,1,\ldots,4\}$ and supports, and defined by

$$\mathsf{WS}_{w,\gamma} = \mathbf{1}_{\mathsf{c}^\gamma=w}\prod_{s=1}^{4}\binom{4}{w_s}^{-1}, \tag{11.11}$$

using the notation defined in (11.7). Letting

$$\mathrm{P}[w,w'] = \left|\{\text{supports } \gamma \text{ such that } \mathsf{c}^\gamma = w \text{ and } \mathsf{d}^\gamma = w'\}\right| = \sum_\gamma \mathbf{1}_{\mathsf{c}^\gamma=w}\mathbf{1}_{\mathsf{d}^\gamma=w'}, \tag{11.12}$$

we see that for all $w, w' \in \{0, 1, \dots, 4\}^4$ the transition matrices WS and SW are such that

$$(\mathsf{WS} \times \mathsf{SW})_{w,w'} = \mathrm{P}[w, w'] \prod_{s=1}^{4} \binom{4}{w_s}^{-1} = \frac{\mathrm{P}[w, w']}{\sum_{w''} \mathrm{P}[w, w'']}.$$

In the rest of this section, we let W be the $5^4 \times 5^4$ matrix indexed by 4-tuple of weights in $\{0, 1, \dots, 4\}$ and defined by

$$\mathsf{W}_{w,w'} = (\mathsf{WS} \times \mathsf{SW})_{w,w'} = \frac{\mathrm{P}[w, w']}{\sum_{w''} \mathrm{P}[w, w'']}. \tag{11.13}$$

Lemma 11.4 *Let* $\overline{\mathsf{M}}$ *be* $2^{16} \times 2^{16}$ *matrix indexed by supports, such that there exists a* $5^4 \times 5^4$ *matrix* $\overline{\overline{\mathsf{M}}}$ *indexed by 4-tuple of weights in* $\{0, 1, \dots, 4\}$ *verifying*

$$\overline{\mathsf{M}} = \mathsf{SW} \times \overline{\overline{\mathsf{M}}} \times \mathsf{WS}$$

where SW *and* WS *are defined in* (11.10) *and in* (11.11) *respectively. Then*

$$|||\overline{\mathsf{M}}|||_\infty = |||\overline{\overline{\mathsf{M}}}|||_\infty.$$

Proof. By definition of the $||| \cdot |||_\infty$ norm, of SW in (11.10) and of WS in (11.11) we have

$$\begin{aligned} |||\overline{\mathsf{M}}|||_\infty &= \max_\gamma \sum_{\gamma'} \left| \sum_{w,w'} \mathbf{1}_{\mathsf{d}^\gamma = w} \overline{\overline{\mathsf{M}}}_{w,w'} \mathbf{1}_{\mathsf{c}^{\gamma'} = w'} \prod_{s=1}^{4} \binom{4}{w'_s}^{-1} \right| \\ &= \max_\gamma \sum_{\gamma'} \left| \overline{\overline{\mathsf{M}}}_{\mathsf{d}^\gamma, \mathsf{c}^{\gamma'}} \right| \prod_{s=1}^{4} \binom{4}{\mathsf{c}_s^{\gamma'}}^{-1} \\ &= \max_\gamma \sum_{w'} \left| \overline{\overline{\mathsf{M}}}_{\mathsf{d}^\gamma, w'} \right| \underbrace{\left(\prod_{s=1}^{4} \binom{4}{w'_s}^{-1} \right) \sum_{\gamma'} \mathbf{1}_{w' = \mathsf{c}^{\gamma'}}}_{=1} \end{aligned}$$

since the sum on the supports γ' counts the number of supports with given column weights $w' = (w'_1, w'_2, w'_3, w'_4)$. This leads to

$$|||\overline{\mathsf{M}}|||_\infty = \max_\gamma \sum_{w'} \left| \overline{\overline{\mathsf{M}}}_{\mathsf{d}^\gamma, w'} \right| = \max_{w \in \{0,\dots,4\}^4} \sum_{w'} \left| \overline{\overline{\mathsf{M}}}_{w,w'} \right| = |||\overline{\overline{\mathsf{M}}}|||_\infty.$$

□

From the expression we obtained for $\overline{\mathsf{L}}$ in Lemma 11.3, it is easy to see that letting $\overline{\overline{\mathsf{L}}}$ be the $5^4 \times 5^4$ matrix indexed by 4-tuple of weights in $\{0, 1, \dots, 4\}$ and defined by

$$\overline{\overline{\mathsf{L}}}_{w,w'} = \prod_{s=1}^{4} \binom{4}{w'_s} \frac{\mathcal{W}_{\mathcal{C}}(w_s + w'_s)}{\binom{8}{w_s+w'_s}} (s-1)^{-w_s},$$

we have

$$\overline{\mathsf{L}} = \mathsf{SW} \times \overline{\overline{\mathsf{L}}} \times \mathsf{WS}. \tag{11.14}$$

Plugging this result in the expression we obtained for $[\mathsf{C}]^2$ in (11.2), we get (for $r \geq 2$)

$$\begin{aligned} [\mathsf{C}]^2 &= \mathsf{PS} \times (\mathsf{SW} \times \overline{\overline{\mathsf{L}}} \times \mathsf{WS})^{r-1} \times \mathsf{SP} \\ &= \mathsf{PS} \times \mathsf{SW} \times (\overline{\overline{\mathsf{L}}} \times \mathsf{W})^{r-2} \times \overline{\overline{\mathsf{L}}} \times \mathsf{WS} \times \mathsf{SP}, \end{aligned} \tag{11.15}$$

where W is the $5^4 \times 5^4$ matrix defined in (11.13).

To conclude we need to show that the distribution matrix of the perfect cipher can also be written in similar way. For all states a, a', b, b' we clearly have

$$[\mathsf{C}^\star]^2_{(a,a'),(b,b')} = \mathbf{1}_{\substack{a\neq a'\\ b\neq b'}} s^{-16}(s^{16}-1)^{-1} + \mathbf{1}_{\substack{a=a'\\ b=b'}} s^{-16}.$$

This can be written in terms of the weights of the diagonals of $\textsc{supp}(a \oplus a')$ and of the weights of the columns of $\textsc{supp}(b \oplus b')$ respectively, and thus expressed as a product $\mathsf{PS} \times \mathsf{SW} \times \cdot \times \mathsf{WS} \times \mathsf{SP}$. First noting that

$$(\mathsf{PS} \times \mathsf{SW})_{(a,a'),w} = \mathbf{1}_{\mathsf{d}^{\textsc{supp}(a\oplus a')}=w}$$

and

$$(\mathsf{WS} \times \mathsf{SP})_{w,(a,a')} = \mathbf{1}_{\mathsf{c}^{\textsc{supp}(a\oplus a')}=w} s^{-16} \prod_{s=1}^{4} (s-1)^{-w_s} \binom{4}{w_s}^{-1},$$

it is easy to show that

$$\begin{aligned} [\mathsf{C}^\star]^2_{(a,a'),(b,b')} &= \mathbf{1}_{\substack{\mathsf{d}^{\textsc{supp}(a\oplus a')}\neq 0\\ \mathsf{c}^{\textsc{supp}(b\oplus b')}\neq 0}} s^{-16}(s^{16}-1)^{-1} + \mathbf{1}_{\substack{\mathsf{d}^{\textsc{supp}(a\oplus a')}=0\\ \mathsf{c}^{\textsc{supp}(b\oplus b')}=0}} s^{-16} \\ &= \sum_{w,w'} \mathbf{1}_{\mathsf{d}^{\textsc{supp}(a\oplus a')}=w} \mathbf{1}_{\mathsf{c}^{\textsc{supp}(b\oplus b')}=w'} \left(\mathbf{1}_{\substack{w\neq 0\\ w'\neq 0}} s^{-16}(s^{16}-1)^{-1} + \mathbf{1}_{\substack{w=0\\ w'=0}} s^{-16} \right) \\ &= \sum_{w,w'} (\mathsf{PS} \times \mathsf{SW})_{(a,a'),w} \overline{\overline{\mathsf{C}^\star}}_{w,w'} (\mathsf{WS} \times \mathsf{SP})_{w',(b,b')} \end{aligned} \tag{11.16}$$

where we let $\overline{\overline{\mathsf{C}^\star}}$ be the $5^4 \times 5^4$ matrix indexed by 4-tuple of weights in $\{0, 1, \ldots, 4\}$ defined by

$$\begin{aligned} \overline{\overline{\mathsf{C}^\star}}_{w,w'} &= \left(\mathbf{1}_{\substack{w\neq 0\\ w'\neq 0}} (s^{16}-1)^{-1} + \mathbf{1}_{\substack{w=0\\ w'=0}} \right) \prod_{s=1}^{4} \binom{4}{w'_s} (s-1)^{w'_s} \\ &= \mathbf{1}_{\substack{w=0\\ w'=0}} + \mathbf{1}_{\substack{w\neq 0\\ w'\neq 0}} (s^{16}-1)^{-1} \prod_{s=1}^{4} \binom{4}{w'_s} (s-1)^{w'_s}. \end{aligned}$$

Since (11.16) means that we can write

$$[\mathsf{C}^\star]^2 = \mathsf{PS} \times \mathsf{SW} \times \overline{\overline{\mathsf{C}^\star}} \times \mathsf{WS} \times \mathsf{SP}, \tag{11.17}$$

we can easily prove the final result which makes it possible to compute the exact advantage of the best 2-limited adversary against C.

Theorem 11.3 *The respective advantages of the best* 2*-limited non-adaptive adversary* A_{na} *and of the best* 2*-limited adaptive adversary* A_{a} *against* $r > 1$ *rounds of* C *are such that*

$$\mathrm{Adv}_{\mathsf{A}_{\mathrm{na}}}(\mathsf{C}, \mathsf{C}^\star) = \mathrm{Adv}_{\mathsf{A}_{\mathrm{a}}}(\mathsf{C}, \mathsf{C}^\star) = \frac{1}{2}|||(\overline{\overline{\mathsf{L}}} \times \mathsf{W})^{r-2} \times \overline{\overline{\mathsf{L}}} - \overline{\overline{\mathsf{C}^\star}}|||_\infty,$$

where $\overline{\overline{\mathsf{L}}}$*,* W*, and* $\overline{\overline{\mathsf{C}^\star}}$ *are three* $5^4 \times 5^4$ *indexed by* 4*-tuple of weights in* $\{0, 1, \ldots, 4\}$ *and respectively defined by*

$$\overline{\overline{\mathsf{L}}}_{w,w'} = \prod_{s=1}^{4} \binom{4}{w'_s} \frac{\mathcal{W}(w_s + w'_s)}{\binom{8}{w_s + w'_s}} (s-1)^{-w_s}$$

with $\mathcal{W}(0) = 1$*,* $\mathcal{W}(i) = 0$ *for* $1 \le i < 5$*, and*

$$\mathcal{W}(i) = \binom{8}{i} \sum_{j=5}^{i} \binom{i}{j} (-1)^{i-j} (s^{j-4} - 1)$$

for $5 \le i \le 8$*,*

$$\mathsf{W}_{w,w'} = \frac{\mathrm{P}[w,w']}{\sum_{w''} \mathrm{P}[w,w'']}$$

where

$$\mathrm{P}[w,w'] = \left|\{\textit{supports } \gamma \textit{ such that } \mathsf{c}^\gamma = w \textit{ and } \mathsf{d}^\gamma = w'\}\right| = \sum_\gamma \mathbf{1}_{\mathsf{c}^\gamma = w} \mathbf{1}_{\mathsf{d}^\gamma = w'},$$

and

$$\overline{\overline{\mathsf{C}^\star}}_{w,w'} = \begin{cases} 1 & \textit{if } w = w' = 0, \\ (s^{16} - 1)^{-1} \prod_{s=1}^{4} \binom{4}{w'_s} (s-1)^{w'_s} & \textit{if } w \neq 0 \textit{ and } w' \neq 0, \\ 0 & \textit{otherwise,} \end{cases}$$

for all $w, w' \in \{0, 1, \ldots, 4\}^4$ *and where* $s = 2^8$*.*

Proof. The equality between the respective advantages of the best non-adaptive and non-adaptive distinguishers was shown in Lemma 11.1. From (11.15) and (11.17) we see that

$$[\mathsf{C}]^2 - [\mathsf{C}^\star]^2 = \mathsf{PS} \times \mathsf{SW} \times \left((\overline{\overline{\mathsf{L}}} \times \mathsf{W})^{r-2} \times \overline{\overline{\mathsf{L}}} - \overline{\overline{\mathsf{C}^\star}}\right) \times \mathsf{WS} \times \mathsf{SP}.$$

Using lemmas 10.5 and 11.4 successively, we obtain

$$|||[\mathsf{C}]^2 - [\mathsf{C}^\star]^2|||_\infty = |||(\overline{\overline{\mathsf{L}}} \times \mathsf{W})^{r-2} \times \overline{\overline{\mathsf{L}}} - \overline{\overline{\mathsf{C}^\star}}|||_\infty.$$

r	1	2	3	4	5	6
$\mathrm{Adv}(\mathsf{C},\mathsf{C}^\star)$	1	1	$2^{-4.0}$	$2^{-23.4}$	$2^{-45.8}$	$2^{-71.0}$
r	7	8	9	10	11	12
$\mathrm{Adv}(\mathsf{C},\mathsf{C}^\star)$	$2^{-126.3}$	$2^{-141.3}$	$2^{-163.1}$	$2^{-185.5}$	$2^{-210.8}$	$2^{-238.9}$

Table 11.1: Exact values of the advantage of the best 2-limited (non-)adaptive distinguisher for several number of rounds r.

Theorem 10.1 allows to conclude. □

Results of our practical computations are reported in Table 11.1. These experiments where programmed in C using the GNU Multiple Precision arithmetic library (GMP) [55] and the MPFR library [115] for multiprecision floating-point computations. All the intermediate computations where done using rational numbers instead of floating point numbers to keep maximum precision.

Security Result 11.1 Seven rounds of C are enough to obtain provable security against 2-limited (non-)adaptive adversaries.

11.3 Consequences for Iterated Attacks of Order 1, Linear and Differential Cryptanalysis

According to Corollary 10.1 and to the results obtained in Table 11.1, 7 rounds of C are enough to ensure provable security against any iterated attack of order 1, provided that the number of queries q is negligible compared to 2^{64}. In the particular case of linear cryptanalysis, the discussion following (10.1) allows to deduce from Table 11.1 that 7 rounds are enough resist linear cryptanalysis (whatever the number of queries granted to the adversary). Equation (10.2) leads to the same conclusion for differential cryptanalysis. In the following section we will derive exact results concerning both linear and differential cryptanalysis instead of upper-bounds.

Security Result 11.2 Seven rounds of C are enough to obtain provable security against iterated attacks of order 1.

11.4 Exact Security against Linear and Differential Cryptanalysis

Security against Linear Cryptanalysis

From Heuristic 8.2, we know that the data complexity of the best linear distinguisher between C and the perfect cipher $\mathsf{C}^\star$ is close to

$$\frac{8\ln 2}{\max_{a,b\neq 0}\mathrm{ELP}_{a,b}(\mathsf{C})}$$

where (according to Definition 8.6)

$$\mathrm{ELP}_{a,b}(\mathsf{C}) = \mathrm{E}_K\left(\mathrm{LP}_{a,b}(\mathsf{C}_K)\right) \quad \text{with} \quad \mathrm{LP}_{a,b}(\mathsf{C}_k) = \left(2\mathrm{Pr}_P[a \bullet P = b \bullet \mathsf{C}_k(P)] - 1\right)^2,$$

the random variable $P \in \{0,1\}^n$ being uniformly distributed and C_k denoting the permutation obtained using the extended key k. As in sections 11.2 and 11.3, we assume that the round keys are independent.

The results obtained so far allow us to easily conclude that any linear distinguisher will eventually fail to distinguish C from the perfect cipher as soon as the best 2-limited distinguisher has a negligible advantage. Indeed, we note that the linear probability can also be expressed as

$$\mathrm{ELP}_{a,b}(\mathsf{C}_k) = 1 - 2\cdot \mathrm{E}_K\left(\mathrm{E}_{P,P'}\left(\mathsf{A}(P,\mathsf{C}_K) \oplus \mathsf{A}(P',\mathsf{C}_K)\right)\right)$$

where

$$\mathsf{A}(P,\mathsf{C}_k) = \mathbf{1}_{a\bullet P \oplus b \bullet \mathsf{C}_k(P)},$$

from which we clearly see that a linear distinguisher can be expressed as a 2-limited distinguisher (and thus has an advantage bounded by that of the best 2-limited distinguisher, which is negligible for 7 rounds and more). In the rest of this section, we use another approach to actually compute the expected linear probability.

The *exact* value of $\mathrm{ELP}_{a,b}(\mathsf{C})$ can be expressed as a function of the ELP's of the individual rounds by means of Nyberg's hull principle (see Theorem 8.2) since C is a Markov cipher [97]:

$$\mathrm{ELP}_{a,b}\left(\mathsf{C}\right) = \sum_{c_1,\ldots,c_{r-1}} \prod_{i=1}^{r} \mathrm{ELP}_{c_{i-1},c_i}\left(\mathsf{R}^{(i)}\right), \tag{11.18}$$

where $c_0 = a$ and $c_r = b$. In general, the complexity of computing the expected linear probability by means of the previous formula is prohibitive since, once input/output masks are given, one has to sum over all possible intermediate masks in order to take into account every possible characteristic. We will see that this is not the case for C.

Lemma 11.5 *Let $s = 2^m$. Let $a, b \in \mathrm{GF}(s)\backslash\{0\}$ be two non-zero input/output masks on the uniformly distributed random substitution box S and let $\sigma = s-1$. The average* LP *value over all possible random S-boxes is independent of* a *and* b*, and is* $\mathrm{E}_S(\mathrm{LP}^S(a,b)) = \sigma^{-1}$*.*

Proof. Similarly to the proof of Lemma 14 in [155], we note that

$$\mathrm{E}_S(\mathrm{LP}^S(a,b)) = 2^{-2m} \sum_{\substack{x1,x2 \\ y_1,y_2}} (-1)^{(x_1 \oplus x_2)\bullet a \oplus (y_1 \oplus y_2)\bullet b} \Pr[(x_1,x_2) \xrightarrow{S} (y_1,y_2)].$$

Since S is uniformly distributed we have

$$\Pr[(x_1,x_2) \xrightarrow{S} (y_1,y_2)] = \begin{cases} 2^{-m} & \text{when } x_1 = x_2 \text{ and } y_1 = y_2, \\ 2^{-m}(2^m-1)^{-1} & \text{when } x_1 \neq x_2 \text{ and } y_1 \neq y_2, \\ 0 & \text{otherwise,} \end{cases}$$

which easily leads to the announced result. □

We note that for any S-box S we have $\mathrm{LP}_{a,0}(S) = \mathrm{LP}_{0,b}(S) = 0$ (for non-zero a and b) and $\mathrm{LP}_{0,0}(S) = 1$. From this and the Piling-up lemma, we derive the expected linear probability over the substitution layer S of C.

Lemma 11.6 *Let* $s = 2^8$ *and* $\sigma = s - 1$*. Let* a *and* b *be two non-zero masks in* $\mathrm{GF}(s)^{16}$*, and let* α *and* β *be their respective supports. We have*

$$\mathrm{ELP}_{a,b}(\mathsf{S}) = \begin{cases} \sigma^{-w(\alpha)} & \textit{if } \alpha = \beta, \\ 0 & \textit{otherwise,} \end{cases}$$

where the mean is taken over all possible uniformly distributed and independent random substitution boxes.

From the previous lemma, it is easy to derive the expected linear probability over one full round of C.

Lemma 11.7 *Let* $s = 2^8$ *and* $\sigma = s - 1$*. Let* a *and* b *be two non-zero masks in* $\mathrm{GF}(s)^{16}$ *of support* α *and* β *respectively. The expected linear probability over one full round* $\mathsf{R}^{(i)}$ *of* C*, for* $1 \leq i < r$*, with input (resp. output) mask* a *(resp.* b*) is given by*

$$\mathrm{ELP}_{a,b}(\mathsf{R}^{(i)}) = \begin{cases} \sigma^{-w(\alpha)} & \textit{if } \alpha = \textsc{supp}(\mathsf{L}^T \times b), \\ 0 & \textit{otherwise.} \end{cases}$$

Similarly, the expected linear probability over the last round is given by

$$\mathrm{ELP}_{a,b}(\mathsf{R}^{(r)}) = \begin{cases} \sigma^{-w(\alpha)} & \textit{if } \alpha = \beta, \\ 0 & \textit{otherwise.} \end{cases}$$

Proof. Since L is linear, then for all x we have $b \bullet (\mathsf{L} \times x) = (\mathsf{L}^T \times b) \bullet x$. For intermediate rounds we thus have

$$\mathrm{ELP}_{a,b}(\mathsf{R}^{(i)}) = \mathrm{ELP}_{a,b}(\mathsf{L} \circ \mathsf{S}^{(i)}) = \mathrm{ELP}_{a,\mathsf{L}^T \times b}(\mathsf{S}^{(i)}).$$

Lemma 11.6 then allows to conclude. □

Theorem 11.4 *Let $s = 2^8$ and $\sigma = s - 1$. Let a and b be two masks in $\mathrm{GF}(s)^{16}$ of support α and β respectively. The expected linear probability over $r > 1$ rounds of C, when a is the input mask and b the output mask is*

$$\mathrm{ELP}_{a,b}(\mathsf{C}) = \sigma^{-w(\beta)} \times \left(\overline{\mathsf{L}}^{r-1}\right)_{\alpha,\beta},$$

where $\overline{\mathsf{L}}$ is the $2^{16} \times 2^{16}$ square matrix, indexed by supports, and defined by

$$\overline{\mathsf{L}}_{\alpha,\beta} = \sigma^{-w(\alpha)} \mathrm{N}[\alpha, \beta],$$

where $\mathrm{N}[\alpha, \beta]$ denotes the number of ways of connecting a support α to a support β through L, i.e.,

$$\mathrm{N}[\alpha, \beta] = \left|\{\textit{supports } a \textit{ such that } \mathrm{SUPP}(a) = \alpha \textit{ and } \mathrm{SUPP}(\mathsf{L} \times a) = \beta\}\right|.$$

Proof. Starting from (11.18), replacing the round's expected linear probabilities by the expression given in Lemma 11.7 and inserting an artificial sum over supports we obtain

$$\mathrm{ELP}_{a,b}(\mathsf{C}) = \sum_{\substack{c_1,\ldots,c_{r-1} \\ \gamma_1,\ldots,\gamma_{r-1}}} \sigma^{-w(\gamma_r)} \mathbf{1}_{\gamma_{r-1}=\gamma_r} \prod_{i=1}^{r-1} \mathbf{1}_{\gamma_i=\mathrm{SUPP}(c_i)} \sigma^{-w(\gamma_{i-1})} \mathbf{1}_{\gamma_{i-1}=\mathrm{SUPP}(\mathsf{L}^T \times c_i)}$$

where $c_0 = a$, $c_r = b$, $\gamma_0 = \mathrm{SUPP}(c_0)$, and $\gamma_r = \mathrm{SUPP}(c_r)$. The previous equality leads to

$$\mathrm{ELP}_{a,b}(\mathsf{C}) = \sum_{\gamma_1,\ldots,\gamma_{r-1}} \sigma^{-w(\gamma_r)} \mathbf{1}_{\gamma_{r-1}=\gamma_r} \prod_{i=1}^{r-1} \sigma^{-w(\gamma_{i-1})} \sum_{c_i} \mathbf{1}_{\gamma_i=\mathrm{SUPP}(c_i)} \mathbf{1}_{\gamma_{i-1}=\mathrm{SUPP}(\mathsf{L}^T \times c_i)}.$$

We now note that the transpose and the inverse of a linear multipermutation still are multipermutations. Furthermore, $\mathrm{N}[\cdot,\cdot]$ only depends on the fact that the underlying linear transformation is a linear multipermutation (which is clear from (11.9)). Consequently, the sum over c_i can be expressed as

$$\sum_{c_i} \mathbf{1}_{\gamma_i=\mathrm{SUPP}(c_i)} \mathbf{1}_{\gamma_{i-1}=\mathrm{SUPP}(\mathsf{L}^T \times c_i)} = \sum_{c'_i} \mathbf{1}_{\gamma_i=\mathrm{SUPP}((\mathsf{L}^T)^{-1} \times c'_i)} \mathbf{1}_{\gamma_{i-1}=\mathrm{SUPP}(c'_i)} = \mathrm{N}[\gamma_{i-1}, \gamma_i].$$

The expected linear probability of C now reads

$$\begin{aligned}\mathrm{ELP}_{a,b}\left(\mathsf{C}\right) &= \sum_{\gamma_1,\ldots,\gamma_{r-1}} \sigma^{-w(\gamma_r)}\mathbf{1}_{\gamma_{r-1}=\gamma_r} \prod_{i=1}^{r-1} \sigma^{-w(\gamma_{i-1})}\mathrm{N}[\gamma_{i-1},\gamma_i] \\ &= \sum_{\gamma_1,\ldots,\gamma_{r-1}} \sigma^{-w(\gamma_r)}\mathbf{1}_{\gamma_{r-1}=\gamma_r} \prod_{i=1}^{r-1} \overline{\mathsf{L}}_{\gamma_{i-1},\gamma_i} \\ &= \sigma^{-w(\gamma_r)} \sum_{\gamma_1,\ldots,\gamma_{r-2}} \prod_{i=1}^{r-1} \overline{\mathsf{L}}_{\gamma_{i-1},\gamma_i}\end{aligned}$$

where $\gamma_{r-1} = \gamma_r$. The definition of the product of square matrices concludes the proof. □

Based on the results we obtained in Section 11.2, it is now easy to derive an expression which makes it possible to compute the exact value of the expected linear probability of C for various number of rounds.

Theorem 11.5 *Let* $s = 2^8$ *and* $\sigma = s - 1$*. The maximum expected linear probability of* $r > 1$ *rounds of* C *over non-zero masks verifies*

$$\max_{a,b\neq 0} \mathrm{ELP}_{a,b}(\mathsf{C}) = \max_{w'\neq 0} \mathsf{U}^{(r)}_{w'} \left(\prod_{s=1}^{4} \binom{4}{w'_s}^{-1} \sigma^{-w'_s} \right)$$

where the max *is taken over* 4*-tuple of weights in* $\{0,1,\ldots,4\}$*, where*

$$\mathsf{U}^{(r)}_{w'} = \max_{w\neq 0} \left((\overline{\overline{\mathsf{L}}} \times \mathsf{W})^{r-2} \times \overline{\overline{\mathsf{L}}} \right)_{w,w'},$$

the matrices $\overline{\overline{\mathsf{L}}}$ *and* W *being two* $5^4 \times 5^4$ *matrices defined in Theorem 11.3.*

Proof. Let α and β respectively denote the supports of a and b. From Theorem 11.4 and (11.14) we have

$$\begin{aligned}\mathrm{ELP}_{a,b}(\mathsf{C}) &= \sigma^{-w(\beta)} \left((\mathsf{SW} \times \overline{\mathsf{L}} \times \mathsf{WS})^{r-1} \right)_{\alpha,\beta} \\ &= \sigma^{-w(\beta)} \Big(\mathsf{SW} \times \underbrace{(\overline{\overline{\mathsf{L}}} \times \mathsf{W})^{r-2} \times \overline{\overline{\mathsf{L}}}}_{\mathsf{M}^{(r)}} \times \mathsf{WS} \Big)_{\alpha,\beta}.\end{aligned}$$

From the definitions of SW and WS, it is easy to show that

$$(\mathsf{SW} \times \mathsf{M}^{(r)} \times \mathsf{WS})_{\alpha,\beta} = \mathsf{M}_{\mathsf{d}^\alpha,\mathsf{c}^\beta} \left(\prod_{s=1}^{4} \binom{4}{\mathsf{c}^\beta_s}^{-1} \right),$$

r	2	3	4	5	6	7	8	9
$\max \mathrm{ELP}$	$2^{-31.9774}$	$2^{-55.9605}$	$2^{-127.9096}$	$2^{-127.9096}$	$2^{-127.9999}$	$2^{-127.9999}$	$2^{-128.0}$	$2^{-128.0}$

Table 11.2: $\max_{a,b\neq 0} \mathrm{ELP}_{a,b}(\mathsf{C})$ for various number of rounds r.

so that

$$\begin{aligned}
\max_{a,b\neq 0} \mathrm{ELP}_{a,b}(\mathsf{C}) &= \max_{\alpha,\beta\neq 0} \sigma^{-w(\beta)} \mathsf{M}^{(r)}_{\mathsf{d}^\alpha,\mathsf{c}^\beta} \left(\prod_{s=1}^{4} \binom{4}{\mathsf{c}_s^\beta}^{-1} \right) \\
&= \max_{w,w'\neq 0} \mathsf{M}^{(r)}_{w,w'} \left(\prod_{s=1}^{4} \binom{4}{w'_s}^{-1} \sigma^{-w'_s} \right) \\
&= \max_{w'\neq 0} \mathsf{U}^{(r)}_{w'} \left(\prod_{s=1}^{4} \binom{4}{w'_s}^{-1} \sigma^{-w'_s} \right).
\end{aligned}$$

□

Results of our practical computations are reported in Table 11.2. These experiments where programmed in C using the GNU Multiple Precision arithmetic library (GMP) [55] and the MPFR library [115] for multiprecision floating-point computations. All the intermediate computations where done using rational numbers instead of floating point numbers to keep maximum precision.

Security Result 11.3 Four rounds of C are enough to obtain provable security against linear cryptanalysis.

Security against Differential Cryptanalysis

Just as the efficiency of linear cryptanalysis can be measured by means of ELP's, the efficiency of differential cryptanalysis can be measured by means of EDP's [124], as discussed on page 148. The computations that we performed on the ELP of C can be applied, with almost no modification, in order to compute the EDP. The major modification is that we do not use the fact that $b\bullet(\mathsf{M}\times x) = (\mathsf{M}^T \times b)\bullet x$ but rather that if the difference between two inputs of a linear transformation M is equal to a, then the output difference is equal to $\mathsf{M} \times a$.

We now follow the steps that lead to the final result on the ELP coefficient and see whether they apply to the EDP coefficient. Lemma 11.5 applies to the EDP coefficient, and therefore, it is also the case for Lemma 11.6 (where we use the independence of the 16 inputs on the S-boxes in order to obtain a product of EDP, instead of the Piling-up Lemma). Because the relation between an input difference on the linear transformation of C and its output difference is not the same as in the case where we considered input/output masks, we must replace L by $(\mathsf{L}^T)^{-1}$ in the definition of $\mathrm{N}[\cdot,\cdot]$. Yet, as already noted in the proof of Theorem 11.4, the actual *values* of $\mathrm{N}[\cdot,\cdot]$ do not depend on which underlying multipermutation is used, it just needs to be one. In other words, replacing L by $(\mathsf{L}^T)^{-1}$ in the definition of $\mathrm{N}[\cdot,\cdot]$ does not change its entries. The computations on the ELP coefficient thus still apply for the EDP coefficient. Lemma 11.7, Theorem 11.4, and Theorem 11.5 apply to the EDP, and thus, the numerical results given in Table 11.2 are also valid for the value of $\max_{a,b\neq 0} \mathrm{EDP}_{a,b}(\mathsf{C})$.

Security Result 11.4 Four rounds of C are enough to obtain provable security against differential cryptanalysis.

11.5 Towards the Perfect Cipher

From the results obtained in the previous section, it is possible to prove a conjecture made by Keliher, Meijer, and Tavares in [82], namely that all ELP's converge towards $1/(2^{128}-1)$ (which corresponds to the ELP of the perfect cipher) as the number of rounds increases. This means that C behaves exactly like the perfect cipher (as far as linear cryptanalysis is concerned) when the number of rounds is high enough.

Clearly, for any non-zero mask c, $\mathrm{LP}_{c,0}(\mathsf{C}) = \mathrm{LP}_{0,x}(\mathsf{C}) = 0$ and $\mathrm{LP}_{0,0}(\mathsf{C}) = 1$. Thus, the $2^{16} \times 2^{16}$ square matrix $\bar{\mathsf{L}}$ of Theorem 11.4 has the following shape

$$\bar{\mathsf{L}} = \left(\begin{array}{c|c} 1 & \mathbf{0} \\ \hline \mathbf{0} & \mathsf{M} \end{array}\right) \tag{11.19}$$

where M is a $(2^{16}-1)\times(2^{16}-1)$ square matrix, indexed by non-zero supports. We can now notice from Theorem 11.4 that

$$\mathrm{ELP}_{a,b}(\mathsf{C}) = \sigma^{-w(\beta)}\mathsf{M}_{\alpha,\beta}$$

for any non-zero a and b of supports α and β respectively. Since $\sum_b \mathrm{ELP}_{a,b}(\mathsf{C}) = 1$ we have

$$1 = \sum_b \sigma^{-w(\beta)}\mathsf{M}_{\alpha,\beta} = \sum_\beta \sigma^{w(\beta)}\sigma^{-w(\beta)}\mathsf{M}_{\alpha,\beta} = \sum_\beta \mathsf{M}_{\alpha,\beta}.$$

We also note that $\mathsf{M}_{\alpha,\beta} \geq 0$ for any α and β.

Lemma 11.8 *The matrix* M *defined by* (11.19) *is the transition matrix of a Markov chain, whose set of states is the set of non-zero supports and whose transition probability from a non-zero support* α *to a non-zero support* β *is given by* $\mathsf{M}_{\alpha,\beta}$.

The transition graph of the Markov chain is the directed graph whose vertices are the σ non-zero supports and such that there is an edge from α to β when $\mathsf{M}_{\alpha,\beta} > 0$. From the study of supports propagation [41] (which is based on the MDS criterion), it clearly appears that from any graph state, there is a path towards the graph state corresponding to the full support $\textsc{supp}_{\text{full}}$ (for example, two steps are required to go from a support of Hamming weight 1 to $\textsc{supp}_{\text{full}}$). Moreover, from the graph state corresponding to $\textsc{supp}_{\text{full}}$ one can reach any graph state. Hence, from each graph state there is a sequence of arrows leading to *any* other graph state. This means that the corresponding Markov chain is *irreducible* [61]. Since there is an arrow from $\textsc{supp}_{\text{full}}$ to itself, one can find a sequence of arrows leading from any graph state to any graph state, of *any* (yet long enough) length. This means the Markov chain is *aperiodic*. We can deduce that there exists exactly one stationary distribution (see for example chapter 5 in [61]), i.e., a $1 \times (2^{16}-1)$ row vector $\pi = (\pi_\alpha)_{\alpha \neq 0}$ indexed by non-zero supports such that $\pi_\alpha \geq 0$ for all non-zero α with $\sum_{\alpha\neq 0}\pi_\alpha = 1$, and such that $\pi\mathsf{M} = \pi$ (which is to say that $\pi_\beta = \sum_{\alpha\neq 0}\pi_\alpha \mathsf{M}_{\alpha,\beta}$ for all non zero β). It is easy to show that the row vector π indexed by non-zero supports such that $\pi_\alpha = \sigma^{w(\alpha)}(s^{16}-1)^{-1}$ is a stationary distribution of the Markov chain described by the transition matrix M. Indeed,

$$\sum_{\alpha\neq 0}\pi_\alpha = \frac{1}{s^{16}-1}\sum_{\alpha\neq 0}\left(\sum_{s=1}^{16}\mathbf{1}_{s=w(\alpha)}\right)\sigma^{w(\alpha)} = \frac{1}{s^{16}-1}\sum_{s=1}^{16}\binom{16}{s}\sigma^s = 1,$$

and therefore π is a probability distribution. Moreover, for any non-zero β,

$$(\pi\mathsf{M})_\beta = \frac{1}{s^{16}-1}\sum_{\alpha\neq 0}\mathrm{N}[\alpha,\beta] = \frac{1}{s^{16}-1}\sigma^{w(\beta)} = \pi_\beta,$$

as the sum is simply the number of non-zero states that can be connected to some non-zero support β through L, which is exactly the number of states of support equal to β, as each state of support β has one and only one preimage through L.

It is well known [60] that $(\mathsf{M}^r)_{\alpha,\beta} \to \pi_\beta$ when $r \to \infty$. As

$$\mathrm{ELP}_{a,b}(\mathsf{C}) = \sigma^{-w(\beta)}(\mathsf{M}^{r-1})_{\alpha,\beta}$$

for non-zero masks a and b of respective supports α and β, we have proved the following theorem (which corresponds to the conjecture in [82]).

Theorem 11.6 *Let $s = 2^8$. Let a and b be two non-zero masks in $\mathrm{GF}(s)^{16}$. Then*

$$\lim_{r\to\infty}\mathrm{ELP}_{a,b}(\mathsf{C}[r]) = \mathrm{ELP}_{a,b}(\mathsf{C}^\star) = \frac{1}{s^{16}-1}, \qquad (11.20)$$

where $\mathsf{C}^\star$ denote the perfect cipher on $\mathrm{GF}(s)^{16}$.

11.6 Provable Security against Impossible Differentials

Impossible Differentials [18] attacks are a variation of differential cryptanalysis. They consist in finding pairs of input/output differences such that for any instance c of

C we have $\mathrm{DP}_{a,b}(\mathsf{c}) = 0$. In other words, an input difference of a can never (i.e., for any input and any key) lead to an output difference of b. In the case of C we can prove that five rounds are enough to have no impossible differential[5], i.e., given any input/output masks a and b, there exists an instance c of $\mathsf{C}[5]$ (i.e., a key defining 80 permutations) such that $\mathrm{DP}_{a,b}(\mathsf{c}) \neq 0$.

Lemma 11.9 *Let $a, b \in \{0,1\}^{128}$ be any two differences of full support. One substitution layer S is enough to ensure that there exists an instance s of S such that $\mathrm{DP}_{a,b}(\mathsf{s}) \neq 0$.*

Proof. Considering the two plaintexts 0 and a, we can choose the 16 substitution boxes $\mathsf{s}_0, \ldots, \mathsf{s}_{15}$ of one round such that $\mathsf{s}_i(0) = 0$ and $\mathsf{s}_i(a_i) = b_i$ (where $a = (a_0, a_1, \ldots, a_{15})$ and $b = (b_0, b_1, \ldots, b_{15})$). As both a_i and b_i are non-zero (a and b are of full support), both conditions can be verified without being inconsistent with the fact that s_i is a permutation. □

Lemma 11.10 *Let $a \in \{0,1\}^{128}$ be a non-zero difference of arbitrary support. Considering a two full rounds version of C (i.e., $\mathsf{C} = \mathsf{R}^{(2)} \circ \mathsf{R}^{(1)} = \mathsf{L}^{(2)} \circ \mathsf{S}^{(2)} \circ \mathsf{L}^{(1)} \circ \mathsf{S}^{(1)}$), there exists a difference $b \in \{0,1\}^{128}$ of full support and an instance c of C such that $\mathrm{DP}_{a,b}(\mathsf{c}) \neq 0$.*

Proof. For simplicity reasons, we restrict ourselves to the case where the support of a is of weight 1 (the other cases can be treated in a similar way). Without loss of generality, assume $a_0 \neq 0$ while $a_i = 0$ for $i = 1, \ldots, 15$. We consider the two plaintexts to be 0 and a. Letting $\mathsf{S}_i^{(1)}(0) = 0$ for all i, we have $\mathsf{L}^{(1)} \circ \mathsf{S}^{(1)}(0) = 0$. By carefully choosing $\mathsf{S}_0^{(1)}(a_0)$, we can make sure that $\mathsf{L}^{(1)} \circ \mathsf{S}^{(1)}(a)$ has a support of weight 4 (on the first columns of the four by four array). Proceeding in the same manner in the second round, we can make sure that $\mathsf{C}(0) = 0$ and $b = \mathsf{C}(a)$ is of full support. □

Consider any two differences $a, b \in \{0,1\}^{128}$ and a five round version of $\mathsf{C} = \mathsf{S}^{(5)} \circ \mathsf{L}^{(4)} \circ \mathsf{S}^{(4)} \circ \mathsf{L}^{(3)} \circ \mathsf{S}^{(3)} \circ \mathsf{L}^{(2)} \circ \mathsf{S}^{(2)} \circ \mathsf{L}^{(1)} \circ \mathsf{S}^{(1)}$. From Lemma 11.10, there exists an instance $\mathsf{c}_{\mathrm{start}}$ of the first two rounds $\mathsf{L}^{(2)} \circ \mathsf{S}^{(2)} \circ \mathsf{L}^{(1)} \circ \mathsf{S}^{(1)}$ and a difference d of full support such that

$$\mathrm{DP}_{a,d}(\mathsf{c}_{\mathrm{start}}) \neq 0.$$

Starting from the end, there exists an instance $\mathsf{c}_{\mathrm{end}}$ of $\mathsf{S}^{(5)} \circ \mathsf{L}^{(4)} \circ \mathsf{S}^{(4)} \circ \mathsf{L}^{(3)}$ and a difference e of full support such that

$$\mathrm{DP}_{b,e}(\mathsf{c}_{\mathrm{end}}^{-1}) \neq 0, \quad \text{so that} \quad \mathrm{DP}_{e,b}(\mathsf{c}_{\mathrm{end}}) \neq 0.$$

From Lemma 11.9, there exists an instance $\mathsf{c}_{\mathrm{mid}}$ of $\mathsf{S}^{(3)}$ such that

$$\mathrm{DP}_{d,e}(\mathsf{c}_{\mathrm{mid}}) \neq 0.$$

Consequently,

$$\mathrm{DP}_{a,b}(\mathsf{c}_{\mathrm{end}} \circ \mathsf{c}_{\mathrm{mid}} \circ \mathsf{c}_{\mathrm{start}}) \neq 0.$$

[5] There exists an impossible differential on 4 rounds of the AES leading to an attack on 6 rounds [33].

Security Result 11.5 Five rounds of C are enough to ensure that no impossible differential exists.

11.7 Taking the Key-Schedule into Account

We assumed in sections 11.2 to 11.6 that the random substitution boxes were independent and uniformly distributed. When choosing these boxes by means of the key-schedule algorithm described in Section 11.1 this assumption does not hold anymore. Yet, we will show that under a certain intractability assumption, the keyed C is not less secure than the version of C studied in the previous sections.

All Substitution Boxes of C are Indistinguishable from Independent Perfectly Random Permutations

A pseudo-random bit generator (PRBG) is said to be *cryptographically secure* if no polynomial-time statistical test can distinguish an output sequence of this generator from a uniformly distributed random bit string of the same length with a significant advantage [166]. Such a generator can always be distinguished if the length of the bit string is longer than the generator's period. We need to prove that the Blum-Blum-Shub generator (BBS) we use has a period long enough to generate a complete extended key.

We know from the original paper [29] that the period of the x_i's sequence of the BBS generator divides $\lambda(\lambda(n))$ (where λ denotes the Carmichael function) if both p and q are strong-primes and both p and q are Blum integers. Obviously, the period of the bit string output by BBS divides the period of the x_i's. By making sure that $\lambda(\lambda(n))$ does not contain small factors, we can prove that this length will be large enough. This can be done by choosing strong-strong-primes p and q. In such a case we can write $p = 2p_1 + 1 = 4p_2 + 3$ and $q = 2q_1 + 1 = 4q_2 + 3$, and obtain $\lambda(\lambda(n)) = \lambda(\mathrm{lcm}(2p_1, 2q_1)) = \lambda(2p_1q_1) = \mathrm{lcm}(2p_2, 2q_2) = 2p_2q_2$. Therefore, if the period of the bit string is not 2, it is necessarily long enough to generate a complete extended key as $\min(p_2, q_2) \gg 300\,000$.

It is known that the original Blum-Blum-Shub pseudo-random bit generator is cryptographically secure [29,30]. Vazirani and Vazirani showed that outputting both the least and most significant bits of the quadratic residues produced by the generator is also cryptographically secure [158,159].

Definition 11.3 *Let s_0 and s_1 be two bit strings, such that s_0 is obtained using the BBS pseudo-random generator and s_1 is perfectly random. The advantage of an adversary* A *trying to distinguish s_0 from s_1 is given by*

$$\mathrm{Adv}_{\mathsf{A}}^{\mathsf{BBS}} = \Pr\big[\mathrm{Adv}(s_0) = 0\big] - \Pr\big[\mathrm{Adv}(s_1) = 0\big].$$

Assuming that the problem of deciding the quadratic residuosity modulo n is hard (an assumption we will refer to as the *quadratic residuosity assumption* [56]), we know that $\mathrm{Adv}_{\mathsf{A}}^{\mathsf{BBS}}$ can be made arbitrarily small by increasing the value of n. The key schedule of C relies on the BBS generator and makes sure that the mapping from the set of 2^{128} keys to the set of possible seeds of the pseudo-random generator is injective. Therefore, the pseudo-random sequence produced by the key schedule of C is indistinguishable from a perfectly random binary sequence of the same length. The method we use to convert this binary sequence into substitution boxes makes sure that for an unbiased sequence one obtains an unbiased set of substitution boxes. By choosing a suitable n, the substitution boxes of C can thus be made indistinguishable from independent perfectly random permutations.

The Keyed **C** is not Less Secure than **C** with Independent Boxes

Definition 11.4 *Let k_0 and k_1 be two extended keys of C, such that k_0 is obtained through the key schedule seeded by a perfectly random 128-bit key and k_1 is uniformly distributed. The advantage of an adversary A trying to distinguish k_0 from k_1 is given by*

$$\mathrm{Adv}_{\mathsf{A}}^{\mathsf{key}} = \Pr\big[\mathsf{A}(k_0) = 0\big] - \Pr\big[\mathsf{A}(k_1) = 0\big].$$

Lemma 11.11 *Let k_0 and k_1 be two extended keys as in Definition 11.4 and s_0, s_1 be two bit strings as in Definition 11.3. An adversary A able to distinguish k_0 from k_1 with probability p can distinguish s_0 from s_1 with probability $p' \geq p$, i.e., $\mathrm{Adv}_{\mathsf{A}}^{\mathsf{key}} \leq \mathrm{Adv}_{\mathsf{A}}^{\mathsf{BBS}}$.*

Proof. Given s_b ($b \in \{0, 1\}$), the adversary can derive an acceptable extended key k_b. From this, the adversary has an advantage $\mathrm{Adv}_{\mathsf{A}}^{\mathsf{key}}$ of guessing the correct value of b and thus obtains a distinguisher on BBS with advantage $\mathrm{Adv}_{\mathsf{A}}^{\mathsf{key}}$. □

The strongest notion of security for a block cipher is its indistinguishability from a perfectly random permutation $\mathsf{C}^\star$. Proving the security of C against a distinguishing attack performed by A consists in upper bounding $\mathrm{Adv}_{\mathsf{A}}(\mathsf{C}, \mathsf{C}^\star)$.

Let k_0 and k_1 be two random extended keys of C picked as in Definition 11.4, defining two random instances of C denoted $\mathsf{C}_{\mathsf{key}}$ and $\mathsf{C}_{\mathsf{rand}}$ respectively. Obviously, distinguishing $\mathsf{C}_{\mathsf{key}}$ from $\mathsf{C}_{\mathsf{rand}}$ is harder than distinguishing k_0 from k_1, so that

$$\mathrm{Adv}_{\mathsf{A}}(\mathsf{C}_{\mathsf{key}}, \mathsf{C}_{\mathsf{rand}}) \leq \mathrm{Adv}_{\mathsf{A}}^{\mathsf{key}}.$$

Assume there exists a distinguishing attack on $\mathsf{C}_{\mathsf{key}}$ that does not work on $\mathsf{C}_{\mathsf{rand}}$ such that, for an adversary A using it,

$$\mathrm{Adv}_{\mathsf{A}}(\mathsf{C}_{\mathsf{key}}, \mathsf{C}^\star) \geq 2 \cdot \mathrm{Adv}_{\mathsf{A}}(\mathsf{C}_{\mathsf{rand}}, \mathsf{C}^\star).$$

From the triangular inequality we have

$$\mathrm{Adv}_{\mathsf{A}}(\mathsf{C}_{\mathsf{key}}, \mathsf{C}^\star) - \mathrm{Adv}_{\mathsf{A}}(\mathsf{C}_{\mathsf{rand}}, \mathsf{C}^\star) \leq \mathrm{Adv}_{\mathsf{A}}(\mathsf{C}_{\mathsf{key}}, \mathsf{C}_{\mathsf{rand}})$$

so that

$$\mathrm{Adv}_{\mathsf{A}}(\mathsf{C}_{\mathsf{key}}, \mathsf{C}^\star) \leq 2 \cdot \mathrm{Adv}_{\mathsf{A}}(\mathsf{C}_{\mathsf{key}}, \mathsf{C}_{\mathsf{rand}}) \leq 2 \cdot \mathrm{Adv}_{\mathsf{A}}^{\mathsf{key}}.$$

In conclusion, using Lemma 11.11, any distinguishing attack twice as effective on $\mathsf{C}_{\mathsf{key}}$ than on $\mathsf{C}_{\mathsf{rand}}$ gives an advantage which is bounded by $2 \cdot \mathrm{Adv}_{\mathsf{A}}^{\mathsf{BBS}}$. Under the quadratic residuosity assumption, such an attack cannot be efficient.

Although the quadratic residuosity problem is not equivalent to the problem of factoring $p \cdot q$, the best known attacks require it. The exact cost of this factorization is not obvious. For a given symmetric key size, there are several estimates for an equivalent asymmetric key size [84]. According to the NIST recommendations, a 2048-bit modulus is equivalent to a 112-bit symmetric key [53].

Security Result 11.6 Under the quadratic residuosity assumption, C used with the key schedule described in Section 11.1 is as secure as C used with independent perfectly random substitution boxes.

The Keyed C has no Equivalent Keys

Two block cipher keys are said to be *equivalent* when they define the same permutation. It is easy to build equivalent *extended* keys for C (when *not* using the key schedule). Consider an extended key k_1 defining a set of 160 substitution boxes such that the first 32 are the identity. We consider a second extended key k_2 defining another set of substitution boxes such that the last 128 are identical to that defined by k_1 and such that the first 16 boxes simply xor a constant $a \in \{0,1\}^{128}$ to the plaintext, the remaining boxes (in the second layer) correcting the influence of a by xoring $\mathsf{L}(a)$ to its input. Although they are different, k_1 and k_2 define the same permutation. Such a property could be a threat to the security of C. If too many such extended keys were equivalent, it could be possible to find equivalent 128-bit keys for $\mathsf{C}_{\mathsf{key}}$.

We can prove that the probability that two 128-bit equivalent keys exist is negligible. Indeed, this probability depends on the number of equivalence classes among the extended keys. Considering a one round version of C, it can be seen that no equivalent extended keys exist. Consequently, there are at least $(2^8!)^{16} \approx 2^{26944}$ equivalence classes. Adding rounds (thus increasing the extended key size) cannot decrease this number of classes. Assuming that the key schedule based on BBS uniformly distributes the extended keys obtained from the 128-bit keys among these classes, the probability that two keys fall into the same class can be upper bounded by

$$1 - e^{-(2^{128})^2/(2*2^{26944})} \approx 2^{-26689}.$$

Security Result 11.7 The probability that two 128-bit keys lead to the same instance of C is upper bounded by 2^{-26689}.

11.8 Unproved Security against other Attacks

C is (not that) Resistant to Saturation Attacks

Saturation attacks [69, 93] are chosen-plaintext attacks on byte-oriented ciphers. An attack on four rounds of the AES can be performed [42] by choosing a set of 2^8 plaintexts equal on all but one byte. After 3 rounds of the AES, the xor of all the corresponding ciphertexts is 0. This makes it easy to guess the key of the fourth round, as all round key bytes can be guessed independently.

In our case, the property on the third round output still holds. Nevertheless, it only allows to exclude 255 out of 256 keys for each substitution box. This was enough for the AES, but in our case an adversary would still be left with 255! valid substitution boxes, so that a more subtle approach is needed.

In [26], Biryukov and Shamir present an attack on SASAS, a generic construction with three rounds of random key-dependent substitution boxes linked by random key-dependent affine layers. Following their approach, the saturation attacks on the AES can be adapted to C but with a non-negligible cost. In this approach, an exhaustive search on 8 bits (as necessary with the AES) is replaced by a linear algebra step which requires 2^{24} operations. The additional workload is thus of the order of 2^{16}. This overhead implies that any attack with a complexity higher than 2^{112} becomes infeasible. In particular the saturation attacks on 7 rounds of the AES [51] should not apply to C.

We believe that saturation-like attacks are the biggest threat for reduced rounds versions of C. Chances that such attacks apply to 10 rounds are however very low.

C is Resistant to a Wide Variety of Attacks

Algebraic attacks consist in rewriting the whole block cipher as a system of algebraic equations. The solutions of this system correspond to valid plaintext, ciphertext, and key triples. Algebraic attack attempts on the AES take advantage of the simple algebraic structure of the substitution box [36]. In our case, substitution boxes can by no means be described by simple algebraic forms, and thus, algebraic attacks will necessarily be much more complex against C than against the AES. We do believe that they will be more expensive than exhaustive key search.

Slide attacks [27] exploit a correlation between the different round keys of a cipher. These attacks apply for example against ciphers with weak key schedules or against block ciphers with key-dependent substitution boxes and periodic key schedules. C uses independent perfectly random substitution boxes, so that all rounds are independent from each other. Slide attacks cannot apply here.

The boomerang attack [162] is a special type of differential cryptanalysis. It needs to find a differential characteristic on half the rounds of the cipher. Four rounds of C being sufficient to be provably secure against differential cryptanalysis, 10 rounds are necessarily sufficient to resist the boomerang attack. Similarly, neither differential-linear cryptanalysis [20, 99] nor the rectangle attack [19] apply to C.

11.9 A Fast Variant of C without Security Compromise

The main drawback in the design of C is the huge amount of pseudo-random bits required for the key schedule. Having to generate hundreds of thousands of bits with the Blum-Blum-Shub generator is unacceptable for many applications. We propose here an adaptation of C, enjoying the same security proofs, but requiring much less pseudo-random bits.

Using Order 2 Decorrelated Substitutions Boxes. One can note that the security results obtained in sections 11.2, 11.3, 11.4, 11.5, and 11.7 do *not* require from the substitution boxes to be perfectly random permutations. In reality, one only needs to have order 2 decorrelated substitution boxes.

Suppose we have a family D_2 of order 2 decorrelated substitution boxes. Using the Blum-Blum-Shub generator and the same method as for the standard C key schedule, we can generate a set of 160 substitution boxes from D_2 indistinguishable from 160 randomly chosen D_2 boxes. Again, it is possible to prove that any attack on a keyed C using substitution boxes in D_2 requires to be able to distinguish the output of the Blum-Blum-Shub generator from a perfectly random binary stream.

Hence, apart from the resistance to impossible differentials, all proved security arguments of C remain untouched when using boxes of D_2. However, each time the key schedule required $\log_2 256!$ bits from the Blum-Blum-Shub generator, it only requires $\log_2 |\mathrm{D}_2|$ now.

$\boldsymbol{A \oplus \frac{B}{X}}$: a Good Family of Order 2 Decorrelated Substitution Boxes. From what we have just seen, whatever the family D_2 we use, most of the security results will still hold. For optimal efficiency, we need to select the smallest possible such family. It was shown in [4] that any family of the form

$$\mathrm{D}_2 = \left\{X \mapsto A \oplus B \cdot S(X) \ : \ A, B \in \{0,1\}^8, B \neq 0\right\}$$

where S is any *fixed* permutation of $\mathrm{GF}(2^8)$ (and where $\cdot$ represents a product in $\mathrm{GF}(2^8)$) is decorrelated at order 2. We propose to use the family

$$\mathrm{D}_2 = \left\{X \mapsto A \oplus \frac{B}{X} \ : \ A, B \in \{0,1\}^8, B \neq 0\right\}.$$

This family contains 2^{16} elements and the substitution boxes can be chosen uniformly in D_2 from 16 bits of the Blum-Blum-Shub generator. The first 8 bits define A, the last

8 define B. So, the whole key schedule for ten rounds of C only requires 2 560 pseudo-random bits and should be about 100 times faster than an unmodified C with perfectly random permutations. One may believe that this construction is very similar to that of the AES (assuming that the round keys are independent and perfectly random). Nevertheless, deriving the AES construction from ours requires to set $B = 1$. The family obtained in this case is no longer decorrelated at order 2, so that, unfortunately, none of the security results we obtained for C directly applies to the AES.

Security Considerations. Even if this might not be the case for any order 2 decorrelated family of substitution boxes, it is interesting to note that C built on the family D_2 we chose is also resistant to impossible differentials. As for perfectly random permutations, lemmas 11.9 and 11.10 can both be proved for boxes of the form $A \oplus \frac{B}{X}$.

None of the security results we obtained requires using perfectly random permutations and substitution boxes of the form $A \oplus \frac{B}{X}$ are enough. We believe that achieving the same security level with perfectly random permutations is possible with fewer rounds. More precisely, it may be possible to obtain a trade-off between the number of rounds and the level of decorrelation of the random substitution boxes. Fewer rounds lead to fast encryption/decryption procedures but require a higher level of decorrelation. In this case, more pseudo-random bits are necessary to generate each substitution box, and this may lead to a (very) slow key schedule. The best choice depends on the application.

11.10 Implementation and Performances

Implementation. As seen in Section 11.1, before being able to use the Blum-Blum-Shub generator, one needs to generate two strong-strong-primes p and q, which is not an easy operation: it has a complexity of $O((\log p)^6)$. For primes of length 1024, this takes one million times more operations than generating a prime of the same size. Some optimizations exist to improve the constant factor in the prime number generation [71] and can become very useful for strong-strong-prime numbers.

When implementing C, the same optimizations as for the AES are possible. In particular, one round of C can be turned into 16 table look-ups and 12 xors. Basically, the output can be split in four 32-bit blocks, each of which only depends on four bytes of the input. However, all the tables of C are different from each other. This is the only reason why encrypting/decrypting with C could be slower than with the AES. Considering standard 32-bit computers, this has little influence in practice as the 160 tables still fit in the cache of the CPU. The required memory is $160 \cdot 256 \cdot 4 = 160$kBytes. This however becomes an issue when implementing C on a smartcard (but who wants to implement Blum-Blum-Shub on a smartcard anyway?) or on a CPU with 128kBytes of cache.

We programmed C in C using GMP [55] for the key schedule operations. On a 3.0 GHz Pentium D, we obtain encryption/decryption speeds of 500 Mbits/s. Generating the 160 substitution boxes from the 128-bit secret key takes 2.5s when using perfectly random permutations and 25ms when using the $A \oplus \frac{B}{X}$ construction. Note that to decrypt, it is also necessary to invert the substitution boxes. This takes a negligible time compared to the generation of the extended key, which is the most expensive step of the key schedule.

Applications. Given the timings we obtained, it appears that using C for encryption purpose is practical, in particular with the shortened key schedule. Of course, a key schedule of 25ms is much slower than most existing key schedules but is still acceptable in a large majority of applications. This can become negligible when the amount of data to encrypt becomes large.

The 2.5s obtained for the "most secure" version using perfectly random substitution boxes is suitable for only a few very specific applications. However, we believe that in the case where a very high security level is required, this price is not that high. This might not be an issue in certain cases when the key schedule is run in parallel with some other slow operation, like for hard disk drive encryption (for which the key schedule is performed only once during a boot sequence which already takes several seconds).

In some other circumstances however, C is not usable at all. For example, when using it as a compression function in a Merkle-Damgård construction, as one key schedule has to be performed for each block (hashing a 1 MByte message would take more than one day).

Further Improvements. It is known that outputting $\alpha(n) = O(\log \log n)$ bits at each iteration of the Blum-Blum-Shub generator is cryptographically secure [159]. However, for a modulus n of given bit length, no explicit range for $\alpha(n)$ was ever given in the literature [114]. Finding such a constant could considerably improve the speed of the key schedule of C.

Another possible improvement to the key schedule would be to rely on some other cryptographically secure pseudo-random generator. The pseudo-random generator on which the stream cipher QUAD [13, 14] is based may be a good candidate: it offers provable security results and achieves speeds up to 5.7Mbits/s. Using such a construction would certainly improve the key schedule time by an important factor, so that the "most secure" version of C might compare to the current version using derandomized substitution boxes.

C vs. the Vernam Cipher. Since we need to assume the independence of the round key bits in our security proof, we have to use a cryptographically secure pseudo-random bit generator to fill the gap between theory and practice. Yet, in the best case, we need to generate approximately 3 000 key bits, which is more than the $2 \cdot 128 = 256$ bits that can be encrypted in a provably secure way. Obviously, one can wonder why not use the Vernam cipher in that case. We note that once the security of the Vernam cipher starts to decrease, it does exponentially. In contrast, it is not clear that 24 queries (that corresponds to more than 3 000 bits) could allow one to distinguish C from C*. Furthermore, even in the case where C is used with a fast key schedule that provides no security guarantee, it is still true in general that an attack that would hold on C with this key schedule but not on C with perfectly random key bits could easily be avoided by simply choosing a stronger key schedule.

11.11 Summary

We have introduced C, a block cipher provably secure against a wide range of attacks. It is as fast as the AES for encryption on a standard workstation. Provable security requires a cryptographically secure key schedule. Consequently, the key schedule of C is too slow for some applications.

As far as we know, C is the first practical block cipher to provide tight security proofs that do take into account the key schedule. It is proved that C resists:

- 2-limited adaptive distinguishers,
- linear cryptanalysis (taking into account the possible cumulative effects of a linear hull),
- differential cryptanalysis (similarly considering cumulative effects of differentials),
- iterated attacks of order 1
- and impossible differentials.

We also give strong evidence that it also resists: algebraic attacks, slide attacks, the boomerang attack, the rectangle attack, differential-linear cryptanalysis, and, to some extent, saturation attacks. From our point of view, the most significant improvement that could be made on C would be to give a bound on the advantage of the best d-limited adversary for $d > 2$.

"Mind you, even I didn't think of that one... extraordinary."
Chief Insp. Hubbard

Chapter 12

KFC: the Krazy Feistel Cipher

In the previous chapter, we presented C, a block cipher construction provably resistant to (among others) linear and differential cryptanalysis (where the linear hull [125] and differentials [97] effects are taken into account, which is unfortunately not usual in typical security proofs of block ciphers), several of their variants, 2-limited distinguishers and thus, all iterated attacks of order 1. Our aim in this chapter, is to design a block cipher based on the same principles as C but provably secure against q-limited distinguishers for large values of q. We call this construction KFC as it is based on a Feistel scheme [50]. KFC is practical in the sense that it can be implemented and reaches a throughput of a few Mbits/s. This is clearly too low for most applications, but maybe not for all of them. Our objective here is to give more weight to the security proofs than to the throughput of the final implementation. Consequently, just as the typical security proofs of block ciphers do not compare to those that KFC enjoys, the encryption speed reached by KFC does not compare to those of nowadays block ciphers.

Instead of first describing KFC and then review all features and security results that we could prove, we use in this chapter a different approach, closer to the time succession of the questions and issues that we raised (and hopefully solved most of the time) during our research. In the first section, we give some hints about why we choose to use a Feistel scheme [50] for KFC. A description of the structure of the random functions we use in the Feistel scheme is then given in Section 12.2 along with the intuitive reason why we choose this one in particular. The exact advantage of the best 2-limited distinguisher is computed in Section 12.3, and in Section 12.4, we show how to bound the advantage of higher order adversaries. Sections 12.5 and 12.6 give implementations results and conclude this chapter.

Throughout this chapter, a *perfectly random function* denotes a random function uniformly distributed among all possible functions on the appropriate sets. Consequently, when referring to a *random function*, nothing is assumed about its distribution. Also, we will not define any key schedule algorithm for KFC. The reasons are twofold. First of all, one could easily adapt the key schedule of C to KFC (for reasons that will be obvious by the end of Section 12.2), except that much more random bits will be necessary, as we will see. Secondly, we hope that the ideas on which the design of KFC relies will lead to new, more effective constructions. We do not expect KFC to be used as-is, although it could be of course. For this last reason, we assume in the whole chapter that all the random functions (F-boxes) and the random permutations (S-boxes) are mutually independent.

12.1 From the SPN of C to the Feistel Network of KFC

The block cipher C (introduced in the previous chapter) achieves goals similar to those we want to achieve with KFC: being resistant to 2-limited adversaries, it is secure against all iterated attacks of order 1. These results are obtained by exploiting strong symmetries (induced by intrinsic symmetries of the confusion and diffusion layers) in the order 2 distribution matrix of C. Unfortunately, we are not able to exhibit similar symmetries for higher orders. It appears that layers of perfectly random permutations are suitable for proving security results at order 2, not above.

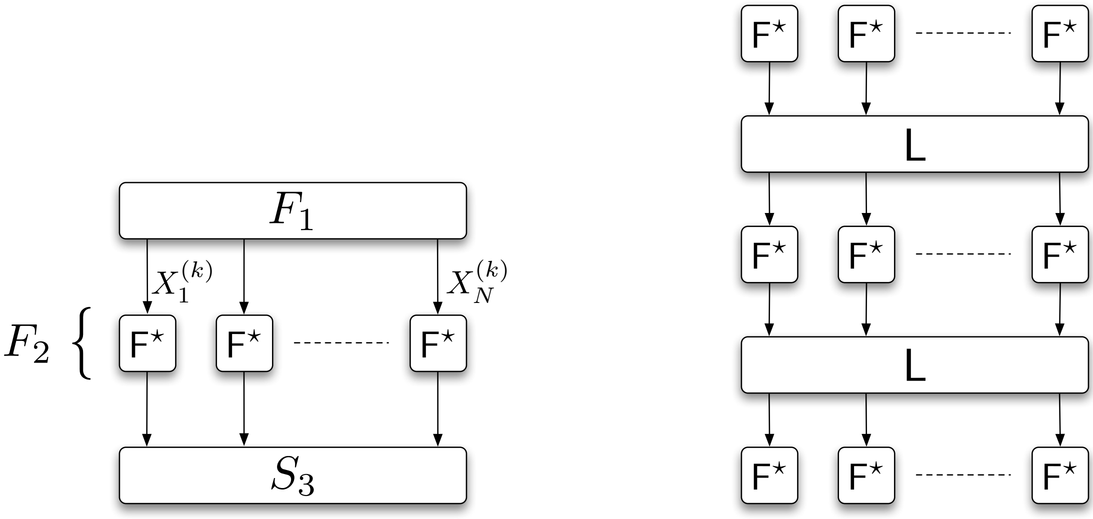
Instead of explicitly computing the advantage of a q-limited distinguisher we will try to bound it by a function of the advantage of the best $(q-1)$-limited distinguisher, and apply this bound recursively down to order 2 (which we know how to compute). This seems clearly impossible with layers of random permutations as two distinct inputs will always lead to two *correlated* outputs. However, this is not the case anymore when considering a layer of mutually independent and perfectly random *functions*. For instance, two distinct inputs of a perfectly random function yield two independent outputs. Similarly, if the two inputs of a layer of functions are distinct on each function input, the outputs are independent. This extends well to a set of q texts: if one text is different from *all* the others on *all* function inputs, the corresponding output is independent from all other outputs. A formal treatment of this idea is given in Section 12.2.

However, layers of random functions cannot always be inverted and thus do not fit in a classical SPN structure. The straightforward solution is to use these functions as the round functions of a Feistel scheme [50]. Moreover, decorrelation results on the round functions of a Feistel scheme extend well to the whole construction. Indeed, Theorem 10.5 shows that if we can instantiate *independent* random functions secure against all q-limited distinguishers, we can obtain a block cipher provably secure against any q-limited distinguisher. In the following sections, we focus on building a round function $\mathsf{F}_{\mathsf{KFC}}$ following the ideas we have introduced here.

12.2 A Good Round Function for the Feistel Scheme

To analyze the behavior of a layer of random functions, we consider the construction $\mathsf{F}_{\mathsf{sff}} : \{0,1\}^n \rightarrow \{0,1\}^n$ defined on binary strings of length $n > 0$ by

$$\mathsf{F}_{\mathsf{sff}} = \mathsf{S}_3 \circ \mathsf{F}_2 \circ \mathsf{F}_1,$$



(a) (b)

Figure 12.1: Increasing the decorrelation order using a layer made of small *independent* and *perfectly random* functions

where $\mathsf{F}_1 : \{0,1\}^n \to \{0,1\}^n$ is a random function, $\mathsf{S}_3 : \{0,1\}^n \to \{0,1\}^n$ is a random permutation, and $\mathsf{F}_2 : \{0,1\}^n \to \{0,1\}^n$ is a layer made of small independent and perfectly random functions on m bits (see Figure 12.1(a)). We therefore assume that $m|n$, i.e., there exists $\ell > 0$ such that $n = \ell \cdot m$. We assume that F_1, F_2, and S_3 are mutually independent. Let A_q denote the best q-limited (adaptive or non-adaptive) distinguisher between $\mathsf{H}_0 : \mathsf{F} = \mathsf{F}^\star$ and $\mathsf{H}_1 : \mathsf{F} = \mathsf{F}_{\mathsf{sff}}$ in the Luby-Rackoff model (see Section 10.1), where $\mathsf{F}^\star : \{0,1\}^n \to \{0,1\}^n$ denotes the uniformly distributed random function from $\{0,1\}^n$ to $\{0,1\}^n$. We obtain an interesting property, making it possible to relate $\mathrm{Adv}_{\mathsf{A}_q}(\mathsf{H}_0, \mathsf{H}_1)$ to $\mathrm{Adv}_{\mathsf{A}_{q-1}}(\mathsf{H}_0, \mathsf{H}_1)$. We consider a set of q inputs of the function $\mathsf{F}_{\mathsf{sff}}$ and denote the corresponding random outputs of F_1 by $X^{(1)}, \ldots, X^{(q)}$, where $X^{(k)} = (X_1^{(k)}, \ldots, X_\ell^{(k)})$ for $k = 1, \ldots, q$. Let e be the event

$$\left\{\exists k \in \{1,\ldots,q\} \text{ s.t. } \forall j \in \{1,\ldots,\ell\} \ : \ X_j^{(k)} \notin \left\{X_j^{(1)}, \ldots, X_j^{(k-1)}, X_j^{(k+1)}, \ldots, X_j^{(1)}\right\}\right\},$$

that is, e is the event that one of the q inputs is different from all the others on the ℓ blocks. If e occurs, at least one of the outputs of the functions layer is a uniformly distributed random variable independent from the others. More formally, we have the following lemma.

Lemma 12.1 *With the notation introduced in this section we have, for all permutations* S_3,

$$\mathrm{Adv}_{\mathsf{A}_q}(\mathsf{H}_0, \mathsf{H}_1) \leq \mathrm{Adv}_{\mathsf{A}_{q-1}}(\mathsf{H}_0, \mathsf{H}_1) + \Pr[\bar{\mathsf{e}}]. \qquad (12.1)$$

Proof. We first note that conditioning the expression of the advantage $\mathrm{Adv}_{\mathsf{A}_q}(\mathsf{H}_0, \mathsf{H}_1)$ by the event e leads to

$$\begin{aligned}&\mathrm{Adv}_{\mathsf{A}_q}(\mathsf{H}_0, \mathsf{H}_1)\\ &= |(\mathrm{Pr}_{\mathsf{H}_1}[\mathsf{A}_q = 1|\mathsf{e}] - \mathrm{Pr}_{\mathsf{H}_0}[\mathsf{A}_q = 1|\mathsf{e}])\mathrm{Pr}[\mathsf{e}] + (\mathrm{Pr}_{\mathsf{H}_1}[\mathsf{A}_q = 1|\bar{\mathsf{e}}] - \mathrm{Pr}_{\mathsf{H}_0}[\mathsf{A}_q = 1|\bar{\mathsf{e}}])\mathrm{Pr}[\bar{\mathsf{e}}]|\\ &\leq |\mathrm{Pr}_{\mathsf{H}_1}[\mathsf{A}_q = 1|\mathsf{e}] - \mathrm{Pr}_{\mathsf{H}_0}[\mathsf{A}_q = 1|\mathsf{e}]| + \mathrm{Pr}[\bar{\mathsf{e}}].\end{aligned}$$

Without loss of generality, we can assume that the adversary does not make the same query twice (as this would not increase its advantage) and that the event e is true for the qth query x_q. This means that $(\mathsf{F}_1(x_q))_j$ is different from all $(\mathsf{F}_1(x_i))_j$ for $i < q$ and $1 \leq j \leq N$ and thus, $(\mathsf{F}_2 \circ \mathsf{F}_1)(x_q)$ is a uniformly distributed random variable independent of $(\mathsf{F}_2 \circ \mathsf{F}_1)(x_i)$ for all $i < q$. As S_3 is a permutation, this property is still true for $(\mathsf{S}_3 \circ \mathsf{F}_2 \circ \mathsf{F}_1)(x_q) = \mathsf{F}_{\mathsf{sff}}(x_q)$. Denoting by Y this random variable we have:

$$\begin{aligned}\mathrm{Pr}[\mathsf{F}_{\mathsf{sff}}(x_1) = y_1, \ldots, \mathsf{F}_{\mathsf{sff}}(x_q) = y_q|\mathsf{e}] &= \mathrm{Pr}[\mathsf{F}_{\mathsf{sff}}(x_1) = y_1, \ldots, \mathsf{F}_{\mathsf{sff}}(x_{q-1}) = y_{q-1}, Y = y_d]\\ &= 2^{-n}\mathrm{Pr}[\mathsf{F}_{\mathsf{sff}}(x_1) = y_1, \ldots, \mathsf{F}_{\mathsf{sff}}(x_{q-1}) = y_{q-1}].\end{aligned}$$

Let $A = |\mathrm{Pr}_{\mathsf{H}_1}[\mathsf{A}_q = 1|\mathsf{e}] - \mathrm{Pr}_{\mathsf{H}_0}[\mathsf{A}_q = 1|\mathsf{e}]|$. Similarly to the proof of Theorem 10 in [155] we know that:

$$A = \frac{1}{2}\max_{x_1}\sum_{y_1}\cdots\max_{x_q}\sum_{y_q}\left|\mathrm{Pr}[\mathsf{F}_{\mathsf{sff}}(x_1) = y_1, \ldots, \mathsf{F}_{\mathsf{sff}}(x_q) = y_q|\mathsf{e}] - 2^{-d\cdot n}\right|.$$

From the two previous equations we obtain that:

$$\begin{aligned}A &= \frac{1}{2}\max_{x_1}\sum_{y_1}\cdots\max_{x_q}\sum_{y_q}2^{-n}\left|\mathrm{Pr}[\mathsf{F}_{\mathsf{sff}}(x_1) = y_1, \ldots, \mathsf{F}_{\mathsf{sff}}(x_{q-1}) = y_{q-1}] - 2^{-(d-1)\cdot n}\right|\\ &= \frac{1}{2}\max_{x_1}\sum_{y_1}\cdots\max_{x_{q-1}}\sum_{y_{q-1}}\left|\mathrm{Pr}[\mathsf{F}_{\mathsf{sff}}(x_1) = y_1, \ldots, \mathsf{F}_{\mathsf{sff}}(x_{q-1}) = y_{q-1}] - 2^{-(d-1)\cdot n}\right|\\ &= \mathrm{Adv}_{\mathsf{A}_{q-1}}(\mathsf{H}_0, \mathsf{H}_1).\end{aligned}$$

□

Why Lemma 12.1 is not Enough. From the previous inequality, it seems natural to consider a substitution-permutation-like construction made of an alternation of layers of independent and perfectly random functions and layers of linear diffusion (as shown on Figure 12.1(b)). Intuitively, one could think that (as it is the case when iterating random permutations) iterating random functions is sufficient to decrease the advantage of a distinguisher. However, this is definitely *not* the case. Indeed, consider a 2-limited attack where the two plaintexts are equal on $\ell - 1$ blocks and different on the last block. There is a non-negligible probability 2^{-m} that, after the first layer of functions, both outputs are completely equal, thus leading to a distinguisher with advantage 2^{-m}. For practical values of m (e.g., $m = 8$), this is not acceptable. Intuitively, this means that we need a good resistance to 2-limited adversaries to initialize the recurrence relation of equation (12.1).

The Sandwich Technique. As proved in Chapter 11, a substitution-permutation network (SPN) made of layers of mutually independent and perfectly random permutations and of well chosen linear diffusion is efficient against 2-limited distinguishers. Intuitively, this means that any set of q inputs will lead to a set of q *pairwise independent* outputs. As we will see in Section 12.4, pairwise independence is exactly what we need to apply the recursive relation (12.1).

For these reasons the construction we choose for $\mathsf{F_{KFC}}$ consists in sandwiching the construction sketched on Figure 12.1(b) between two SPN using layers of mutually independent and perfectly random permutations.

Description of $\mathsf{F_{KFC}}$. The round function $\mathsf{F_{KFC}} : \{0,1\}^n \to \{0,1\}^n$ used in the Feistel scheme defining KFC is based on three different layers:

- a substitution layer S made of ℓ mutually independent and perfectly random m-bit permutations (and thus $n = \ell \cdot m$),
- a function layer F made of ℓ mutually independent and perfectly random m-bit functions,
- a linear layer L which is a $\ell \times \ell$ matrix of elements in $\mathrm{GF}(2^m)$ defining an MDS code (for optimal diffusion), which requires $\ell \leq 2^{m-1}$.

Let r_1 and r_2 be two integers. The round function $\mathsf{F_{KFC}}$ of KFC is defined as:

$$\mathsf{F_{KFC}} = \mathsf{F}_{\mathsf{KFC}[r_1,r_2]} = \mathsf{S} \circ (\mathsf{L} \circ \mathsf{F}^{(r_2)} \circ \cdots \circ \mathsf{L} \circ \mathsf{F}^{(1)}) \circ (\mathsf{L} \circ \mathsf{S}^{(r_1)} \circ \cdots \circ \mathsf{L} \circ \mathsf{S}^{(1)}),$$

where the $\mathsf{S}, \mathsf{S}^{(r_1)}, \ldots, \mathsf{S}^{(1)}, \mathsf{F}^{(r_2)}, \ldots, \mathsf{F}^{(1)}$ are mutually independent.

Description of KFC. The block cipher $\mathsf{KFC} : \{0,1\}^{2n} \to \{0,1\}^{2n}$ is a 3 rounds Feistel scheme where each round function is an independent function corresponding to $\mathsf{F_{KFC}}$.

12.3 Exact Security of $\mathsf{F_{KFC}}$ against 2-limited Adversaries

Shrinking $[\mathsf{F_{KFC}}]^2$

As all layers of $\mathsf{F_{KFC}}$ are mutually independent, then according to Lemma 10.1 the 2-wise distribution matrix $[\mathsf{F_{KFC}}]^2$ can be expressed as

$$\begin{aligned}[\mathsf{F_{KFC}}]^2 &= [\mathsf{S} \circ (\mathsf{L} \circ \mathsf{F}^{(r_2)} \circ \cdots \mathsf{L} \circ \mathsf{F}^{(1)}) \circ (\mathsf{L} \circ \mathsf{S}^{(r_1)} \circ \cdots \mathsf{L} \circ \mathsf{S}^{(1)})]^2 \\ &= ([\mathsf{S}]^2 \times [\mathsf{L}]^2)^{r_1} \times ([\mathsf{F}]^2 \times [\mathsf{L}]^2)^{r_2} \times [\mathsf{S}]^2. \qquad (12.2)\end{aligned}$$

Each of these matrices is a $2^{2n} \times 2^{2n}$ square matrix, which makes direct computations impossible for practical parameters. In the rest of this section we will exploit symmetries in order to reduce the computation to a product of $(\ell+1) \times (\ell+1)$ square matrices.

In the rest of this chapter we considers each element of $\{0,1\}^n$ as a ℓ-tuple of elements in $\{0,1\}^m$. Similarly to what we had in Section 10.5, the support of $a \in \{0,1\}^n$ is the binary ℓ-tuple with 1's at the non-zero positions of a and 0 elsewhere. It is denoted $\textsc{supp}(a)$. The *weight* of the support, denoted $w(\textsc{supp}(a))$ or $w(a)$, is the Hamming weight of the support. When considering a pair $x, x' \in \{0,1\}^n$, the support of the pair is $\textsc{supp}(x \oplus x')$.

Distribution matrices at order 2 are indexed by pairs of texts. Using symmetries at two levels, we will first shrink them to $2^\ell \times 2^\ell$ matrices indexed by supports of pairs and then to $(\ell+1) \times (\ell+1)$ matrices indexed by weights.

Since we assume that the random substitution boxes of the S layers are independent and uniformly distributed, we note that S exactly corresponds to the S-box layer studied in Section 10.5. Using the two matrices PS and SP respectively defined in (10.7) and (10.8), and which verify (according to Lemma 10.3)

$$\mathsf{SP} \times \mathsf{PS} = \mathsf{Id} \quad \text{and} \quad \mathsf{PS} \times \mathsf{SP} = [\mathsf{S}]^2,$$

we note that (12.2) can be written as

$$\begin{aligned}[][\mathsf{F}_{\mathsf{KFC}}]^2 &= (\mathsf{PS} \times \mathsf{SP} \times [\mathsf{L}]^2)^{r_1} \times ([\mathsf{F}]^2 \times [\mathsf{L}]^2)^{r_2} \times [\mathsf{S}]^2 \\ &= \mathsf{PS} \times \overline{\mathsf{L}}^{r_1-1} \times \mathsf{SP} \times [\mathsf{L}]^2 \times ([\mathsf{F}]^2 \times [\mathsf{L}]^2)^{r_2} \times \mathsf{PS} \times \mathsf{SP} \qquad (12.3)\end{aligned}$$

where

$$\overline{\mathsf{L}} = \mathsf{SP} \times [\mathsf{L}]^2 \times \mathsf{PS}. \qquad (12.4)$$

Note that $\overline{\mathsf{L}}$ is a $2^\ell \times 2^\ell$ matrix indexed by supports.

Similarly since we assume that the random function boxes of the F layers are independent and uniformly distributed, we note that F exactly corresponds to the F-box layer studied in Section 10.5. According to Lemma 10.4 we can then write (12.3) as

$$\begin{aligned}[][\mathsf{F}_{\mathsf{KFC}}]^2 &= \mathsf{PS} \times \overline{\mathsf{L}}^{r_1-1} \times \mathsf{SP} \times [\mathsf{L}]^2 \times (\mathsf{PS} \times \overline{\mathsf{F}} \times \mathsf{SP} \times [\mathsf{L}]^2)^{r_2} \times \mathsf{PS} \times \mathsf{SP} \\ &= \mathsf{PS} \times \overline{\mathsf{L}}^{r_1-1} \times \mathsf{SP} \times [\mathsf{L}]^2 \times \mathsf{PS} \times (\overline{\mathsf{F}} \times \mathsf{SP} \times [\mathsf{L}]^2 \times \mathsf{PS})^{r_2} \times \mathsf{SP} \\ &= \mathsf{PS} \times \overline{\mathsf{L}}^{r_1} \times (\overline{\mathsf{F}} \times \overline{\mathsf{L}})^{r_2} \times \mathsf{SP} \qquad (12.5)\end{aligned}$$

where $\overline{\mathsf{F}}$ is a $2^\ell \times 2^\ell$ matrix indexed by supports and defined by

$$\overline{\mathsf{F}}_{\gamma,\gamma'} = \mathbf{1}_{\gamma' \subseteq \gamma} s^{-w(\gamma)} (s-1)^{w(\gamma')},$$

where $s = 2^m$. To simplify (12.5) we will now take advantage of the fact that L is a linear multipermutation. Starting from the definition of $\overline{\mathsf{L}}$ in (12.4) we have

$$\begin{aligned}\overline{\mathsf{L}}_{\gamma,\gamma'} &= \sum_{(x,x')} \sum_{(y,y')} \mathsf{SP}_{\gamma,(x,x')} [\mathsf{L}]^2_{(x,x'),(y,y')} \mathsf{PS}_{(y,y'),\gamma'} \\ &= s^{-\ell} (s-1)^{-w(\gamma)} \sum_{(x,x')} \mathbf{1}_{\gamma = \textsc{supp}(x \oplus x')} \mathbf{1}_{\gamma' = \textsc{supp}(\mathsf{L}(x \oplus x'))} \\ &= (s-1)^{-w(\gamma)} \sum_{x} \mathbf{1}_{\gamma = \textsc{supp}(x)} \mathbf{1}_{\gamma' = \textsc{supp}(\mathsf{L}(x))}.\end{aligned}$$

The sum in this equation is the number of texts of a given support γ that are mapped by the MDS linear layer L on a text of support γ'. According to Theorem 11.2, the number of codewords with given supports can be explicitly computed for any MDS code and, amazingly, only depends on the weights of the supports γ and γ'. If we let $s = 2^m$ and denote by $\mathcal{C}$ the MDS code defined by the linear diffusion of KFC, we obtain that

$$\overline{\mathsf{L}}_{\gamma,\gamma'} = (s-1)^{-w(\gamma)} \frac{\mathcal{W}_{\mathcal{C}}(w(\gamma)+w(\gamma'))}{\binom{2\ell}{w(\gamma)+w(\gamma')}}, \tag{12.6}$$

where $\mathcal{W}_{\mathcal{C}}(i) = \binom{2\ell}{i} \sum_{j=\ell+1}^{i} \binom{i}{j} (-1)^{i-j} (s^{j-\ell}-1)$ for $i > \ell$, $\mathcal{W}_{\mathcal{C}}(0) = 1$, and $\mathcal{W}_{\mathcal{C}}(i) = 0$ for $0 < i \leq \ell$. As the previous equation only depends on the weights of γ and γ', we naturally define to new transition matrices[1] WS and SW from support to weight and from weight to support respectively:

$$\mathsf{SW}_{\gamma,w} = \mathbf{1}_{w(\gamma)=w} \quad \text{and} \quad \mathsf{WS}_{w,\gamma} = \mathbf{1}_{w(\gamma)=w} \binom{\ell}{w}^{-1} \tag{12.7}$$

where $\gamma \in \{0,1\}^\ell$ and $w \in \{0,1,\ldots,\ell\}$. Note that

$$\mathsf{WS} \times \mathsf{SW} = \mathsf{Id}. \tag{12.8}$$

Lemma 12.2 *Let* $\overline{\mathsf{M}}$ *be* $2^{16} \times 2^{16}$ *matrix indexed by supports, such that there exists a* $5^4 \times 5^4$ *matrix* $\overline{\overline{\mathsf{M}}}$ *indexed by 4-tuple of weights in* $\{0,\ldots,4\}$ *verifying*

$$\overline{\mathsf{M}} = \mathsf{SW} \times \overline{\overline{\mathsf{M}}} \times \mathsf{WS}$$

where SW *and* WS *are defined in* (12.7)*. Then*

$$|||\overline{\mathsf{M}}|||_\infty = |||\overline{\overline{\mathsf{M}}}|||_\infty.$$

Proof. By definition,

$$|||\overline{\mathsf{M}}|||_\infty = \max_\gamma \sum_{\gamma'} \sum_{w,w'} \mathbf{1}_{w(\gamma)=w} \mathbf{1}_{w(\gamma')=w'} \overline{\overline{\mathsf{M}}}_{w,w'} \binom{\ell}{w'}^{-1} = \max_\gamma \sum_{w'} \overline{\overline{\mathsf{M}}}_{w(\gamma),w'}$$

from which we easily conclude. □

Starting from the expression we obtained for $\overline{\mathsf{L}}$ in (12.6), and using the two new transition matrices, it is easy to see that

$$\begin{aligned} \overline{\mathsf{L}}_{\gamma,\gamma'} &= \sum_{w,w'} \mathbf{1}_{w(\gamma)=w} \mathbf{1}_{w(\gamma')=w'} (s-1)^{-w} \frac{\mathcal{W}_{\mathcal{C}}(w+w')}{\binom{2\ell}{w+w'}} \\ &= \sum_{w,w'} \mathsf{SW}_{\gamma,w} (s-1)^{-w} \frac{\mathcal{W}_{\mathcal{C}}(w+w')}{\binom{2\ell}{w+w'}} \binom{\ell}{w'} \mathsf{WS}_{w',\gamma'}. \end{aligned}$$

[1] Note that even though the notations are the same than the transition matrices used in the previous chapter, the definition differ.

Defining the $(\ell+1)\times(\ell+1)$ matrix $\overline{\overline{\mathsf{L}}}$ by

$$\overline{\overline{\mathsf{L}}}_{w,w'} = (s-1)^{-w}\frac{\mathcal{W}_{\mathcal{C}}(w+w')}{\binom{2\ell}{w+w'}}\binom{\ell}{w'} \tag{12.9}$$

for all $w, w' \in \{0,\ldots,\ell\}$, the previous expression reads

$$\overline{\mathsf{L}} = \mathsf{SW}\times\overline{\overline{\mathsf{L}}}\times\mathsf{WS}. \tag{12.10}$$

Noting that using (12.8) we have

$$\overline{\mathsf{L}}^{r_1} = (\mathsf{SW}\times\overline{\overline{\mathsf{L}}}\times\mathsf{WS})^{r_1} = \mathsf{SW}\times(\overline{\overline{\mathsf{L}}}\times\underbrace{\mathsf{WS}\times\mathsf{SW}}_{=\mathsf{Id}})^{r_1-1}\times\overline{\overline{\mathsf{L}}}\times\mathsf{WS} = \mathsf{SW}\times\overline{\overline{\mathsf{L}}}^{r_1}\times\mathsf{WS},$$

we can deduce from the discussion on $\overline{\overline{\mathsf{L}}}$ that (12.5) can written as

$$\begin{aligned}[\mathsf{F}_{\mathsf{KFC}}]^2 &= \mathsf{PS}\times\mathsf{SW}\times\overline{\overline{\mathsf{L}}}^{r_1}\times\mathsf{WS}\times(\overline{\mathsf{F}}\times\mathsf{SW}\times\overline{\overline{\mathsf{L}}}\times\mathsf{WS})^{r_2}\times\mathsf{SP}\\ &= \mathsf{PS}\times\mathsf{SW}\times\overline{\overline{\mathsf{L}}}^{r_1}\times(\overline{\overline{\mathsf{F}}}\times\overline{\overline{\mathsf{L}}})^{r_2}\times\mathsf{WS}\times\mathsf{SP},\end{aligned}$$

where $\overline{\overline{\mathsf{F}}}$ is the $(\ell+1)\times(\ell+1)$ matrix indexed by weights such that

$$\overline{\overline{\mathsf{F}}} = \mathsf{WS}\times\overline{\mathsf{F}}\times\mathsf{SW}.$$

We can obtain a closed formula for $\overline{\overline{\mathsf{F}}}$ since for all $w, w' \in \{0,\ldots,\ell\}$ we have

$$\overline{\overline{\mathsf{F}}}_{w,w'} = \sum_{\gamma,\gamma'}\mathsf{WS}_{w,\gamma}\overline{\mathsf{F}}_{\gamma,\gamma'}\mathsf{SW}_{\gamma',w'} = \binom{\ell}{w}^{-1}s^{-w}(s-1)^{w'}\sum_{\gamma,\gamma'}\mathbf{1}_{w(\gamma)=w}\mathbf{1}_{w(\gamma')=w'}\mathbf{1}_{\gamma'\subseteq\gamma}$$

where

$$\sum_{\gamma,\gamma'}\mathbf{1}_{w(\gamma)=w}\mathbf{1}_{w(\gamma')=w'}\mathbf{1}_{\gamma'\subseteq\gamma} = \mathbf{1}_{w'\leq w}\sum_{\gamma}\mathbf{1}_{w(\gamma)=w}\sum_{\gamma'}\mathbf{1}_{w(\gamma')=w'}\mathbf{1}_{\gamma'\subseteq\gamma} = \mathbf{1}_{w'\leq w}\binom{\ell}{w}\binom{w}{w'}.$$

We summarize our results in the following theorem.

Theorem 12.1 *With the notations used in this section, the* 2*-wise distribution matrix* $[\mathsf{F}_{\mathsf{KFC}}]^2$ *of the round function* $\mathsf{F}_{\mathsf{KFC}}$ *can be written as*

$$[\mathsf{F}_{\mathsf{KFC}}]^2 = \mathsf{PS}\times\mathsf{SW}\times\overline{\overline{\mathsf{L}}}^{r_1}\times(\overline{\overline{\mathsf{F}}}\times\overline{\overline{\mathsf{L}}})^{r_2}\times\mathsf{WS}\times\mathsf{SP}, \tag{12.11}$$

where PS*,* SP*,* SW*, and* SW *and the four transition matrices respectively defined in* (10.7)*,* (10.8)*, and* (12.7)*, and where both* $\overline{\overline{\mathsf{F}}}$ *and* $\overline{\overline{\mathsf{L}}}$ *are* $(\ell+1)\times(\ell+1)$ *matrices indexed by weights and respectively defined by*

$$\overline{\overline{\mathsf{F}}}_{w,w'} = s^{-w}(s-1)^{w'}\mathbf{1}_{w'\leq w}\binom{w}{w'}$$

and

$$\overline{\overline{\mathsf{L}}}_{w,w'} = (s-1)^{-w}\frac{\mathcal{W}_{\mathcal{C}}(w+w')}{\binom{2\ell}{w+w'}}\binom{\ell}{w'}$$

for all $w, w' \in \{0,\ldots,\ell\}$*, where* $\mathcal{W}_{\mathcal{C}}(i) = \binom{2\ell}{i}\sum_{j=\ell+1}^{i}\binom{i}{j}(-1)^{i-j}(s^{j-\ell}-1)$ *for* $i > \ell$*,* $\mathcal{W}_{\mathcal{C}}(0) = 1$*, and* $\mathcal{W}_{\mathcal{C}}(i) = 0$ *for* $0 < i \leq \ell$*.*

Practical Computation of the Best Advantage

The expression we just obtained for $[\mathsf{F_{KFC}}]^2$ leads to a simple practical expression for $\|[\mathsf{F_{KFC}}]^2 - [\mathsf{F}^\star]^2\|_a$. We first note that

$$\begin{aligned}
[\mathsf{F}^\star]^2_{(x,x'),(y,y')} &= s^{-\ell}(\mathbf{1}_{x\oplus x'\neq 0}s^{-\ell} + \mathbf{1}_{x\oplus x'=0}\mathbf{1}_{y\oplus y'=0}) \\
&= s^{-\ell}(\mathbf{1}_{w(x\oplus x')\neq 0}s^{-\ell} + \mathbf{1}_{w(x\oplus x')=0}\mathbf{1}_{w(y\oplus y')=0}) \\
&= \sum_{\gamma,\gamma'} \mathsf{PS}_{(x,x'),\gamma}\underbrace{(\mathbf{1}_{w(\gamma)\neq 0}s^{-\ell} + \mathbf{1}_{w(\gamma)=0}\mathbf{1}_{w(\gamma')=0})(s-1)^{w(\gamma')}}_{\overline{\mathsf{M}}_{\gamma,\gamma'}}\mathsf{SP}_{\gamma',(y,y')}
\end{aligned}$$

where $\overline{\mathsf{M}}_{\gamma,\gamma'}$ further simplifies to

$$\begin{aligned}
\overline{\mathsf{M}}_{\gamma,\gamma'} &= \sum_{w,w'} \mathbf{1}_{w=w(\gamma)}\mathbf{1}_{w'=w(\gamma')}(\mathbf{1}_{w\neq 0}s^{-\ell} + \mathbf{1}_{w=0}\mathbf{1}_{w'=0})(s-1)^{w'} \\
&= \sum_{w,w'} \mathsf{SW}_{\gamma,w}\underbrace{(\mathbf{1}_{w\neq 0}s^{-\ell} + \mathbf{1}_{w=0}\mathbf{1}_{w'=0})(s-1)^{w'}\binom{\ell}{w'}}_{\overline{\overline{\mathsf{M}}}_{w,w'}}\mathsf{WS}_{w',\gamma'},
\end{aligned}$$

so that

$$[\mathsf{F}^\star]^2 = \mathsf{PS}\times\overline{\mathsf{M}}\times\mathsf{SP} = \mathsf{PS}\times\mathsf{SW}\times\overline{\overline{\mathsf{M}}}\times\mathsf{WS}\times\mathsf{SP}. \tag{12.12}$$

Theorem 12.2 *Let r_1 and r_2 be two positive integers. The respective advantages of the best 2-limited non-adaptive adversary A_{na} and of the best 2-limited adaptive adversary A_{a} against $\mathsf{F_{KFC}} = \mathsf{F_{KFC[r_1,r_2]}} = \mathsf{S}\circ(\mathsf{L}\circ\mathsf{F})^{r_2}\circ(\mathsf{L}\circ\mathsf{S})^{r_1}$ are*

$$\mathrm{Adv}_{\mathsf{A}_{\mathrm{na}}}(\mathsf{F_{KFC}},\mathsf{F}^\star) = \mathrm{Adv}_{\mathsf{A}_{\mathrm{a}}}(\mathsf{F_{KFC}},\mathsf{F}^\star) = \frac{1}{2}|||\overline{\overline{\mathsf{L}}}^{r_1}\times(\overline{\overline{\mathsf{F}}}\times\overline{\overline{\mathsf{L}}})^{r_2} - \overline{\overline{\mathsf{M}}}|||_\infty,$$

where $\overline{\overline{\mathsf{L}}}$, $\overline{\overline{\mathsf{F}}}$, and $\overline{\overline{\mathsf{M}}}$ are three $(\ell+1)\times(\ell+1)$ matrices indexed by weights and respectively defined for all w, $w'\in\{0,\ldots,\ell\}$ by

$$\overline{\overline{\mathsf{L}}}_{w,w'} = (s-1)^{-w}\frac{\mathcal{W}_{\mathcal{C}}(w+w')}{\binom{2\ell}{w+w'}}\binom{\ell}{w'}, \quad \overline{\overline{\mathsf{F}}}_{w,w'} = s^{-w}(s-1)^{w'}\mathbf{1}_{w'\leq w}\binom{w}{w'},$$

and

$$\overline{\overline{\mathsf{M}}}_{w,w'} = \mathbf{1}_{w\neq 0}s^{-\ell}(s-1)^{w'}\binom{\ell}{w'} + \mathbf{1}_{w=0}\mathbf{1}_{w'=0},$$

where $s = 2^m$, $\mathcal{W}_{\mathcal{C}}(i) = \binom{2\ell}{i}\sum_{j=\ell+1}^{i}\binom{i}{j}(-1)^{i-j}(s^{j-\ell}-1)$ for $i>\ell$, $\mathcal{W}_{\mathcal{C}}(0)=1$, and $\mathcal{W}_{\mathcal{C}}(i)=0$ for $0<i\leq\ell$.

Proof. According to Theorem 10.1 we known that

$$\mathrm{Adv}_{\mathrm{a}}(\mathsf{F_{KFC}},\mathsf{F}^\star) = \frac{1}{2}\|[\mathsf{F_{KFC}}]^2 - [\mathsf{F}^\star]^2\|_2 \quad\text{and}\quad \mathrm{Adv}_{\mathrm{na}}(\mathsf{F_{KFC}},\mathsf{F}^\star) = \frac{1}{2}|||[\mathsf{F_{KFC}}]^2 - [\mathsf{F}^\star]^2|||_\infty.$$

	$\ell = 8$ and $s = 2^8$				$\ell = 8$ and $s = 2^{16}$				$\ell = 16$ and $s = 2^8$			
$r_1 \backslash r_2$	0	1	10	100	0	1	10	100	0	1	10	100
0	1	2^{-5}	2^{-8}	2^{-8}	1	2^{-13}	2^{-16}	2^{-16}	1	2^{-4}	2^{-8}	2^{-8}
1	2^{-5}	2^{-50}	2^{-52}	2^{-49}	2^{-13}	2^{-114}	2^{-116}	2^{-113}	2^{-4}	2^{-95}	2^{-104}	2^{-103}
2	2^{-46}	2^{-53}	2^{-52}	2^{-49}	2^{-110}	2^{-117}	2^{-116}	2^{-113}	2^{-87}	2^{-104}	2^{-104}	2^{-103}
3	2^{-62}	2^{-53}	2^{-52}	2^{-49}	2^{-128}	2^{-117}	2^{-116}	2^{-113}	2^{-120}	2^{-104}	2^{-104}	2^{-103}

Table 12.1: Advantage of the best 2-limited distinguisher against F_{KFC}.

From the expression (12.11) we obtained for $[\mathsf{F}_{\mathsf{KFC}}]^2$ in Theorem 12.1 and the expression of $[\mathsf{F}^\star]^2$ in (12.12) we see that

$$[\mathsf{F}_{\mathsf{KFC}}]^2 - [\mathsf{F}^\star]^2 = \mathsf{PS} \times \mathsf{SW} \times \left(\overline{\overline{\mathsf{L}}}^{r_1} \times (\overline{\overline{\mathsf{F}}} \times \overline{\overline{\mathsf{L}}})^{r_2} - \overline{\overline{\mathsf{M}}}\right) \times \mathsf{WS} \times \mathsf{SP}.$$

We can easily deduce the announced result from lemmas 10.5 and 12.2. □

Explicit values of this advantage for some typical values of ℓ, s, r_1 and r_2 are given in Table 12.1 and were computed using Maple [106]. We note that $r_1 = 3$ is enough (at least for these parameters). Moreover, the advantage increases with the value of r_2. The reason is that the more F layers there is, the higher is the probability of an internal collision.

12.4 Bounding the Security of $\mathsf{F}_{\mathsf{KFC}}$ against Adversaries of Higher Order

Replacing F by F ∘ S

To simplify the proofs, we will replace each F layer of $\mathsf{F}_{\mathsf{KFC}}$ by $\mathsf{F} \circ \mathsf{S}$. Both constructions are completely equivalent in the sense that any decorrelation result holding for the latter also holds for the original construction, as shown in the following lemma.

Lemma 12.3 *Let $q > 0$ be a positive integer. Letting S and F respectively be the substitution and the function layers of* KFC*, we have*

$$[\mathsf{F} \circ \mathsf{S}]^q = [\mathsf{F}]^q.$$

Proof. For any $x = (x_1, \ldots, x_q), y = (y_1, \ldots, y_q) \in \{0,1\}^{nq}$ we have:

$$\begin{aligned} [\mathsf{F} \circ \mathsf{S}]^q_{(x,y)} &= \Pr[(x_1, \ldots, x_q) \xrightarrow{\mathsf{F} \circ \mathsf{S}} (y_1, \ldots, y_q)] \\ &= \prod_{i=1}^{\ell} \Pr[(x_{1,i}, \ldots, x_{q,i}) \xrightarrow{\mathsf{F}^\star \circ \mathsf{C}^\star} (y_{1,i}, \ldots, y_{q,i})] \end{aligned}$$

where $\mathsf{F}^\star$ (resp. $\mathsf{C}^\star$) denotes the uniformly distributed random function (resp. permutation) from $\{0,1\}^m$ to $\{0,1\}^m$. Consequently, we have

$$\begin{aligned}[\mathsf{F} \circ \mathsf{S}]^q_{(x,y)} &= \prod_{i=1}^{\ell} \frac{1}{2^m!} \sum_{\mathsf{c}} \Pr[(\mathsf{c}(x_{1,i}), \ldots, \mathsf{c}(x_{q,i})) \xrightarrow{\mathsf{F}^\star} (y_{1,i}, \ldots, y_{q,i})] \\ &= \prod_{i=1}^{\ell} \Pr[(x_{1,i}, \ldots, x_{q,i}) \xrightarrow{\mathsf{F}^\star} (y_{1,i}, \ldots, y_{q,i})]\end{aligned}$$

since the probability that $(u_1, u_2, \ldots, u_q) \xrightarrow{\mathsf{F}^\star} (v_1, v_2, \ldots, v_q)$ does not depend on the particular values of the u_i's but on how many distinct values there are in the set, which is not changed by applying a permutation c. It follows that

$$[\mathsf{F} \circ \mathsf{S}]^q_{(x,y)} = \Pr[(x_1, \ldots, x_q) \xrightarrow{\mathsf{F}} (y_1, \ldots, y_q)] = [\mathsf{F}]^q_{(x,y)}.$$

□

From now on, we study the construction $\mathsf{F}_{\mathsf{KFC}} = \mathsf{F}_{\mathsf{KFC}[r_1,r_2]}$ defined by

$$\mathsf{F}_{\mathsf{KFC}} = \mathsf{S} \circ (\mathsf{L} \circ \mathsf{F}^{(r_2)} \circ \mathsf{S}^{(r_1+r_2)} \circ \cdots \circ \mathsf{L} \circ \mathsf{F}^{(1)} \circ \mathsf{S}^{(r_1+1)}) \circ (\mathsf{L} \circ \mathsf{S}^{(r_1)} \circ \cdots \circ \mathsf{L} \circ \mathsf{S}^{(1)}), \quad (12.13)$$

which is completely equivalent (in terms of security) than the original definition of $\mathsf{F}_{\mathsf{KFC}}$.

Assumption 12.1 For $r_1 > 2$, any $i \in \{0, \ldots, r_2\}$ and any 2-limited distinguisher $\mathbb{A}_2$, we have

$$\mathrm{Adv}_{\mathbb{A}_2}(\mathsf{F}_{\mathsf{KFC}[r_1,r_2]}, \mathsf{F}^\star) \geq \mathrm{Adv}_{\mathbb{A}_2}(\mathsf{F}_{\mathsf{KFC}[r_1,i]}, \mathsf{F}^\star).$$

This assumption seems natural from Table 12.1, although it might prove wrong in the general case (in particular, the threshold for r_1 might be different for other values of ℓ and s). However, we experimentally verified it for all values of the parameters we consider in the rest of this chapter.

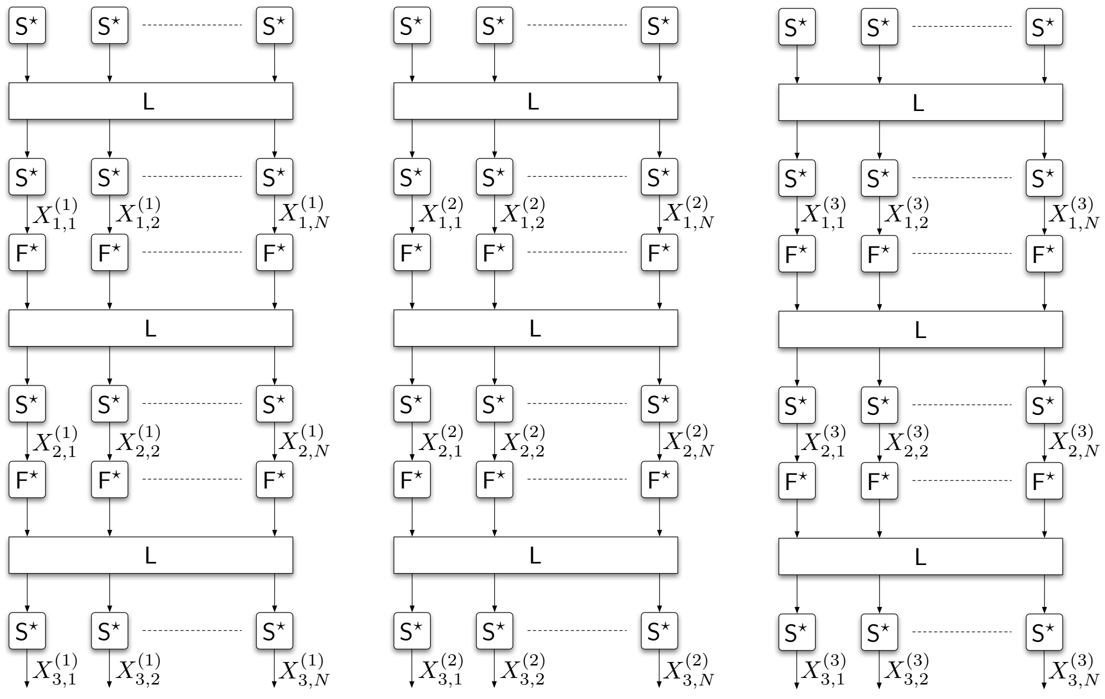
In practice, when the advantage of the best 2-limited distinguisher against $\mathsf{F}_{\mathsf{KFC}}$ is negligible, Assumption 12.1 means that this is also the case before any F layer. The inputs of any F layer can thus almost be considered as *pairwise independent*.

Taking Advantage of the Pairwise Independence

For all $i \in \{0, \ldots, r_2\}$ we denote by

$$X_i^{(k)} = (X_{i,1}^{(k)}, X_{i,2}^{(k)}, \ldots, X_{i,\ell}^{(k)}) \quad (12.14)$$

the output of $\mathsf{F}_{\mathsf{KFC}[r_1,i]}$ when the input corresponds to the kth query (so that $k = 1, 2, \ldots, q$). This notation is illustrated on Figure 12.2. Under this notation, we obtain the following lemma, which is a direct consequence of Lemma 12.1 in Section 12.2.


S⋆
L
F⋆
$X_{1,1}^{(1)}$ $X_{1,2}^{(1)}$ $X_{1,N}^{(1)}$
$X_{2,1}^{(1)}$ $X_{2,2}^{(1)}$ $X_{2,N}^{(1)}$
$X_{3,1}^{(1)}$ $X_{3,2}^{(1)}$ $X_{3,N}^{(1)}$
$X_{1,1}^{(2)}$ $X_{1,2}^{(2)}$ $X_{1,N}^{(2)}$
$X_{2,1}^{(2)}$ $X_{2,2}^{(2)}$ $X_{2,N}^{(2)}$
$X_{3,1}^{(2)}$ $X_{3,2}^{(2)}$ $X_{3,N}^{(2)}$
$X_{1,1}^{(3)}$ $X_{1,2}^{(3)}$ $X_{1,N}^{(3)}$
$X_{2,1}^{(3)}$ $X_{2,2}^{(3)}$ $X_{2,N}^{(3)}$
$X_{3,1}^{(3)}$ $X_{3,2}^{(3)}$ $X_{3,N}^{(3)}$

First query Second query Third query

Figure 12.2: Illustration of the notation in (12.14) for $r_1 = 1$, $r_2 = 2$, and $q = 3$

Lemma 12.4 *Let $r_1, r_2 > 0$ be two positive integers and let $\mathsf{F}_{\mathsf{KFC}[r_1,r_2]}$ be as in* (12.13). *For all $q > 1$ and all $i \in \{1, \ldots, r_2\}$ we have*

$$\mathrm{Adv}_{\mathsf{A}_q}(\mathsf{F}_{\mathsf{KFC}[r_1,i]}, \mathsf{F}^\star) \leq \mathrm{Adv}_{\mathsf{A}_{q-1}}(\mathsf{F}_{\mathsf{KFC}[r_1,i]}, \mathsf{F}^\star) + \Pr[\bar{\mathsf{e}}_i],$$

where e_i is the event that one of the q inputs is different from all others on the ℓ blocks at the output of $\mathsf{F}_{\mathsf{KFC}[r_1,i-1]}$, i.e.,

$$\mathsf{e}_i = \Big\{ \exists k \in \{1, \ldots, q\} \text{ s.t. } \forall j \in \{1, \ldots, \ell\} \ : \\ X_{i,j}^{(k)} \notin \Big\{ X_{i,j}^{(1)}, \ldots, X_{i,j}^{(k-1)}, X_{i,j}^{(k+1)}, \ldots, X_{i,j}^{(1)} \Big\} \Big\}.$$

Proof. Let $i \in \{1, \ldots, r_2\}$. Applying Lemma 12.1 with $\mathsf{e} = \mathsf{e}_i$, $\mathsf{F}_1 = \mathsf{F}_{\mathsf{KFC}[r_1,i-1]}$, $\mathsf{F}_2 = \mathsf{F}$, and $\mathsf{S}_3 = \mathsf{S} \circ \mathsf{L}$ allows to conclude. □

From the previous lemma, we see that bounding $\Pr[\bar{\mathsf{e}}_i]$ will allow us to recursively bound $\mathrm{Adv}_{\mathsf{A}_q}(\mathsf{F}_{\mathsf{KFC}[r_1,i]}, \mathsf{F}^\star)$. Before we try to bound $\Pr[\bar{\mathsf{e}}_i]$, we first prove a very simple lemma, which shows that if the advantage of the best 2-limited adversary against $\mathsf{F}_{\mathsf{KFC}}$ is negligible, then the probability that two given m-bit outputs (among the N possible) are equal is close to 2^{-m}.

Lemma 12.5 *Let $r_1, r_2 > 0$ be two positive integers, let $i \in \{0, \dots, r_2\}$ and let $\mathsf{F}_{\mathsf{KFC}[r_1,i]}$ be as in* (12.13)*. Let*

$$\epsilon = \max_{0 \le i \le r_2} \mathrm{BestAdv}^2(\mathsf{F}_{\mathsf{KFC}[r_1,i]}, \mathsf{F}^\star)$$

be the maximum advantage over all rounds of the best 2-limited adversary. With the notations used in (12.14) *we have for all $k \neq k'$, $i \in \{0, \dots, r_2\}$, and $j \in \{1, \dots, N\}$:*

$$\Pr[X_{i,j}^{(k)} = X_{i,j}^{(k')}] \le \frac{1}{2^m} + \epsilon.$$

Proof. Let A be the 2-limited distinguisher against $\mathsf{F}_{\mathsf{KFC}[r_1,i]}$ which outputs 1 when $X_{i,j}^{(k)} = X_{i,j}^{(k')}$ and 0 otherwise. By assumption, its advantage

$$\mathrm{Adv}_{\mathsf{A}}(\mathsf{F}_{\mathsf{KFC}[r_1,i]}, \mathsf{F}^\star) = \left|\Pr_{\mathsf{F}=\mathsf{KFC}[r_1,i]}[\mathsf{A} = 1] - \Pr_{\mathsf{F}=\mathsf{F}^\star}[\mathsf{A} = 1]\right|$$

is bounded by ϵ. By definition we have $\Pr_{\mathsf{F}=\mathsf{KFC}[r_1,i]}[\mathsf{A} = 1] = \Pr[X_{i,j}^{k} = X_{i,j}^{k'}]$ and, on the other hand, $\Pr_{\mathsf{F}=\mathsf{F}^\star}[\mathsf{A} = 1] = \Pr[U = U']$ where U and U' are two independent m-bit uniformly distributed random strings. This allows to conclude. □

Lemma 12.6 *With the notations of lemmas 12.4 and 12.5, we have that*

$$\Pr[\bar{\mathsf{e}}_i] \le 1 - \left(1 - (q-1)\left(\frac{1}{2^m} + \epsilon\right)\right)^\ell \tag{12.15}$$

for all $i \in \{0, \dots, r_2\}$.

Proof. For all $i \in \{0, \dots, r_2\}$, let λ_i denote the number of $X_i^{(k)}$'s different from all other texts on all ℓ blocks, i.e.,

$$\lambda_i = \sum_{k=1}^{q} \prod_{j=1}^{\ell} \prod_{\substack{k'=1 \\ k' \neq k}}^{q} \mathbf{1}_{X_{i,j}^{(k)} \neq X_{i,j}^{(k')}}.$$

Using the linearity of the mean and the independence between the ℓ blocks we obtain

$$\mathrm{E}(\lambda_i) = q \cdot \left(\Pr[X_{i,1}^{(1)} \notin \{X_{i,1}^{(2)}, X_{i,1}^{(3)}, \dots, X_{i,1}^{(q)}\}]\right)^\ell.$$

Letting $P_q = \Pr[X_{i,1}^{(1)} \notin \{X_{i,1}^{(2)}, X_{i,1}^{(3)}, \dots, X_{i,1}^{(q)}\}]$, it is easy to show by induction on q that $P_q \ge 1 - (q-1)(\frac{1}{2^m} + \epsilon)$. For $q = 1$ the result is trivial. Assume that the result is true for some arbitrary q. Using the result of Lemma 12.5:

$$\begin{aligned} P_{q+1} &= P_q - \Pr[X_{i,1}^{(1)} \notin \{X_{i,1}^{(2)}, X_{i,1}^{(3)}, \ldots, X_{i,1}^{(q)}\}, X_{i,1}^{(1)} = X_{i,1}^{(q+1)}] \\ &\geq P_q - \Pr[X_{i,1}^{(1)} = X_{i,1}^{(q+1)}] \geq P_q - \left(\frac{1}{2^m} + \epsilon\right), \end{aligned}$$

from which we conclude that $P_q \geq 1 - (q-1)\left(\frac{1}{2^m} + \epsilon\right)$ for all $q > 0$. From this and from the expression we obtained for $\mathrm{E}(\lambda_i)$ we deduce that

$$\mathrm{E}(\lambda_i) \geq q \cdot \left(1 - \frac{q-1}{2^m}\right)^{\ell}. \qquad (12.16)$$

Since we also have that

$$\mathrm{E}(\lambda_i) = \sum_{k=1}^{q} k \Pr[\lambda_i = k] \leq q \cdot \Pr[\lambda_i \neq 0] = q \cdot \Pr[\mathsf{e}_i],$$

we deduce from (12.16) that

$$\Pr[\overline{\mathsf{e}}_i] \leq 1 - \left(1 - (q-1)\left(\frac{1}{2^m} + \epsilon\right)\right)^{\ell}.$$

□

Based on lemmas 12.4 and 12.6 it is now possible to upper-bound the advantage of the best q-limited distinguisher for $q > 2$.

Theorem 12.3 *Let $r_1, r_2 > 0$ be two positive integers. For any positive integer $q > 1$, the advantage of the best q-limited distinguisher A_q between $\mathsf{F}_{\mathsf{KFC}[r_1,r_2]}$ and $\mathsf{F}^\star$ is such that*

$$\mathrm{Adv}_{\mathsf{A}_q}(\mathsf{F}_{\mathsf{KFC}[r_1,r_2]}, \mathsf{F}^\star) \leq \epsilon + \sum_{i=2}^{q-1} \left(1 - \left(1 - i\left(\tfrac{1}{2^m} + \epsilon\right)\right)^{\ell}\right),$$

where $\epsilon = \max_{0 \leq i \leq r_2} \mathrm{BestAdv}^2(\mathsf{F}_{\mathsf{KFC}[r_1,i]}, \mathsf{F}^\star)$.

Proof. Using the results obtained in lemmas 12.4 and 12.6 successively, we get

$$\begin{aligned} \mathrm{Adv}_{\mathsf{A}_q}(\mathsf{F}_{\mathsf{KFC}[r_1,r_2]}, \mathsf{F}^\star) &\leq \mathrm{Adv}_{\mathsf{A}_{q-1}}(\mathsf{F}_{\mathsf{KFC}[r_1,r_2]}, \mathsf{F}^\star) + \Pr[\overline{\mathsf{e}}_{r_2}] \\ &\leq \mathrm{Adv}_{\mathsf{A}_{q-1}}(\mathsf{F}_{\mathsf{KFC}[r_1,r_2]}, \mathsf{F}^\star) + 1 - \left(1 - (q-1)\left(\tfrac{1}{2^m} + \epsilon\right)\right)^{\ell}. \end{aligned}$$

Applying the same two steps recursively we get

$$\mathrm{Adv}_{\mathsf{A}_q}(\mathsf{F}_{\mathsf{KFC}[r_1,r_2]}, \mathsf{F}^\star) \leq \mathrm{Adv}_{\mathsf{A}_2}(\mathsf{F}_{\mathsf{KFC}[r_1,r_2]}, \mathsf{F}^\star) + \sum_{i=2}^{q-1} \left(1 - \left(1 - i\left(\tfrac{1}{2^m} + \epsilon\right)\right)^{\ell}\right).$$

We conclude using the assumption that $\mathrm{Adv}_{\mathsf{A}_2}(\mathsf{F}_{\mathsf{KFC}[r_1,r_2]}, \mathsf{F}^\star) \leq \epsilon$. □

Obviously, the bound on $\Pr[\overline{\mathsf{e}}_i]$ we obtained in Lemma 12.6 cannot be used directly to obtain a meaningful bound on the advantage of high order distinguishers since the bound obtained in Theorem 12.3 is not tight enough. We address this problem in the following subsection.

Considering Several Rounds at the Same Time

We can improve the previous approach by considering t successive e_i events and give an upper bound on the probability that *none* of them occurs.

Lemma 12.7 *Let $r_1, r_2 > 0$ be two positive integers and let $\mathsf{F}_{\mathsf{KFC}[r_1,r_2]}$ be as in* (12.13)*. For all $q > 1$, all $i \in \{1, \ldots, r_2\}$, and all $t \leq i$ we have*

$$\mathrm{Adv}_{\mathsf{A}_q}(\mathsf{F}_{\mathsf{KFC}[r_1,i]}, \mathsf{F}^\star) \leq \mathrm{Adv}_{\mathsf{A}_{q-1}}(\mathsf{F}_{\mathsf{KFC}[r_1,i]}, \mathsf{F}^\star) + \Pr[\bar{\mathsf{e}}_1, \bar{\mathsf{e}}_2, \ldots, \bar{\mathsf{e}}_t],$$

where e_i is the event that one of the q inputs is different from all others on the ℓ blocks at the output of $\mathsf{F}_{\mathsf{KFC}[r_1,i-1]}$, i.e.,

$$\mathsf{e}_i = \Big\{ \exists k \in \{1, \ldots, q\} \text{ s.t. } \forall j \in \{1, \ldots, \ell\} \ : \\ X_{i,j}^{(k)} \notin \Big\{ X_{i,j}^{(1)}, \ldots, X_{i,j}^{(k-1)}, X_{i,j}^{(k+1)}, \ldots, X_{i,j}^{(1)} \Big\} \Big\}.$$

Proof. Let $i \in \{1, \ldots, r_2\}$, let $t \leq i$, and let $\mathsf{H}_0 : \mathsf{F} = \mathsf{F}^\star$ and $\mathsf{H}_1 : \mathsf{F} = \mathsf{F}_{\mathsf{KFC}[r_1,i]}$. We denote by e the event $\mathsf{e}_1 \cup \mathsf{e}_2 \cup \cdots \cup \mathsf{e}_t$, so that $\bar{\mathsf{e}} = \bar{\mathsf{e}}_1 \cap \bar{\mathsf{e}}_2 \cap \cdots \cap \bar{\mathsf{e}}_t$. We have

$$\begin{aligned}
&\mathrm{Adv}_{\mathsf{A}_q}(\mathsf{H}_0, \mathsf{H}_1) \\
&= |(\Pr_{\mathsf{H}_1}[\mathsf{A}_q = 1|\mathsf{e}] - \Pr_{\mathsf{H}_0}[\mathsf{A}_q = 1|\mathsf{e}])\Pr[\mathsf{e}] + (\Pr_{\mathsf{H}_1}[\mathsf{A}_q = 1|\bar{\mathsf{e}}] - \Pr_{\mathsf{H}_0}[\mathsf{A}_q = 1|\bar{\mathsf{e}}])\Pr[\bar{\mathsf{e}}]| \\
&\leq |\Pr_{\mathsf{H}_1}[\mathsf{A}_q = 1|\mathsf{e}] - \Pr_{\mathsf{H}_0}[\mathsf{A}_q = 1|\mathsf{e}]| + \Pr[\bar{\mathsf{e}}].
\end{aligned}$$

Using the same approach than that of the proof of Lemma 12.1, it is easy to see that

$$|\Pr_{\mathsf{H}_1}[\mathsf{A}_q = 1|\mathsf{e}] - \Pr_{\mathsf{H}_0}[\mathsf{A}_q = 1|\mathsf{e}]| \leq \mathrm{Adv}_{\mathsf{A}_{q-1}}(\mathsf{H}_0, \mathsf{H}_1),$$

from which we deduce the announced result. □

Lemma 12.8 *With the notations of Lemma 12.7, we have that for all $i \in \{0, \ldots, r_2\}$ and all $t \leq i$:*

$$\Pr[\bar{\mathsf{e}}_1, \bar{\mathsf{e}}_2, \ldots, \bar{\mathsf{e}}_t] \leq \Big(1 - \big(1 - (q-1)\big(\tfrac{1}{2^m} + \epsilon\big)\big)^\ell\Big)^t. \qquad (12.17)$$

Proof. For $t = 0$ the result is trivial and for $t = 1$ it corresponds to that of Lemma 12.6. Assume that $t \geq 2$. We first have

$$\Pr[\bar{\mathsf{e}}_1, \bar{\mathsf{e}}_2, \ldots, \bar{\mathsf{e}}_t] = \Pr[\bar{\mathsf{e}}_t|\bar{\mathsf{e}}_1, \bar{\mathsf{e}}_2, \ldots, \bar{\mathsf{e}}_{t-1}] \cdot \Pr[\bar{\mathsf{e}}_1, \bar{\mathsf{e}}_2, \ldots, \bar{\mathsf{e}}_{t-1}].$$

As the bound (12.16) on $\mathrm{E}(\lambda_i)$ in the proof of Lemma 12.6 only relies on the pairwise independence of the inputs of the i-th round, the bound given by equation (12.15) on $\Pr[\bar{\mathsf{e}}_i]$ can also be proved for $\Pr[\bar{\mathsf{e}}_t|\bar{\mathsf{e}}_1, \bar{\mathsf{e}}_2, \ldots, \bar{\mathsf{e}}_{t-1}]$. Iterating, we finally obtain that

$$\Pr[\bar{\mathsf{e}}_1, \bar{\mathsf{e}}_2, \ldots, \bar{\mathsf{e}}_t] \leq \Big(1 - \big(1 - (q-1)\big((\tfrac{1}{2^m} + \epsilon\big)\big)^\ell\Big)^t.$$

□

Based on lemmas 12.7 and 12.8 it is now possible to give a *meaningful* upper-bound on the advantage of the best q-limited distinguisher for $q > 2$.

Theorem 12.4 *Assume that the advantage of the best 2-limited distinguisher against* $\mathsf{F}_{\mathsf{KFC}[r_1,r_2]}$ *is bounded by* ϵ. *For any positive* $q > 2$ *and set of integers* $\{t_2, \ldots, t_{q-1}\}$ *such that*

$$\sum_{i=2}^{q-1} t_i \leq r_2,$$

the advantage of the best q*-limited distinguisher* A_q *against* $\mathsf{F}_{\mathsf{KFC}[r_1,r_2]}$ *is such that*

$$\mathrm{Adv}_{\mathsf{A}_q}(\mathsf{F}_{\mathsf{KFC}[r_1,r_2]}, \mathsf{F}^\star) \leq \epsilon + \sum_{i=2}^{q-1} \left(1 - \left(1 - i\left(\tfrac{1}{2^m} + \epsilon\right)\right)^\ell\right)^{t_i}.$$

Proof. Using the results obtained in lemmas 12.7 and 12.8 successively, we get

$$\begin{aligned}\mathrm{Adv}_{\mathsf{A}_q}(\mathsf{F}_{\mathsf{KFC}[r_1,r_2]}, \mathsf{F}^\star) &\leq \mathrm{Adv}_{\mathsf{A}_{q-1}}(\mathsf{F}_{\mathsf{KFC}[r_1,r_2]}, \mathsf{F}^\star) + \Pr[\bar{\mathsf{e}}_1, \bar{\mathsf{e}}_2, \ldots, \bar{\mathsf{e}}_{t_{q-1}}] \\ &\leq \mathrm{Adv}_{\mathsf{A}_{q-1}}(\mathsf{F}_{\mathsf{KFC}[r_1,r_2]}, \mathsf{F}^\star) + \left(1 - \left(1 - (q-1)\left(\tfrac{1}{2^m} + \epsilon\right)\right)^\ell\right)^{t_{q-1}}.\end{aligned}$$

Applying the same two steps recursively we get

$$\mathrm{Adv}_{\mathsf{A}_q}(\mathsf{F}_{\mathsf{KFC}[r_1,r_2]}, \mathsf{F}^\star) \leq \mathrm{Adv}_{\mathsf{A}_2}(\mathsf{F}_{\mathsf{KFC}[r_1,r_2]}, \mathsf{F}^\star) + \sum_{i=2}^{q-1} \left(1 - \left(1 - i\left(\tfrac{1}{2^m} + \epsilon\right)\right)^\ell\right)^{t_i}.$$

We conclude using the assumption that $\mathrm{Adv}_{\mathsf{A}_2}(\mathsf{F}_{\mathsf{KFC}[r_1,r_2]}, \mathsf{F}^\star) \leq \epsilon$. □

Fixing $r_1 = 3$, the previous theorem bounds, for any value of $q > 2$, the advantage of the best q-limited distinguisher against a given number of rounds r_2 of $\mathsf{F}_{\mathsf{KFC}}$. In Table 12.2 we give the best bounds we obtain for various values of r_2, q, ℓ, and m. If one aims at a specific value of q and wants to select r_2 in order to bound the advantage of the best q-limited distinguisher, the best choice is probably to select the t_i's such that $\Pr[\bar{\mathsf{e}}_1, \ldots, \bar{\mathsf{e}}_{t_i}] < \epsilon$, which bounds the advantage by $(q-1)\cdot\epsilon$. The following theorem generalizes this idea.

Theorem 12.5 *Assume that the advantage of the best 2-limited distinguisher against* $\mathsf{F}_{\mathsf{KFC}[r_1,r_2]}$ *is bounded by* ϵ. *Let*

$$t_q(\beta) = \min_t \{\Pr[\bar{\mathsf{e}}_1, \ldots, \bar{\mathsf{e}}_t] < \beta \cdot \epsilon\} = \left\lceil \frac{\log(\beta\cdot\epsilon)}{\log\left(1 - \left(1-(q-1)\left(\frac{1}{2^m} + \epsilon\right)\right)^\ell\right)} \right\rceil.$$

For any q *such that*

$$\sum_{i=2}^{q-1} t_i(\beta) \leq r_2,$$

	$N = 8$ and $m = 8$						
$r_2 \diagdown q$	2	3	4	8	16	32	64
10	2^{-52}	2^{-40}	2^{-17}	2^{-2}	1	1	1
100	2^{-49}	2^{-49}	2^{-49}	2^{-46}	2^{-11}	1	1
250	2^{-48}	2^{-48}	2^{-48}	2^{-48}	2^{-33}	2^{-5}	1
1000	2^{-46}	2^{-46}	2^{-46}	2^{-46}	2^{-46}	2^{-35}	2^{-2}

	$N = 8$ and $m = 16$						
$r_2 \diagdown q$	2	3	4	8	16	32	64
10	2^{-116}	2^{-116}	2^{-57}	2^{-11}	1	1	1
100	2^{-113}	2^{-113}	2^{-113}	2^{-113}	2^{-66}	2^{-23}	2^{-5}
250	2^{-112}	2^{-112}	2^{-112}	2^{-112}	2^{-112}	2^{-69}	2^{-25}
1000	2^{-110}	2^{-110}	2^{-110}	2^{-110}	2^{-110}	2^{-110}	2^{-110}

	$N = 16$ and $m = 8$						
$r_2 \diagdown q$	2	3	4	8	16	32	64
10	2^{-104}	2^{-31}	2^{-12}	1	1	1	1
100	2^{-103}	2^{-103}	2^{-103}	2^{-31}	2^{-5}	1	1
250	2^{-103}	2^{-103}	2^{-103}	2^{-81}	2^{-18}	1	1
1000	2^{-102}	2^{-102}	2^{-102}	2^{-102}	2^{-82}	2^{-12}	1

Table 12.2: Bounds on $\mathrm{Adv}_{\mathsf{A}_q}(\mathsf{F}_{\mathsf{KFC}[r_1,r_2]}, \mathsf{F}^\star)$ for $r_1 = 3$ and various parameters.

the advantage of the best q-limited distinguisher A_q *against* $\mathsf{F}_{\mathsf{KFC}[r_1,r_2]}$ *is such that:*

$$\mathrm{Adv}_{\mathsf{A}_d}(\mathsf{F}_{\mathsf{KFC}[r_1,r_2]}, \mathsf{F}^\star) \leq \epsilon + \sum_{i=2}^{q-1} \Big(1 - \big(1 - i\big(\tfrac{1}{2^m} + \epsilon\big)\big)^{\ell}\Big)^{t_i(\beta)} \leq (1 + (q-2)\cdot\beta)\cdot\epsilon.$$

12.5 KFC in Practice

At this time, no key schedule has been specified for KFC. Of course, one can apply the exact same trick as the one used for the key schedule of C (see Section 11.1), i.e., use a key schedule based on a cryptographically secure pseudo-random generator (for example the good old BBS [29] or a faster generator like QUAD [13, 14]). This way, all the results we have proved assuming the mutual independence of the random functions and permutations remain valid when implementing KFC in practice with a 128-bit secret key. We propose two sets of parameters:

Regular KFC: $\ell = 8$, $m = 8$, $r_1 = 3$, $r_2 = 100$. These parameters lead to provable security against 8-limited adaptive distinguishers. Consequently, Regular KFC is resistant to iterated attacks of order 4, which include linear and differential cryptanalysis, the boomerang attack and others. Based on existing implementation results on C, we estimate the encryption speed of Regular KFC to 15-25 Mbits/s on a Pentium IV 2GHz. The key schedule needs to generate approximately 2^{22} cryptographically secure pseudo-random bits.

Extra Crispy KFC: $\boldsymbol{\ell = 8}$, $\boldsymbol{m = 16}$, $\boldsymbol{r_1 = 3}$, $\boldsymbol{r_2 = 1000}$. Using these quite *extreme* parameters, we manage to obtain provable security against 70-limited adaptive adversaries, but encryption rate could probably never reach more than 1 Mbit/s. Also, the key schedule should produce 2^{35} pseudo random bits, which means that Extra Crispy KFC requires at least 4 GB of memory.

12.6 Further Improvements

We introduced KFC, a block cipher based on a three round Feistel scheme. Each of the three round functions has an SPN-like structure for which we can either compute or bound the advantage of the best q-limited adaptive adversary, for any value of q. Using results from the Decorrelation Theory, we extend these results to the whole KFC construction.

To the best of our knowledge, KFC is the first practical block cipher to propose tight security proofs of resistance to large classes of attacks, including most classical cryptanalysis (such as linear and differential cryptanalysis, taking hull effect in consideration in both cases, higher order differential cryptanalysis, the boomerang attack, differential-linear cryptanalysis, or the rectangle attack). Of course, this security guarantee has a price in terms of encryption rates which can be up to 500 times smaller than that of the AES. Yet, KFC can certainly be improved in several ways.

For example, as a consequence of results by Naor and Reingold [120], it might be possible to reduce the 3-round Feistel scheme to a 2-round Feistel scheme plus an initial well chosen random permutation (and reduced-round versions of C could be excellent candidates). We informally introduce PDB [148] for which we believe that strong security results can be proved and which is certainly more efficient than KFC:

$$\mathsf{PDB}[r_0, r_1, r_2] = \Psi(\mathsf{F}_{\mathsf{KFC}[r_1,r_2]}, \mathsf{F}'_{\mathsf{KFC}[r_1,r_2]}) \circ \mathsf{C}[r_0].$$

Several similar constructions can be thought off, using for example results obtained by Lucks in [103] or by Maurer et. al in [113].

Another possible improvement could be to guarantee security not only against chosen plaintext attacks (CPA) but against chosen ciphertext attacks (CCA), which might be easily done on the construction $\mathsf{C}[r_0] \circ \mathsf{PDB}[r_0, r_1, r_2]$, still according to results in [120].

Last but not least, the bound we obtain in Lemma 12.6 (and in Lemma 12.8) essentially relies on Markov's inequality since we exclusively focus on the mean $\mathrm{E}(\lambda_i)$ to bound $\Pr[\lambda_i \neq 0]$. Using the Bienaymé-Chebyshev inequality (for example) might result in a tighter bound but would render the proofs more complex.

Chapter 13

Conclusion and Future Work

Since the publication of the Diffie-Hellman key exchange, every new public key cryptographic scheme has to provide strong arguments (if not a rigorous proof) of its security in order to get a chance to be even considered by the academic community. By contrast, the Advanced Encryption Standard does not give any strong security guarantee against the most basic statistic attacks, and though this fact does not seem to be much of a concern. This situation is absurd since, in practice, the end-user is essentially interested by the security of the global cryptographic system, such as hybrid encryption for example, which at least requires that of its individual parts.

Yet, provably secure symmetric schemes exist, the Vernam cipher being certainly the most representative example. In that particular case though, we note that once the security starts to decrease, it does exponentially. For example, it is clear that if the same key is used twice in the Vernam cipher, then the exclusive-or of the two ciphertexts is equal to the exclusive-or of the two plaintexts, an unfortunate feature which results from the fact that only linear operations are used within the encryption process. The Luby-Rackoff construction is another example of a provably secure symmetric construction; but it suffers from the drawback of being completely impractical due to the huge quantity of randomness it requires to instantiate the round functions. These two examples illustrate the fact that in the struggle for provably secure symmetric schemes, cryptanalysts often end up with either unpractical schemes or designs whose practical security collapse as soon as the security proofs' hypotheses are violated.

For these reasons, most modern block cipher designs only focus on *heuristic security arguments* which are essentially intuitive arguments that strengthen the plausibility of the hypothesis that none of the already known cryptanalytic attacks applies to the new design. Naturally, each new block cipher is considered as a new challenge by the cryptographic community, leading to the discovery of new attack methods. For example the block cipher SAFER+ [85] lead to the invention of integral cryptanalysis [69] which was then applied on the block ciphers SQUARE [40], IDEA [97,118], AES [41,51] and FOX [76,165], and afterwards extended to attack Twofish [104,136].

The block cipher DFC [54,59] started to fill the gap between the perfect security of the impractical Vernam cipher and the absence of security guarantee of practical block ciphers. Our proposals C and (to some extent) KFC are fast block ciphers which provide *rigorous* security proofs of resistance to a wider range cryptanalytic attacks than the DFC. Furthermore, they both differ from previous provably secure constructions in as far as they build their foundations on the same principles that drive all standard designs: they avoid the use of algebraic decorrelation modules that surely facilitate the security proofs but also certainly weaken any security argument against high order adversaries. When designing C and KFC we assumed that the best strategy was probably to take the best from both worlds: strong arguments from provable security and intuitive arguments from heuristic security.

There is still plenty of space for improvements. In particular, we believe that C is secure against higher order adversaries. More precisely, there might be a link between the decorrelation order of the substitution boxes and that of the whole construction, so there might be an elegant trade-off between the amount of randomness within the substitution boxes and the level of security provided by C. Similarly, the security proofs of KFC can definitely be improved, for example by working on the principle that not only one round input might be different from all the others on all boxes, but by assuming that this can happen for several inputs at the same time.

Implementation considerations apart, designing a block cipher essentially boils down to select a subset of the permutations defined by the perfect cipher. With security in mind, making sure that the distribution matrix of the block cipher considered is as close as possible to that of the perfect cipher appears to be, in itself, a very natural thing to aim to and a quite desirable feature. This task is highly challenging since we furthermore have to cope with implementation issues, which seem to restrict the possible permutation subsets to those containing permutations which are all based on some specific structure (such as substitution-permutation networks or Feistel schemes). Yet, being close to the perfect cipher is simply the *exact* assumption that is made on the underlying block cipher in the *ideal cipher model* [12] which is considered in most operation modes security proofs [11]. Consequently, we do not see any reason why we should not expect that kind of guarantee from future constructions. To quote Jacques Stern,

> "*[...] the methodology of provable security has become unavoidable in designing, analyzing and evaluating new schemes*" [144].

This statement initially concerns public key schemes. We hope that this thesis makes a significant step towards its extension to block ciphers.

Part IV

Appendixes

Appendix A

A Proof of Sanov's Theorem

Lemma 1.1 *Let $\mathcal{Z}$ be a finite set. We have $|\mathcal{P}_q| \leq (q+1)^{|\mathcal{Z}|}$.*

Proof. An element of $\mathcal{P}_q$ is a vector with $|\mathcal{Z}|$ components, each of which can take at most $q+1$ values. □

Theorem 1.1 *Let $\mathcal{Z}$ be a finite set. For any type $\mathsf{P} \in \mathcal{P}_q(\mathcal{Z})$ we have*

$$\frac{1}{(q+1)^{|\mathcal{Z}|}} 2^{qH(\mathsf{P})} \leq |T_q(\mathsf{P})| \leq 2^{qH(\mathsf{P})}.$$

Proof. As $\mathsf{P} \in \mathcal{P}_q$, we can write $\mathsf{P} = (\frac{n_1}{q}, \frac{n_2}{q}, \ldots, \frac{n_{|\mathcal{Z}|}}{q})$ for some $0 \leq n_1, n_2, \ldots, n_{|\mathcal{Z}|} \leq q$ where $\sum_i n_i = q$. Clearly, $|T_q(\mathsf{P})|$ is equal to

$$\binom{q}{n_1}\binom{q-n_1}{n_2}\cdots\binom{q-\sum_{i\neq|\mathcal{Z}|} n_i}{n_{|\mathcal{Z}|}} = \frac{q!}{n_1! n_2! \cdots n_{|\mathcal{Z}|}!} = \binom{q}{n_1, n_2, \ldots, n_{|\mathcal{Z}|}}.$$

To bound $|T_q(\mathsf{P})|$, we thus need to bound a multinomial coefficient. We generalize a simple trick used in the binary case in [37, p.284]. Denoting $p_i = \frac{n_i}{q}$, we have that $\sum_{i=1}^{|\mathcal{Z}|} p_i = 1$, so that, using the multinomial theorem,

$$1 = (p_1 + p_2 + \cdots + p_{|\mathcal{Z}|})^q = \sum_{k_1, k_2, \ldots, k_{|\mathcal{Z}|}} \binom{q}{k_1, k_2, \ldots, k_{|\mathcal{Z}|}} p_1^{k_1} p_2^{k_2} \cdots p_{|\mathcal{Z}|}^{k_{|\mathcal{Z}|}},$$

where the sum runs over all integer indices $k_1, k_2, \ldots, k_{|\mathcal{Z}|}$ such that $\sum_{i=1}^{|\mathcal{Z}|} k_i = q$. As all the terms in the sum are positive, we obtain that for the $n_1, n_2, \ldots, n_{|\mathcal{Z}|}$ term

$$1 \geq \binom{q}{n_1, n_2, \ldots, n_{|\mathcal{Z}|}} p_1^{n_1} p_2^{n_2} \cdots p_{|\mathcal{Z}|}^{n_{|\mathcal{Z}|}}. \tag{A.1}$$

Moreover,

$$\prod_i p_i^{n_i} = \prod_i \left(\frac{n_i}{q}\right)^{n_i} = 2^{\sum_i n_i \log \frac{n_i}{q}} = 2^{-qH(\mathsf{P})}. \tag{A.2}$$

From (A.1) and (A.2) we conclude that

$$|T_q(\mathsf{P})| = \binom{q}{n_1, n_2, \dots, n_{|\mathcal{Z}|}} \leq 2^{qH(\mathsf{P})}.$$

On the other hand, since the sum in the multinomial theorem runs over less than $(q+1)^{|\mathcal{Z}|}$ terms,

$$1 \leq (q+1)^{|\mathcal{Z}|} \max_{k_1,k_2,\dots,k_{|\mathcal{Z}|}} \binom{q}{k_1, k_2, \dots, k_{|\mathcal{Z}|}} p_1^{k_1} p_2^{k_2} \cdots p_{|\mathcal{Z}|}^{k_{|\mathcal{Z}|}}.$$

The term we need to maximize corresponds to the probability mass function of a multinomial distribution with outcome probabilities $p_1, p_2, \dots, p_{|\mathcal{Z}|}$, so that the maximum value is obtained for the most probable outcome, which is obtained for $k_1 = qp_1 = n_1$, $k_2 = qp_2 = n_2$,. . . ,$k_{|\mathcal{Z}|} = qp_{|\mathcal{Z}|} = n_{|\mathcal{Z}|}$. Therefore, using (A.2),

$$\begin{aligned} 1 &\leq (q+1)^{|\mathcal{Z}|} \binom{q}{n_1, n_2, \dots, n_{|\mathcal{Z}|}} p_1^{n_1} p_2^{n_2} \cdots p_{|\mathcal{Z}|}^{n_{|\mathcal{Z}|}} \\ &= (q+1)^{|\mathcal{Z}|} \binom{q}{n_1, n_2, \dots, n_{|\mathcal{Z}|}} 2^{-qH(\mathsf{P})} \\ &= (q+1)^{|\mathcal{Z}|} \, |T_q(\mathsf{P})| \, 2^{-qH(\mathsf{P})}. \end{aligned}$$

Which leads to the lower bound on $|T_q(\mathsf{P})|$. □

Theorem 1.2 *Let $\mathcal{Z}$ be a finite set and P be a probability distribution on $\mathcal{Z}$. Let $Z_1, Z_2, \dots, Z_q$ be q i.i.d. samples drawn according to P. For any $\mathsf{P}' \in \mathcal{P}_q(\mathcal{Z})$ we have*

$$\frac{1}{(q+1)^{|\mathcal{Z}|}} 2^{-q\mathrm{D}(\mathsf{P}'\|\mathsf{P})} \leq \Pr[\mathsf{P}_{\mathbf{Z}^q} = \mathsf{P}'] \leq 2^{-q\mathrm{D}(\mathsf{P}'\|\mathsf{P})}.$$

Proof. On the one hand we have

$$\Pr[\mathsf{P}_{\mathbf{Z}^q} = \mathsf{P}'] = \sum_{\mathbf{z}^q \in T_q(\mathsf{P}')} \Pr_{\mathsf{P}^q}[\mathbf{z}^q] = \sum_{\mathbf{z}^q \in T_q(\mathsf{P}')} 2^{-q(H(\mathsf{P}_{\mathbf{z}^q}) + \mathrm{D}(\mathsf{P}_{\mathbf{z}^q}\|\mathsf{P}))},$$

using Lemma 6.1. Since we sum over $\mathbf{z}^q \in T_q(\mathsf{P}')$, we have $\mathsf{P}_{\mathbf{z}^q} = \mathsf{P}'$, so that we obtain

$$\Pr[\mathsf{P}_{\mathbf{Z}^q} = \mathsf{P}'] = \left|T_q(\mathsf{P}')\right| 2^{-q(H(\mathsf{P}') + \mathrm{D}(\mathsf{P}'\|\mathsf{P}))}.$$

The bounds obtained on $|T_q(\mathsf{P}')|$ in Theorem 1.1 lead to the announced result. □

Theorem 1.3 ***(Sanov's theorem)*** *Let* P *be a probability distribution over a finite set* $\mathcal{Z}$*,* $\mathcal{Z}'$ *be a non-empty subset of* $\mathcal{Z}$*, and* Π *be a set of probability distributions of full support over* $\mathcal{Z}'$*. If* $Z_1, Z_2, \ldots, Z_q$ *are* q *i.i.d. random variables drawn according to the distribution* P*, we have*

$$\Pr[\mathsf{P}_{\mathbf{Z}^q} \in \Pi] \leq (q+1)^{|\mathcal{Z}|} 2^{-q\mathrm{D}(\Pi\|\mathsf{P})},$$

where $\mathrm{D}(\Pi\|\mathsf{P}) = \inf_{\mathsf{P}'\in\Pi} \mathrm{D}(\mathsf{P}'\|\mathsf{P})$*. Moreover, if the closure of* $\Pi \subset \mathcal{P}(\mathcal{Z}')$ *is equal to the closure of its interior, i.e., if* $\overline{\Pi} = \overline{\overset{\circ}{\Pi}}$ *under the topology of probability distributions over* $\mathcal{Z}'$*, then*

$$\Pr[\mathsf{P}_{\mathbf{Z}^q} \in \Pi] \doteq 2^{-q\mathrm{D}(\Pi\|\mathsf{P})}.$$

Proof. On the one hand we have

$$\Pr[\mathsf{P}_{\mathbf{Z}^q} \in \Pi] = \sum_{P\in\Pi} \Pr[\mathsf{P}_{\mathbf{Z}^q} = P] = \sum_{P\in\Pi\cap\mathcal{P}_q(\mathcal{Z})} \Pr[\mathsf{P}_{\mathbf{Z}^q} = P], \tag{A.3}$$

since $\Pr[\mathsf{P}_{\mathbf{Z}^q} = P] = 0$ when $P \notin \mathcal{P}_q(\mathcal{Z})$. Using Theorem 1.2 (which we can do, as $P \in \mathcal{P}_q(\mathcal{Z})$) we obtain

$$\begin{aligned}
\Pr[\mathsf{P}_{\mathbf{Z}^q} \in \Pi] &\leq \sum_{P\in\Pi\cap\mathcal{P}_q(\mathcal{Z})} 2^{-q\mathrm{D}(P\|\mathsf{P})} \\
&\leq |\Pi \cap \mathcal{P}_q(\mathcal{Z})| \max_{P\in\Pi\cap\mathcal{P}_q(\mathcal{Z})} 2^{-q\mathrm{D}(P\|\mathsf{P})} \\
&\leq |\Pi \cap \mathcal{P}_q(\mathcal{Z})| \sup_{P\in\Pi} 2^{-q\mathrm{D}(P\|\mathsf{P})} \\
&= |\Pi \cap \mathcal{P}_q(\mathcal{Z})|\, 2^{-q\inf_{P\in\Pi}\mathrm{D}(P\|\mathsf{P})}.
\end{aligned}$$

As by definition $\inf_{P\in\Pi} \mathrm{D}(P\|\mathsf{P}) = \mathrm{D}(\Pi\|\mathsf{P})$ and as $|\Pi \cap \mathcal{P}_q(\mathcal{Z})| \leq |\mathcal{P}_q(\mathcal{Z})| \leq (q+1)^{|\mathcal{Z}|}$ by Lemma 1.1, we conclude that

$$\Pr[\mathsf{P}_{\mathbf{Z}^q} \in \Pi] \leq (q+1)^{|\mathcal{Z}|} 2^{-q\mathrm{D}(\Pi\|\mathsf{P})}, \tag{A.4}$$

which is the upper bound we wanted to obtain. We now consider the case where the closure of $\Pi \subset \mathcal{P}_q(\mathcal{Z}')$ is equal to the closure of its interior. If $\mathrm{D}(\Pi\|\mathsf{P}) = +\infty$ (which happens when $\mathsf{P}(a) = 0$ for some $a \in \mathcal{Z}'$), the theorem is obviously true since $\Pr[\mathsf{P}_{\mathbf{Z}^q} \in \Pi] = 0$ in this case. We now assume that $\mathrm{D}(\Pi\|\mathsf{P}) < \infty$ (i.e., $\mathsf{P}(a) > 0$ for all $a \in \mathcal{Z}'$). For all $P \in \mathcal{P}$ and $d > 0$, we denote by $B_\infty(P, \delta) \subset \mathcal{P}$ the set of all distributions $P' \in \mathcal{P}$ such that $\|P' - P\|_\infty < \delta$. Let $\epsilon > 0$. Since $\overline{\Pi} = \overline{\overset{\circ}{\Pi}}$ under the topology of distributions over $\mathcal{Z}'$, there exists $\mathsf{P}' \in \overset{\circ}{\Pi}$ (thus of full support over $\mathcal{Z}'$) such that

$$\left|\mathrm{D}(\mathsf{P}'\|\mathsf{P}) - \mathrm{D}(\Pi\|\mathsf{P})\right| < \epsilon \tag{A.5}$$

and $\epsilon' > 0$ such that $B_\infty(\mathsf{P}', \epsilon') \subset \overset{\circ}{\Pi}$. For all positive integer q such that $q \geq \left(\frac{\epsilon'}{2}\right)^{-1}$ there exists $\mathsf{P}_q \in B_\infty(\mathsf{P}', \epsilon') \cap \mathcal{P}_q(\mathcal{Z}')$. Since $P \mapsto \mathrm{D}(P\|\mathsf{P})$ is a continuous function over $\mathcal{P}(\mathcal{Z}')$ (as we assumed that $\mathsf{P}(a) > 0$ for all $a \in \mathcal{Z}'$), this means that

$$\left|\mathrm{D}(\mathsf{P}_q\|\mathsf{P}) - \mathrm{D}(\mathsf{P}'\|\mathsf{P})\right| < \epsilon \tag{A.6}$$

when ϵ' is chosen small enough (i.e., when q is large enough). Starting from (A.3) we have

$$\Pr[\mathsf{P}_{\mathbf{Z}^q} \in \Pi] = \sum_{Q \in \Pi \cap \mathcal{P}_q(\mathcal{Z})} \Pr[\mathsf{P}_{\mathbf{Z}^q} = Q] \geq \Pr[\mathsf{P}_{\mathbf{Z}^q} = \mathsf{P}_q] \geq \frac{1}{(q+1)^{|\mathcal{Z}|}} 2^{-q\mathrm{D}(\mathsf{P}_q\|\mathsf{P})},$$

using Theorem 1.2. Consequently,

$$\begin{aligned} \frac{1}{q} \log \frac{\Pr[\mathsf{P}_{\mathbf{Z}^q} \in \Pi]}{2^{-q\mathrm{D}(\Pi\|\mathsf{P})}} &= \mathrm{D}(\Pi\|\mathsf{P}) + \frac{1}{q} \log \Pr[\mathsf{P}_{\mathbf{Z}^q} \in \Pi] \\ &\geq \mathrm{D}(\Pi\|\mathsf{P}) - \mathrm{D}(\mathsf{P}_q\|\mathsf{P}) - |\mathcal{Z}| \frac{\log(q+1)}{q}, \end{aligned}$$

which is (according to (A.5) and (A.6)) greater than -3ϵ when q is large enough. This holds for any $\epsilon > 0$, so that

$$\lim_{q \to \infty} \frac{1}{q} \log \frac{\Pr[\mathsf{P}_{\mathbf{Z}^q} \in \Pi]}{2^{-q\mathrm{D}(\Pi\|\mathsf{P})}} \geq 0.$$

This, combined with (A.4) allows us to conclude. □

Appendix B

Proof of Lemma 6.6

We first prove two lemmas.

Lemma 2.1 *Let* P_0 *and* P_1 *be two probability distributions with finite support* $\mathcal{Z}$*. Let* $p_z = \mathsf{P}_0[z]$ *and* $\epsilon = (\epsilon_z)_{z\in\mathcal{Z}}$ *where* $\epsilon_z = \frac{\mathsf{P}_1[z]-\mathsf{P}_0[z]}{p_z}$*. Assuming that* $|\epsilon_z| \leq \frac{1}{2}$ *for all* $z \in \mathcal{Z}$*, we have*

$$\left|\frac{1}{\ln 2} - \frac{1}{\ln 2}\sum_{z\in\mathcal{Z}} p_z\sqrt{1+\epsilon_z} - \mathrm{B}(\mathsf{P}_0,\mathsf{P}_1)\right| \leq \frac{5}{96\ln 2}\|\mathsf{P}_0\|_\infty^2\|\epsilon\|_2^4.$$

Proof. Since $|\epsilon_z| \leq \frac{1}{2}$ for all $z \in \mathcal{Z}$ we have

$$\sum_{z\in\mathcal{Z}} p_z\sqrt{1+\epsilon_z} \geq \sum_{z\in\mathcal{Z}} p_z\sqrt{1-|\epsilon_z|} \geq \sum_{z\in\mathcal{Z}} p_z\sqrt{1-\frac{1}{2}} \geq \frac{1}{2}.$$

Denoting $u = 1 - \sum_{z\in\mathcal{Z}} p_z\sqrt{1+\epsilon_z}$, we thus have $0 \leq u \leq \frac{1}{2}$, $\mathrm{B}(\mathsf{P}_0,\mathsf{P}_1) = -\log(1-u)$, and want to bound $\left|\frac{u}{\ln 2} + \log(1-u)\right|$. We have

$$\log(1-u) = -\frac{1}{\ln 2}\sum_{k=1}^{\infty}\frac{u^k}{k}$$

so that

$$\begin{aligned}\left|\frac{u}{\ln 2} + \log(1-u)\right| &= \frac{1}{\ln 2}\sum_{k=2}^{\infty}\frac{u^k}{k}\\ &= \frac{u^2}{2\ln 2} + \frac{1}{\ln 2}\sum_{k=3}^{\infty}\frac{u^k}{k}\\ &\leq \frac{u^2}{2\ln 2} + \frac{1}{3\ln 2}\sum_{k=3}^{\infty}u^k\\ &= \frac{u^2}{2\ln 2} + \frac{1}{3\ln 2}\cdot\frac{u^3}{1-u}.\end{aligned}$$

Since $u \leq \frac{1}{2}$, we have $\frac{u^3}{1-u} \leq u^2$ and thus

$$\left|\frac{u}{\ln 2} + \log(1-u)\right| \leq \frac{5}{6\ln 2}u^2. \tag{B.1}$$

We will now bound u^2:

$$\begin{aligned} u^2 &= \left(1 - \sum_z p_z \sqrt{1+\epsilon_z}\right)^2 \\ &= \left(\sum_z p_z \sum_{k=2}^{\infty} \frac{\frac{1}{2}(\frac{1}{2}-1)\cdots(\frac{1}{2}-k+1)}{k!}\epsilon_z^k\right)^2 \\ &\leq \left(\sum_z p_z \sum_{k=2}^{\infty} \left|\frac{\frac{1}{2}(\frac{1}{2}-1)\cdots(\frac{1}{2}-k+1)}{k!}\right| |\epsilon_z|^k\right)^2. \end{aligned}$$

Since

$$\frac{\frac{1}{2}(1-\frac{1}{2})}{2!} = \frac{1}{8} \quad \text{and} \quad \left|\frac{\frac{1}{2}\cdots(\frac{1}{2}-k)}{(k+1)!}\right| \leq \left|\frac{\frac{1}{2}\cdots(\frac{1}{2}-k+1)}{k!}\right|$$

for all $k \geq 1$, we have

$$\begin{aligned} u^2 &\leq \frac{1}{64}\left(\sum_z p_z \sum_{k=2}^{\infty} |\epsilon_z|^k\right)^2 \\ &\leq \frac{\|\mathsf{P}_0\|_\infty^2}{64}\left(\sum_z \sum_{k=2}^{\infty} |\epsilon_z|^k\right)^2 \\ &= \frac{\|\mathsf{P}_0\|_\infty^2}{64}\left(\sum_z |\epsilon_z|^2 + \sum_z \frac{|\epsilon_z|^3}{1-|\epsilon_z|}\right)^2. \end{aligned}$$

Since we assumed that $|\epsilon_z| \leq \frac{1}{2}$ for all $z \in \mathcal{Z}$, $\frac{|\epsilon_z|^3}{1-|\epsilon_z|} \leq |\epsilon_z|^2$ and we finally obtain

$$u^2 \leq \frac{1}{16}\|\mathsf{P}_0\|_\infty^2 \|\epsilon\|_2^4.$$

From the previous inequality and (B.1) we conclude that

$$\left|\frac{u}{\ln 2} + \log(1-u)\right| \leq \frac{5}{96\ln 2}\|\mathsf{P}_0\|_\infty^2 \|\epsilon\|_2^4$$

which concludes the proof. □

Lemma 2.2 *Let P_0 and P_1 be two probability distributions over a finite set $\mathcal{Z}$. Let $p_z = \mathsf{P}_0[z]$ and $\epsilon = (\epsilon_z)_{z\in\mathcal{Z}}$ where $\epsilon_z = \frac{\mathsf{P}_1[z]-\mathsf{P}_0[z]}{p_z}$. Assuming that $|\epsilon_z| \leq \frac{1}{2}$ for all $z \in \mathcal{Z}$, we have*

$$\left|\frac{1}{\ln 2}\sum_{z\in\mathcal{Z}} p_z\sqrt{1+\epsilon_z} - \frac{1}{\ln 2} + \frac{1}{8\ln 2}\sum_{z\in\mathcal{Z}} p_z\epsilon_z^2\right| \leq \frac{1}{8\ln 2}\sqrt{|\mathcal{Z}|}\|\mathsf{P}_0\|_\infty\|\epsilon\|_2^3.$$

Proof. Expanding $\sqrt{1+\epsilon_z}$ in Taylor series we have

$$\sqrt{1+\epsilon_z} = 1 + \frac{1}{2}\epsilon_z - \frac{\frac{1}{2}(1-\frac{1}{2})}{2}\epsilon_z^2 + \sum_{k=3}^{\infty} \frac{\frac{1}{2}(\frac{1}{2}-1)\cdots(\frac{1}{2}-k+1)}{k!}\epsilon_z^k,$$

so that, since $\sum_z p_z \epsilon_z = 0$,

$$\sum_z p_z \sqrt{1+\epsilon_z} = 1 - \frac{\frac{1}{2}(1-\frac{1}{2})}{2}\sum_z p_z \epsilon_z^2 + \sum_z \sum_{k=3}^{\infty} \frac{\frac{1}{2}(\frac{1}{2}-1)\cdots(\frac{1}{2}-k+1)}{k!} p_z \epsilon_z^k,$$

and thus

$$\left| \frac{1}{\ln 2}\sum_z p_z \sqrt{1+\epsilon_z} - \frac{1}{\ln 2} + \frac{\frac{1}{2}(1-\frac{1}{2})}{2\ln 2}\sum_z p_z \epsilon_z^2 \right| = \left| \frac{1}{\ln 2}\sum_z \sum_{k=3}^{\infty} \frac{\frac{1}{2}\cdots(\frac{1}{2}-k+1)}{k!} p_z \epsilon_z^k \right|.$$

We will now bound the right-hand side of the previous equation, which we denote by R. We have

$$\begin{aligned} R &= \left| \sum_z \frac{\frac{1}{2}(\frac{1}{2}-1)(\frac{1}{2}-2)}{3!\ln 2} p_z \epsilon_z^3 + \sum_z \sum_{k=4}^{\infty} \frac{\frac{1}{2}\cdots(\frac{1}{2}-k+1)}{k!\ln 2} p_z \epsilon_z^k \right| \\ &\leq \frac{\frac{1}{2}(1-\frac{1}{2})(2-\frac{1}{2})}{3!\ln 2}\sum_z p_z |\epsilon_z|^3 + \sum_z \sum_{k=4}^{\infty} \left| \frac{\frac{1}{2}\cdots(\frac{1}{2}-k+1)}{k!\ln 2} \right| p_z |\epsilon_z|^k . \end{aligned}$$

It is easy to see that

$$\frac{\frac{1}{2}(1-\frac{1}{2})(2-\frac{1}{2})}{3!} = \frac{1}{16} \quad \text{and} \quad \left| \frac{\lambda^\star\cdots(\lambda^\star-k)}{(k+1)!} \right| \leq \left| \frac{\lambda^\star\cdots(\lambda^\star-k+1)}{k!} \right|$$

for all $k \geq 1$ so that

$$\begin{aligned} R &\leq \frac{1}{16\ln 2}\sum_z p_z |\epsilon_z|^3 + \frac{1}{16\ln 2}\sum_z \sum_{k=4}^{\infty} p_z |\epsilon_z|^k \\ &\leq \frac{1}{16\ln 2}\sum_z p_z |\epsilon_z|^3 + \frac{1}{16\ln 2}\sum_z p_z |\epsilon_z|^4 \sum_{k=0}^{\infty} |\epsilon_z|^k \\ &= \frac{1}{16\ln 2}\sum_z p_z |\epsilon_z|^3 + \frac{1}{16\ln 2}\sum_z p_z \frac{|\epsilon_z|^4}{1-|\epsilon_z|}. \end{aligned}$$

As we assumed that $|\epsilon_z| \leq \frac{1}{2}$ for all $z \in \mathcal{Z}$, we have $\frac{|\epsilon_z|^4}{1-|\epsilon_z|} \leq |\epsilon_z|^3$, which leads to

$$R \leq \frac{1}{8\ln 2}\sum_{z\in\mathcal{Z}} p_z |\epsilon_z|^3 \leq \frac{1}{8\ln 2}\|\mathsf{P}_0\|_\infty \|\epsilon\|_3^3. \tag{B.2}$$

Using a classical extension of Cauchy's inequality (see [64]) we have

$$\|\epsilon\|_3^3 \leq \sqrt{|\mathcal{Z}|}\|\epsilon\|_6^3 = \sqrt{|\mathcal{Z}|}\sqrt{\sum_i \left(|x_i|^2\right)^3} \leq \sqrt{|\mathcal{Z}|}\sqrt{\left(\sum_i |x_i|^2\right)^3} = \sqrt{|\mathcal{Z}|}\|\epsilon\|_2^3.$$

From this last inequality and (B.2) we obtain

$$R \leq \frac{1}{8\ln 2}\sqrt{|\mathcal{Z}|}\|\mathsf{P}_0\|_\infty\|\epsilon\|_2^3.$$

□

Lemmas 2.1 and 2.2 easily lead to Lemma 6.6.

Appendix C

Proofs of the Lemmas Used in Example 7.3

Lemma 3.1 *Let $\mathbf{u} = u_1 u_2 \ldots, u_n$ be a n-bit binary string and let w denote its Hamming weight. We have*

$$\sum_{1 \leq j < k \leq n} u_j u_k = \frac{w(w-1)}{2}.$$

Proof. It is easy to see that

$$\sum_{1 \leq j < k \leq n} u_j u_k = \sum_{j=1}^{n-1} u_j \sum_{k=j+1}^{n} u_k = (w-1) + (w-2) + \cdots 1 = \sum_{\ell=1}^{w-1} \ell = \frac{w(w-1)}{2}.$$

□

Lemma 3.2 *For any integer $n > 2$ such that 4 divides $n+1$, we have:*

$$\begin{aligned} \sum_{k=0}^{(n-3)/4} \binom{n}{4k} &= \sum_{k=0}^{(n-3)/4} \binom{n}{4k+3} &= \frac{1}{4}(2^n + (1+i)^n + (1-i)^n), \\ \sum_{k=0}^{(n-3)/4} \binom{n}{4k+1} &= \sum_{k=0}^{(n-3)/4} \binom{n}{4k+2} &= \frac{1}{4}(2^n - i(1+i)^n + i(1-i)^n), \end{aligned}$$

where i is the complex imaginary unit.

Proof. The equalities between the sums of binomial coefficients can be shown by letting $\ell = \frac{n-3}{4} - k$ in both cases, using the fact that $\binom{n}{u} = \binom{n}{n-u}$ for all $u = 0, 1, \ldots, n$.

From the binomial theorem we easily obtain

$$2^n + (1+i)^n + (1-i)^n = \sum_{\ell=0}^{n} \binom{n}{\ell} (1 + i^\ell + (-i)^\ell) = \sum_{\ell=0}^{n} \binom{n}{\ell} f(\ell),$$

where we let $f(\ell) = 1 + i^\ell + (-i)^\ell$. Obviously, for all integer $k \geq 0$, $f(4k) = 3$, $f(4k+1) = 1$, $f(4k+2) = -1$, and $f(4k+3) = 1$. This leads to

$$\begin{aligned}
&2^n + (1+i)^n + (1-i)^n \\
&= \sum_{k=0}^{\frac{n-3}{4}} \left(3\binom{n}{4k} + \binom{n}{4k+1} - \binom{n}{4k+2} + \binom{n}{4k+3} \right) \\
&= 4\sum_{k=0}^{\frac{n-3}{4}} \binom{n}{4k} + \sum_{n=0}^{\frac{n-3}{4}} \left(-\binom{n}{4k} + \binom{n}{4k+1} - \binom{n}{4k+2} + \binom{n}{4k+3} \right) \\
&= 4\sum_{k=0}^{\frac{n-3}{4}} \binom{n}{4k} + (1-1)^n = 4\sum_{k=0}^{\frac{n-3}{4}} \binom{n}{4k},
\end{aligned}$$

which proves the first equality. The proof of the second one is similar:

$$2^n - i(1+i)^n + i(1-i)^n = \sum_{\ell=0}^{n} \binom{n}{\ell} (1 - i^{\ell+1} - (-i)^\ell) = \sum_{\ell=0}^{n} \binom{n}{\ell} g(\ell),$$

where we let $g(\ell) = 1 - i^{\ell+1} - (-i)^\ell$. For all integer $k \geq 0$ we have $g(4k) = 1$, $g(4k+1) = 3$, $g(4k+2) = 1$, and $g(4k+3) = -1$. This leads to

$$\begin{aligned}
&2^n - i(1+i)^n + i(1-i)^n \\
&= \sum_{k=0}^{\frac{n-3}{4}} \left(\binom{n}{4k} + 3\binom{n}{4k+1} + \binom{n}{4k+2} - \binom{n}{4k+3} \right) \\
&= 4\sum_{k=0}^{\frac{n-3}{4}} \binom{n}{4k+1} + \sum_{k=0}^{\frac{n-3}{4}} \left(\binom{n}{4k} - \binom{n}{4k+1} + \binom{n}{4k+2} - \binom{n}{4k+3} \right) \\
&= 4\sum_{k=0}^{\frac{n-3}{4}} \binom{n}{4k+1} + (1-1)^n = 4\sum_{k=0}^{\frac{n-3}{4}} \binom{n}{4k+1}.
\end{aligned}$$

□

Appendix D

The Substitution Box of DEAN27.

Tables D.1 and D.2 describe the fixed substitution box that we suggest for DEAN27.

0 832	1 931	2 235	3 823	4 171	5 434	6 569	7 138	8 911	9 737	10 749	11 72	12 436	13 498	14 487	15 427	16 946	17 284
18 127	19 11	20 172	21 225	22 142	23 496	24 428	25 312	26 242	27 101	28 876	29 181	30 297	31 564	32 407	33 19	34 553	35 675
36 41	37 793	38 920	39 165	40 305	41 461	42 729	43 709	44 497	45 471	46 973	47 125	48 865	49 565	50 680	51 502	52 227	53 874
54 278	55 336	56 289	57 508	58 599	59 450	60 453	61 331	62 414	63 329	64 23	65 908	66 813	67 268	68 895	69 53	70 70	71 462
72 394	73 224	74 115	75 309	76 292	77 113	78 704	79 514	80 900	81 768	82 986	83 639	84 200	85 119	86 930	87 527	88 492	89 808
90 756	91 786	92 500	93 555	94 644	95 685	96 110	97 511	98 720	99 122	100 385	101 468	102 383	103 794	104 892	105 951	106 430	107 724
108 554	109 523	110 837	111 649	112 667	113 314	114 534	115 236	116 330	117 633	118 711	119 582	120 516	121 134	122 898	123 469	124 63	125 399
126 610	127 921	128 400	129 841	130 617	131 587	132 562	133 918	134 820	135 38	136 277	137 678	138 576	139 725	140 97	141 179	142 367	143 174
144 302	145 85	146 797	147 293	148 334	149 867	150 741	151 949	152 810	153 313	154 791	155 796	156 981	157 533	158 538	159 870	160 705	161 765
162 653	163 71	164 998	165 659	166 137	167 294	168 557	169 815	170 5	171 954	172 104	173 822	174 351	175 636	176 985	177 977	178 30	179 742
180 202	181 69	182 435	183 608	184 893	185 228	186 204	187 250	188 666	189 362	190 626	191 552	192 648	193 117	194 392	195 859	196 398	197 875
198 107	199 381	200 812	201 173	202 355	203 739	204 784	205 241	206 785	207 884	208 646	209 887	210 402	211 304	212 75	213 56	214 335	215 79
216 501	217 222	218 844	219 445	220 748	221 504	222 616	223 901	224 194	225 231	226 442	227 55	228 803	229 849	230 420	231 763	232 581	233 631
234 456	235 995	236 728	237 560	238 690	239 240	240 58	241 189	242 248	243 943	244 346	245 579	246 826	247 752	248 856	249 299	250 755	251 798
252 333	253 247	254 601	255 730	256 180	257 996	258 111	259 493	260 175	261 413	262 78	263 861	264 620	265 451	266 269	267 907	268 913	269 266
270 774	271 133	272 8	273 291	274 14	275 62	276 868	277 324	278 154	279 126	280 717	281 962	282 366	283 243	284 689	285 970	286 83	287 848
288 858	289 459	290 586	291 692	292 219	293 606	294 282	295 899	296 544	297 651	298 455	299 513	300 286	301 520	302 374	303 50	304 358	305 182
306 904	307 630	308 112	309 160	310 732	311 98	312 207	313 164	314 483	315 571	316 128	317 449	318 215	319 187	320 482	321 448	322 424	323 991
324 106	325 577	326 838	327 679	328 67	329 405	330 288	331 863	332 417	333 103	334 670	335 378	336 267	337 642	338 489	339 779	340 280	341 339
342 64	343 775	344 695	345 337	346 528	347 789	348 198	349 753	350 327	351 316	352 938	353 708	354 909	355 213	356 35	357 613	358 474	359 596
360 712	361 747	362 184	363 974	364 792	365 51	366 10	367 221	368 118	369 480	370 760	371 26	372 615	373 340	374 551	375 997	376 229	377 934
378 150	379 421	380 714	381 491	382 252	383 733	384 61	385 719	386 477	387 919	388 703	389 814	390 510	391 387	392 191	393 764	394 463	395 885
396 426	397 568	398 296	399 706	400 230	401 44	402 148	403 688	404 348	405 389	406 782	407 672	408 589	409 271	410 923	411 130	412 404	413 982
414 851	415 910	416 759	417 530	418 694	419 393	420 68	421 121	422 546	423 349	424 994	425 408	426 647	427 593	428 969	429 441	430 612	431 96

Table D.1: The fixed substitution box on $\mathbf{Z}_{10}^{3}$ of DEAN27 (part 1)

432 82	433 505	434 31	435 945	436 254	437 486	438 319	439 783	440 81	441 162	442 879	443 936	444 183	445 193	446 444	447 49	448 944	449 464
450 490	451 264	452 488	453 563	454 9	455 457	456 643	457 976	458 84	459 295	460 7	461 116	462 377	463 984	464 864	465 59	466 955	467 507
468 1	469 761	470 906	471 657	472 758	473 806	474 210	475 261	476 941	477 750	478 380	479 972	480 169	481 933	482 136	483 926	484 698	485 674
486 244	487 738	488 237	489 472	490 146	491 338	492 176	493 270	494 811	495 149	496 556	497 524	498 123	499 668	500 669	501 548	502 574	503 506
504 429	505 24	506 629	507 446	508 987	509 354	510 368	511 332	512 246	513 623	514 959	515 559	516 303	517 539	518 743	519 773	520 443	521 836
522 700	523 990	524 583	525 185	526 206	527 52	528 306	529 147	530 795	531 883	532 388	533 894	534 2	535 950	536 888	537 124	538 567	539 641
540 638	541 272	542 716	543 734	544 561	545 419	546 135	547 199	548 396	549 693	550 216	551 167	552 942	553 359	554 619	555 683	556 476	557 166
558 992	559 77	560 361	561 958	562 105	563 458	564 186	565 431	566 595	567 665	568 233	569 740	570 16	571 89	572 821	573 360	574 925	575 352
576 522	577 143	578 701	579 975	580 42	581 780	582 274	583 141	584 76	585 301	586 256	587 854	588 245	589 29	590 819	591 830	592 418	593 308
594 889	595 846	596 223	597 707	598 188	599 20	600 673	601 33	602 36	603 787	604 584	605 32	606 853	607 109	608 73	609 54	610 542	611 828
612 214	613 357	614 287	615 891	616 603	617 645	618 158	619 829	620 598	621 177	622 310	623 503	624 824	625 727	626 129	627 34	628 957	629 201
630 234	631 585	632 372	633 263	634 877	635 573	636 93	637 454	638 805	639 499	640 344	641 253	642 549	643 495	644 37	645 916	646 590	647 713
648 963	649 897	650 91	651 376	652 878	653 655	654 258	655 767	656 48	657 699	658 966	659 226	660 550	661 470	662 731	663 605	664 635	665 547
666 478	667 276	668 788	669 108	670 769	671 323	672 903	673 262	674 161	675 197	676 239	677 772	678 86	679 341	680 369	681 650	682 364	683 345
684 771	685 411	686 710	687 438	688 373	689 410	690 687	691 722	692 94	693 988	694 328	695 406	696 205	697 356	698 43	699 872	700 634	701 375
702 871	703 609	704 259	705 494	706 353	707 371	708 621	709 917	710 702	711 57	712 238	713 425	714 968	715 594	716 831	717 866	718 543	719 825
720 409	721 592	722 735	723 102	724 912	725 940	726 847	727 517	728 961	729 343	730 386	731 475	732 283	733 588	734 718	735 625	736 661	737 663
738 290	739 852	740 195	741 325	742 681	743 558	744 279	745 628	746 212	747 927	748 307	749 979	750 637	751 350	752 881	753 232	754 65	755 800
756 537	757 423	758 15	759 850	760 935	761 799	762 656	763 391	764 440	765 217	766 403	767 953	768 999	769 980	770 363	771 781	772 575	773 676
774 842	775 151	776 660	777 914	778 627	779 140	780 120	781 967	782 540	783 156	784 6	785 762	786 211	787 640	788 370	789 896	790 275	791 220
792 746	793 776	794 924	795 816	796 192	797 25	798 485	799 384	800 22	801 273	802 602	803 12	804 518	805 691	806 298	807 532	808 572	809 47
810 770	811 915	812 802	813 664	814 80	815 460	816 757	817 416	818 611	819 465	820 790	821 835	822 512	823 4	824 618	825 139	826 479	827 0
828 322	829 964	830 285	831 412	832 95	833 545	834 203	835 965	836 251	837 153	838 379	839 60	840 658	841 766	842 686	843 342	844 671	845 145
846 778	847 833	848 845	849 452	850 529	851 260	852 415	853 326	854 971	855 624	856 804	857 948	858 255	859 721	860 993	861 597	862 190	863 466
864 827	865 922	866 983	867 395	868 818	869 843	870 114	871 320	872 862	873 90	874 21	875 654	876 390	877 715	878 978	879 525	880 87	881 18
882 614	883 99	884 952	885 809	886 481	887 315	888 157	889 536	890 168	891 218	892 929	893 178	894 566	895 17	896 249	897 159	898 92	899 3
900 622	901 365	902 591	903 437	904 422	905 321	906 74	907 947	908 882	909 447	910 100	911 937	912 600	913 257	914 956	915 960	916 939	917 318
918 439	919 467	920 682	921 604	922 531	923 578	924 632	925 473	926 66	927 152	928 902	929 839	930 857	931 736	932 39	933 45	934 163	935 432
936 860	937 840	938 27	939 382	940 509	941 684	942 580	943 570	944 880	945 209	946 526	947 928	948 754	949 40	950 777	951 801	952 401	953 347
954 170	955 311	956 662	957 989	958 886	959 515	960 521	961 869	962 196	963 855	964 144	965 890	966 541	967 265	968 677	969 88	970 131	971 697
972 317	973 208	974 535	975 751	976 834	977 807	978 652	979 607	980 873	981 696	982 817	983 726	984 745	985 397	986 46	987 723	988 132	989 28
990 484	991 932	992 13	993 433	994 905	995 744	996 281	997 519	998 300	999 155								

Table D.2: The fixed substitution box on $\mathbf{Z}_{10}^{3}$ of DEAN27 (part 2).

Appendix E

Complementary Informations on the Generalized Linear Cryptanalysis of SAFER

5.1 List of Some of the Possible Successions of Patterns on the Linear Layer

1 → 1	$[0000000*] \xrightarrow{1} [*0000000]$		
1 → 2	$[000*0000] \xrightarrow{1} [**000000]$	$[00000*00] \xrightarrow{1} [*000*000]$	$[000000*0] \xrightarrow{1} [*0*00000]$
2 → 1	$[000*000*] \xrightarrow{1} [0*000000]$	$[00000*0*] \xrightarrow{1} [0000*000]$	$[000000**] \xrightarrow{1} [00*00000]$
1 → 3	None.		
2 → 2	$[0*0*0000] \xrightarrow{1} [0000**00]$	$[0*000*00] \xrightarrow{1} [0*000*00]$	$[00**0000] \xrightarrow{1} [00**0000]$
	$[00*000*0] \xrightarrow{1} [0*0*0000]$	$[000*0*00] \xrightarrow{1} [0*00*000]$	$[000*00*0] \xrightarrow{1} [0**00000]$
	$[0000**00] \xrightarrow{1} [00*000*0]$	$[0000*0*0] \xrightarrow{1} [0000*0*0]$	$[00000**0] \xrightarrow{1} [00*0*000]$
3 → 1	None.		
1 → 4	$[0*000000] \xrightarrow{1} [**00**00]$	$[00*00000] \xrightarrow{1} [****0000]$	$[0000*000] \xrightarrow{1} [*0*0*0*0]$
	$[0000000*] \xrightarrow{2} [***0*000]$		
2 → 3	$[0*00000*] \xrightarrow{2} [*0*00*00]$	$[0*00000*] \xrightarrow{1} [0*00**00]$	$[00*0000*] \xrightarrow{2} [*00**000]$
	$[00*0000*] \xrightarrow{1} [0***0000]$	$[000*000*] \xrightarrow{2} [0*0*0*00]$	$[000*000*] \xrightarrow{2} [*0*0*000]$
	$[0000*00*] \xrightarrow{2} [**0000*0]$	$[0000*00*] \xrightarrow{1} [00*0*0*0]$	$[00000*0*] \xrightarrow{2} [0000***0]$
	$[00000*0*] \xrightarrow{2} [***00000]$	$[000000**] \xrightarrow{2} [00**00*0]$	$[000000**] \xrightarrow{2} [**00*000]$
3 → 2	$[0***0000] \xrightarrow{2} [**000000]$	$[0*0*0*00] \xrightarrow{1} [*0000*00]$	$[0*00**00] \xrightarrow{2} [*000*000]$
	$[0*0000**] \xrightarrow{2} [*0000*00]$	$[00**00*0] \xrightarrow{1} [*00*0000]$	$[00*0*0*0] \xrightarrow{2} [*0*00000]$
	$[00*00*0*] \xrightarrow{2} [*00*0000]$	$[000**00*] \xrightarrow{2} [*00000*0]$	$[000*0*0*] \xrightarrow{2} [*0*00000]$
	$[000*00**] \xrightarrow{2} [*000*000]$	$[0000***0] \xrightarrow{1} [*00000*0]$	$[00000***] \xrightarrow{2} [**000000]$
4 → 1	$[0*0*0*0*] \xrightarrow{1} [00000*00]$	$[00**00**] \xrightarrow{1} [000*0000]$	$[000*0***] \xrightarrow{2} [*0000000]$
	$[0000****] \xrightarrow{1} [000000*0]$		
1 → 5	None.		

Table E.1: List of possible succession of patterns on the linear layer of SAFER.

2 → 4	[**000000] $\xrightarrow{254}$ [**00**00]	[**000000] $\xrightarrow{255}$ [00**00**]	[*0*00000] $\xrightarrow{254}$ [****0000]
	[*0*00000] $\xrightarrow{255}$ [0000****]	[*00*0000] $\xrightarrow{254}$ [**0000**]	[*000*000] $\xrightarrow{254}$ [*0*0*0*0]
	[*000*000] $\xrightarrow{255}$ [0*0*0*0*]	[*0000*00] $\xrightarrow{254}$ [*00**00*]	[*00000*0] $\xrightarrow{254}$ [*0*00*0*]
	[0**00000] $\xrightarrow{255}$ [00****00]	[0*0*0000] $\xrightarrow{254}$ [****0000]	[0*0*0000] $\xrightarrow{254}$ [0000****]
	[0*00*000] $\xrightarrow{255}$ [0**00**0]	[0*000*00] $\xrightarrow{254}$ [*0*0*0*0]	[0*000*00] $\xrightarrow{254}$ [0*0*0*0*]
	[0*0000*0] $\xrightarrow{1}$ [0**0**00]	[0*00000*] $\xrightarrow{252}$ [*0*00*0*]	[00**0000] $\xrightarrow{254}$ [**00**00]
	[00**0000] $\xrightarrow{254}$ [00**00**]	[00*0*000] $\xrightarrow{255}$ [0*0**0*0]	[00*00*00] $\xrightarrow{1}$ [0****000]
	[00*000*0] $\xrightarrow{254}$ [*0*0*0*0]	[00*000*0] $\xrightarrow{254}$ [0*0*0*0*]	[00*0000*] $\xrightarrow{252}$ [*00**00*]
	[000**000] $\xrightarrow{1}$ [0**0*0*0]	[000*0*00] $\xrightarrow{254}$ [0*0**0*0]	[000*00*0] $\xrightarrow{254}$ [0**00**0]
	[000*000*] $\xrightarrow{2}$ [**0*0*00]	[000*000*] $\xrightarrow{252}$ [*0*0*0*0]	[000*000*] $\xrightarrow{252}$ [0*0*0*0*]
	[0000**00] $\xrightarrow{254}$ [**00**00]	[0000**00] $\xrightarrow{254}$ [00**00**]	[0000*0*0] $\xrightarrow{254}$ [****0000]
	[0000*0*0] $\xrightarrow{254}$ [0000****]	[0000*00*] $\xrightarrow{252}$ [**0000**]	[00000**0] $\xrightarrow{254}$ [00****00]
	[00000*0*] $\xrightarrow{252}$ [****0000]	[00000*0*] $\xrightarrow{2}$ [*000***0]	[00000*0*] $\xrightarrow{252}$ [0000****]
	[000000**] $\xrightarrow{252}$ [**00**00]	[000000**] $\xrightarrow{2}$ [*0**00*0]	[000000**] $\xrightarrow{252}$ [00**00**]
3 → 3	[0*0*00*0] $\xrightarrow{2}$ [**0*0000]	[0*0*000*] $\xrightarrow{1}$ [*000**00]	[0*0*000*] $\xrightarrow{2}$ [0*0**000]
	[0*000**0] $\xrightarrow{2}$ [*000*0*0]	[0*000*0*] $\xrightarrow{1}$ [**000*00]	[0*000*0*] $\xrightarrow{2}$ [0*00*0*0]
	[0*0000**] $\xrightarrow{4}$ [**00*000]	[00**0*00] $\xrightarrow{2}$ [**000*00]	[00**000*] $\xrightarrow{1}$ [*0**0000]
	[00**000*] $\xrightarrow{2}$ [0**00*00]	[00*00**0] $\xrightarrow{2}$ [*0*000*0]	[00*00*0*] $\xrightarrow{4}$ [***00000]
	[00*000**] $\xrightarrow{1}$ [**0*0000]	[00*000**] $\xrightarrow{2}$ [0**000*0]	[000***00] $\xrightarrow{2}$ [*000**00]
	[000**0*0] $\xrightarrow{2}$ [*0**0000]	[000**00*] $\xrightarrow{4}$ [*0*0*000]	[000*0*0*] $\xrightarrow{1}$ [**00*000]
	[000*0*0*] $\xrightarrow{4}$ [*0*00*00]	[000*00**] $\xrightarrow{1}$ [***00000]	[000*00**] $\xrightarrow{4}$ [*00**000]
	[0000**0*] $\xrightarrow{1}$ [*0*000*0]	[0000**0*] $\xrightarrow{2}$ [00*0**00]	[0000*0**] $\xrightarrow{1}$ [*000*0*0]
	[0000*0**] $\xrightarrow{2}$ [00***000]	[00000***] $\xrightarrow{4}$ [**0000*0]	[00000***] $\xrightarrow{1}$ [*0*0*000]
4 → 2	[****0000] $\xrightarrow{252}$ [**000000]	[****0000] $\xrightarrow{254}$ [00**0000]	[****0000] $\xrightarrow{254}$ [0000**00]
	[****0000] $\xrightarrow{255}$ [000000**]	[**00**00] $\xrightarrow{252}$ [*000*000]	[**00**00] $\xrightarrow{254}$ [0*000*00]
	[**00**00] $\xrightarrow{254}$ [00*000*0]	[**00**00] $\xrightarrow{255}$ [000*000*]	[**0000**] $\xrightarrow{252}$ [*0000*00]
	[**0000**] $\xrightarrow{254}$ [00*0000*]	[*0*0*0*0] $\xrightarrow{252}$ [*0*00000]	[*0*0*0*0] $\xrightarrow{254}$ [0*0*0000]
	[*0*0*0*0] $\xrightarrow{254}$ [0000*0*0]	[*0*0*0*0] $\xrightarrow{255}$ [00000*0*]	[*0*00*0*] $\xrightarrow{252}$ [*00*0000]
	[*0*00*0*] $\xrightarrow{254}$ [0000*00*]	[*00**00*] $\xrightarrow{252}$ [*00000*0]	[*00**00*] $\xrightarrow{254}$ [0*00000*]
	[0***000*] $\xrightarrow{2}$ [**000000]	[0**00**0] $\xrightarrow{254}$ [00*0*000]	[0**00**0] $\xrightarrow{255}$ [000*0*00]
	[0*0**0*0] $\xrightarrow{254}$ [0**00000]	[0*0**0*0] $\xrightarrow{255}$ [00000**0]	[0*0*0**0] $\xrightarrow{1}$ [00*00*00]
	[0*0*0*0*] $\xrightarrow{252}$ [*0*00000]	[0*0*0*0*] $\xrightarrow{254}$ [0*0*0000]	[0*0*0*0*] $\xrightarrow{254}$ [0000*0*0]
	[0*0*0*0*] $\xrightarrow{254}$ [00000*0*]	[0*00**0*] $\xrightarrow{2}$ [*000*000]	[00****00] $\xrightarrow{254}$ [0*00*000]
	[00****00] $\xrightarrow{255}$ [000*00*0]	[00**0**0] $\xrightarrow{1}$ [000**000]	[00**00**] $\xrightarrow{252}$ [*000*000]
	[00**00**] $\xrightarrow{254}$ [0*000*00]	[00**00**] $\xrightarrow{254}$ [00*000*0]	[00**00**] $\xrightarrow{254}$ [000*000*]
	[00*0*0**] $\xrightarrow{2}$ [*0*00000]	[000****0] $\xrightarrow{1}$ [0*0000*0]	[0000****] $\xrightarrow{252}$ [**000000]
	[0000****] $\xrightarrow{254}$ [00**0000]	[0000****] $\xrightarrow{254}$ [0000**00]	[0000****] $\xrightarrow{254}$ [000000**]
5 → 1	None.		
1 → 6	[000*0000] $\xrightarrow{2}$ [*****00]	[00000*00] $\xrightarrow{2}$ [***0***0]	[000000*0] $\xrightarrow{2}$ [*****0*0]
2 → 5	[*000000*] $\xrightarrow{252}$ [***0*00*]	[*000000*] $\xrightarrow{254}$ [*00*0***]	[0*0000*0] $\xrightarrow{254}$ [*0**0**0]
	[0*0000*0] $\xrightarrow{254}$ [0**0**0*]	[00*00*00] $\xrightarrow{254}$ [*00****0]	[00*00*00] $\xrightarrow{254}$ [0****00*]
	[000**000] $\xrightarrow{254}$ [**0*0**0]	[000**000] $\xrightarrow{254}$ [0**0*0**]	[000*0*00] $\xrightarrow{2}$ [****0*00]
	[000*0*00] $\xrightarrow{254}$ [**0**0*0]	[000*0*00] $\xrightarrow{2}$ [*0*0***0]	[000*00*0] $\xrightarrow{254}$ [***00**0]
	[000*00*0] $\xrightarrow{2}$ [**0***00]	[000*00*0] $\xrightarrow{2}$ [*0***0*0]	[000*000*] $\xrightarrow{252}$ [***0*0*0]
	[000*000*] $\xrightarrow{252}$ [**0*0*0*]	[00000**0] $\xrightarrow{2}$ [****00*0]	[00000**0] $\xrightarrow{2}$ [**00***0]
	[00000**0] $\xrightarrow{254}$ [*0****00]	[00000*0*] $\xrightarrow{252}$ [*****000]	[00000*0*] $\xrightarrow{252}$ [*000****]
	[000000**] $\xrightarrow{252}$ [***0**00]	[000000**] $\xrightarrow{252}$ [*0**00**]	

Table E.2: List of possible succession of patterns on the linear layer of SAFER (continued).

$3 \rightarrow 4$	$[***00000] \xrightarrow{255} [**0000**]$	$[**0*0000] \xrightarrow{254} [**00**00]$	$[**0*0000] \xrightarrow{254} [00****00]$
	$[**00*000] \xrightarrow{255} [*00**00*]$	$[**000*00] \xrightarrow{254} [**00**00]$	$[**000*00] \xrightarrow{254} [0**00**0]$
	$[**0000*0] \xrightarrow{1} [*00*00**]$	$[**00000*] \xrightarrow{254} [**00**00]$	$[**00000*] \xrightarrow{2} [0**0**00]$
	$[*0**0000] \xrightarrow{254} [****0000]$	$[*0**0000] \xrightarrow{254} [00****00]$	$[*0*0*000] \xrightarrow{255} [*0*00*0*]$
	$[*0*00*00] \xrightarrow{1} [*0000***]$	$[*0*000*0] \xrightarrow{254} [****0000]$	$[*0*000*0] \xrightarrow{254} [0*0**0*0]$
	$[*0*0000*] \xrightarrow{254} [****0000]$	$[*0*0000*] \xrightarrow{2} [0****000]$	$[*00**000] \xrightarrow{1} [*00*0*0*]$
	$[*00*000*] \xrightarrow{254} [**0000**]$	$[*000**00] \xrightarrow{254} [*0*0*0*0]$	$[*000**00] \xrightarrow{254} [0**00**0]$
	$[*000*0*0] \xrightarrow{254} [*0*0*0*0]$	$[*000*0*0] \xrightarrow{254} [0*0**0*0]$	$[*000*00*] \xrightarrow{254} [*0*0*0*0]$
	$[*000*00*] \xrightarrow{2} [0**0*0*0]$	$[*0000*0*] \xrightarrow{254} [*00**00*]$	$[*00000**] \xrightarrow{254} [*0*00*0*]$
	$[0***0000] \xrightarrow{252} [****0000]$	$[0***0000] \xrightarrow{252} [**00**00]$	$[0***0000] \xrightarrow{252} [**0000**]$
	$[0***0000] \xrightarrow{254} [00****00]$	$[0**0*000] \xrightarrow{1} [*00*0**0]$	$[0**00*00] \xrightarrow{1} [*0**0*00]$
	$[0**000*0] \xrightarrow{1} [*00***00]$	$[0**0000*] \xrightarrow{2} [**0*0*00]$	$[0*0**000] \xrightarrow{1} [*0*00**0]$
	$[0*0*0*00] \xrightarrow{254} [*0*00*0*]$	$[0*0*00*0] \xrightarrow{252} [****0000]$	$[0*0*00*0] \xrightarrow{1} [*0*0**00]$
	$[0*0*00*0] \xrightarrow{2} [0**0*0*0]$	$[0*0*000*] \xrightarrow{254} [****0000]$	$[0*0*000*] \xrightarrow{2} [***00*00]$
	$[0*0*000*] \xrightarrow{2} [**0**000]$	$[0*0*000*] \xrightarrow{252} [0*0**0*0]$	$[0*00**00] \xrightarrow{252} [**00**00]$
	$[0*00**00] \xrightarrow{252} [*0*0*0*0]$	$[0*00**00] \xrightarrow{252} [*00**00*]$	$[0*00**00] \xrightarrow{254} [0**00**0]$
	$[0*00*0*0] \xrightarrow{1} [**000**0]$	$[0*00*00*] \xrightarrow{2} [*000***0]$	$[0*000**0] \xrightarrow{1} [***00*00]$
	$[0*000**0] \xrightarrow{252} [*0*0*0*0]$	$[0*000**0] \xrightarrow{2} [0****000]$	$[0*000*0*] \xrightarrow{2} [**00*0*0]$
	$[0*000*0*] \xrightarrow{2} [*0*0**00]$	$[0*000*0*] \xrightarrow{254} [*0*0*0*0]$	$[0*000*0*] \xrightarrow{252} [0*0**0*0]$
	$[0*0000**] \xrightarrow{4} [***0*000]$	$[0*0000**] \xrightarrow{248} [**00**00]$	$[0*0000**] \xrightarrow{252} [*0*00*0*]$
	$[00***000] \xrightarrow{1} [*00**0*0]$	$[00**0*00] \xrightarrow{252} [**00**00]$	$[00**0*00] \xrightarrow{1} [*0***000]$
	$[00**0*00] \xrightarrow{2} [0**0*0*0]$	$[00**00*0] \xrightarrow{254} [*00**00*]$	$[00**000*] \xrightarrow{2} [***00*00]$
	$[00**000*] \xrightarrow{2} [**0**000]$	$[00**000*] \xrightarrow{254} [**00**00]$	$[00**000*] \xrightarrow{252} [0**00**0]$
	$[00*0**00] \xrightarrow{1} [**0*00*0]$	$[00*0*0*0] \xrightarrow{252} [****0000]$	$[00*0*0*0] \xrightarrow{252} [*0*0*0*0]$
	$[00*0*0*0] \xrightarrow{252} [*0*00*0*]$	$[00*0*0*0] \xrightarrow{254} [0*0**0*0]$	$[00*0*00*] \xrightarrow{2} [*0**00*0]$
	$[00*00**0] \xrightarrow{1} [**0**000]$	$[00*00**0] \xrightarrow{252} [*0*0*0*0]$	$[00*00**0] \xrightarrow{2} [0**0**00]$
	$[00*00*0*] \xrightarrow{248} [****0000]$	$[00*00*0*] \xrightarrow{4} [***0*000]$	$[00*00*0*] \xrightarrow{252} [*00**00*]$
	$[00*000**] \xrightarrow{2} [***000*0]$	$[00*000**] \xrightarrow{2} [*0***000]$	$[00*000**] \xrightarrow{254} [*0*0*0*0]$
	$[00*000**] \xrightarrow{252} [0**00**0]$	$[000***00] \xrightarrow{1} [***000*0]$	$[000***00] \xrightarrow{252} [**00**00]$
	$[000***00] \xrightarrow{2} [0****000]$	$[000**0*0] \xrightarrow{252} [****0000]$	$[000**0*0] \xrightarrow{1} [**00*0*0]$
	$[000**0*0] \xrightarrow{2} [0**0**00]$	$[000**00*] \xrightarrow{4} [***0*000]$	$[000**00*] \xrightarrow{252} [**0000**]$
	$[000**00*] \xrightarrow{248} [*0*0*0*0]$	$[000*0**0] \xrightarrow{9} [***0*000]$	$[000*0**0] \xrightarrow{2} [**0*0*00]$
	$[000*0**0] \xrightarrow{2} [*0**00*0]$	$[000*0**0] \xrightarrow{2} [*000***0]$	$[000*0*0*] \xrightarrow{252} [****0000]$
	$[000*0*0*] \xrightarrow{4} [***00*00]$	$[000*0*0*] \xrightarrow{4} [*0*0**00]$	$[000*0*0*] \xrightarrow{252} [*0*0*0*0]$
	$[000*0*0*] \xrightarrow{248} [*0*00*0*]$	$[000*0*0*] \xrightarrow{2} [0*0***00]$	$[000*0*0*] \xrightarrow{254} [0*0**0*0]$
	$[000*0*0*] \xrightarrow{2} [0*00***0]$	$[000*00**] \xrightarrow{4} [**0**000]$	$[000*00**] \xrightarrow{252} [**00**00]$
	$[000*00**] \xrightarrow{4} [*0***000]$	$[000*00**] \xrightarrow{252} [*0*0*0*0]$	$[000*00**] \xrightarrow{248} [*00**00*]$
	$[000*00**] \xrightarrow{2} [0***0*00]$	$[000*00**] \xrightarrow{2} [0***00*0]$	$[000*00**] \xrightarrow{254} [0**00**0]$
	$[0000***0] \xrightarrow{254} [**0000**]$	$[0000**0*] \xrightarrow{254} [**00**00]$	$[0000**0*] \xrightarrow{2} [**00*0*0]$
	$[0000**0*] \xrightarrow{2} [*0*0**00]$	$[0000**0*] \xrightarrow{252} [00****00]$	$[0000*0**] \xrightarrow{254} [****0000]$
	$[0000*0**] \xrightarrow{2} [***000*0]$	$[0000*0**] \xrightarrow{2} [*0***000]$	$[0000*0**] \xrightarrow{252} [00****00]$
	$[00000***] \xrightarrow{252} [****0000]$	$[00000***] \xrightarrow{4} [***000*0]$	$[00000***] \xrightarrow{252} [**00**00]$
	$[00000***] \xrightarrow{4} [**00*0*0]$	$[00000***] \xrightarrow{248} [**0000**]$	$[00000***] \xrightarrow{254} [00****00]$
	$[00000***] \xrightarrow{2} [00***0*0]$	$[00000***] \xrightarrow{2} [00*0***0]$	

Table E.3: List of possible succession of patterns on the linear layer of SAFER (continued).

4 → 3	$[***0000*] \xrightarrow{1} [0*0000**]$	$[**00*00*] \xrightarrow{1} [000**00*]$	$[**0000**] \xrightarrow{248} [**00*000]$
	$[**0000**] \xrightarrow{252} [*0*00*00]$	$[**0000**] \xrightarrow{254} [*0*0000*]$	$[**0000**] \xrightarrow{254} [0*00**00]$
	$[**0000**] \xrightarrow{252} [00**00*0]$	$[**0000**] \xrightarrow{255} [000*00**]$	$[*0*0*00*] \xrightarrow{1} [00*00*0*]$
	$[*0*00*0*] \xrightarrow{248} [***00000]$	$[*0*00*0*] \xrightarrow{252} [*00**000]$	$[*0*00*0*] \xrightarrow{254} [*000*00*]$
	$[*0*00*0*] \xrightarrow{254} [0***0000]$	$[*0*00*0*] \xrightarrow{252} [0000***0]$	$[*0*00*0*] \xrightarrow{255} [00000***]$
	$[*00**00*] \xrightarrow{252} [**0000*0]$	$[*00**00*] \xrightarrow{254} [**00000*]$	$[*00**00*] \xrightarrow{248} [*0*0*000]$
	$[*00**00*] \xrightarrow{252} [0*0*0*00]$	$[*00**00*] \xrightarrow{254} [00*0*0*0]$	$[*00**00*] \xrightarrow{255} [000*0*0*]$
	$[0***0*00] \xrightarrow{2} [**00*000]$	$[0***00*0] \xrightarrow{2} [***00000]$	$[0***000*] \xrightarrow{2} [*00*0*00]$
	$[0***000*] \xrightarrow{2} [0**0*000]$	$[0**0*00*] \xrightarrow{1} [000*0**0]$	$[0**00**0] \xrightarrow{254} [*0*0*000]$
	$[0**00**0] \xrightarrow{252} [*0*000*0]$	$[0**00**0] \xrightarrow{252} [*000*0*0]$	$[0**00**0] \xrightarrow{254} [0*0*0*00]$
	$[0**00**0] \xrightarrow{254} [0*0*000*]$	$[0**00**0] \xrightarrow{254} [0*000*0*]$	$[0**00*0*] \xrightarrow{1} [00**0*00]$
	$[0**000**] \xrightarrow{1} [000***00]$	$[0*0***00] \xrightarrow{2} [**00*000]$	$[0*0**0*0] \xrightarrow{254} [***00000]$
	$[0*0**0*0] \xrightarrow{252} [**0*0000]$	$[0*0**0*0] \xrightarrow{252} [*0**0000]$	$[0*0**0*0] \xrightarrow{254} [0000***0]$
	$[0*0**0*0] \xrightarrow{254} [0000**0*]$	$[0*0**0*0] \xrightarrow{254} [0000*0**]$	$[0*0**00*] \xrightarrow{1} [00*00**0]$
	$[0*0*0**0] \xrightarrow{2} [*0000*0*]$	$[0*0*0**0] \xrightarrow{2} [0**000*0]$	$[0*0*0**0] \xrightarrow{2} [00***000]$
	$[0*0*0*0*] \xrightarrow{252} [***00000]$	$[0*0*0*0*] \xrightarrow{254} [**0*0000]$	$[0*0*0*0*] \xrightarrow{252} [*0*0*000]$
	$[0*0*0*0*] \xrightarrow{248} [*0*00*00]$	$[0*0*0*0*] \xrightarrow{254} [*000*0*0]$	$[0*0*0*0*] \xrightarrow{254} [*0000*0*]$
	$[0*0*0*0*] \xrightarrow{252} [0*0**000]$	$[0*0*0*0*] \xrightarrow{252} [0*0*0*00]$	$[0*0*0*0*] \xrightarrow{252} [0*00*0*0]$
	$[0*0*0*0*] \xrightarrow{254} [0*000*0*]$	$[0*0*0*0*] \xrightarrow{252} [0000***0]$	$[0*0*0*0*] \xrightarrow{254} [0000**0*]$
	$[0*0*00**] \xrightarrow{2} [**0*0000]$	$[0*0*00**] \xrightarrow{4} [**00*000]$	$[0*0*00**] \xrightarrow{2} [**000*00]$
	$[0*0*00**] \xrightarrow{1} [00*0**00]$	$[0*00***0] \xrightarrow{2} [*0*0*000]$	$[0*00**0*] \xrightarrow{2} [*0000**0]$
	$[0*00**0*] \xrightarrow{2} [0**0*000]$	$[0*00*0**] \xrightarrow{1} [0*000**0]$	$[0*000***] \xrightarrow{4} [**00*000]$
	$[0*000***] \xrightarrow{2} [*000**00]$	$[0*000***] \xrightarrow{2} [*000*0*0]$	$[0*000***] \xrightarrow{1} [0**00*00]$
	$[00****00] \xrightarrow{254} [**00*000]$	$[00****00] \xrightarrow{252} [**000*00]$	$[00****00] \xrightarrow{252} [*000**00]$
	$[00****00] \xrightarrow{254} [00**00*0]$	$[00****00] \xrightarrow{254} [00**000*]$	$[00****00] \xrightarrow{254} [00*000**]$
	$[00***0*0] \xrightarrow{2} [***00000]$	$[00***00*] \xrightarrow{1} [000**0*0]$	$[00**0**0] \xrightarrow{2} [*00*000*]$
	$[00**0**0] \xrightarrow{2} [0*00*0*0]$	$[00**0**0] \xrightarrow{2} [00*0**00]$	$[00**0*0*] \xrightarrow{4} [***00000]$
	$[00**0*0*] \xrightarrow{2} [**0*0000]$	$[00**0*0*] \xrightarrow{2} [**000*00]$	$[00**0*0*] \xrightarrow{1} [00***000]$
	$[00**00**] \xrightarrow{252} [**00*000]$	$[00**00**] \xrightarrow{254} [**000*00]$	$[00**00**] \xrightarrow{252} [*0*0*000]$
	$[00**00**] \xrightarrow{254} [*0*000*0]$	$[00**00**] \xrightarrow{248} [*00**000]$	$[00**00**] \xrightarrow{254} [*00*000*]$
	$[00**00**] \xrightarrow{252} [0**00*00]$	$[00**00**] \xrightarrow{252} [0**000*0]$	$[00**00**] \xrightarrow{252} [0*0*0*00]$
	$[00**00**] \xrightarrow{254} [0*0*000*]$	$[00**00**] \xrightarrow{252} [00**00*0]$	$[00**00**] \xrightarrow{254} [00**000*]$
	$[00*0***0] \xrightarrow{2} [*0*0*000]$	$[00*0**0*] \xrightarrow{1} [0*0*00*0]$	$[00*0*0**] \xrightarrow{2} [*00*00*0]$
	$[00*0*0**] \xrightarrow{2} [0**0*000]$	$[00*00***] \xrightarrow{4} [***00000]$	$[00*00***] \xrightarrow{2} [*0**0000]$
	$[00*00***] \xrightarrow{2} [*0*000*0]$	$[00*00***] \xrightarrow{1} [0*0**000]$	$[000****0] \xrightarrow{2} [*00000**]$
	$[000****0] \xrightarrow{2} [0**00*00]$	$[000****0] \xrightarrow{2} [0*0**000]$	$[000***0*] \xrightarrow{4} [*0*0*000]$
	$[000***0*] \xrightarrow{2} [*000**00]$	$[000***0*] \xrightarrow{2} [*000*0*0]$	$[000***0*] \xrightarrow{1} [0**000*0]$
	$[000**0**] \xrightarrow{2} [*0**0000]$	$[000**0**] \xrightarrow{4} [*0*0*000]$	$[000**0**] \xrightarrow{2} [*0*000*0]$
	$[000**0**] \xrightarrow{1} [0*00*0*0]$	$[000*0***] \xrightarrow{4} [**0000*0]$	$[000*0***] \xrightarrow{4} [*0*00*00]$
	$[000*0***] \xrightarrow{4} [*00**000]$	$[000*0***] \xrightarrow{9} [0**0*000]$	$[0000****] \xrightarrow{252} [***00000]$
	$[0000****] \xrightarrow{252} [**00*000]$	$[0000****] \xrightarrow{248} [**0000*0]$	$[0000****] \xrightarrow{254} [*0**0000]$
	$[0000****] \xrightarrow{254} [*000**00]$	$[0000****] \xrightarrow{254} [*00000**]$	$[0000****] \xrightarrow{252} [00***000]$
	$[0000****] \xrightarrow{252} [00**00*0]$	$[0000****] \xrightarrow{252} [00*0**00]$	$[0000****] \xrightarrow{254} [00*000**]$
	$[0000****] \xrightarrow{252} [0000***0]$	$[0000****] \xrightarrow{254} [0000*0**]$	

Table E.4: List of possible succession of patterns on the linear layer of SAFER (continued).

$5 \to 2$	[****000*] $\xrightarrow{252}$ [**000000]	[***0*00*] $\xrightarrow{254}$ [*000000*]	[**0*0**0] $\xrightarrow{254}$ [00*00*00]
	[**00**0*] $\xrightarrow{252}$ [*000*000]	[*0**0**0] $\xrightarrow{254}$ [000**000]	[*0*0*0**] $\xrightarrow{252}$ [*0*00000]
	[*00****0] $\xrightarrow{254}$ [0*0000*0]	[*00*0***] $\xrightarrow{252}$ [*000000*]	[0****00*] $\xrightarrow{254}$ [0*0000*0]
	[0***0*0*] $\xrightarrow{2}$ [0**00000]	[0***00**] $\xrightarrow{2}$ [0*00*000]	[0**0**0*] $\xrightarrow{254}$ [000**000]
	[0**0*0**] $\xrightarrow{254}$ [00*00*00]	[0**00***] $\xrightarrow{254}$ [00*0*000]	[0*0***0*] $\xrightarrow{2}$ [00*0*000]
	[0*0**0**] $\xrightarrow{254}$ [0**00000]	[0*0*0***] $\xrightarrow{252}$ [*0*00000]	[0*00****] $\xrightarrow{2}$ [0*00*000]
	[00****0*] $\xrightarrow{254}$ [0*00*000]	[00***0**] $\xrightarrow{2}$ [00*0*000]	[00**0***] $\xrightarrow{252}$ [*000*000]
	[00*0****] $\xrightarrow{2}$ [0**00000]	[000*****] $\xrightarrow{252}$ [**000000]	
$6 \to 1$	[0***0***] $\xrightarrow{2}$ [0*000000]	[0*0*****] $\xrightarrow{2}$ [0000*000]	[00******] $\xrightarrow{2}$ [00*00000]
$6 \to 2$	[******00] $\xrightarrow{254}$ [*0000*00]	[******00] $\xrightarrow{255}$ [00*0000*]	[*****0*0] $\xrightarrow{254}$ [*00*0000]
	[*****0*0] $\xrightarrow{255}$ [0000*00*]	[****0*0*] $\xrightarrow{252}$ [0**00000]	[****0*0*] $\xrightarrow{254}$ [0000**00]
	[****0*0*] $\xrightarrow{254}$ [00000**0]	[****00**] $\xrightarrow{252}$ [0*00*000]	[****00**] $\xrightarrow{254}$ [00**0000]
	[****00**] $\xrightarrow{254}$ [000*00*0]	[***0***0] $\xrightarrow{254}$ [*00000*0]	[***0***0] $\xrightarrow{255}$ [0*00000*]
	[**0***0*] $\xrightarrow{254}$ [0*000*00]	[**0***0*] $\xrightarrow{252}$ [00*0*000]	[**0***0*] $\xrightarrow{254}$ [000*0*00]
	[**0*0***] $\xrightarrow{252}$ [*0000*00]	[**0*0***] $\xrightarrow{254}$ [00*00*00]	[**00****] $\xrightarrow{252}$ [0*00*000]
	[**00****] $\xrightarrow{254}$ [00*000*0]	[**00****] $\xrightarrow{254}$ [000*00*0]	[*0***0**] $\xrightarrow{254}$ [0*0*0000]
	[*0***0**] $\xrightarrow{252}$ [00*0*000]	[*0***0**] $\xrightarrow{254}$ [000*0*00]	[*0**0***] $\xrightarrow{252}$ [*00*0000]
	[*0**0***] $\xrightarrow{254}$ [000**000]	[*0*0****] $\xrightarrow{252}$ [0**00000]	[*0*0****] $\xrightarrow{254}$ [0000*0*0]
	[*0*0****] $\xrightarrow{254}$ [00000**0]	[*00*****] $\xrightarrow{252}$ [*00000*0]	[*00*****] $\xrightarrow{254}$ [0*0000*0]
	[0***0***] $\xrightarrow{2}$ [**000000]	[0***0***] $\xrightarrow{252}$ [*00000*0]	[0***0***] $\xrightarrow{252}$ [0*0*0000]
	[0***0***] $\xrightarrow{252}$ [0*000*00]	[0***0***] $\xrightarrow{252}$ [0*00000*]	[0***0***] $\xrightarrow{254}$ [000*0*00]
	[0*0*****] $\xrightarrow{252}$ [*00*0000]	[0*0*****] $\xrightarrow{2}$ [*000*000]	[0*0*****] $\xrightarrow{252}$ [0000**00]
	[0*0*****] $\xrightarrow{252}$ [0000*0*0]	[0*0*****] $\xrightarrow{252}$ [0000*00*]	[0*0*****] $\xrightarrow{254}$ [00000**0]
	[00******] $\xrightarrow{2}$ [*0*00000]	[00******] $\xrightarrow{252}$ [*0000*00]	[00******] $\xrightarrow{252}$ [00**0000]
	[00******] $\xrightarrow{252}$ [00*000*0]	[00******] $\xrightarrow{252}$ [00*0000*]	[00******] $\xrightarrow{254}$ [000*00*0]

Table E.5: List of possible succession of patterns on the linear layer of SAFER (continued).

5.2 Sequences of Three Weights

1 → 1 → 1	∅	1 → 1 → 2	∅	1 → 1 → 3	∅	1 → 1 → 4	∅	1 → 1 → 5	∅	1 → 1 → 6	∅
1 → 2 → 1	∅	1 → 2 → 2	∅	1 → 2 → 3	∅	1 → 2 → 4	✓	1 → 2 → 5	∅	1 → 2 → 6	
1 → 3 → 1	∅	1 → 3 → 2	∅	1 → 3 → 3	∅	1 → 3 → 4	∅	1 → 3 → 5	∅	1 → 3 → 6	∅
1 → 4 → 1	∅	1 → 4 → 2	✓	1 → 4 → 3	∅	1 → 4 → 4		1 → 4 → 5		1 → 4 → 6	
1 → 5 → 1	∅	1 → 5 → 2	∅	1 → 5 → 3	∅	1 → 5 → 4	∅	1 → 5 → 5	∅	1 → 5 → 6	∅
1 → 6 → 1	∅	1 → 6 → 2		1 → 6 → 3		1 → 6 → 4		1 → 6 → 5		1 → 6 → 6	
2 → 1 → 1	∅	2 → 1 → 2	∅	2 → 1 → 3	∅	2 → 1 → 4	0	2 → 1 → 5	∅	2 → 1 → 6	∅
2 → 2 → 1	∅	2 → 2 → 2	0	2 → 2 → 3	∅	2 → 2 → 4	✓	2 → 2 → 5	∅	2 → 2 → 6	
2 → 3 → 1	∅	2 → 3 → 2	0	2 → 3 → 3	∅	2 → 3 → 4	✓	2 → 3 → 5		2 → 3 → 6	
2 → 4 → 1	✓	2 → 4 → 2	✓	2 → 4 → 3		2 → 4 → 4		2 → 4 → 5		2 → 4 → 6	
2 → 5 → 1	∅	2 → 5 → 2		2 → 5 → 3		2 → 5 → 4		2 → 5 → 5		2 → 5 → 6	
2 → 6 → 1		2 → 6 → 2		2 → 6 → 3		2 → 6 → 4		2 → 6 → 5		2 → 6 → 6	
3 → 1 → 1	∅	3 → 1 → 2	∅	3 → 1 → 3	∅	3 → 1 → 4	∅	3 → 1 → 5	∅	3 → 1 → 6	∅
3 → 2 → 1	∅	3 → 2 → 2	∅	3 → 2 → 3	∅	3 → 2 → 4		3 → 2 → 5	∅	3 → 2 → 6	
3 → 3 → 1	∅	3 → 3 → 2	∅	3 → 3 → 3	∅	3 → 3 → 4		3 → 3 → 5		3 → 3 → 6	
3 → 4 → 1	∅	3 → 4 → 2		3 → 4 → 3		3 → 4 → 4		3 → 4 → 5		3 → 4 → 6	
3 → 5 → 1	∅	3 → 5 → 2		3 → 5 → 3		3 → 5 → 4		3 → 5 → 5		3 → 5 → 6	
3 → 6 → 1		3 → 6 → 2		3 → 6 → 3		3 → 6 → 4		3 → 6 → 5		3 → 6 → 6	
4 → 1 → 1	∅	4 → 1 → 2	0	4 → 1 → 3	∅	4 → 1 → 4	∅	4 → 1 → 5	∅	4 → 1 → 6	
4 → 2 → 1	✓	4 → 2 → 2	✓	4 → 2 → 3		4 → 2 → 4	✓	4 → 2 → 5		4 → 2 → 6	
4 → 3 → 1	∅	4 → 3 → 2		4 → 3 → 3		4 → 3 → 4		4 → 3 → 5		4 → 3 → 6	
4 → 4 → 1		4 → 4 → 2		4 → 4 → 3		4 → 4 → 4		4 → 4 → 5		4 → 4 → 6	
4 → 5 → 1	∅	4 → 5 → 2		4 → 5 → 3		4 → 5 → 4		4 → 5 → 5		4 → 5 → 6	
4 → 6 → 1		4 → 6 → 2		4 → 6 → 3		4 → 6 → 4		4 → 6 → 5		4 → 6 → 6	
5 → 1 → 1	∅	5 → 1 → 2	∅	5 → 1 → 3	∅	5 → 1 → 4	∅	5 → 1 → 5	∅	5 → 1 → 6	∅
5 → 2 → 1	∅	5 → 2 → 2	∅	5 → 2 → 3	∅	5 → 2 → 4		5 → 2 → 5		5 → 2 → 6	
5 → 3 → 1	∅	5 → 3 → 2		5 → 3 → 3		5 → 3 → 4		5 → 3 → 5		5 → 3 → 6	
5 → 4 → 1		5 → 4 → 2		5 → 4 → 3		5 → 4 → 4		5 → 4 → 5		5 → 4 → 6	
5 → 5 → 1	∅	5 → 5 → 2		5 → 5 → 3		5 → 5 → 4		5 → 5 → 5		5 → 5 → 6	
5 → 6 → 1		5 → 6 → 2		5 → 6 → 3		5 → 6 → 4		5 → 6 → 5		5 → 6 → 6	
6 → 1 → 1	∅	6 → 1 → 2	∅	6 → 1 → 3	∅	6 → 1 → 4		6 → 1 → 5	∅	6 → 1 → 6	∅
6 → 2 → 1		6 → 2 → 2		6 → 2 → 3		6 → 2 → 4		6 → 2 → 5		6 → 2 → 6	
6 → 3 → 1	∅	6 → 3 → 2		6 → 3 → 3		6 → 3 → 4		6 → 3 → 5		6 → 3 → 6	
6 → 4 → 1		6 → 4 → 2		6 → 4 → 3		6 → 4 → 4		6 → 4 → 5		6 → 4 → 6	
6 → 5 → 1	∅	6 → 5 → 2		6 → 5 → 3		6 → 5 → 4		6 → 5 → 5		6 → 5 → 6	
6 → 6 → 1		6 → 6 → 2		6 → 6 → 3		6 → 6 → 4		6 → 6 → 5		6 → 6 → 6	

Table E.6: List of all possible succession of weights for patterns on two rounds.

5.3 Complexities of the Attacks against 3, 4, and 5 Rounds

Reduced hull	$\min_{\mathbf{a}_0,\mathbf{a}_3} \frac{8 \ln 2}{(d-1)\mathrm{ELP}^{\mathsf{H}(3)}(\mathbf{a}_0,\mathbf{a}_3)}$	2^{n_p}	2^{n_k}	Complexity
$\texttt{[000*0000]} \xrightarrow{1} \texttt{[**000000]} \xrightarrow{255} \texttt{[00**00**]} \xrightarrow{1} \texttt{[000*0000]}$	$2^{39.18}$	2^{16}	$2^8/2^{16}$	$2^{39.18}/2^{39.18}$
$\texttt{[00000*00]} \xrightarrow{1} \texttt{[*000*000]} \xrightarrow{255} \texttt{[0*0*0*0*]} \xrightarrow{1} \texttt{[00000*00]}$	$2^{39.07}$	2^{16}	$2^8/2^{16}$	$2^{39.07}/2^{39.07}$
$\texttt{[000000*0]} \xrightarrow{1} \texttt{[*0*00000]} \xrightarrow{255} \texttt{[0000****]} \xrightarrow{1} \texttt{[000000*0]}$	$2^{39.18}$	2^{16}	$2^8/2^{16}$	$2^{39.18}/2^{39.18}$
$\texttt{[0*000000]} \xrightarrow{1} \texttt{[**00**00]} \xrightarrow{255} \texttt{[000*000*]} \xrightarrow{1} \texttt{[0*000000]}$	$2^{38.75}$	2^{16}	$2^8/2^{16}$	$\mathbf{2^{38.75}/2^{38.75}}$
$\texttt{[0000*000]} \xrightarrow{1} \texttt{[*0*0*0*0]} \xrightarrow{255} \texttt{[00000*0*]} \xrightarrow{1} \texttt{[0000*000]}$	$2^{39.18}$	2^{16}	$2^8/2^{16}$	$2^{39.18}/2^{39.18}$
$\texttt{[00*00000]} \xrightarrow{1} \texttt{[****0000]} \xrightarrow{255} \texttt{[000000**]} \xrightarrow{1} \texttt{[00*00000]}$	$2^{39.18}$	2^{16}	$2^8/2^{16}$	$2^{39.18}/2^{39.18}$

Table E.7: Reduced hull on three diffusion layers and attack complexities against three rounds of SAFER K/SK.

Reduced hull	$\min \frac{8 \ln 2}{(d-1)\mathrm{ELP}}$	2^{n_p}	2^{n_k}	Complexity
$[000*0000] \xrightarrow{1} [**000000] \xrightarrow{254} [**00**00] \xrightarrow{255} [000*000*] \xrightarrow{1} [0*000000]$	$2^{49.22}$	2^{16}	$2^{16}/2^{24}$	$2^{49.22}/\mathbf{2^{49.22}}$
$[000*0000] \xrightarrow{1} [**000000] \xrightarrow{255} [00**00**] \xrightarrow{254} [000*000*] \xrightarrow{1} [0*000000]$	0			
$[00000*00] \xrightarrow{1} [*000*000] \xrightarrow{254} [*0*0*0*0] \xrightarrow{255} [00000*0*] \xrightarrow{1} [0000*000]$	$2^{49.82}$	2^{16}	$2^{8}/2^{8}$	$2^{49.82}/2^{49.82}$
$[00000*00] \xrightarrow{1} [*000*000] \xrightarrow{255} [0*0*0*0*] \xrightarrow{254} [00000*0*] \xrightarrow{1} [0000*000]$	0			
$[000000*0] \xrightarrow{1} [*0*00000] \xrightarrow{254} [****0000] \xrightarrow{255} [000000**] \xrightarrow{1} [00*00000]$	$2^{50.56}$	2^{16}	$2^{24}/2^{24}$	$2^{50.56}/2^{50.56}$
$[000000*0] \xrightarrow{1} [*0*00000] \xrightarrow{255} [0000****] \xrightarrow{254} [000000**] \xrightarrow{1} [00*00000]$	0			
$[0*0*0000] \xrightarrow{1} [0000**00] \xrightarrow{254} [**00**00] \xrightarrow{255} [000*000*] \xrightarrow{1} [0*000000]$	$2^{49.18}$	2^{24}	$2^{16}/2^{32}$	$\mathbf{2^{49.18}}/2^{56}$
$[0*0*0000] \xrightarrow{1} [0000**00] \xrightarrow{254} [00**00**] \xrightarrow{254} [000*000*] \xrightarrow{1} [0*000000]$	0			
$[0*000*00] \xrightarrow{1} [0*000*00] \xrightarrow{254} [0*0*0*0*] \xrightarrow{254} [00000*0*] \xrightarrow{1} [0000*000]$	0			
$[00**0000] \xrightarrow{1} [00**0000] \xrightarrow{254} [00**00**] \xrightarrow{254} [000*000*] \xrightarrow{1} [0*000000]$	0			
$[00*000*0] \xrightarrow{1} [0*0*0000] \xrightarrow{254} [0000****] \xrightarrow{254} [000000**] \xrightarrow{1} [00*00000]$	0			
$[0000**00] \xrightarrow{1} [00*000*0] \xrightarrow{254} [0*0*0*0*] \xrightarrow{254} [00000*0*] \xrightarrow{1} [0000*000]$	0			
$[0000*0*0] \xrightarrow{1} [0000*0*0] \xrightarrow{254} [****0000] \xrightarrow{255} [000000**] \xrightarrow{1} [00*00000]$	$2^{50.82}$	2^{24}	$2^{24}/2^{32}$	$2^{50.82}/2^{56}$
$[0000*0*0] \xrightarrow{1} [0000*0*0] \xrightarrow{254} [0000****] \xrightarrow{254} [000000**] \xrightarrow{1} [00*00000]$	0			
$[000*0000] \xrightarrow{1} [**000000] \xrightarrow{254} [**00**00] \xrightarrow{254} [0*000*00] \xrightarrow{1} [0*000*00]$	$2^{49.39}$	2^{24}	$2^{24}/2^{32}$	$2^{49.39}/2^{56}$
$[000*0000] \xrightarrow{1} [**000000] \xrightarrow{254} [**00**00] \xrightarrow{254} [00*000*0] \xrightarrow{1} [0*0*0000]$	$2^{49.78}$	2^{24}	$2^{16}/2^{24}$	$2^{49.78}/2^{49.78}$
$[000*0000] \xrightarrow{1} [**000000] \xrightarrow{255} [00**00**] \xrightarrow{254} [0*000*00] \xrightarrow{1} [0*000*00]$	$2^{49.82}$	2^{24}	$2^{24}/2^{32}$	$2^{49.82}/2^{54}$
$[000*0000] \xrightarrow{1} [**000000] \xrightarrow{255} [00**00**] \xrightarrow{254} [00*000*0] \xrightarrow{1} [0*0*0000]$	$2^{50.17}$	2^{24}	$2^{16}/2^{24}$	$2^{50.17}/2^{50.17}$
$[00000*00] \xrightarrow{1} [*000*000] \xrightarrow{254} [*0*0*0*0] \xrightarrow{254} [0*0*0000] \xrightarrow{1} [0000**00]$	$2^{49.82}$	2^{24}	$2^{8}/2^{16}$	$2^{49.82}/2^{49.82}$
$[00000*00] \xrightarrow{1} [*000*000] \xrightarrow{254} [*0*0*0*0] \xrightarrow{254} [0000*0*0] \xrightarrow{1} [0000*0*0]$	$2^{50.08}$	2^{24}	$2^{16}/2^{8}$	$2^{50.08}/2^{50.08}$
$[00000*00] \xrightarrow{1} [*000*000] \xrightarrow{255} [0*0*0*0*] \xrightarrow{254} [0*0*0000] \xrightarrow{1} [0000**00]$	0			
$[00000*00] \xrightarrow{1} [*000*000] \xrightarrow{255} [0*0*0*0*] \xrightarrow{254} [0000*0*0] \xrightarrow{1} [0000*0*0]$	$2^{49.74}$	2^{24}	$2^{16}/2^{8}$	$2^{49.74}/2^{49.74}$
$[000000*0] \xrightarrow{1} [*0*00000] \xrightarrow{254} [****0000] \xrightarrow{254} [00**0000] \xrightarrow{1} [00**0000]$	$2^{50.83}$	2^{24}	$2^{16}/2^{16}$	$2^{50.83}/2^{50.83}$
$[000000*0] \xrightarrow{1} [*0*00000] \xrightarrow{254} [****0000] \xrightarrow{254} [0000**00] \xrightarrow{1} [00*000*0]$	$2^{50.83}$	2^{24}	$2^{24}/2^{24}$	$2^{50.83}/2^{50.83}$
$[000000*0] \xrightarrow{1} [*0*00000] \xrightarrow{255} [0000****] \xrightarrow{254} [00**0000] \xrightarrow{1} [00**0000]$	0			
$[000000*0] \xrightarrow{1} [*0*00000] \xrightarrow{255} [0000****] \xrightarrow{254} [0000**00] \xrightarrow{1} [00*000*0]$	$2^{50.82}$	2^{24}	$2^{24}/2^{24}$	$2^{50.82}/2^{50.82}$

Table E.8: Reduced hull on four diffusion layers and attack complexities against four rounds of SAFER K/SK.

Reduced hull	$\min \frac{8 \ln 2}{(d-1)\mathrm{ELP}}$	2^{n_p}	2^{n_k}
$[000*0000] \xrightarrow{1} [**000000] \xrightarrow{254} [**00**00] \xrightarrow{254} [0*000*00] \xrightarrow{1} [0*000*00] \xrightarrow{254} [*0*0*0*0]$	0		
$[000*0000] \xrightarrow{1} [**000000] \xrightarrow{254} [**00**00] \xrightarrow{254} [0*000*00] \xrightarrow{1} [0*000*00] \xrightarrow{254} [0*0*0*0*]$	$2^{53.21}$	2^{40}	$2^{16}/2^{24}$
$[000*0000] \xrightarrow{1} [**000000] \xrightarrow{255} [00**00**] \xrightarrow{254} [0*000*00] \xrightarrow{1} [0*000*00] \xrightarrow{254} [*0*0*0*0]$	0		
$[000*0000] \xrightarrow{1} [**000000] \xrightarrow{255} [00**00**] \xrightarrow{254} [0*000*00] \xrightarrow{1} [0*000*00] \xrightarrow{254} [0*0*0*0*]$	$2^{53.65}$	2^{40}	$2^{16}/2^{24}$
$[000*0000] \xrightarrow{1} [**000000] \xrightarrow{254} [**00**00] \xrightarrow{254} [00*000*0] \xrightarrow{1} [0*0*0000] \xrightarrow{254} [****0000]$	0		
$[000*0000] \xrightarrow{1} [**000000] \xrightarrow{254} [**00**00] \xrightarrow{254} [00*000*0] \xrightarrow{1} [0*0*0000] \xrightarrow{254} [0000****]$	$2^{54.93}$	2^{40}	$2^{24}/2^{24}$
$[000*0000] \xrightarrow{1} [**000000] \xrightarrow{255} [00**00**] \xrightarrow{254} [00*000*0] \xrightarrow{1} [0*0*0000] \xrightarrow{254} [****0000]$	0		
$[000*0000] \xrightarrow{1} [**000000] \xrightarrow{255} [00**00**] \xrightarrow{254} [00*000*0] \xrightarrow{1} [0*0*0000] \xrightarrow{254} [0000****]$	$2^{55.32}$	2^{40}	$2^{24}/2^{24}$
$[00000*00] \xrightarrow{1} [*000*000] \xrightarrow{255} [0*0*0*0*] \xrightarrow{254} [0*0*0000] \xrightarrow{1} [0000**00] \xrightarrow{254} [**00**00]$	0		
$[00000*00] \xrightarrow{1} [*000*000] \xrightarrow{255} [0*0*0*0*] \xrightarrow{254} [0*0*0000] \xrightarrow{1} [0000**00] \xrightarrow{254} [00**00**]$	0		
$[00000*00] \xrightarrow{1} [*000*000] \xrightarrow{254} [*0*0*0*0] \xrightarrow{254} [0*0*0000] \xrightarrow{1} [0000**00] \xrightarrow{254} [**00**00]$	$2^{54.71}$	2^{40}	$2^{24}/2^{24}$
$[00000*00] \xrightarrow{1} [*000*000] \xrightarrow{254} [*0*0*0*0] \xrightarrow{254} [0*0*0000] \xrightarrow{1} [0000**00] \xrightarrow{254} [00**00**]$	$2^{54.97}$	2^{40}	$2^{24}/2^{32}$
$[00000*00] \xrightarrow{1} [*000*000] \xrightarrow{254} [*0*0*0*0] \xrightarrow{254} [0000*0*0] \xrightarrow{1} [0000*0*0] \xrightarrow{254} [****0000]$	$2^{54.97}$	2^{40}	$2^{24}/2^{32}$
$[00000*00] \xrightarrow{1} [*000*000] \xrightarrow{254} [*0*0*0*0] \xrightarrow{254} [0000*0*0] \xrightarrow{1} [0000*0*0] \xrightarrow{254} [0000****]$	$2^{55.23}$	2^{40}	$2^{24}/2^{24}$
$[00000*00] \xrightarrow{1} [*000*000] \xrightarrow{255} [0*0*0*0*] \xrightarrow{254} [0000*0*0] \xrightarrow{1} [0000*0*0] \xrightarrow{254} [****0000]$	$2^{54.63}$	2^{40}	$2^{24}/2^{32}$
$[00000*00] \xrightarrow{1} [*000*000] \xrightarrow{255} [0*0*0*0*] \xrightarrow{254} [0000*0*0] \xrightarrow{1} [0000*0*0] \xrightarrow{254} [0000****]$	$2^{54.89}$	2^{40}	$2^{24}/2^{24}$
$[000000*0] \xrightarrow{1} [*0*00000] \xrightarrow{254} [****0000] \xrightarrow{254} [00**0000] \xrightarrow{1} [00**0000] \xrightarrow{254} [**00**00]$	0		
$[000000*0] \xrightarrow{1} [*0*00000] \xrightarrow{254} [****0000] \xrightarrow{254} [00**0000] \xrightarrow{1} [00**0000] \xrightarrow{254} [00**00**]$	$2^{55.98}$	2^{40}	$2^{24}/2^{32}$
$[000000*0] \xrightarrow{1} [*0*00000] \xrightarrow{255} [0000****] \xrightarrow{254} [00**0000] \xrightarrow{1} [00**0000] \xrightarrow{254} [**00**00]$	0		
$[000000*0] \xrightarrow{1} [*0*00000] \xrightarrow{255} [0000****] \xrightarrow{254} [00**0000] \xrightarrow{1} [00**0000] \xrightarrow{254} [00**00**]$	0		
$[000000*0] \xrightarrow{1} [*0*00000] \xrightarrow{254} [****0000] \xrightarrow{254} [0000**00] \xrightarrow{1} [00*000*0] \xrightarrow{254} [*0*0*0*0]$	0		
$[000000*0] \xrightarrow{1} [*0*00000] \xrightarrow{254} [****0000] \xrightarrow{254} [0000**00] \xrightarrow{1} [00*000*0] \xrightarrow{254} [0*0*0*0*]$	$2^{54.65}$	2^{40}	$2^{24}/2^{32}$
$[000000*0] \xrightarrow{1} [*0*00000] \xrightarrow{255} [0000****] \xrightarrow{254} [0000**00] \xrightarrow{1} [00*000*0] \xrightarrow{254} [*0*0*0*0]$	0		
$[000000*0] \xrightarrow{1} [*0*00000] \xrightarrow{255} [0000****] \xrightarrow{254} [0000**00] \xrightarrow{1} [00*000*0] \xrightarrow{254} [0*0*0*0*]$	$2^{54.65}$	2^{40}	$2^{24}/2^{32}$

Table E.9: Reduced hull on five diffusion layers and attack complexities against five rounds of SAFER K/SK.

Bibliography

[1] Carlisle M. Adams, Howard M. Heys, Stafford E. Tavares, and Michael J. Wiener. CAST256: a submission for the advanced encryption standard, 1998. First AES Candidate Conference (AES1).

[2] Ross J. Anderson, editor. *Fast Software Encryption, Cambridge Security Workshop, Cambridge, UK, December 9-11, 1993, Proceedings*, volume 809 of *LNCS*. Springer-Verlag, 1994.

[3] Kazumaro Aoki, Tetsuya Ichikawa, Masayuki Kanda, Mitsuru Matsui, Shiho Moriai, Junko Nakajima, and Toshio Tokita. Camellia: A 128-bit block cipher suitable for multiple platforms - design and analysis. In Stinson and Tavares [146], pages 39–56.

[4] Kazumaro Aoki and Serge Vaudenay. On the use of GF-inversion as a cryptographic primitive. In Mitsuru Matsui and Robert J. Zuccherato, editors, *Selected Areas in Cryptography, 10th Annual International Workshop, SAC 2003, Ottawa, Canada, August 14-15, 2003, Revised Papers*, volume 3006 of *LNCS*, pages 234–247. Springer-Verlag, 2004.

[5] Thomas Baignères and Matthieu Finiasz. KFC - the Krazy Feistel Cipher. In Xuejia Lai and Kefei Chen, editors, *Advances in Cryptology - ASIACRYPT 2006, 12th International Conference on the Theory and Application of Cryptology and Information Security, Shanghai, China, December 3-7, 2006, Proceedings*, volume 4284 of *LNCS*, pages 380–395. Springer-Verlag, 2006.

[6] Thomas Baignères and Matthieu Finiasz. Dial C for Cipher. In Biham and Youssef [25], pages 76–95.

[7] Thomas Baignères, Pascal Junod, and Serge Vaudenay. How far can we go beyond linear cryptanalysis? In Pil Joong Lee, editor, *Advances in Cryptology - ASIACRYPT 2004, 10th International Conference on the Theory and Application of Cryptology and Information Security, Jeju Island, Korea, December 5-9, 2004, Proceedings*, volume 3329 of *LNCS*, pages 432–450. Springer-Verlag, 2004.

[8] Thomas Baignères, Jacques Stern, and Serge Vaudenay. Linear cryptanalysis of non binary ciphers. In Carlisle M. Adams, Ali Miri, and Michael J. Wiener,

editors, *Selected Areas in Cryptography, 14th International Workshop, SAC 2007, Ottawa, Canada, August 16-17, 2007, Revised Selected Papers*, volume 4876 of *LNCS*, pages 184–211. Springer-Verlag, 2007.

[9] Thomas Baignères and Serge Vaudenay. Proving the security of AES substitution-permutation network. In Bart Preneel and Stafford E. Tavares, editors, *Selected Areas in Cryptography, 12th International Workshop, SAC 2005, Kingston, ON, Canada, August 11-12, 2005, Revised Selected Papers*, volume 3897 of *LNCS*, pages 65–81. Springer-Verlag, 2006.

[10] Thomas Baignères and Serge Vaudenay. The complexity of distinguishing distributions. In Rei Safavi-Naini, editor, *The 2nd International Conference on Information Theoretic Security (ICITS)*, LNCS. Springer-Verlag, 2008. To be published.

[11] Mihir Bellare, Anand Desai, E. Jokipii, and Phillip Rogaway. A concrete security treatment of symmetric encryption. In *38th Symposium on Foundations of Computer Science*. IEEE, 1997.

[12] Mihir Bellare, David Pointcheval, and Phillip Rogaway. Authenticated key exchange secure against dictionary attacks. In Bart Preneel, editor, *Advances in Cryptology - EUROCRYPT 2000, International Conference on the Theory and Application of Cryptographic Techniques, Bruges, Belgium, May 14-18, 2000, Proceeding*, volume 1807 of *LNCS*, pages 139–155. Springer-Verlag, 2000.

[13] Côme Berbain, Olivier Billet, and Henri Gilbert. Efficient implementations of multivariate quadratic systems. In Biham and Youssef [25], pages 174–187.

[14] Côme Berbain, Henri Gilbert, and Jacques Patarin. QUAD: a practical stream cipher with provable security. In Vaudenay [156], pages 109–128.

[15] Irenée-Jules Bienaymé. Mémoire sur la probabilité des résultats moyens des observations; démonstration directe de la règle de Laplace. *Mémoires de l'Académie des Sciences de l'Institut de France, Paris, Série Etrangers*, 5:513–558, 1838.

[16] Eli Biham. On Matsui's linear cryptanalysis. In De Santis [45], pages 341–355.

[17] Eli Biham, editor. *Fast Software Encryption, 4th International Workshop, FSE '97, Haifa, Israel, January 20-22, 1997, Proceedings*, volume 1267 of *LNCS*. Springer-Verlag, 1997.

[18] Eli Biham, Alex Biryukov, and Adi Shamir. Cryptanalysis of Skipjack reduced to 31 rounds using impossible differentials. In Stern [143], pages 12–23.

[19] Eli Biham, Orr Dunkelman, and Nathan Keller. The rectangle attack - rectangling the Serpent. In Pfitzmann [130], pages 340–357.

[20] Eli Biham, Orr Dunkelman, and Nathan Keller. Enhancing differential-linear cryptanalysis. In Zheng [168], pages 254–266.

[21] Eli Biham and Adi Shamir. Differential cryptanalysis of DES-like cryptosystems. *Journal of Cryptology*, 4:3–72, 1991.

[22] Eli Biham and Adi Shamir. Differential cryptanalysis of DES-like cryptosystems (extended abstract). In Alfred Menezes and Scott A. Vanstone, editors, *Advances in Cryptology - CRYPTO '90, 10th Annual International Cryptology Conference, Santa Barbara, California, USA, August 11-15, 1990, Proceedings*, volume 537 of *LNCS*, pages 2–21. Springer-Verlag, 1991.

[23] Eli Biham and Adi Shamir. *Differential Cryptanalysis of the Data Encryption Standard*. Springer-Verlag, 1993.

[24] Eli Biham and Adi Shamir. Differential cryptanalysis of the full 16-round DES. In Ernest F. Brickell, editor, *Advances in Cryptology - CRYPTO '92, 12th Annual International Cryptology Conference, Santa Barbara, California, USA, August 16-20, 1992, Proceedings*, volume 740 of *LNCS*, pages 487–496. Springer-Verlag, 1993.

[25] Eli Biham and Amr M. Youssef, editors. *Selected Areas in Cryptography, 13th International Workshop, SAC 2006, Montreal, Canada, August 17-18, 2006 Revised Selected Papers*, volume 4356 of *LNCS*. Springer-Verlag, 2007.

[26] Alex Biryukov and Adi Shamir. Structural cryptanalysis of SASAS. In Pfitzmann [130], pages 394–405.

[27] Alex Biryukov and David Wagner. Slide attacks. In Knudsen [89], pages 245–259.

[28] G. R. Blakley and David Chaum, editors. *Advances in Cryptology, Proceedings of CRYPTO '84, Santa Barbara, California, USA, August 19-22, 1984, Proceedings*, volume 196 of *LNCS*. Springer-Verlag, 1985.

[29] Lenore Blum, Manuel Blum, and Mike Shub. Comparison of two pseudo-random number generators. In David Chaum, Ronald L. Rivest, and Alan T. Sherman, editors, *Advances in Cryptology: Proceedings of CRYPTO '82*, pages 61–78, New York, 1983. Plenum.

[30] Lenore Blum, Manuel Blum, and Mike Shub. A simple unpredictable pseudo-random number generator. *SIAM J. Comput.*, 15(2):364–383, May 1986.

[31] Karl Brincat and Alko Meijer. On the SAFER cryptosystem. In Michael Darnell, editor, *Cryptography and Coding, 6th IMA International Conference*, volume 1355 of *LNCS*, pages 59–68, 1997.

[32] Florent Chabaud and Serge Vaudenay. Links between differential and linear cryptanalysis. In De Santis [45], pages 356–365.

[33] Jung Hee Cheon, MunJu Kim, Kwangjo Kim, Jung-Yeun Lee, and SungWoo Kang. Improved impossible differential cryptanalysis of Rijndael and Crypton.

In *Information Security and Cryptology ICISC'01*, volume 2288 of *LNCS*, pages 39–49. Springer-Verlag, 2002.

[34] Herman Chernoff. *Sequential Analysis and Optimal Design*, volume 8 of *CBMS-NSF Regional Conference Series in Applied Mathematics.* SIAM, 1972.

[35] Don Coppersmith, Shai Halevi, and Charanjit S. Jutla. Cryptanalysis of stream ciphers with linear masking. In Moti Yung, editor, *Advances in Cryptology - CRYPTO 2002, 22nd Annual International Cryptology Conference, Santa Barbara, California, USA, August 18-22, 2002, Proceedings*, volume 2442 of *LNCS*, pages 515–532. Springer-Verlag, 2002.

[36] Nicolas Courtois and Josef Pieprzyk. Cryptanalysis of block ciphers with overdefined systems of equations. In Zheng [168], pages 267–287.

[37] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory.* Wiley Series in Telecommunications. John Wiley & Sons, 1991.

[38] Ronald Cramer and Victor Shoup. A practical public key cryptosystem provably secure against adaptive chosen ciphertext attack. In Hugo Krawczyk, editor, *Advances in Cryptology - CRYPTO '98, 18th Annual International Cryptology Conference, Santa Barbara, California, USA, August 23-27, 1998, Proceedings*, volume 1462 of *LNCS*, pages 13–25. Springer-Verlag, 1998.

[39] Noel Cressie and Timothy R.C. Read. Multinomial goodness-of-fit tests. *Journal of the Royal Statistical Society. Series B (Methodological)*, 46(3):440–464, 1984.

[40] Joan Daemen, Lars R. Knudsen, and Vincent Rijmen. The block cipher SQUARE. In Biham [17], pages 149–165.

[41] Joan Daemen and Vincent Rijmen. AES proposal: Rijndael. NIST AES Proposal, 1998.

[42] Joan Daemen and Vincent Rijmen. *The Design of Rijndael.* Information Security and Cryptography. Springer-Verlag, 2002.

[43] Donald W. Davies. Some regular properties of the DES. In Allen Gersho, editor, *Advances in Cryptology: a report on CRYPTO'81, IEEE Workshop on Communication Security, Santa Barbara, August 24-26, 1981. U.C. Santa Barbara, Dept. of Elec. and Computer Eng., ECE Report No 82-84*, page 41, 1982.

[44] Donald W. Davies, editor. *Advances in Cryptology - EUROCRYPT '91, Workshop on the Theory and Application of of Cryptographic Techniques, Brighton, UK, April 8-11, 1991, Proceedings*, volume 547 of *LNCS.* Springer-Verlag, 1991.

[45] Alfredo De Santis, editor. *Advances in Cryptology - EUROCRYPT '94, Workshop on the Theory and Application of Cryptographic Techniques, Perugia, Italy, May 9-12, 1994, Proceedings*, volume 950 of *LNCS.* Springer-Verlag, 1995.

[46] Yvo Desmedt, editor. *Advances in Cryptology - CRYPTO '94, 14th Annual International Cryptology Conference, Santa Barbara, California, USA, August 21-25, 1994, Proceedings*, volume 839 of *LNCS*. Springer-Verlag, 1994.

[47] Whitfield Diffie and Martin E. Hellman. New directions in cryptography. *IEEE Transactions on Information Theory*, 22:644–654, 1976.

[48] Mostafa El-Khamy and Robert J. McEliece. The partition weight enumerator of MDS codes and its applications. In *IEEE International Symposium on Information Theory, ISIT 2005*. IEEE, 2005. Available on `http://arxiv.org/pdf/cs.IT/0505054`.

[49] Taher ElGamal. A public key cryptosystem and signature scheme based on discrete logarithms. In Blakley and Chaum [28], pages 10–18.

[50] Horst Feistel. Cryptography and computer privacy. *Scientific American*, 228:15–23, 1973.

[51] Niels Ferguson, John Kelsey, Stefan Lucks, Bruce Schneier, Michael Stay, David Wagner, and Doug Whiting. Improved cryptanalysis of Rijndael. In Schneier [135], pages 213–230.

[52] Murray F. Freeman and John W. Tukey. Transformations related to the angular and the square root. *The Annals of Mathematical Statistics*, 21(4):607–611, 1950.

[53] Christian Gehrmann and Mats Näslund. Ecrypt yearly report on algorithms and keysizes (2005). Technical report, Ecrypt, January 2006.

[54] Henri Gilbert, Marc Girault, Philippe Hoogvorst, Fabrice Noilhan, Thomas Pornin, Guillaume Poupard, Jacques Stern, and Serge Vaudenay. Decorrelated Fast Cipher: an AES candidate (extended abstract). In *Proceedings from the First Advanced Encryption Standard Candidate Conference*. National Institute of Technology (NIST), August 1998.

[55] GMP. GNU Multiple Precision arithmetic library. `http://www.swox.com/gmp`.

[56] Shafi Goldwasser and Silvio Micali. Probabilistic Encryption. *Journal of Computer and System Sciences*, 28(2):270–299, 1984.

[57] Dieter Gollmann, editor. *Fast Software Encryption, Third International Workshop, Cambridge, UK, February 21-23, 1996, Proceedings*, volume 1039 of *LNCS*. Springer-Verlag, 1996.

[58] Louis Granboulan, Éric Levieil, and Gilles Piret. Pseudorandom permutation families over Abelian groups. In Matthew J. B. Robshaw, editor, *Fast Software Encryption, 13th International Workshop, FSE 2006, Graz, Austria, March 15-17, 2006, Revised Selected Papers*, volume 4047 of *LNCS*, pages 57–77. Springer-Verlag, 2006.

[59] Louis Granboulan, Phong Q. Nguyen, Fabrice Noilhan, and Serge Vaudenay. DFCv2. In Stinson and Tavares [146], pages 57–71.

[60] Geoffrey Grimmett and David Stirzaker. *Probability and Random Processes.* Oxford University Press, 3d edition, 2001.

[61] Olle Häggström. *Finite Markov Chains and Algorithmic Applications.* London Mathematical Society Student Texts. Cambridge University Press, 2002.

[62] Helena Handschuh and Henri Gilbert. χ^2 cryptanalysis of the SEAL encryption algorithm. In Biham [17], pages 1–12.

[63] Helena Handschuh and M. Anwar Hasan, editors. *Selected Areas in Cryptography, 11th International Workshop, SAC 2004, Waterloo, Canada, August 9-10, 2004, Revised Selected Papers*, volume 3357 of *LNCS*. Springer-Verlag, 2004.

[64] G. Hardy, J.E. Littlewood, and G. Pólya. *Inequalities.* Cambridge University Press, 2nd edition, 1952.

[65] Carlo Harpes, Gerhard G. Kramer, and James L. Massey. A generalization of linear cryptanalysis and the applicability of Matsui's piling-up lemma. In Louis C. Guillou and Jean-Jacques Quisquater, editors, *Advances in Cryptology - EUROCRYPT'95: International Conference on the Theory and Application of Cryptographic Techniques, Saint-Malo, France, May 1995, Proceedings*, volume 921 of *LNCS*, pages 24–38. Springer-Verlag, 1995.

[66] Martin E. Hellman. A cryptanalysis time-memory trade off. *IEEE Transactions on Information Theory*, IT-26:401–406, 1980.

[67] Alfred Hitchcock. Dial M for Murder, 1954.

[68] Susan Dadakis Horn. Goodness-of-fit tests for discrete data: A review and an application to a health impairment scale. *Biometrics*, 33(1):237–247, 1977.

[69] Yupu Hu, Yuqing Zhang, and Guozhen Xiao. Integral cryptanalysis of SAFER+. *IEE Electronics Letters*, 35(17):1458–1459, 1999.

[70] Norman L. Johnson, Samuel Kotz, and N. Balakrishnan. *Continuous Univariate Distributions*, volume 1 of *Wiley Series in Probability and Mathematical Statistics.* John Wiley & Sons, 2 edition, 1994.

[71] Marc Joye, Pascal Paillier, and Serge Vaudenay. Efficient generation of prime numbers. In Çetin K. Koç and Christof Paar, editors, *Cryptographic Hardware and Embedded Systems - CHES 2000, Second International Workshop, Worcester, MA, USA, August 17-18, 2000, Proceedings*, volume 1965 of *LNCS*, pages 340–354. Springer-Verlag, 2000.

[72] Pascal Junod. On the complexity of Matsui's attack. In Serge Vaudenay and Amr M. Youssef, editors, *Selected Areas in Cryptography, 8th Annual International Workshop, SAC 2001 Toronto, Ontario, Canada, August 16-17, 2001, Revised Papers*, volume 2259 of *LNCS*, pages 199–211. Springer-Verlag, 2001.

[73] Pascal Junod. *Statistical Cryptanalysis of Block Ciphers.* PhD thesis, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, 2005.

[74] Pascal Junod. Yet another proof of the PRP/PRF switching lemma. Presented at the rump session of EUROCRYPT'05, 2005.

[75] Pascal Junod and Serge Vaudenay. Optimal key ranking procedures in a statistical cryptanalysis. In Thomas Johansson, editor, *Fast Software Encryption, 10th International Workshop, FSE 2003, Lund, Sweden, February 24-26, 2003, Revised Papers*, volume 2887 of *LNCS*, pages 235–246. Springer-Verlag, 2003.

[76] Pascal Junod and Serge Vaudenay. FOX: a new family of block ciphers. In Handschuh and Hasan [63], pages 114–129.

[77] Pascal Junod and Serge Vaudenay. Perfect diffusion primitives for block ciphers. In Handschuh and Hasan [63], pages 84–99.

[78] David Kahn. *The Code-Breakers.* Scribner, second edition, 1996. Original edition published on September 27, 1967.

[79] Liam Keliher. *Linear Cryptanalysis of Substitution-Permutation Networks.* PhD thesis, Quenn's University, Kingston, Ontario, Canada, October 2003. Available on `http://mathcs.mta.ca/faculty/lkeliher`.

[80] Liam Keliher. Refined analysis of bounds related to linear and differential cryptanalysis for the AES. In Hans Dobbertin, Vincent Rijmen, and Aleksandra Sowa, editors, *Advanced Encryption Standard - AES, 4th International Conference, AES 2004, Bonn, Germany, May 10-12, 2004, Revised Selected and Invited Papers*, volume 3373 of *LNCS*, pages 42–57. Springer-Verlag, 2005.

[81] Liam Keliher, Henk Meijer, and Stafford E. Tavares. New method for upper bounding the maximum average linear hull probability for SPNs. In Pfitzmann [130], pages 420–436.

[82] Liam Keliher, Henk Meijer, and Stafford E. Tavares. Toward the true random cipher: On expected linear probability values for SPNs with randomly selected S-boxes. In V. Bhargava, H.V. Poor, V. Tarokh, and S. Yoon, editors, *Communication, Information and Network Security*, pages 123–146. Kluwer Academic Publishers, 2003.

[83] Auguste Kerckhoffs. La cryptographie militaire. *Journal des sciences militaires*, 9:161–191, February 1883.

[84] Keylength.com. `http://www.keylength.com`.

[85] G.H. Khachatrian, M.K. Kuregian, and James L. Massey. Nomination of SAFER+ as candidate algorithm for the Advanced Encryption Standard (AES), June 1998.

[86] G.H. Khachatrian, M.K. Kuregian, and James L. Massey. Nomination of SAFER++ as candidate algorithm for the New European Schemes for Signatures, Integrity, and Encryption (NESSIE), September 2000.

[87] Lars R. Knudsen. A key-schedule weakness in SAFER K-64. In Don Coppersmith, editor, *Advances in Cryptology - CRYPTO '95, 15th Annual International Cryptology Conference, Santa Barbara, California, USA, August 27-31, 1995, Proceedings*, volume 963 of *LNCS*, pages 274–286. Springer-Verlag, 1995.

[88] Lars R. Knudsen. Truncated and higher order differentials. In Preneel [131], pages 196–211.

[89] Lars R. Knudsen, editor. *Fast Software Encryption, 6th International Workshop, FSE '99, Rome, Italy, March 24-26, 1999, Proceedings*, volume 1636 of *LNCS*. Springer-Verlag, 1999.

[90] Lars R. Knudsen. A detailed analysis of SAFER K. *Journal of Cryptology*, 13(4):417–436, 2000.

[91] Lars R. Knudsen and Thomas A. Berson. Truncated differentials of SAFER. In Gollmann [57], pages 15–26.

[92] Lars R. Knudsen and Vincent Rijmen. On the decorrelated fast cipher (DFC) and its theory. In Knudsen [89], pages 81–94.

[93] Lars R. Knudsen and David Wagner. Integral cryptanalysis (extended abstract). In Joan Daemen and Vincent Rijmen, editors, *Fast Software Encryption, 9th International Workshop, FSE 2002, Leuven, Belgium, February 4-6, 2002, Revised Papers*, volume 2365 of *LNCS*, pages 112–127. Springer-Verlag, 2002.

[94] Xuejia Lai. On the design and security of block ciphers. In James L. Massey, editor, *ETH Series in Information Processing*, volume 1. Hartung-Gorre Verlag Konstanz, Technische Hochschule (Zurich), 1992.

[95] Xuejia Lai. Higher order derivatives and differential cryptanalysis. In Kluwer Academic Publishers, editor, *Symposium on Communication, Coding and Cryptography*, pages 227–233, 1994.

[96] Xuejia Lai and James L. Massey. A proposal for a new block encryption standard. In Ivan Damgård, editor, *Advances in Cryptology - EUROCRYPT '90, Workshop on the Theory and Application of of Cryptographic Techniques, Aarhus, Denmark, May 21-24, 1990, Proceedings*, volume 473 of *LNCS*, pages 389–404. Springer-Verlag, 1991.

[97] Xuejia Lai, James L. Massey, and Sean Murphy. Markov ciphers and differential cryptanalysis. In Davies [44], pages 17–38.

[98] Henry O. Lancaster. Forerunners of the Pearson χ^2. *Australian Journal of Statistics*, 8:117–126, 1966.

[99] Susan K. Langford and Martin E. Hellman. Differential-linear cryptanalysis. In Desmedt [46], pages 17–25.

[100] Steven Levy. *Crypto*. Penguin Books, 2000.

[101] Chae Hoon Lim. A revised version of CRYPTON: CRYPTON V1.0. In Knudsen [89], pages 31–45.

[102] Michael Luby and Charles Rackoff. How to construct pseudorandom permutations from pseudorandom functions. *SIAM J. Comput.*, 17(2):373–386, 1988.

[103] Stefan Lucks. Faster Luby-Rackoff ciphers. In Gollmann [57], pages 189–203.

[104] Stefan Lucks. The saturation attack - a bait for Twofish. In Mitsuru Matsui, editor, *Fast Software Encryption, 8th International Workshop, FSE 2001 Yokohama, Japan, April 2-4, 2001, Revised Papers*, volume 2355 of *LNCS*, pages 1–15. Springer-Verlag, 2002.

[105] F.J. MacWilliams and N.J.A. Sloane. *The Theory of Error-Correcting Codes*, volume 16 of *North-Holland Mathematical Library.* North-Holland, 1977. Tenth impression.

[106] Maplesoft. Maple 11. `http://www.maplesoft.com/`.

[107] James L. Massey. SAFER-K64: a byte-oriented block-ciphering algorithm. In Anderson [2], pages 1–17.

[108] James L. Massey. SAFER-K64: one year later. In Preneel [131], pages 212–241.

[109] James L. Massey. Strengthened key schedule for the cipher SAFER. Posted on USENET newsgroup sci.crypt, September 9, 1995.

[110] Mitsuru Matsui. Linear cryptanalysis method for DES cipher. In Tor Helleseth, editor, *Advances in Cryptology - EUROCRYPT'93: Workshop on the Theory and Application of Cryptographic Techniques, Lofthus, Norway, May 1993, Proceedings*, volume 765 of *LNCS*, pages 386–397. Springer-Verlag, 1993.

[111] Mitsuru Matsui. The first experimental cryptanalysis of the Data Encryption Standard. In Desmedt [46], pages 1–11.

[112] Mitsuru Matsui. New structure of block ciphers with provable security against differential and linear cryptanalysis. In Gollmann [57], pages 205–218.

[113] Ueli M. Maurer, Yvonne Anne Oswald, Krzysztof Pietrzak, and Johan Sjödin. Luby-Rackoff ciphers from weak round functions? In Vaudenay [156], pages 391–408.

[114] Alfred Menezes, Paul C. van Oorschot, and Scott A. Vanstone. *Handbook of applied cryptography.* The CRC Press series on discrete mathematics and its applications. CRC-Press, 1997.

[115] MPFR. MPFR C library for multiple-precision floating-point computations with correct rounding. `http://www.mpfr.org/`.

[116] Sean Murphy. An analysis of SAFER. *Journal of Cryptology*, 11(4):235–251, 1998.

[117] David Naccache and Jacques Stern. A new public-key cryptosystem. In Walter Fumy, editor, *Advances in Cryptology - EUROCRYPT '97, International Conference on the Theory and Application of Cryptographic Techniques, Konstanz, Germany, May 11-15, 1997, Proceeding*, volume 1233 of *LNCS*, pages 27–36. Springer-Verlag, 1997.

[118] Jorge Nakahara, Paulo Barreto, Bart Preneel, and Joos Vandewalle. Square attacks on reduced-round PES and IDEA block ciphers. In B. Macq and Jean-Jacques Quisquater, editors, *Proceedings of the 23rd Symposium on Information Theory in the Benelux*, Werkgemeenschap voor Informatie en Communicatietheorie, pages 187–195, 2002.

[119] Jorge Nakahara, Bart Preneel, and Joos Vandewalle. Linear cryptanalysis of reduced-round versions of the SAFER block cipher family. In Schneier [135], pages 244–261.

[120] Moni Naor and Omer Reingold. On the construction of pseudorandom permutations: Luby-Rackoff revisited. *Journal of Cryptology*, 12(1):29–66, 1999.

[121] Melvyn B. Nathanson. *Elementary Methods in Number Theory.* Graduate Texts in Mathematics. Springer-Verlag, 2000.

[122] National Bureau of Standards, U. S. Department of Commerce. *Data Encryption Standard*, 1977.

[123] Jerzy Neyman and Egon S. Pearson. On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 231:289–337, 1933.

[124] Kaisa Nyberg. Perfect nonlinear S-boxes. In Davies [44], pages 378–386.

[125] Kaisa Nyberg. Linear approximation of block ciphers. In De Santis [45], pages 439–444.

[126] Luke O'Connor. Properties of linear approximation tables. In Preneel [131], pages 131–136.

[127] Philippe Oechslin. Making a faster cryptanalytic time-memory trade-off. In Dan Boneh, editor, *Advances in Cryptology - CRYPTO 2003, 23rd Annual International Cryptology Conference, Santa Barbara, California, USA, August 17-21, 2003, Proceedings*, volume 2729 of *LNCS*, pages 617–630. Springer-Verlag, 2003.

[128] Pascal Paillier. Public-key cryptosystems based on composite degree residuosity classes. In Stern [143], pages 223–238.

[129] Karl Pearson. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can reasonably be supposed to have arisen from random sampling. *Philosophy Magazines Series*, 50(5):157–172, 1900.

[130] Birgit Pfitzmann, editor. *Advances in Cryptology - EUROCRYPT 2001, International Conference on the Theory and Application of Cryptographic Techniques, Innsbruck, Austria, May 6-10, 2001, Proceeding*, volume 2045 of *LNCS*. Springer-Verlag, 2001.

[131] Bart Preneel, editor. *Fast Software Encryption: Second International Workshop. Leuven, Belgium, 14-16 December 1994, Proceedings*, volume 1008 of *LNCS*. Springer-Verlag, 1995.

[132] Ronald L. Rivest, Adi Shamir, and Leonard Adleman. A method for obtaining digital signatures and public-key cryptosystems. *Communications of the ACM*, 21:120–126, 1978.

[133] Ivan N. Sanov. On the probability of large deviations of random variables. *Mat. Sbornik*, 42:11–44, 1957.

[134] Bruce Schneier. Description of a new variable-length key, 64-bit block cipher (Blowfish). In Anderson [2], pages 191–204.

[135] Bruce Schneier, editor. *Fast Software Encryption, 7th International Workshop, FSE 2000, New York, NY, USA, April 10-12, 2000, Proceedings*, volume 1978 of *LNCS*. Springer-Verlag, 2001.

[136] Bruce Schneier, John Kelsey, Doug Whiting, David Wagner, Chris Hall, and Niels Ferguson. Twofish: A 128-bit block cipher. Available on `http://www.schneier.com`, 1998.

[137] Claus-Peter Schnorr and Serge Vaudenay. Black box cryptanalysis of hash networks based on multipermutations. In De Santis [45], pages 47–57.

[138] Rich Schroeppel. Hasty pudding cipher specification, June 1998. Available on `http://www.cs.arizona.edu/~rcs/hpc/hpc-spec`.

[139] Claude E. Shannon. Communication theory of secrecy systems. *Bell Systems Technical Journal*, 28(4):656–715, October 1949. Re-edited in *Claude Elwood Shannon - Collected Papers*. IEEE Press, New York, 1993.

[140] Victor Shoup. Sequences of games: A tool for taming complexity of security proofs, 2006. Available on `http://shoup.net`.

[141] Simon Singh. *The Code Book: The Science of Secrecy from Ancient Egypt to Quantum Cryptography*. Anchor, 2000.

[142] Jacques Stern. *La Science du Secret*. Odile Jacob, 1998.

[143] Jacques Stern, editor. *Advances in Cryptology - EUROCRYPT '99, International Conference on the Theory and Application of Cryptographic Techniques, Prague, Czech Republic, May 2-6, 1999, Proceeding*, volume 1592 of *LNCS*. Springer-Verlag, 1999.

[144] Jacques Stern. Why provable security matters? In Eli Biham, editor, *Advances in Cryptology - EUROCRYPT 2003, International Conference on the Theory and Applications of Cryptographic Techniques, Warsaw, Poland, May 4-8, 2003, Proceedings*, volume 2656 of *LNCS*, pages 449–461. Springer-Verlag, 2003.

[145] Jacques Stern and Serge Vaudenay. CS-Cipher. In Serge Vaudenay, editor, *Fast Software Encryption, 5th International Workshop, FSE '98, Paris, France, March 23-25, 1998, Proceedings*, volume 1372 of *LNCS*, pages 189–204. Springer-Verlag, 1998.

[146] Douglas R. Stinson and Stafford E. Tavares, editors. *Selected Areas in Cryptography, 7th Annual International Workshop, SAC 2000, Waterloo, Ontario, Canada, August 14-15, 2000, Proceedings*, volume 2012 of *LNCS*. Springer-Verlag, 2001.

[147] Anne Tardy-Corfdir and Henri Gilbert. A known plaintext attack of FEAL-4 and FEAL-6. In Joan Feigenbaum, editor, *Advances in Cryptology - CRYPTO '91, 11th Annual International Cryptology Conference, Santa Barbara, California, USA, August 11-15, 1991, Proceedings*, volume 576 of *LNCS*, pages 172–182. Springer-Verlag, 1992.

[148] `http://en.wikipedia.org/wiki/Poulet_de_Bresse`.

[149] Serge Vaudenay. Plaquette du département de Mathématiques et d'Informatique de l'Ecole Normale Superieure, 45 rue d'Ulm, 75005 Paris.

[150] Serge Vaudenay. On the need for multipermutations: Cryptanalysis of MD4 and SAFER. In Preneel [131], pages 286–297.

[151] Serge Vaudenay. An experiment on DES statistical cryptanalysis. In *3rd ACM Conference on Computer and Communications Security*, pages 139–147. ACM Press, 1996.

[152] Serge Vaudenay. Provable security for block ciphers by decorrelation. In *STACS'98*, volume 1373 of *LNCS*, pages 249–275. Springer-Verlag, 1998.

[153] Serge Vaudenay. On the Lai-Massey scheme. In Kwok-Yan Lam, Eiji Okamoto, and Chaoping Xing, editors, *Advances in Cryptology - ASIACRYPT '99, International Conference on the Theory and Applications of Cryptology and Information Security, Singapore, November 14-18, 1999, Proceedings*, volume 1716 of *LNCS*, pages 8–19. Springer-Verlag, 1999.

[154] Serge Vaudenay. Resistance against general iterated attacks. In Stern [143], pages 255–271.

[155] Serge Vaudenay. Decorrelation: a theory for block cipher security. *Journal of Cryptology*, 16(4):249–286, 2003.

[156] Serge Vaudenay, editor. *Advances in Cryptology - EUROCRYPT 2006, 25th Annual International Conference on the Theory and Applications of Cryptographic Techniques, St. Petersburg, Russia, May 28 - June 1, 2006, Proceedings*, volume 4004 of *LNCS*. Springer-Verlag, 2006.

[157] Serge Vaudenay. *A Classical Introduction to Cryptography: Applications for Communications Security*. Springer-Verlag, 2006. Website of the book: `http://www.vaudenay.ch/crypto/`.

[158] Umesh V. Vazirani and Vijay V. Vazirani. Efficient and secure pseudo-random number generation. In Blakley and Chaum [28], pages 193–202.

[159] Umesh V. Vazirani and Vijay V. Vazirani. Efficient and secure pseudo-random number generation (extended abstract). In *Proceedings of FOCS'84*, pages 458–463. IEEE, 1985.

[160] Gilbert S. Vernam. Cipher printing telegraph systems for secret wire and radio telegraphic communications. *Journal of the IEEE*, 55:109–115, 1926.

[161] J. von zur Gathen and J. Gerhard. *Modern Computer Algebra*. Cambridge University Press, 2nd edition, 2003. First published 1999.

[162] David Wagner. The boomerang attack. In Knudsen [89], pages 156–170.

[163] David Wagner. Towards a unifying view of block cipher cryptanalysis. In Bimal K. Roy and Willi Meier, editors, *Fast Software Encryption, 11th International Workshop, FSE 2004, Delhi, India, February 5-7, 2004, Revised Papers*, volume 3017 of *LNCS*, pages 16–33. Springer-Verlag, 2004.

[164] Hongjun Wu, Feng Bao, Robert H. Deng, and Qin-Zhong Ye. Improved truncated differential attacks on SAFER. In Kazuo Ohta and Dingyi Pei, editors, *Advances in Cryptology - ASIACRYPT '98, International Conference on the Theory and Applications of Cryptology and Information Security, Beijing, China, October*

18-22, 1998, Proceedings, volume 1514 of *LNCS*, pages 133–147. Springer-Verlag, 1998.

[165] Wenling Wu, Wentao Zhang, and Dengguo Feng. Integral cryptanalysis of reduced FOX block cipher. In Dongho Won and Seungjoo Kim, editors, *Information Security and Cryptology - ICISC 2005, 8th International Conference, Seoul, Korea, December 1-2, 2005, Revised Selected Papers*, volume 3935 of *LNCS*, pages 229–241. Springer-Verlag, 2006.

[166] Andrew Chi-Chih Yao. Theory and applications of trapdoor functions (extended abstract). In *23rd Annual Symposium on Foundations of Computer Science, 3-5 November 1982, Chicago, Illinois, USA*, pages 80–91, 1982.

[167] Amr M. Youssef and Stafford E. Tavares. Resistance of balanced S-boxes to linear and differential cryptanalysis. *Information Processing Letters*, 56:249–252, 1995.

[168] Yuliang Zheng, editor. *Advances in Cryptology - ASIACRYPT 2002, 8th International Conference on the Theory and Application of Cryptology and Information Security, Queenstown, New Zealand, December 1-5, 2002, Proceedings*, volume 2501 of *LNCS*. Springer-Verlag, 2002.

Curriculum Vitæ

Education

EPFL (Ecole Polytechnique Fédérale de Lausanne), LAUSANNE, SWITZERLAND 2003-2008
PhD DEGREE IN CRYPTOGRAPHY
PhD Thesis: Quantitative Security of Block Ciphers: Designs and Cryptanalysis Tools. Degree expected mid. 2008.
Supervisor: Prof. Serge Vaudenay.
Fellowship from the Swiss National Science Foundation (SNSF).

Lectures attended at EPFL's Doctoral School:

- Prof. Arikan's lectures on Quantum Computation and Quantum Information (Grade: 6/6)
- Prof. Shokrollahi's lectures on Algorithmic Number Theory (Grade: 6/6)
- Prof. Vaudenay's lectures on Selected Topics in Cryptography (Grade: 6/6)

EPFL, LAUSANNE, SWITZERLAND 1998-2003
MASTER DEGREE IN COMMUNICATIONS SYSTEMS
Master Thesis: A Generalization of Linear Cryptanalysis. (Grade: 6/6)
Semester Project: Factorisation de Grands Nombres à l'Aide de Courbes Elliptiques. (Grade: 6/6)
Semester Project: Attaque à Textes Chiffrés Choisis contre PKCS#1. (Grade: 6/6)

UPC (Universitat Politècnica de Catalunya), BARCELONA, SPAIN 2000-2001
Erasmus Exchange Program

Professional Experience

JOINT GENERAL CHAIRMAN OF THE 2008 EDITION OF FAST SOFTWARE ENCRYPTION (FSE) WORKSHOP
FSE 2008 is the 15th annual Fast Software Encryption workshop. It is sponsored by the International Association for Cryptologic Research (IACR) and is the world's most important event focusing on symmetric cryptographic primitives.

PROGRAMMING OF iCHAIR, A COMPREHENSIVE SUBMISSION/REVIEW SERVER SOFTWARE (2006)
iChair is designed to help the programme chairman of a conference with submission collection, assignments of articles to reviewers, gathering of reviews, discussions, mailing. Since the end of

2006, iChair handled the submission/review process of more than 40 conferences and workshops worldwide.

SYSTEM ADMINISTRATOR (2005-2008)
Administrator of the network and of the file/web servers of the Security & Cryptography laboratory (Linux & Mac OS environments).

NAGRA - KUDELSKI GROUP (2002)
Summer Job in 2002. Software Programming in Java.

Academic Experience

PUBLICATIONS WITH PEER REVIEW & TALKS

- Linear Cryptanalysis of Non Binary Ciphers (with an Application to SAFER)
  Thomas Baignères, Jacques Stern, and Serge Vaudenay
  To be Published in the Proceedings of SAC 07 (Ottawa, Canada)
- KFC - The Krazy Feistel Cipher
  Thomas Baignères and Matthieu Finiasz
  Published in the Proceedings of Asiacrypt'06 (Shanghai, China)
- Dial C for Cipher
  Thomas Baignères and Matthieu Finiasz
  Published in the Proceedings of SAC 06 (Montreal, Canada)
- Proving the Security of the AES Substitution-Permutation Network
  Thomas Baignères and Serge Vaudenay
  Published in the Proceedings of SAC 05 (Kingston, Canada)
- How Far Can We Go Beyond Linear Cryptanalysis?
  Thomas Baignères, Pascal Junod, and Serge Vaudenay
  Published in the Proceedings of Asiacrypt'04 (Jeju Island, Korea)

BOOK
A Classical Introduction to Cryptography: Exercise Book
Thomas Baignères, Pascal Junod, Lu Yi, Jean Monnerat, and Serge Vaudenay
Published by Springer-Verlag.

OTHER PUBLICATIONS, TALKS & UNPUBLISHED WORK

- The Complexity of Distinguishing Distributions
  Thomas Baignères and Serge Vaudenay
  To be Published in the Proceedings of ICITS 08 (Calgary, Canada)
- Practical Decorrelation
  Thomas Baignères
  Invited talk at ESC'08 (Echternach, Luxembourg)
- Provable Security in Cryptography
  Thomas Baignères
  Lectures given at EPFL in 2007
- Cryptosystems and LLL
  Thomas Baignères
- Quantum Cryptography: On the Security of the BB84 Key-Exchange Protocol
  Thomas Baignères

TEACHING ACTIVITY

- Assistant lecturer at the master's Cryptography & Security lectures from 2003-2006.
- Guest lecturer of Prof. Vaudenay's Doctoral course on Selected Topics in Cryptography, EPFL, Summer semester 2007: Provable Security in Cryptography.
- Supervision of eight master's student semester projects.

EXTERNAL REFEREE

Acted as an external referee for more than 20 conferences, including Crypo, Eurocrypt, and Asiacrypt which are the top three annual events in the field.

Scientific Skills

MY PhD THESIS IN TWENTY SECONDS

Two main research achievements:

- Development of quantitative methods to break cryptographic protocols. This involves the description of realistic mathematical security models which allow to formally state what it means for a cryptographic system to be secure, illustration of how real-life security problems translate in these models and derivation of optimal statistical algorithms that solve them, and model checking by implementing practical attacks against existing cryptographic algorithms. This last step involves C programming and requires to manage a huge amount of data to obtain meaningful statistics.
- Design of two new block ciphers (i.e., symmetric encryption algorithm) providing strong mathematical security proofs of their security. None of the previous publicly known designs provides similar irrefutable evidence of security. One of the two designs reaches encryption speeds meeting today's industrial needs and could thus be used in practical situations requiring an exceptional security level (e.g., wire transfers, mail encryption, etc.).

EXPERIENCE OF SEVERAL PROGRAMMING LANGUAGES

- Excellent knowledge of C: some of my thesis practical experiments required to run test programs for more than one month and thus efficient and error-free programs.
- Excellent knowledge of PHP: iChair is entirely programmed in PHP, including some SQL database calls, XML and XSLT documents, and W3C valid XHTML and CSS pages.
- Good knowledge of Java; programming of a pay TV protocol simulation software, including encrypted video streams, video player, and user interfaces.
- Experience in C++ (use of NTL library for number theory), Perl, Maple, Matlab.

Languages

French	English	Italian & Spanish
Mother tongue	Written and spoken	Spoken only

Personal Situation

30 years old, married, one child, French and Swiss nationalities. No military obligation.