University of Trento

Department of Mathematics
Doctoral School - XXX cycle

Ph.D. Thesis

Differential attacks using alternative operations and block cipher design

Advisors:
Prof. Massimiliano Sala
(University of Trento)
Dr. Céline Blondeau
(Aalto University)

Ph.D. Candidate:
Roberto Civino

MSC 2010: 94A60, 68P25, 20B35, 20B15

Academic Year 2016/2017

Differential attacks using alternative operations and block cipher design

Roberto Civino

Assessment Committee

Massimiliano Sala	*University of Trento*	Advisor
Christina Boura	*Versailles University*	Referee
Lars R. Knudsen	*Technical University of Denmark*	Referee
Éric Filiol	*ESIEA*	Examiner
Norberto Gavioli	*University of L'Aquila*	Examiner

So when I say that I know me, how can I know that?
What kind of spider understands arachnophobia?
I have my senses and my sense of having senses.
Do I guide them? Or they me?

Overview

The thesis is divided into three parts, each of which is described in the corresponding section of this overview. Part I mainly serves as an introduction to the contents of Part II and Part III which are, in turn, mutually independent.

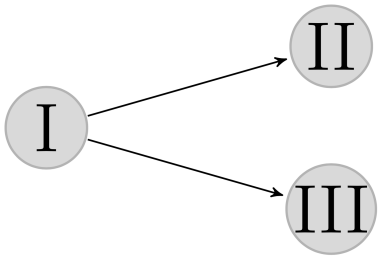

Figure 1: Logical dependence of the parts of this thesis

Preliminaries Block ciphers are the main subject of this work and are introduced in Chapter 1. After having described our model for the frameworks of Substitution-Permutation Networks (SPN) and Feistel Networks (FN), attention is given to the theme of security, with a particular focus on differential (Chapter 2) and algebraic (Chapter 3) attacks. These are indeed the families of cryptanalytic techniques the novel contributions of this work belong to.

Differential cryptanalysis using alternative operations Chapter 4 and Chapter 5 deal with the following problem: is it possible that a block cipher apparently immune to classical differential cryptanalysis can be attacked considering a different operation on the message space? In [Ber92], Berson introduced the modular difference to study the MD/SHA family of hash functions. In [AS11], the authors tried to use a similar method to cryptanalyse the block cipher PRESENT [BKL+07], featuring a bit-wise round-key addition. Even though this attempt has been unsuccessful, the idea of using alternative difference operations is for the first time taken into consideration and used in block ciphers with a bit-wise key addition. More recently, Calderini and Sala showed how to effectively compute alternative operations on a vector space which can serve as message space for a block cipher such that the resulting structure is still a vector space [CS17]. The authors used those operations to mount a linearisation attack against a toy cipher. Here we study similar operations in the differential context, investigating how alternative operations interact with the layers of an SPN, and show how they influence the differential probabilities, when the difference taken into consideration is different from the usual bit-wise addition modulo two. In particular, in Chapter 4, we study constraints coming from the combination of the bit-wise key addition with these operations, by introducing and studying the *key distribution table*. Moreover, we study the differential uniformity, with respect to other operations, of some non-linear permutations such as the classical cubic function. In Chapter 5 we designed a 15-bit block cipher, which presents some similarities with the block ciphers PRINTcipher [KLPR10] and PRINCE [BCG+12], and represents an example of SPN which is resistant against the classical differential attack, with XOR differences, but it is not resistant against a differential attack which makes use of alternative difference coming from another operation defined on the message space.

On the design of wave ciphers Round functions used as building blocks for iterated block ciphers, both in the case of Substitution-Permutation Networks and Feistel Networks, are often obtained as the composition of different layers which provide confusion and diffusion, and key additions. The

bijectivity of any encryption function, crucial in order to make the decryption possible, is guaranteed by the use of invertible layers or by the Feistel structure. In Chapter 6 a new family of ciphers, called *wave ciphers*, is introduced. In wave ciphers, round functions feature *wave functions*, which are vectorial Boolean functions obtained as the composition of *non-invertible* layers, where the confusion layer enlarges the message which returns to its original size after the diffusion layer is applied. This is motivated by the fact that relaxing the requirement that all the layers are invertible allows to consider more functions which are optimal with regard to non-linearity. In particular it allows to consider injective APN S-boxes even in cases where no APN permutations have been found, e.g. the cases of a number of variables equals to four or eight, which are optimal for implementation needs. In order to guarantee efficient decryption we propose to use wave functions in Feistel Networks. With regard to security, in Chapter 7 we investigate the immunity from some algebraic attacks. In particular, we focus on the security from the group-theoretical attack described in [Pat99], where the author designed a DES-like cipher, resistant to both linear and differential cryptanalysis, whose encryption functions generate an imprimitive group, and showed how the knowledge of this trapdoor can be turned into an efficient attack to the cipher. In this work it is shown how to avoid that the group generated by the round functions of a wave cipher acts imprimitively, by giving conditions on the Boolean functions composing the layers of the wave-shaped round functions.

Short contents

Contents

Notation

The following notation and terminology will be used throughout all this work.

Functions We use the postfix notation for every function evaluation, i.e. if f is a function and x an element in the domain of f, we denote by xf the evaluation of f in x. We denote by $\operatorname{Im} f$ the range of f and by Yf^{-1} the pre-image of a set Y. A *vectorial Boolean function* is a function from $(\mathbb{F}_2)^n$ to $(\mathbb{F}_2)^m$, where n and m are integers, and $\mathbb{F}_2$ denotes the finite field with two elements.

Vectors We denote by V a finite vector space over $\mathbb{F}_2$ of dimension $n \in \mathbb{N}$. We assume $\dim(V) = n = s \times b$ and write $V = V_1 \oplus V_2 \oplus \ldots \oplus V_b$, where $\dim(V_j) = s$ for $1 \leq j \leq b$, and $\oplus$ represents the direct sum of subspaces, called *bricks*. When $x \in V$, for each $1 \leq j \leq b$ we denote by $x^{[j]}$ the s components of x in the j^{th} brick. The vector $x^{[j]} \in (\mathbb{F}_2)^s$ is also called a brick. The canonical basis for V is denoted by $\{e_1, e_2, \ldots, e_n\}$. For each $1 \leq j \leq b$ the map $\pi_j : V \to V$ denotes the canonical projection on V_j. The *Hamming weight* of a vector $x \in V$, i.e. the number of non-zero coordinates of x with respect to the canonical decomposition, is denoted by $\operatorname{weight}(x)$. Each vector in V can be interpreted as a binary number, most significant bit first, and then represented using the hexadecimal notation. For example, $(0, 0, 0, 1) = \mathtt{1}_{\mathrm{x}}$ and $(1, 1, 1, 1) = \mathtt{F}_{\mathrm{x}}$.

Fields We denote by $\mathbb{F}_{2^n}$ the finite field with 2^n elements. For each $\alpha \in \mathbb{F}_{2^n}$, the *trace* of α is defined as $\mathrm{Tr}(\alpha) \stackrel{\text{def}}{=} \alpha + \alpha^2 + \ldots + \alpha^{2^{n-1}}$. The function $\mathrm{Tr} : \mathbb{F}_{2^n} \to \mathbb{F}_2$ is a linear function, where both $\mathbb{F}_{2^n}$ and $\mathbb{F}_2$ are viewed as vector spaces over $\mathbb{F}_2$. Let us also recall that for each $\alpha \in \mathbb{F}_{2^n}$ it holds $\mathrm{Tr}(\alpha) = \mathrm{Tr}(\alpha^2)$.

Groups If G is any finite group acting on V, for each $g \in G$ and $v \in V$ we denote the action of g on v as vg. The group is called an *abelian* group if and only if $g_1 g_2 = g_2 g_1$ for each $g_1, g_2 \in G$; moreover G is 2-*elementary* if for each $g \in G$ it holds $g^2 = \mathbb{1}_G$, where $\mathbb{1}_G$ denotes the neutral element in the group. The action of G on V is said to be *transitive* if for each $v_1, v_2 \in V$ there exixts $g \in G$ such that $v_1 g = v_2$. If for any v_1, v_2 such g is unique, the action of G on V is said *regular*. If $H \subset G$, we denote $H < G$ to mean that H is a subgroup of G.
The identity matrix over $\mathbb{F}_2$ of size ℓ is denoted by $\mathbb{1}_\ell$, and the zero matrix $\mathbb{F}_2$ of size $\ell \times h$ is denoted by $\mathbb{0}_{\ell,h}$. The *symmetric group acting on* V, i.e. the group of all the permutations on the space V, is denoted by $\mathrm{Sym}(V)$. The subgroup of $\mathrm{Sym}(V)$ generated by the even permutations, i.e. the permutations obtainable from an even number of two-element swaps, is called the *alternating group* and is denoted by $\mathrm{Alt}(V)$. The group of all the affine permutations of $(V, +)$, which is a primitive maximal subgroup of $\mathrm{Sym}(V)$, is denoted by $\mathrm{AGL}(V, +)$. The group of all the linear permutations of $(V, +)$ is denoted by $\mathrm{GL}(V, +)$. For any linear map $\lambda \in \mathrm{GL}(V, +)$ we denote by $\mathrm{Ker}(\lambda)$ the kernel of λ. The group of the translations of V is denoted by $T_+ \stackrel{\text{def}}{=} \{\sigma_a \mid x \mapsto x + a, a \in V\} < \mathrm{Sym}(V)$. Sometimes, when it is important to highlight the dimension of the vector space the translations are acting on, we denote T_+ by T_n.

Part I

Preliminaries

1 Introduction to block ciphers

In this first chapter the main subjects of this work, i.e. block ciphers, are introduced. Some preliminary results are shown and the notations used along the thesis are explained. This presentation as well as the one of Chapter 2 is inspired, among others, by the following references [Rij97, DR13, KR11, LMM91].

1.1 Block ciphers

Block ciphers start to play a role in the security of the today's communications between two parties whenever the parties agree on a secret and shared key by means of whatever asymmetric cryptosystem. A block cipher is a symmetric primitive which, taking as input a fixed-length block of a message and a parameter called *key*, transforms the former into a string of the same length n, in such a way that only authorised parties can access it. More precisely, a block cipher is a set of permutations defined on a message space $\mathcal{M}$, each of which is indexed by the key parameter, called *encryption functions*. The operation of transforming the message, also called *plaintext*, into the output of the parametrised encryption function, called *ciphertext*, is called *encryption*, whereas the reverse process is called *decryption*. In order to describe a block cipher, it is required to define the process for obtaining the encryption function once the key is chosen in the key space $\mathcal{K}$. This is done, according to the Kerckhoffs's principle [Ker83], by making public

all the procedures and keeping secret only the key, since *"it should not be a problem if it* [the cipher's description] *falls into enemy hands"*. Moreover, for the above-mentioned procedures, a relatively simple description is required, and the following properties need to be satisfied:

efficiency there exists an efficiently computable procedure which, for any message in the message space and any key in the key space, provides the encryption of the given message with the current encryption function, preferably on a wide range of platforms / devices. The same should hold for the decryption process;

security the $\#\mathcal{K}$ key-induced permutations should look like being chosen uniformly at random in the set of all the possible permutations of the message space, in such a way it is not possible, given a text, to predict whether it is a ciphertext of a randomly-generated string of the same length.

The second ideal requirement incidentally means that, parties not entitled to access the encrypted data recover no information on the plaintext when the ciphertext is given but the key used for the encryption is unknown. Such parties are called *attackers* or *cryptanalysts*, usually depending on whether one wants to put the accent on their bad or good intentions, respectively.

The following is a very general definition of block cipher.

Definition 1.1.1. Let $\mathcal{M}$ and $\mathcal{K}$ be non-empty sets. A *block cipher* Φ is an injective function $\mathcal{K} \to \mathrm{Sym}(\mathcal{M})$. The set $\mathcal{M}$ is called the *message space*, $\mathcal{K}$ the *key space*, and the permutation $E_K \stackrel{\text{def}}{=} K\Phi$ is called the *encryption function induced by the key* K. The set $\{E_K \mid K \in \mathcal{K}\}$ is called the *set of the encryption functions*. It is common to identify

$$\Phi \equiv \mathcal{K}\Phi = \{E_K \mid K \in \mathcal{K}\} \subset \mathrm{Sym}(\mathcal{M}).$$

1.1.1 Perfect secrecy

Is his seminal work [Sha49], Shannon gave a formal definition of security for block ciphers, which basically requires that the security of the cryptosystem

does not rely on assumptions on the computational capabilities of the adversary, i.e. the cipher is unbreakable even if the computational power of the adversary is not bounded. Before recalling Shannon's definition of perfect secrecy, let us consider the plaintext P and the key K as random variables. Let us assume that a probability distribution $\mathbb{P}_{\mathcal{M}}$ on $\mathcal{M}$ is given, in such a way we can denote by $\mathbb{P}_{\mathcal{M}}[P = p]$ the probability that the plaintext $p \in \mathcal{M}$ occurs. Moreover, let us consider a probability distribution $\mathbb{P}_{\mathcal{K}}$ on $\mathcal{K}$, independent on P, such that $\mathbb{P}_{\mathcal{K}}[K = k]$ denotes the probability that the key $k \in \mathcal{K}$ is chosen. Let us consider the random variable $C : \mathcal{M} \times \mathcal{K} \to \mathcal{M}$ defined by $(p, k)C = pE_k$, where $\mathcal{M} \times \mathcal{K}$ is equipped with the probability distribution $\mathbb{P}_{\mathcal{M},\mathcal{K}} = \mathbb{P}_{\mathcal{M}} \times \mathbb{P}_{\mathcal{K}}$.

Definition 1.1.2. A block cipher Φ is said to have *perfect secrecy* if for each $p, c \in \mathcal{M}$ it holds

$$\mathbb{P}_{\mathcal{M},\mathcal{K}}[P = p \mid C = c] = \mathbb{P}_{\mathcal{M}}[P = p].$$

The notion of perfect secrecy means that the *a posteriori* distribution of the plaintext P when the value of ciphertext C is known, is equal to the *a priori* distribution of the plaintext, i.e. that an attacker obtains no more information on the plaintext when the ciphertext is known.

The following characterisation of the perfect secrecy is due to Shannon.

Theorem 1.1.3. *Let Φ be a block cipher and let us assume that $\#\mathcal{M} = \#\mathcal{K}$. Then Φ has perfect secrecy if and only if for each $p, c \in \mathcal{M}$ there exists a unique $k \in \mathcal{K}$ such that $pE_k = c$ and $\mathbb{P}_{\mathcal{K}}$ is uniform, i.e. for each $k \in \mathcal{K}$ it holds $\mathbb{P}_{\mathcal{K}}[K = k] = 1/\#\mathcal{K}$.* □

A well-known example of cipher with perfect secrecy is the one-time pad [Mil82], which is however important only from a theoretical point of view, since it is not easy to give a practical implementation of the cipher.

1.2 Iterated block ciphers

In the second part of this work, Shannon introduced the concepts of *confusion* and *diffusion* to guide the design of practical ciphers. Those are today considered cardinal principles for obtaining the properties of security and efficiency discussed before:

confusion the ciphertext statistics should depend on the plaintext statistics in a manner too complicated to be exploited by the cryptanalyst;

diffusion the encryption spreads out of the influence of a single plaintext digit over many ciphertext digits so as to hide the statistical structure of the plaintext.

Since a cipher should not only be difficult to break, but it must also be easy to use (i.e. to encrypt and decrypt when the secret key is known), a very common approach for creating diffusion and confusion is to use a *product cipher*, i.e., a cipher that can be implemented as a succession of simple ciphers, each of which adds its modest share to the overall large amount of diffusion and confusion [Mas88]. This idea leads to the definition of an iterated cipher, which is the most common framework nowadays for block ciphers. What follows is a general definition of iterated cipher.

Definition 1.2.1. The block cipher $\Phi = \{E_K \mid K \in \mathcal{K}\}$ is called an iterated block cipher if there exists $R \in \mathbb{N}$ such that for each $K \in \mathcal{K}$ the encryption function E_K is the composition of R functions, i.e. $E_K = \varepsilon_{1,K}\, \varepsilon_{2,K} \ldots \varepsilon_{R,K}$. For each $1 \leq i \leq R$, the function $\varepsilon_{i,K}$ is called the *round function for the* i^{th} *round* of the encryption function E_K.

To provide efficiency, each round function is the composition of a public component provided by the designers, and a private component derived from the user-provided key by means of a public procedure known as *key-schedule*. As we will discuss later in detail, each component of the cipher is designed to fulfil its specific purpose. In this first informal description of the block ciphers' fundamentals, we can describe the components of each round function, also called *layers*, in this way:

confusion layer is designed to provide the Shannon's principle of confusion by replacing certain blocks of bits of the state with other blocks of bits, following a specific rule. The relation between the input and the output of the confusion layer is designed to be as complex as possible, hence look-up tables are usually required in implementation;

diffusion layer is designed to provide the Shannon's principle of diffusion by rearranging the bits of the block in such a way that a change in one bit of the state affects as much bits as possible. Such a layer usually operates simple manupulations on the bits of the block;

key-addition layer is the only layer of the cipher whose input is not public. It makes all the bits of the block dependent on a user-selected *key* in the manner intended by the designers of the cipher.

In the theory of modern iterated block ciphers, two frameworks are mainly considered: *Substitution-Permutation Networks (SPN)* (see e.g. AES [DR13], PRESENT [BKL+07], SERPENT [BAK98]) and *Feistel Networks (FN)* (see e.g. DES [Pub77], Camelia [AIK+00], GOST [Dol10]). Figure 1.1 depicts the more general framework of SPNs, FNs and their round functions; one can note that inside the round function of an FN, a function called F-function is applied to a half of the state. In both cases, the principles of confusion and diffusion suggested by Shannon [Sha49] are implemented by considering each round function / F-function as the composition of key-induced permutations as well as non-linear confusion layers and linear diffusion layers, which are invertible in the case of SPNs and preferably (but not necessarily) invertible in the case of FNs. These two families are briefly discussed in the following sections.

1.2.1 Substitution-Permutation Networks

The framework of SPNs has been widely studied in the last years, and nowadays is known especially as the basic structure of the current U.S. encryption standard AES [DR13]. In an SPN the block is divided into multiple smaller bricks, each brick becomes the input of a non-linear function (S-box), then

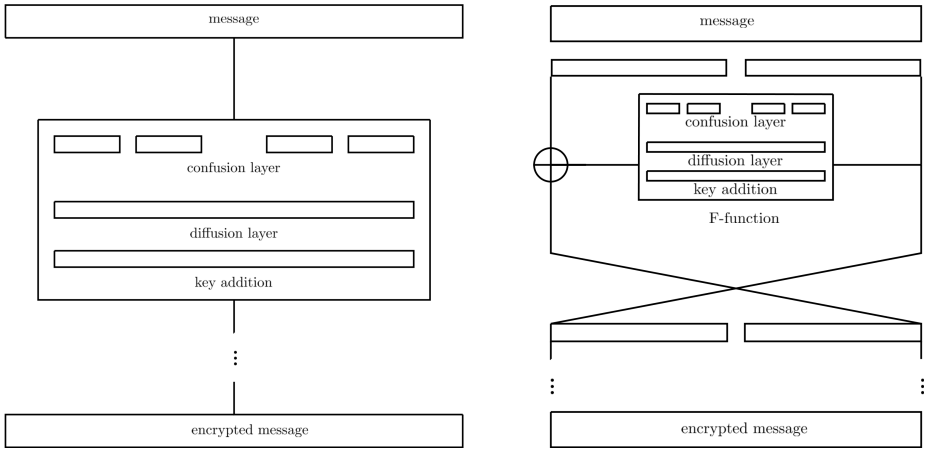

Figure 1.1: Round function of an SPN and of an FN

the bits of the block are mixed by means of a linear function. The key addition may occur before or after these two operations. The following definition gives a mathematical description of SPNs in the model we will be using throughout this work. In this case we are assuming $\mathcal{M} = V$.

Definition 1.2.2. An R-round iterated cipher Φ is called a *Substitution-Permutation Network (SPN)* if Φ is a family of encryption functions $\{ E_K \mid K \in \mathcal{K}\} \subset \mathrm{Sym}(V)$ such that for each $K \in \mathcal{K}$ the map E_K is the composition of R round functions, i.e. $E_K = \varepsilon_{1,K}\, \varepsilon_{2,K} \ldots \varepsilon_{R,K}$, where $\varepsilon_{i,K} = \gamma\lambda\sigma_{k_i}$ and

- $\gamma \in \mathrm{Sym}(V)$ is a non-linear bricklayer transformation which acts in a parallel way on each V_j, i.e.

  $$(x_1, x_2, \ldots, x_n)\gamma = \big((x_1, \ldots, x_s)\gamma_1, \ldots, (x_{s(b-1)+1}, \ldots, x_n)\gamma_b\big)\,.$$

  The maps $\gamma_j \in \mathrm{Sym}(V_j)$ for each $1 \leq j \leq b$ are traditionally called *S-boxes*;

- $\lambda \in \mathrm{Sym}(V)$ is a linear map;

- $\sigma_{k_i} : V \to V, x \mapsto x + k_i$ represents the key addition, where $+$ is the usual bit-wise XOR on $\mathbb{F}_2$. The *round keys* $k_i \in V$ are usually derived from the master key K by means of a public algorithm, called *key-schedule*. Using the terminology developed later in this work, we say that the key addition defines a translation of k_i to the vector x.

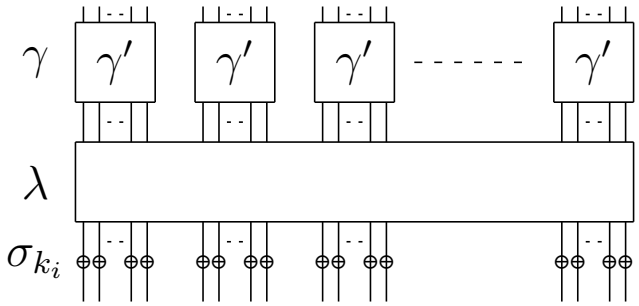

Figure 1.2: Example of 1-round encryption of an SPN

The function $\rho \stackrel{\text{def}}{=} \gamma\lambda$ is called the *generating function* of the SPN. For each $1 \leq r \leq R$ we denote by $E_K^{(r)}$ the composition of the first r round functions of the encryption function E_K. In particular $E_K = E_K^{(R)}$. Figure 1.2 displays the structure of a round function for an SPN.

For example, for the classical AES-128 cipher a 10-round encryption is performed and 8-bit S-boxes are implemented. The state is arranged into a 4×4 matrix of bytes and the diffusion layer is a combination of ShiftRows and MixColumns: in the former the last three rows of the state are shifted cyclically, whereas in the latter a mixing operation which operates on the columns of the state combining the four bytes in each column is applied [DR13]. Instead PRESENT, briefly described also in Section 1.2.2, manipulates 4-bit S-boxes; the diffusion is granted by a permutation matrix. The designers have estimated that the iteration of 31 rounds gives a sufficient margin of security to the cipher [BKL+07].

Remark 1.2.3. Notice that a bit-wise XOR addition is not the only possible way to define a key-addition layer. In many modern ciphers, the key addition is performed e.g. by considering a modular addition. However, for the purposes of this work, when studying SPNs, only key-addition layers induced by the bit-wise sum modulo 2 will be considered. Moreover, Definition 1.2.2 does not include, for sake of simplicity, atypical rounds of the cipher, since it is out of the scope of this work. However, it is worth mentioning that in many ciphers the first and the last round, for efficiency and security reasons, may be different from the others. For example, it is common that in the first or in the last round only the key-addition layer is applied (*whitening*

x	0_x	1_x	2_x	3_x	4_x	5_x	6_x	7_x	8_x	9_x	A_x	B_x	C_x	D_x	E_x	F_x
xS	C_x	5_x	6_x	B_x	9_x	0_x	A_x	D_x	3_x	E_x	F_x	8_x	4_x	7_x	1_x	2_x

Figure 1.3: The S-box S of PRESENT

or *post-whitening* process) and in several ciphers the diffusion layer is not applied in the last round.

Notice that, since every round function is composed by invertible layers, the decryption can be performed by applying the inverse of the layers to the ciphertext, in reverse order.

1.2.2 PRESENT

The block cipher PRESENT [BKL+07] has been designed in 2007, for the purpose of being a *lightweight* cipher, i.e. a cipher suitable for hardware implementation in low-power devices or constrained environments such as RFID tags and sensor networks. It is an example of SPN and consists of 31 rounds. The block length is 64 bits and two key lengths of 80 and 128 bits are supported. The encryption functions are induced by 32 round-keys, 31 of which are used in the standard rounds and the remaining one is used in the last and atypical round for post-whitening. Typically the round function is unique and is obtained as the composition of the following three layers:

addRoundKey the 64-bit round key, derived by the designed key-schedule [BKL+07], is XOR-ed to the partial state;

sBoxlayer the 64-bit block is split into 16 4-dimensional bricks, each of these is substituted in accordance with the S-box $S \in \mathrm{Sym}(\mathbb{F}_2)^4$ displayed in Fig. 1.3;

pLayer with the convention that the left-most bit of the block is in position 0, and the right-most in position 63, for each $0 \leq i \leq 63$ the bit of the state in position i is moved to position $P(i)$, where the permutation P

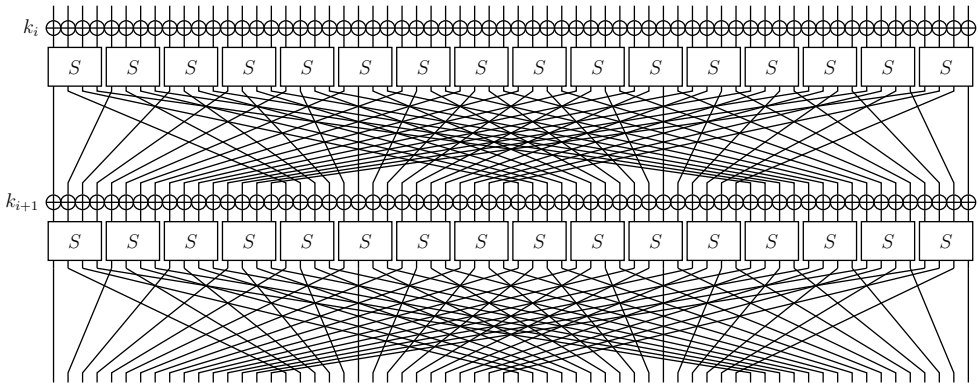

Figure 1.4: A 2-round encryption of PRESENT

is defined as follows

$$P(i) \stackrel{\text{def}}{=} \begin{cases} 16 \times i \mod 63 & 0 \leq i < 63 \\ 63 & i = 63. \end{cases}$$

Figure 1.4 displays the composition of two typical round functions of PRESENT.

1.2.3 Feistel Networks

Besides SPNs, many modern ciphers and their precursors are based on the framework of Feistel Networks, which became popular when the U.S. Federal Government adopted the DES as the standard encryption algorithm for the protection of sensitive, unclassified electronic government data. This cipher, derived from the block cipher Lucifer designed by Horst Feistel and Don Coppersmith in 1973, has been withdrawn in the end of the nineties, mainly due to its short key-length (56 bits). However, its study has highly influenced the advancement of modern cryptography, both in the direction of understanding the properties of FNs and developing new cryptanalytic techniques.

Similarly to SPNs, a $2n$-bit Feistel Network consists of the repetition of R rounds of an identical structure. This repeated structure is realised by means of the so-called F-function and a swap operation. The F-function maps a n-bit input into a n-bit output under the action of a set of round

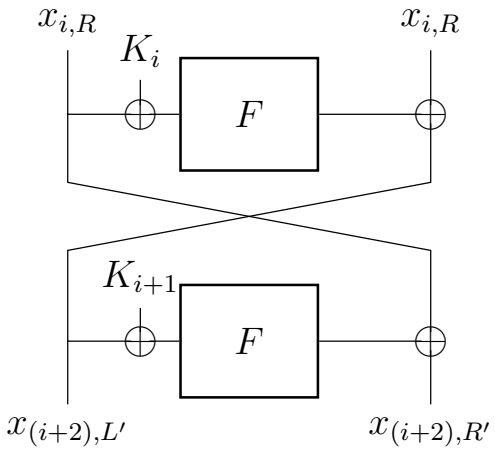

Figure 1.5: Example of 2-round encryption of a FN

keys. Such a function encrypts the right half of the state, which is then XOR-ed to the remaining part of the message. The two halves of the current state are then swapped and the round is repeated. Feistel Networks usually include an atypical last round, where no swap is performed. An example of 2-round encryption of an FN is illustrated in Fig. 1.5.

Let us now describe more formally this procedure.

Definition 1.2.4. Let $f : (\mathbb{F}_2)^n \to (\mathbb{F}_2)^n$ be a vectorial Boolean function. We denote by $\overline{f}$ the formal operator $\overline{f} : (\mathbb{F}_2)^{2n} \to (\mathbb{F}_2)^{2n}$

$$\overline{f} \stackrel{\text{def}}{=} \begin{pmatrix} \mathbb{0}_n & \mathbb{1}_n \\ \mathbb{1}_n & f \end{pmatrix},$$

which for any $(x_1, x_2) \in (\mathbb{F}_2)^n \times (\mathbb{F}_2)^n$ acts as $(x_1, x_2)\overline{f} \stackrel{\text{def}}{=} (x_2, x_1 + x_2 f)$. The operator $\overline{f}$ is called the *Feistel operator induced by* f.

The operator previously defined allows to give an algebraic description of FNs, which are defined on the message space $\mathcal{M} = V \times V$.

Definition 1.2.5. An R-round iterated cipher Φ is called a *Feistel Network (FN)* if Φ is a family of encryption functions $\{ E_K \mid K \in \mathcal{K}\} \subset \mathrm{Sym}(V \times V)$ such that for each $K \in \mathcal{K}$ the map E_K is the composition of R key-dependent

Feistel operators, i.e. $E_K = \overline{\varepsilon_{1,K}}\,\overline{\varepsilon_{2,K}}\ldots\overline{\varepsilon_{R,K}}$, where $\varepsilon_{i,K} : V \to V$ for $1 \leq i \leq R$.

Feistel operators are usually designed in accordance to the Shannon's principles, i.e. contain layers providing confusion and diffusion. The key-addition layers may be designed in several ways, more frequently they are induced from the XOR or from a modular addition. Notice that, as in the case of Definition 1.2.2, we do not include possible atypical rounds in the definition of the cipher.

Remark 1.2.6. One advantage of the Feistel Network is that the decryption process is identical to encryption, provided the round keys are taken in reverse order. Moreover, notice that a Feistel operator $\bar{f}$ is always invertible, regardless the invertibility of f, and it has the following inverse

$$\overline{f}^{-1} = \begin{pmatrix} f & \mathbb{1}_n \\ \mathbb{1}_n & \mathbb{0}_n \end{pmatrix}.$$

It is indeed easy to check that

$$(x_2, x_1 + x_2 f) \begin{pmatrix} f & \mathbb{1}_n \\ \mathbb{1}_n & \mathbb{0}_n \end{pmatrix} = (x_1, x_2).$$

The cipher presented in the following section is a FN with an easy description.

1.2.4 GOST 28147-89

Developed in the 1970s, the cryptosystem GOST has been kept secret since the dissolution of the USSR, when it was declassified and it was released to the public in 1994. The cipher was a soviet alternative to the U.S. standard DES, with whom it shares a similar structure. It is a 64-bit FN with a key-length of 256 bits. Each Feistel operator applies to the 32-bit right half of the state a key-dependent F-function where confusion is provided by means of a parallel layer made up by 4-bit S-boxes, diffusion is provided by a left rotation of 11 bits, whereas the key is added to the state by means of an addition modulo 2^{32}. The full description of the variant GOST 28147-89 may be found e.g. in [Dol10].

1.3 Classical round functions

After having presented the main differences between SPNs and FNs, we introduce the notion of *classical round function*, which allows to describe formally both cipher families in a unified way, provided the round key is used as a translation (i.e., the key addition is the usual XOR). The family of classical round functions for iterated block ciphers of our model is large enough to include the round functions of well-established SPNs e.g. AES, PRESENT, SERPENT, and the F-function of FNs like Camelia. Notice that, for sake of simplicity, atypical rounds are again not considered in this description.

Definition 1.3.1. For each $k \in V$, a *classical round function* induced by k is a map $\varepsilon_k \in \mathrm{Sym}(V)$ of the type $\varepsilon_k = \gamma\lambda\sigma_k$, where

- $\gamma : V \to V$ is a non-linear permutation (parallel S-box) which acts in a parallel way on each V_j, i.e.

$$(x_1, x_2, \ldots, x_n)\gamma = \big((x_1, \ldots, x_s)\gamma_1, \ldots, (x_{s(b-1)+1}, \ldots, x_n)\gamma_b\big)\,,$$

  applying the S-box $\gamma_j : V_j \to V_j$ to the j^{th} brick;

- $\lambda \in \mathrm{Sym}(V)$ is a linear map,

- $\sigma_k : V \to V, x \mapsto x + k$ represents the addition with the round key k.

When used inside block ciphers, the round keys in V are derived by the designer-provided key-scheduling function from the master key $K \in \mathcal{K}$. Since, as we will discuss later in detail, studying the role of the key-schedule is out of the scope of this work, one can simply assume that round keys are randomly-generated vectors in V.

It is important to recall here that, even though the terms "SPN" and "FN" refer to a larger variety of ciphers (i.e. different key-addition or a different arranging of the layers may be considered), for the purposes of this work we choose to focus only on ciphers with a XOR-based key addition. For this reason, saying SPN we refer to any cipher $\{E_K \mid K \in \mathcal{K}\} \subseteq \mathrm{Sym}(\mathcal{M})$ having an SPN-like structure with $\mathcal{M} = V$ and having classical round functions on V

as round functions, and saying FN to any cipher $\{E_K \mid K \in \mathcal{K}\} \subseteq \mathrm{Sym}(\mathcal{M})$ having an FN-like structure with $\mathcal{M} = V \times V$ and having classical round functions on V as F-functions. Moreover, we assume that the Feistel operators defining an FN are of the type of $\overline{\varepsilon_{i,K}}$, where $\varepsilon_{i,K}$ is a classical round function as in Definition 1.3.1. Hence, as in the case of SPNs, we refer to $\rho \stackrel{\text{def}}{=} \gamma\lambda$ by saying the *generating function* of the Feistel Network.

1.4 Cryptanalysis

When dealing with a new cryptosystem, one of the main issues is to define how its security can be evaluated. A first classification of the security of a block cipher, which was introduced in [Knu94a], can be made considering the possible outcomes of the attacks, here listed in ascending order of dangerousness:

key recovery the attacker finds the secret key K used for the encryption;

global deduction the attacker finds a function which is equivalent to the encryption function E_K, without knowing the key K;

local deduction the attacker manages to encrypt or decrypt one message, which he did not obtain from the legitimate sender;

information deduction the attacker recovers some bits of the key or of the plaintext, which he did not get directly from the sender and which he did not have before the attack;

distinguishing attack the attacker can effectively distinguish between two black boxes, one containing the block cipher with a randomly chosen encryption key and the other containing a randomly chosen permutation over the same space.

A second classification is usually done in terms of the capabilities of the attacker to collect information. The aim of the attacker being to recover (partial) information on the key, the following scenarios are considered:

ciphertext-only the attacker has access to some ciphertexts, and has no access to the corresponding plaintexts. This information is the easiest an attacker can gain access to, since the attacker only needs to intercept the messages of an encrypted conversation. However, today's cipher are very unlikely vulnerable to this kind of attack;

known-plaintext the attacker has access to a number of pairs of plaintexts and the corresponding ciphertexts, encrypted with the unknown key;

chosen-plaintext the attacker has access to an encryption oracle which can provide the encryption with the same key of a set of messages provided by the attacker;

adaptive chosen-plaintext the attacker can behave several time as in the chosen-plaintext scenario. After viewing and analysing the output of the oracle, the attacker can make new queries;

related-key the attacker has access to the encryption of the same plaintext using unknown keys which are related to the target key in some mathematical way;

side-channel it is not an attack to the mathematical structure of the cipher but rather to its implementation. The attacker exploit external facts related to the encryption / decryption process, such as electric noise, power consumption, computation time etc., to recover (partial) information of the plaintext or on the key used.

The attacks are listed here in ascending order of data required to the attacker, i.e. from the most practical to the most impractical. Since a proof of the security of a block cipher from every attack can almost never be given, focusing on rather impractical attacks, as the lasts in the list, gives a sufficiently reasonable margin of safety.

To conclude, the success of a cryptanalytic attack can be measured in terms of the effort required for the attack to be performed. In particular, we focus is usually put on the following parameters:

time complexity the time needed to perform an attack, measured in terms of operations the attacker needs to perform;

memory complexity the storage needed to perform the attack;

data complexity the amount of data to obtain in order to perform the attack.

2 Differential cryptanalysis

In this chapter, the attack of differential cryptanalysis is described. Later in this work, we generalise the attack to the case of alternative difference operators.

2.1 Description of the attack

Differential cryptanalysis was publicly introduced by Biham and Shamir in the beginning of the 90's [BS91a] as a powerful tool to cryptanalyse some cryptographic primitives, including mainly block ciphers. It is today known that the IBM designers of DES were already aware of the differential attack before it was published, and consequently designed the cipher in way to be resistant against the attack. However, for a matter of national security, they were asked by the NSA to keep the cryptanalytic technique secret. After the attack was published, many block ciphers were cryptanalysed using this method [BS91b, BS91c, BAB93, RP94]. The idea has later been widely generalised and many variants have been introduced in recent years [Knu94b, BBS99, Knu98]. We will refer to the Biham's and Shamir's attack by saying the *classical* differential attack.

Informal description Let us recall that we denote by V a plaintext / ciphertext space, where $V = (\mathbb{F}_2)^n$. The classical differential cryptanalysis is a chosen-plaintext attack, where the *difference* between plaintexts is fixed.

For any x and y in V, the *difference between x and y with respect to $*$* is defined as $\text{diff}(x, y) \stackrel{\text{def}}{=} x * y^{-1}$, where $*$ is any group operation on V, and y^{-1} is the inverse of y with respect to $*$. In order a cipher to be secure against the larger class of attacks, we expect its encryption functions to destroy patterns in the plaintexts. For example, let P be a set of pairs $P \stackrel{\text{def}}{=} \{(x_1, x_2) \mid x_1, x_2 \in V, \text{diff}(x_1, x_2) = \Delta\}$, where all the corresponding elements in a pair have difference fixed to a given value $\Delta \in V$, i.e.

$$\#\{\text{diff}(x_1, x_2) \mid (x_1, x_2) \in P\} = 1.$$

Let now E be an encryption function of the cipher and let C be the set obtained by encrypting every pair in P with the same encryption function E, i.e. $C \stackrel{\text{def}}{=} \{(x_1E, x_2E) \mid (x_1, x_2) \in P\}$. Since the function E is an encryption function, we expect that it is not possible to predict the difference of corresponding message in a pair after the encryption is performed (output difference), i.e. we expect that

$$\#\{\text{diff}\,(y_1, y_2) \mid (y_1, y_2) \in C\} = \#\{\text{diff}\,(x_1E, x_2E) \mid (x_1, x_2) \in P\} \sim \#P.$$

Notice that, when P contains all the possible pairs having difference fixed to Δ, then $\#P = 2^{n-1}$. If it happens that there exists a suitable difference $\Delta \in V$ such that, even if the key is unknown, it is possible to predict the difference after the encryption with a relatively high probability, then the cipher is vulnerable to differential cryptanalysis. Stated differently, if some differences propagate with unusually high or low probability during the encryption process, this leads to a non-uniform distribution of the output differences which may be used to show a non-random behavior of the cipher.

In order to attack an R-round iterated cipher, in the classical setting, the cryptanalyst needs to choose an input difference Δ_I which, after a partial encryption of $R - 1$ rounds, corresponds to an output difference Δ_O with a significantly high probability p. Then he can proceed as follows:

- he generates a set of pairs $P \stackrel{\text{def}}{=} \{x \mid x \in V^2\}$ with the property that for each $x = (x_1, x_2) \in P$ it holds $\text{diff}\,(x_1, x_2) = \Delta_I$;

- he submit P to an encryption oracle, which encrypts each pair with a target key K, and obtains the set $C \stackrel{\text{def}}{=} \{\text{diff}\,(x_1E_K, x_2E_K) \mid (x_1, x_2) \in P\}$;
- for each pair in C he performs a partial decryption for each message of the pair, trying to guess the round key used in the last round, and he increments a counter every time for the selected round key the difference of the partially decrypted messages equals the expected value Δ_O;
- after the previous step is performed for all the round keys of interest, he chooses the round key candidate with the highest value in its counter.

Assuming that if messages are partially decrypted with the wrong round key then their differences are uniformly distributed, if the probability p is sufficiently high the cryptanalyst succeeds in recovering some bits of the last round key.

2.2 Classical differential cryptanalysis

Although differences can be computed with respect to every operation $*$ such that $(V, *)$ is a group, in general the difference taken into consideration depends on the operation that is used to perform the key addition. In the case of the SPNs, for example, this operation is usually the XOR, since the key addition is performed by XOR-ing the round key bits with the bits of the message. However, keeping in mind that the scope of this work is to cryptanalyse an SPN using an operation different from the XOR, we will describe classical differential cryptanalysis with respect to a general group operation $*$. In order to keep the notation lighter, when it is clear from the context, we will not write explicitly the dependence on the operation $*$.

Let us consider $\Phi = \{E_K \mid K \in \mathcal{K}\} < \text{Sym}(V)$ an R-round iterated block cipher, let $1 \leq r \leq R$ and let $*$ be a group operation on V. Let us denote by diff the difference operator induced by $*$.

Definition 2.2.1. A pair $(\Delta_I, \Delta_O) \in V^2$ is called a *differential*. The elements Δ_I and Δ_O are called *input difference* and *output difference*, respectively.

Actually we are not interested in the concept of the differential itself, but rather in its probability.

Definition 2.2.2. Let $f \in \mathrm{Sym}(V)$, and let $(\Delta_I, \Delta_O) \in V^2$ be a differential. The *differential probability of the differential* (Δ_I, Δ_O) *with respect to* f *and to the operation* $*$ is

$$p^f_{\Delta_I}(\Delta_O) \stackrel{\text{def}}{=} \frac{\#\{x \mid x \in V,\, \mathrm{diff}\,(xf, \mathrm{diff}\,(x, \Delta_I)f) = \Delta_O\}}{\#V},$$

which represents the probability that given any two messages with input difference Δ_I, i.e. x and $\mathrm{diff}\,(x, \Delta_I)$, it holds that the corresponding output difference is Δ_O, i.e. $\mathrm{diff}\,(xf, \mathrm{diff}\,(x, \Delta_I)f) = \Delta_O$, where x is uniformly distributed on V. If $p^f_{\Delta_I}(\Delta_O) > 0$ we say that the differential (Δ_I, Δ_O) is *admissible* over f. Moreover, if $p^f_{\Delta_I}(\Delta_O) = 1$, we say that the differential $p^f_{\Delta_I}(\Delta_O)$ is *deterministic* over f, and if $p^f_{\Delta_I}(\Delta_O) = 0$ that the differential (Δ_I, Δ_O) is *impossible* over f.

Remark 2.2.3. Let $f \in \mathrm{Sym}(V)$ and let (Δ_I, Δ_O) be an admissible differential over f. Then, since f is bijective it holds

$$\Delta_I = 0 \Leftrightarrow \Delta_O = 0.$$

Definition 2.2.4. Let $f \in \mathrm{Sym}(V)$, $(\Delta_I, \Delta_O) \in V^2$ be a differential, and let $x \in V$. We say that the pair $(x, \mathrm{diff}\,(x, \Delta_I))$ *follows the differential* (Δ_I, Δ_O) *with respect to* f if $\mathrm{diff}\,(xf, \mathrm{diff}\,(x, \Delta_I)f) = \Delta_O$.

As already mentioned, in a real case scenario, the key used for the encryption is unknown to the attacker. For this reason, we need to extend the notion of differential probability to the case when the function used is unknown.

Definition 2.2.5. Let (Δ_I, Δ_O) be a differential. The *differential probability of the differential* (Δ_I, Δ_O) *with respect to* $\Phi^{(r)}$ is the expected value of the differential probability of (Δ_I, Δ_O) with respect to $E^{(r)}_K$, assuming that the keys are uniformly distributed on $\mathcal{K}$, i.e.

$$p^{\Phi^{(r)}}_{\Delta_I}(\Delta_O) \stackrel{\text{def}}{=} \sum_{K \in \mathcal{K}} p^{E^{(r)}_K}_{\Delta_I}(\Delta_O) \cdot \frac{1}{\#\mathcal{K}}.$$

When looking at a differential (Δ_I, Δ_O) with the intention of studying its differential probability with respect to $\Phi^{(r)}$, we will also call (Δ_I, Δ_O) an *r-round differential*. If it is clear that (Δ_I, Δ_O) denotes an r-round differential, then $p_{\Delta_I}^{\Phi^{(r)}}(\Delta_O)$ is also written simply as $p_{\Delta_I}(\Delta_O)$.

Notice that, in the context of a differential attack, only plaintexts can be chosen by the attacker, whereas the key used for the encryption in unknown. In our definition of differential probabilities, however, we are assuming that the plaintexts and the keys are independent and uniformly random. These probabilities will be used to determine the best differential suitable to perform the attack. Hence, we are tacitly assuming that, for a given differential (Δ_I, Δ_O), the differential probability for a chosen key equals its expected value $p_{\Delta_I}(\Delta_O)$. This hypothesis is known with the name of *hypothesis of stochastic equivalence* [LMM91].

Definition 2.2.6. The *best differential probability over* $\Phi^{(r)}$ is defined as

$$p_{\max} \stackrel{\text{def}}{=} \max_{\Delta_I \neq 0, \Delta_O} p_{\Delta_I}^{\Phi^{(r)}}(\Delta_O).$$

Each differential (Δ_I, Δ_O) such that $p_{\Delta_I}^{\Phi^{(r)}}(\Delta_O) = p_{\max}$ is called the *best differential over* $\Phi^{(r)}$.

The size of the value $p_{\max}$ reflects the security of the cipher in terms of resistance against the standard differential attack. In particular, the cipher Φ is secure with respect to the classical differential cryptanalysis if it is not possible to detect a relevant bias in the distribution of all the possible r-round differentials, when r is close to the actual number of rounds R of the cipher. More precisely, we can assume that the cipher is secure if no r-round differential (r close to R) has probability different enough from 2^{-n} so that it is possible to distinguish the set of parametrised permutations from a random one. However, for a real-size cipher, the problem of determining the best r-round differential cannot practically be solved. For this reason, an approximation of this value is required.

Definition 2.2.7. Every sequence $(\Delta_0, \Delta_1, \ldots, \Delta_r) \in V^{r+1}$ is called an *r-round differential trail*. Given a differential (Δ_I, Δ_O), we denote by $\mathcal{D}_{(\Delta_I, \Delta_O), r}$ the set of all the r-round differential trails from Δ_I to Δ_O.

Definition 2.2.8. Let $(\Delta_0, \Delta_1, \ldots, \Delta_r) \in V^{r+1}$ be an r-round differential trail, and let $x \in V$. We say that the pair $(x, \mathrm{diff}\,(x, \Delta_0))$ *follows the differential trail* $(\Delta_0, \Delta_1, \ldots, \Delta_r)$ *with respect to* $\Phi^{(r)}$ if for each $1 \leq i \leq r$ it holds $\mathrm{diff}\left(xE_K^{(i)}, \mathrm{diff}\,(x + \Delta_0)E_K^{(i)}\right) = \Delta_i$.

A differential trail $(\Delta_0, \Delta_1, \ldots, \Delta_r)$ is hence a sequence of intermediate differences at each round, starting and ending respectively at the extremities of the r-round differential (Δ_0, Δ_r). As in the case of differentials, we are interested in the probability that a given differential trail holds, which is defined as the probability that a pair follows the differential trail.

Definition 2.2.9. Let $(\Delta_0, \Delta_1, \ldots, \Delta_r) \in V^{r+1}$ be an r-round differential trail, and $K \in \mathcal{K}$ be a key of the cipher. Then the *probability of the differential trail* $(\Delta_0, \Delta_1, \ldots, \Delta_r)$ *with respect to* E_K is

$$p_{(\Delta_0,\Delta_1,\ldots,\Delta_r),K} \stackrel{\mathrm{def}}{=} \frac{\#\left\{x \mid x \in V, \forall 1 \leq i \leq r \ \mathrm{diff}\left(xE_K^{(i)}, \mathrm{diff}\,(x, \Delta_0)E_K^{(i)}\right) = \Delta_i\right\}}{\#V}.$$

It should be clear that if a pair $(x, \mathrm{diff}\,(x, \Delta_I))$ follows an r-round differential (Δ_I, Δ_O), then it is uniquely determined an r-round differential trail $(\Delta_0, \Delta_1, \ldots, \Delta_r)$ such that $\Delta_0 = \Delta_I$, $\Delta_r = \Delta_O$, and the difference of the partial states at the stage i equals Δ_i. Conversely, each pair following an $r-$round differential trail $(\Delta_I, \Delta_1, \ldots, \Delta_{r-1}, \Delta_O)$ follows the differential (Δ_I, Δ_O). Hence the following result holds.

Proposition 2.2.10. *Let* (Δ_I, Δ_O) *be an* r*-round differential. For any encryption function* E_K*, it holds*

$$p_{\Delta_I}^{E_K^{(r)}}(\Delta_O) = \sum_{\substack{(\Delta_I,\Delta_1,\ldots,\Delta_{r-1},\Delta_O) \\ \in \mathcal{D}_{(\Delta_I,\Delta_O),r}}} p_{(\Delta_I,\Delta_1,\ldots,\Delta_{r-1},\Delta_O),K}.$$

□

The probability of a given r-round differential (Δ_I, Δ_O) with respect to a fixed key K is then obtained as the sum of the probabilities of all the possible paths having length $r + 1$ and going from Δ_I to Δ_O. However, this result does not simplify the problem of computing the probability of a differential

for a fixed key. Indeed, the number of paths from Δ_I to Δ_O having length $r+1$, $\#\mathcal{D}_{(\Delta_I,\Delta_O),r}$, increases so rapidly with r that it is not possible to list all of them. Therefore, computing the probability of an r-round differential trail is still a difficult task, which can be simplified assuming the following hypothesis.

Definition 2.2.11. An iterated block cipher Φ is called a *Markov cipher* if the probability of an output difference of any encryption function, once the input difference is given, is independent on the chosen message if the keys are uniformly distributed in the key space.

In such an hypothesis the following result holds.

Proposition 2.2.12. *Let us assume that Φ it is a Markov cipher. Then the probability of an r-round differential trail $(\Delta_0, \Delta_1, \ldots, \Delta_r)$ with respect to the key K is*

$$p_{(\Delta_0,\Delta_1,\ldots,\Delta_r),K} = \prod_{i=0}^{r-1} p_{\Delta_i}^{E_{K_{i+1}}}(\Delta_{i+1}).$$

□

The Markov condition hence allows to calculate the probability of an r-round differential trail as the product of probabilities for 1-round trails. Under this assumption, the probability of Proposition 2.2.10, i.e. the probability of a given r-round differential (Δ_I, Δ_O) for a fixed key, can be seen as the sum of the probabilities with respect to the fixed key of all the r-round differential trails from Δ_I to Δ_O, each of these is the product of 1-round differential trails. Naturally it is not possible, in general, to verify that the Markov condition holds. Nevertheless, it is believed that if the key-schedule is not extremely bad designed, calculating the probability of a differential by multiplying 1-round differential probabilities gives in general a reasonable approximation for practical purposes [BS91a]. However, an attacker is not interested in these key-dependent values in general, since the key used for the encryption is unknown. The average of this values turns out to be a more interesting marker.

Definition 2.2.13. Let $(\Delta_0, \Delta_1, \ldots, \Delta_r) \in V^{r+1}$ be an r-round differential trail. Then the *probability of the differential trail* $(\Delta_0, \Delta_1, \ldots, \Delta_r)$ *with respect to* $\Phi^{(r)}$ is

$$p_{(\Delta_0,\Delta_1,\ldots,\Delta_r)} = \sum_{K \in \mathcal{K}} p_{(\Delta_0,\Delta_1,\ldots,\Delta_r),K} \cdot \frac{1}{\#\mathcal{K}}.$$

The following intuitive result links the probability of an r-round differential with the probability of all the r-round trails which compose the differential.

Theorem 2.2.14. *Let* (Δ_I, Δ_O) *be a differential. Then*

$$p_{\Delta_I}^{\Phi^{(r)}}(\Delta_O) = \sum_{\substack{(\Delta_I,\Delta_1,\ldots,\Delta_{r-1},\Delta_O) \\ \in \mathcal{D}_{r,\Delta_I,\Delta_O}}} p_{(\Delta_I,\Delta_1,\ldots,\Delta_{r-1},\Delta_O)}.$$

□

These values are still out of reach for a real-size cipher, therefore nowadays the best methods to provide evidence that a given cipher is secure / insecure with respect to the classical attack are based on the assumption that it is possible to estimate the best $(R-1)$-round differential (Δ_I, Δ_O) by determining several high-probability $(R-1)$-differential trails from Δ_I to Δ_O.

2.3 The case of SPNs

As already mentioned is the previous chapter, a cipher which belongs to the family of the SPNs is a set of encryption functions, each of which is the composition of R round functions. Each of these round functions in turn is the composition of three different layers: a confusion layer which is public and XOR-non-linear, a diffusion layer which is public and XOR-linear, and a key addition layer which acts as a XOR-translation of the round key derived from the user-selected master key.
In an SPN, if differentials are computed with respect to the XOR, the input difference to the key-addition layer always equals its output difference, since for each $x, k, \Delta \in V$, $(x + k) + (x + \Delta + k) = \Delta$. This suggests, in the case of the classical differential attack, to use the XOR as the operation inducing

the difference operator, hence to consider $\mathrm{diff}\,(x, y) \stackrel{\text{def}}{=} x + y$.

In this context, the following trivial result shows that affine maps admit deterministic differentials.

Lemma 2.3.1. *Let $f \in \mathrm{AGL}(V, +)$, and let $A \in \mathrm{GL}(V, +)$ and $a \in V$ such that $xf = xA + a$ for each $x \in V$. Then for each $\Delta_I \in V$, $(\Delta_I, \Delta_I A)$ is the only admissible differential over f. In particular if $f \in T_n$, then $p^f_{\Delta_I}(\Delta_I) = 1$.*

Proof. Let $\Delta_I \in V$, then for each $x \in V$ it holds $xf + (x + \Delta_I)f = xA + a + (x + \Delta_I)A + a = xA + a + xA + \Delta_I A + a = \Delta_I A$, hence for each $\Delta \in V$

$$p^f_{\Delta_I}(\Delta) = \begin{cases} 1 & \text{if } \Delta = \Delta_I A \\ 0 & \text{if } \Delta \neq \Delta_I A \end{cases}.$$

In particular, if f is a translation, then $A = \mathbb{1}_n$, hence the desired holds. □

Rephrased in the language of SPNs, the previous result means that, when studying the propagation of differentials through the layers of a SPN, diffusion and key-addition layer have deterministic differentials, hence only the confusion layer requires a probabilistic analysis.

Corollary 2.3.2. *Let Φ be an SPN and let λ be its diffusion layer. To each input difference $\Delta \in V$ to the diffusion layer corresponds the output difference $\Delta\lambda$ for each $x \in V$. Moreover, to each input difference Δ to the key-addition layer corresponds the output difference Δ for each $x \in V$.* □

Remark 2.3.3. In the light of the previous result, the probability of any 1-round differential does not depend on the used round-key.

Under the assumption that Φ is a Markov SPN, Proposition 2.2.12 can be restated. The probability of an r-round differential trail for a given key can be expressed as the product of key-independent 1-round-trail probabilities. In particular the probability of an r-round differential trail is independent on the key used.

Proposition 2.3.4. *Let us assume that Φ is a Markov SPN. Then the probability of an r-round differential trail $(\Delta_0, \Delta_1, \ldots, \Delta_r)$ with respect to $\Phi^{(r)}$ is*

$$p_{(\Delta_0,\Delta_1,\ldots,\Delta_r)} = \prod_{i=0}^{r-1} p^{\gamma}_{\Delta_i}(\Delta_{i+1}\lambda^{-1}).$$

Proof. Let us fix a key $K \in \mathcal{K}$. Then

$$p_{(\Delta_0,\Delta_1,\ldots,\Delta_r),K} = \prod_{i=0}^{r-1} p^{E_{K_{i+1}}}_{\Delta_i}(\Delta_{i+1})$$

by virtue of Proposition 2.2.12. The right side of the equation does not depends of the actual value of the round key derived from the actual K for Remark 2.3.3. Moveover, from Corollary 2.3.2, each round function of an SPN sends the input difference Δ_I to the output difference Δ_O with probability p if and only if the confusion layer γ sends Δ_I to $\Delta_O\lambda^{-1}$ with probability p. Indeed, if this happens, then the diffusion layer sends $\Delta_O\lambda^{-1}$ to Δ_O with probability 1, and the key-addition layer keeps this difference unchanged.

□

Since the confusion layer γ is a parallel map, then the probability of a 1-round differential can be expressed as the product of the probability that the bricks of the difference pass through the corresponding S-boxes. In this count, thanks to Remark 2.2.3, S-boxes entered with a zero difference can be ignored. The latter are important in the context of a differential attack, therefore they are worth a more formal definition.

Definition 2.3.5. Let $(\Delta_0, \Delta_1, \ldots, \Delta_r)$ be an r-round differential trail. Let $1 \leq j \leq b$ and let $1 \leq \ell \leq r$. The S-box γ_j is said to be *active* at the round ℓ with respect to $(\Delta_0, \Delta_1, \ldots, \Delta_r)$ if $\Delta_{\ell-1}\pi_j \neq 0$, i.e. if the coordinates of the j^{th} brick of the input difference at the round $(\ell - 1)$ is non-zero. If an S-box is not active, then it is called *non-active*.

In the following corollary of Proposition 2.3.4, the probability of passing through a non-active S-boxes is 1.

Proposition 2.3.6. *Let us assume that Φ is a Markov SPN. Then the probability of an r-round differential trail $(\Delta_0, \Delta_1, \ldots, \Delta_r)$ with respect to $\Phi^{(r)}$ is*

$$p_{(\Delta_0,\Delta_1,\ldots,\Delta_r)} = \prod_{i=0}^{r-1} \prod_{j=1}^{b} p_{\Delta_i^{[j]}}^{\gamma_j} \left((\Delta_{i+1}\lambda^{-1})^{[j]} \right).$$

□

In the light of the previous result, the complexity of computing the probability of a single trail is significantly reduced. In terms of memory, indeed, this computation requires storing b matrices with $2^s \times 2^s$ entries, one for each S-box, containing the data required to predict the output difference after the S-box, once the input difference is given. It should be clear now that the number of active S-boxes plays a crucial rule in determining the size of the probability for a differential trail, and consequently in the success of a differential attack: the higher this number, the lower the probability that a given pair follows the differential trail.

Let us recall that in the following three cases the transaction from a difference to another occurs with probability 1:

- when entering with a zero difference any S-box (entering a non-active S-box),
- when passing through the diffusion layer,
- when passing through the key-addition layer.

Having noticed this, security from differential cryptanalysis, i.e. guaranteeing that the probability of each $(R-1)$-round differential is low enough that the cipher cannot be distinguished from a random permutation, is based on the fact that the differential probabilities induced by the S-boxes (i.e. the only non-deterministic layers of the cipher in term of difference propagation) are low. This is mainly possible in two ways:

- the differential probabilities of the S-boxes are low themselves,
- the diffusion layer activates many S-boxes.

We will discuss these design goals in detail in the following sections, keeping in mind that the success of a differential attack is due to a failure in a good design of the confusion or of the diffusion layer, or maybe in their poor interaction.

2.4 Resistance to classical differential cryptanalysis

The problem of determining conditions on the layers of an SPN guaranteeing that a given cipher is vulnerable to differential cryptanalysis is difficult. This is due to the fact that security from differential attacks is based not only on suitable properties of the layers, but also on the way they interact. In this section and in the following ones, some of the most used design criteria for diffusion and confusion layers will be explained. As already mentioned, we will not focus on any key-schedule-related property, since we will assume that all the round keys are randomly generated. It is worth to mention though that also the key-scheduling function may affect, if bad designed, the security of the cipher in terms of differential attack. However, including such a topic would lead us out of the scope we have established for this work.

2.4.1 Non-linearity notions for confusion layers

The non-linearity of the confusion layer is a necessary condition for the security of the cipher against the most common attacks, and by virtue of Lemma 2.3.1 is particularly crucial in the context of differential attacks. In order to understand what is a good confusion layer with respect to differential cryptanalysis, let us assume the most extreme hypothesis: all the encryption functions are affine. Then in that case, fixing $\Delta \in V$ and considering P the set of pairs having a fixed difference Δ, i.e. $P \stackrel{\text{def}}{=} \{(x, x+\Delta) \mid x \in V\}$, after encrypting all these pairs with an affine function E we obtain

$$\#\{x_1E + x_2E \mid (x_1, x_2) \in P\} = 1.$$

Stated alternatively, the function $\partial_\Delta E : x \mapsto xE + (x+\Delta)E$ in constant on V. It is clear now that a possible way to measure the non-linearity of E is

in terms of the number of values that the function $\partial_\Delta E$ assumes, for each $\Delta \in V$. Let us define this in a concrete way.

Definition 2.4.1. Let $f : (\mathbb{F}_2)^s \to (\mathbb{F}_2)^t$ be a vectorial Boolean function and let $u \in (\mathbb{F}_2)^s$. The *derivative of f in the direction u*, denoted by $\partial_u f$, is the function

$$\begin{array}{rccc} \partial_u f : & (\mathbb{F}_2)^s & \to & (\mathbb{F}_2)^t \\ & x & \mapsto & xf + (x+u)f. \end{array}$$

As already noted, whenever f is linear, the derivatives in every direction are constant. Hence, the more the derivatives of f are far from being constant, the more we can assume that f is non-linear. In this sense, the following definitions can give a first estimate of the non-linearity of f [Nyb93].

Definition 2.4.2. Let $f : (\mathbb{F}_2)^s \to (\mathbb{F}_2)^t$, $u \in (\mathbb{F}_2)^s$ and $v \in (\mathbb{F}_2)^t$. Let us define

$$\delta_f(u,v) \stackrel{\text{def}}{=} \#\{x \in (\mathbb{F}_2)^s \mid x\,\partial_u f = v\} = \#\left\{\{v\}(\partial_u f)^{-1}\right\}.$$

The values $\delta_f(u,v)$ previously defined can be stored in a table, which is important to predict how differences propagate through the S-boxes.

Definition 2.4.3. Let $f : (\mathbb{F}_2)^s \to (\mathbb{F}_2)^t$. The *difference distribution table (DDT)* of f is the integer table $\mathrm{DDT}_f \in \mathbb{Z}^{s \times t}$ where

$$\mathrm{DDT}_f[u,v] \stackrel{\text{def}}{=} \delta_f(u,v).$$

Example 2.4.4. Figure 2.1 displays the DDT of the 4×4 S-box S of the cipher PRESENT described in Section 1.2.2. As already noticed in Remark 2.2.3, since the considered S-box is invertible, $x \in (\mathbb{F}_2)^4$ is a solution of $xS + (x+\Delta)S = 0$ if and only if $\Delta = 0$, hence the first row and the first column of DDT_S are null except for the entry $16 = \#(\mathbb{F}_2)^4$ corresponding to the input / output pair $(\mathtt{0}_\mathrm{x}, \mathtt{0}_\mathrm{x})$. The remainder of the table reads as follows: since $\mathrm{DDT}_S[\mathtt{1}_\mathrm{x}, \mathtt{3}_\mathrm{x}] = 4$, it means that the equation $xS + (x + \mathtt{1}_\mathrm{x})S = \mathtt{3}_\mathrm{x}$ admits four solutions in $(\mathbb{F}_2)^4$, hence the probability that the output difference to S is $\mathtt{3}_\mathrm{x}$ given that its input difference is $\mathtt{1}_\mathrm{x}$ is $1/4$.

	0_x	1_x	2_x	3_x	4_x	5_x	6_x	7_x	8_x	9_x	A_x	B_x	C_x	D_x	E_x	F_x
0_x	16	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
1_x	·	·	·	4	·	·	·	4	·	4	·	·	·	4	·	·
2_x	·	·	·	2	·	4	2	·	·	·	2	·	2	2	2	·
3_x	·	2	·	2	2	·	4	2	·	·	2	2	·	·	·	·
4_x	·	·	·	·	·	4	2	2	·	2	2	·	2	·	2	·
5_x	·	2	·	·	2	·	·	·	·	2	2	2	4	2	·	·
6_x	·	·	2	·	·	·	2	·	2	·	·	4	2	·	·	4
7_x	·	4	2	·	·	·	2	·	2	·	·	·	2	·	·	4
8_x	·	·	·	2	·	·	·	2	·	2	·	4	·	2	·	4
9_x	·	·	2	·	4	·	2	·	2	·	·	·	2	·	4	·
A_x	·	·	2	2	·	4	·	·	2	·	2	·	·	2	2	·
B_x	·	2	·	·	2	·	·	·	4	2	2	2	·	2	·	·
C_x	·	·	2	·	·	4	·	2	2	2	2	·	·	·	2	·
D_x	·	2	4	2	2	·	·	2	·	·	2	2	·	·	·	·
E_x	·	·	2	2	·	·	2	2	2	2	·	·	2	2	·	·
F_x	·	4	·	·	4	·	·	·	·	·	·	·	·	·	4	4

Figure 2.1: DDT of the S-box S of PRESENT

The DDT of an S-box is related to its differential probabilities as stated in the following trivial result. Notice that, for sake of simplicity, differentials have been defined as pairs of elements in the same set, and their probability have been defined only with respect to a bijective function. However Definition 2.2.2 can be easily generalised to the case of different sets and non-bijective functions.

Lemma 2.4.5. *Let $f : (\mathbb{F}_2)^s \to (\mathbb{F}_2)^t$. Then for any $\Delta_I \in (\mathbb{F}_2)^s$ and $\Delta_O \in (\mathbb{F}_2)^t$, the probability of the differential (Δ_I, Δ_O) with respect to f is*

$$p^f_{\Delta_I}(\Delta_O) = \frac{\mathrm{DDT}_f[\Delta_I, \Delta_O]}{2^s}.$$

□

Notice that, the number of non-zero element in a row $\mathrm{DDT}_f[\Delta, \cdot]$ corresponds to the number of different values that $\partial_\Delta f$ can assume. The following definition is a way to use these numbers as an indicator for the non-linearity of f.

Definition 2.4.6. Let $f : (\mathbb{F}_2)^s \to (\mathbb{F}_2)^t$. The *differential uniformity* of f is defined as

$$\delta(f) \stackrel{\text{def}}{=} \max_{\substack{u,v \\ u \neq 0}} \mathrm{DDT}_f[u, v],$$

The function f is said to be δ-*differentially uniform* if $\delta = \delta(f)$.

It is straightorward to notice the following properties:

- if x is a solution of $xf + (x + \Delta_I)f = \Delta_O$, also $x + \Delta_I$ is a solution, hence $\mathrm{DDT}_f[\Delta_I, \Delta_O]$ is always even, and so is $\delta(f)$;
- $2 \leq \delta(f) \leq 2^s$;
- for each Δ_I

$$\sum_{\Delta} \mathrm{DDT}_f[\Delta_I, \Delta] = 2^s.$$

Since the lower the entries of $\mathrm{DDT}_f[\Delta_I, \cdot]$, the more the values $\partial_{\Delta_I} f$ assumes, functions f which reach the lower bound $\delta(f) = 2$ are optimal in terms of non-linearity, in the sense of preventing difference propagation.

Definition 2.4.7. Let $f : (\mathbb{F}_2)^s \to (\mathbb{F}_2)^t$. If $\delta(f) = 2$, f is called *almost-perfect non-linear (APN).*

2.4.2 Known APN permutations

APN functions could represent the best choice when designing the confusion layer of a cipher. However, the problem of finding APN permutations seems to be quite hard, specially in some cases which are the most relevant for the applications. It has been for a long time conjectured that no permutation is APN if the function has an even number of variables. This conjecture has been proven false in 2010, when Dillon et al. [BDMW10] showed an example of an APN permutation in 6 variables. However this is the only known example so far of APN permutation taking as input an even number of variables,

Name	Exponent s	Conditions	Reference
quadratic function	$2^{d'}+1$	$1 \leq d' \leq \ell$ $\gcd(d', s) = 1$	[Nyb93, Gol68]
Kasami function	$2^{2d'} - 2^{d'} + 1$	$2 \leq d' \leq \ell$ $\gcd(d', s) = 1$	[Kas71]
Welsh function	$2^{\ell} + 3$		[Dob99b, CCD00]
Niho function	$2^{\ell} + 2^{\ell/2} - 1$ $2^{\ell} + 2^{(3\ell+1)/2} - 1$	ℓ even ℓ odd	[Dob99a, HX01]
inverse function	$2^s - 2$		[Nyb93, BD93]

Table 2.1: Known APN permutations of the type $x \mapsto x^d$ on $(\mathbb{F}_2)^s$, $s = 2\ell+1$

up to equivalence. It has been shown that no permutation in $(\mathbb{F}_2)^s$ is APN when $s = 4$ [BL08, CSV17], and the problem is still open for $s \geq 8$.

In the case s odd instead, APN permutations are known. In what follows we will briefly recall the case of power functions, i.e. $(\mathbb{F}_2)^s$-valued function of the type of $x \mapsto x^d$, $d \in \mathbb{N}$. It is known that an APN power function is a permutation over $(\mathbb{F}_2)^s$ if and only if s is odd [Car10], hence let us focus on the case $s = 2\ell + 1$. Table 2.1 collects some families of known APN power functions in odd dimension which are bijective [Blo11].

2.4.3 Other non-linearity notions

In this section we will quickly discuss other non-linearity notions for Boolean functions. Even if not necessarily related to differential attacks, they will be used in the remainder of this work. The following definitions were introduced in [CDVS09b] for the study of group-theoretical properties of ciphers.

The requirement of Definition 2.4.6 is essentially a condition on the pre-images of the derivatives of f. However, alternative definitions focused on the images of the derivatives of f may be given. The following one is an example.

Definition 2.4.8. Let $f : (\mathbb{F}_2)^s \to (\mathbb{F}_2)^t$, and let $\delta \geq 2$. Then f is called *weakly δ-differentially uniform* if δ is the least integer such that for each $u \in (\mathbb{F}_2)^s \setminus \{0\}$ it holds

$$\#\operatorname{Im}(\partial_u f) > \frac{2^{s-1}}{\delta}.$$

The adjective *weak* is justified by the following result [CDVS09b].

Proposition 2.4.9. *Let $f : (\mathbb{F}_2)^s \to (\mathbb{F}_2)^t$, and let $\delta \geq 2$. If f is δ-differentially uniform, then f is weakly δ-differentially uniform.*

Proof. Let $u \in (\mathbb{F}_2)^s, u \neq 0$. From the definition of δ-differential uniformity follows that for $v \in (\mathbb{F}_2)^t$ it holds $\#v(\partial_u f)^{-1} \leq \delta$. Moreover

$$(\mathbb{F}_2)^s = \bigcup_{v \in \operatorname{Im}(\partial_u f)} v(\partial_u f)^{-1},$$

and from the hypothesis it follows

$$2^s = \# \left(\bigcup_{v \in \operatorname{Im}(\partial_u f)} v(\partial_u f)^{-1} \right) \leq \#\operatorname{Im}(\partial_u f)\delta.$$

Therefore

$$\#\operatorname{Im}(\partial_u f) \geq \frac{2^s}{\delta} > \frac{2^s}{\delta}.$$

□

Another useful notion of non-linearity comes from the following consideration. Let $f \in \operatorname{Sym}(\mathbb{F}_2)^s$ be such that $0f = 0$. If f is linear, every vectorial subspace of $(\mathbb{F}_2)^s$ is sent to a subspace of $(\mathbb{F}_2)^s$ of the same dimension, i.e. the dimension of subspace is invariant under f. When f is not linear, the null space is mapped into itself, every 1-dimensional subspace $\{0, a\}$ is mapped into the 1-dimensional subspace $\{0, af\}$, whereas the 2-dimensional subspace $\{0, a, b, a+b\}$ is mapped into a 2-dimensional subspace if an only if $af + bf = (a+b)f$. It should be clear that the bigger the dimension $l < s$ of the subspace, the more unlikely the image of the subspace under f is still a vector subspace. The following definition [CDVS09b] is given in this sense.

Definition 2.4.10. Let $f \in \operatorname{Sym}(\mathbb{F}_2)^s$ be such that $0f = 0$ and let $0 < \delta < s$. The function f is said to be *δ-non-invariant* if for any subspaces $U, V \leq (\mathbb{F}_2)^s$ such that $Uf = V$ either $U = V = (\mathbb{F}_2)^s$ or $\dim(U) = \dim(V) < s - \delta$.

Being δ-non-invariant then means that the largest proper subspace sent by f into another subspace has co-dimension greater than δ.

Example 2.4.11. As an example, let us consider the so-called *patched inversion* $f \in \mathrm{Sym}(\mathbb{F}_{2^8})$, which maps every non-zero element into its multiplicative inverse. It is well known that a function which is equivalent to f, up to a change of variable, is used as S-box in the AES. The patched inversion is 4-differentially uniform on $\mathbb{F}_{2^8}$ [Nyb93], whereas it is known that, for each $u \in \mathbb{F}_{2^8}$, $u \neq 0$, it holds $\#\,\mathrm{Im}(\partial_u f) = 2^7 - 1 > 2^6$, hence f is weakly 2-differentially uniform. On the other hand, it has been proven in [CDVS09b] that f is 1-non-invariant.

2.4.4 Requirements on the diffusion layer

As we have discussed in previous sections, the confusion layer of an iterated cipher, due mainly to efficiency reasons, is a *local* non-linear transformation, i.e. any output bit depends on only a limited number of input bits. In particular, it does not provide any interaction between the different bricks. This role is played by the diffusion layer that, acting over all the bricks of the block, spreads the information of a single bit also to bricks different to the one the bit belongs to. In our setting, diffusion is always realised by means of a linear map on V. The following definitions were given as minimal requirement for diffusion in [CDVS09b], in the case of a bijective layer, and were used to derive group-theoretical results on SPNs.

Let us recall that $V = V_1 \oplus V_2 \oplus \ldots \oplus V_b$, and that each s-dimensional space V_j is called a brick.

Definition 2.4.12. A *wall* V' of V is a non-trivial and proper sum of bricks of V, i.e. there exists $\emptyset \neq J \subsetneq \{1, 2, \ldots, b\}$ such that $V' = \oplus_{j \in J} V_j$.

We say that a diffusion layer λ is proper if no wall of V is invariant under λ, and that λ is strongly proper if no wall of V is mapped by λ into another wall of V.

Definition 2.4.13. A linear transformation $\lambda \in \mathrm{Sym}(V)$ is a *proper diffusion layer* if for any wall $V' = \bigoplus_{j \in J} V_j$ of V it holds $V'\lambda \neq V'$.

Definition 2.4.14. A linear transformation $\lambda \in \mathrm{Sym}(V)$ is a *strongly proper diffusion layer* if for any $\emptyset \neq J_1, J_2 \subsetneq \{1, 2, \ldots, b\}$ such that $\#J_1 = \#J_2$ it holds

$$\left(\bigoplus_{j \in J_1} V_j\right) \lambda \neq \bigoplus_{j \in J_2} V_j.$$

It is clear that the condition in Definition 2.4.14 implies the one in Definition 2.4.13.

On the other hand, with an eye on the security with respect differential attacks, following the approach of [DR13], the diffusion can be estimated in terms of a lower bound on the number of the active S-box of any 1-round differential trail (see Definition 2.2.7).

Definition 2.4.15. Let $x \in V$, $x = (x^{[1]}, x^{[2]}, \ldots, x^{[b]})$, with $x^{[j]} \in (\mathbb{F}_2)^s$ being the j^{th} brick, for each $1 \leq j \leq b$. The *brick weight* of x is defined as

$$\mathrm{weight}_b(x) \stackrel{\text{def}}{=} \sum_{x^{[j]} \neq 0} 1.$$

Let now consider an iterated cipher Φ whose generating function is $\rho = \gamma\lambda$. Notice that, if $(\Delta_0, \Delta_1, \ldots, \Delta_r)$ is an r-round differential trail, the number of active S-boxes at the round ℓ with respect to $(\Delta_0, \Delta_1, \ldots, \Delta_r)$, defined in Definition 2.3.5, is exactly $\mathrm{weight}_b(\Delta_{\ell-1})$. In the context of evaluating the diffusion properties of λ with respect to differential cryptanalysis, a considerable measure is the minimum number of active bricks at the input and output of ρ, called the branch number of λ, which basically provides a lower bound for the minimum brick weight of any 1-round differential trail. To formalise this, we need the following definition [DR13].

Definition 2.4.16. Let $\lambda \in \mathrm{Sym}(V)$. The *branch number of* λ is defined as

$$\min_{\substack{x,y \in V \\ x \neq y}} \left\{\mathrm{weight}_b(x + y) + \mathrm{weight}_b(x\lambda + y\lambda)\right\}.$$

The branch number always ranges between 2 and $b + 1$, and diffusion layers whose branch number equals $b + 1$, also called *perfect diffusion layers*, can be constructed from MDS codes [DR13]. Intuitively, provided that the

diffusion layer is (at least) proper, the larger the branch number, the less the number of rounds the encryption needs to be iterated. On the other hand, the counterpart for this is that layers with a larger branch number have higher implementation costs. As an illustration for this, let us recall that the proper diffusion layer of AES has branch number equal to five, and the cipher, in its version with a 128-bit key, performs a 10-round encryption. On the other hand, the permutation matrix of PRESENT has branch number equal to two, due to the fact that, for example, the vector $(1, 0, 0, \ldots, 0)$ is fixed. The encryption in this case is iterated for 31 round. The following theorem, due to Daemen and Rijmen [DR13], relates the branch number of λ to a bound on the number of active S-boxes in a differential trail.

Theorem 2.4.17. *Let Φ be an iterated block cipher having generating function $\rho = \gamma\lambda$. Then the number of active S-boxes in any 1-round differential trail is lower bounded by the branch number of λ.* □

3 Group theoretical security

3.1 Algebraic security

Besides statistical attacks, also algebraic attacks might represent serious threats for block ciphers, as we elaborate further in the following chapters. It is possible, indeed, to link some algebraic properties of the generating function and some algebraic weaknesses of the corresponding cipher. In particular, in this work we will focus on group-theoretical attacks, which have attracted the attention of some mathematicians and cryptographers in the last forty years. The pioneers of the study of ciphers from a group-theoretical point of view were Coppersmith and Grossman, which in 1975 considered a set of functions which can be used to define a block cipher and, by studying the permutation group generated by those, opened the way to a new branch of research focused on group-theoretical properties which can reveal weaknesses of the cipher itself [CG75]. As it has been proved later in [KRS03], if the group generated by the encryption functions is too small, then the cipher is vulnerable to birthday-paradox attacks. Recently, in [CS17] the authors proved that if such group is contained in an isomorphic image of the affine group of the message space induced by a hidden sum, then it is possible to embed a dangerous trapdoor on it. More relevant in [Pat99], Paterson built a DES-like cipher whose encryption functions generate an imprimitive group and showed how the knowledge of this trapdoor can be turned into an efficient attack to the cipher. For this reason, a branch of research in

symmetric cryptography is focused on showing that the group generated by the encryption functions of a given cipher is primitive and not of affine type [ACTT16, ACDVS14, ACS17, CDVS09a, CDVS09b, SW08, Wer92, Wer02, Wer10].

The definition of the group under consideration presents some issues, which we discuss in the following section.

3.2 The group generated by the round functions

As already explained in Section 3.1, statistical attacks are just some of the issues that can threaten block ciphers. Several researchers have shown in recent years that also algebraic attacks can be effective. In this work, the focus is on a particular group-theoretical attack, described in [Pat99] and here treated in Section 3.3, based on an undesirable property of the permutation group generated by the round functions of a cipher, the *imprimitivity*.

Let us define now the target of the so-called imprimitivity attack: the group generated by the round functions of the cipher.
Let $\Phi = \{E_K \mid K \in \mathcal{K}\} \subseteq \mathrm{Sym}(\mathcal{M})$ be an R-round iterated block cipher. We have stressed that the group generated by all encryption functions

$$\Gamma(\Phi) \stackrel{\text{def}}{=} \langle E_K \mid K \in \mathcal{K} \rangle \leq \mathrm{Sym}(\mathcal{M})$$

can reveal weaknesses of the cipher. However, since $\Gamma(\Phi)$ is strictly related to the key-scheduling procedure, which is not easy to describe in terms of groups and their action, its algebraic study is not an easy task. For this reason researchers classically focus on a group which is related to $\Gamma(\Phi)$ and which, ignoring the effect of the key-schedule, is easier to study (for a recent example of a key-schedule related study, see [BF17]). The latter can be defined as follows: since each permutation E_K is the composition of R round functions $\varepsilon_{1,K}, \varepsilon_{2,K} \ldots, \varepsilon_{R,K}$, for each $1 \leq r \leq R$, it is possible to define the group

$$\Gamma_r(\Phi) \stackrel{\text{def}}{=} \langle \varepsilon_{r,K} \mid K \in \mathcal{K} \rangle,$$

where all the possible round keys for the round r are considered. This lead to the definition of the group

$$\Gamma_\infty(\Phi) \stackrel{\text{def}}{=} \langle \Gamma_r(\Phi) \mid 1 \leq r \leq R \rangle,$$

which is called *the group generated by the round functions of* Φ, and trivially contains the group generated by the encryption functions as a subgroup.

Example 3.2.1. Let us denote by Φ the block cipher PRESENT defined in Section 1.2.2. As already mentioned, the problem of describing the group $\Gamma(\Phi)$ is out of reach today. However, being ρ the generating function of PRESENT, i.e. the composition of its confusion and diffusion layers, one can easily prove that the group generated by the rounds of the cipher is $\Gamma_\infty(\Phi) = \langle \rho, T_{64} \rangle$. Indeed, the *left-to-right* inclusion is trivial due to the definition of round function. On the other hand, when the null key is considered, we obtain $\rho \in \Gamma_\infty$, hence also $\rho^{-1} \in \Gamma_\infty$, and consequently $T_{64} < \Gamma_\infty$. This proves the *right-to-left* inclusion. The same description holds for a large class of ciphers, called *translation-based ciphers* [CDVS09b], which also includes the block ciphers AES and SERPENT and some lightweight ciphers.

3.3 Imprimitivity attack

Before describing the imprimitivity attack, let us recall some basic notions from permutation group theory. Let G be a finite group acting on the set $\mathcal{M}$. We denote by $vG = \{vg \mid g \in G\}$ the orbit of $v \in \mathcal{M}$ and by $G_v = \{g \in G \mid vg = v\}$ its stabiliser. A partition $\mathcal{X}$ of $\mathcal{M}$ is *trivial* if $\mathcal{X} = \{\mathcal{M}\}$ or $\mathcal{X} = \{\{v\} \mid v \in \mathcal{M}\}$, and *$G$-invariant* if for any $X \in \mathcal{X}$ and $g \in G$ it holds $Xg \in \mathcal{X}$. Any non-trivial and G-invariant partition $\mathcal{X}$ of $\mathcal{M}$ is called a *block system*. In particular any $X \in \mathcal{X}$ is called an *imprimitivity block*. The group G is *primitive* in its action on $\mathcal{M}$ (or G *acts primitively* on $\mathcal{M}$) if G is transitive and there exists no block system. Otherwise, the group G is *imprimitive* in its action on $\mathcal{M}$ (or G *acts imprimitively* on $\mathcal{M}$).

We remind the following well-known results which will be useful in the remainder of the work, and whose proofs may be found e.g. in [Cam99].

Lemma 3.3.1. *A block of imprimitivity is the orbit vH of a proper subgroup $H < G$ that properly contains the stabiliser G_v, for some $v \in \mathcal{M}$.* □

Lemma 3.3.2. *If T is a transitive subgroup of G, then a block system for G is also a block system for T.* □

Lemma 3.3.3. *Let us assume that $\mathcal{M}$ is a finite vector space over $\mathbb{F}_2$ and T its translation group, i.e. $T = \{\sigma_v \mid \sigma_v : \mathcal{M} \to \mathcal{M},\ x \mapsto x + v, v \in \mathcal{M}\}$. Then*

- *T is 2-elementary, abelian and regular;*
- *T is transitive and imprimitive on $\mathcal{M}$;*
- *for any proper and non-trivial subgroup U of $(\mathcal{M}, +)$, $\{U + v \mid v \in \mathcal{M}\}$ is a block system.*

□

Description of the imprimitivity attack

When the group $\Gamma_\infty(\Phi)$ turns out to act imprimitively on $\mathcal{M}$, then it is possible to individuate a non-trivial partition $\mathcal{X}$ of $\mathcal{M}$ which is invariant under the action of $\Gamma_\infty(\Phi)$. Then an attacker can proceed as follows:

Preprocessing

- he chooses a random message (blue dots in Figure 3.1) in each imprimitivity block and encrypts it using a target encryption function E_K, depending on the unknown key $K \in \mathcal{K}$;
- for each selected message, he individuates which block its corresponding ciphertext (red dots in Figure 3.1) belongs to.

Doing so, the attacker obtains a description of the way the target function maps blocks into blocks (see Figure 3.1) by computing $|\mathcal{X}|$ encryptions.

Then the attacker, which has been given a target encrypted message y (red dot in Figure 3.2), performs the following:

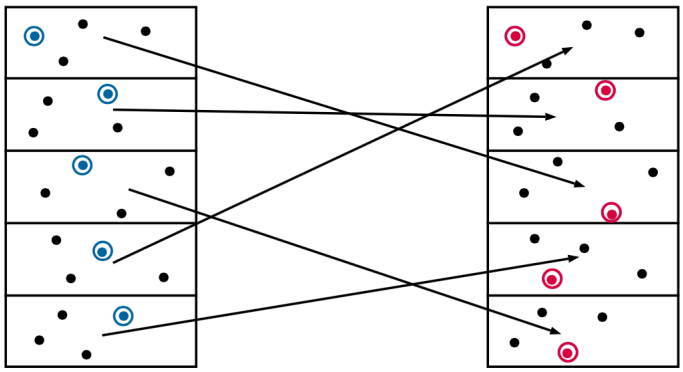

Figure 3.1: Preprocessing of the imprimitivity attack

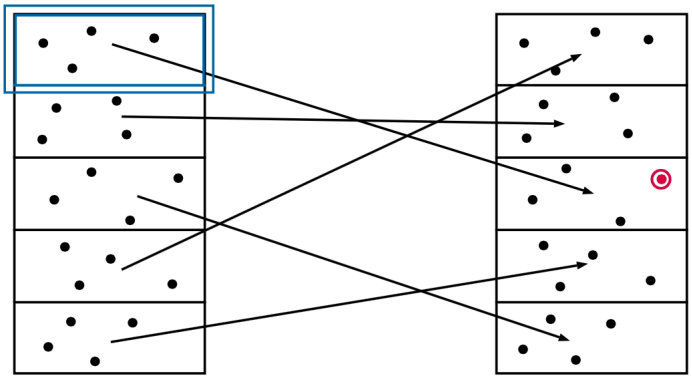

Figure 3.2: Imprimitivity attack

Attack

- he determines which block X' the target encrypted message belongs to;
- he individuates the corresponding imprimitivity block under the action of E_K, i.e. the block X such that $XE_K = X'$ (blue box in Figure 3.2);
- he searches by brute-force for all the meaningful messages in X.

It is now clear that the imprimitivity of the group allows to perform an attack which requires much less than the $|\mathcal{M}|$ operations of a brute-force attack. For this reason, it represents a serious flaw for the cipher Φ, and the primitivity

of $\Gamma_\infty(\Phi)$ is a design goal. An example of a successful attack on a DES-like cipher which makes use of this technique may be found e.g. in [Pat99].

3.4 Resistance to imprimitivity attack

In this section we provide some necessary conditions for the generating function of a cipher which make the group Γ_∞ a primitive group. We will use here the alternative definitions of non-linearity explained in Section 2.4.3. The following result, proved in [CDVS09b] is valid for SPNs.

Theorem 3.4.1. *Let Φ be an R-round SPN, and let $\rho = \gamma\lambda$ its generating function. If λ is proper and there exists $1 \leq \delta < s$ such that for each $1 \leq j \leq b$ the S-box γ_j is*

- *weakly 2^δ-differentially uniform,*
- *δ-non-invariant,*

then $\Gamma_\infty(\Phi))$ is primitive. □

Equivalently, it has been proven in [ACTT16] that the second condition of Theorem 3.4.1 can be weakened as long as the standard notion of differential uniformity is used in the place of the weak one.

Theorem 3.4.2. *Let Φ be an R-round SPN, and let $\rho = \gamma\lambda$ its generating function. If λ is proper and there exists $1 \leq \delta < s$ such that for each $1 \leq j \leq b$ the S-box γ_j is*

- *2^δ-differentially uniform,*
- *$(\delta - 1)$-non-invariant,*

then $\Gamma_\infty(\Phi))$ is primitive. □

To conclude this introduction to group-theoretical security of block ciphers, we list in Table 3.1 all the results concerning the primitivity of established block ciphers and the corresponding references.

Cipher	Γ_∞	Reference
DES [Pub77]	$\mathrm{Alt}(V)$	[Wer92]
SERPENT [BAK98]	$\mathrm{Alt}(V)$	[Wer02]
AES [DR13]	$\mathrm{Alt}(V)$	[Wer10]
KASUMI [Spe07]	$\mathrm{Alt}(V)$	[SW15]
GOST-like [Dol10]	$\mathrm{Alt}(V)$	[ACS17]
PRESENT [BKL+07]	$\mathrm{Alt}(V)$	[ACTT16]
RECTANGLE [ZBL+15]	$\mathrm{Alt}(V)$	[ACTT16]
PRINTcipher [KLPR10]	$\mathrm{Alt}(V)$	[ACTT16]

Table 3.1: Ciphers whose group generated by the round functions is primitive

Part II

Differential cryptanalysis using alternative operations

4

Alternative operations for cryptanalysis

In this part of the thesis the following problem is addressed: is it possible that a block cipher apparently immune to classical differential cryptanalysis can be attacked considering a different operation on the message space?

4.1 Overview and motivation

Differential cryptanalysis was introduced in the beginning of the 90's [BS91a] as a powerful statistical attack targeting first the block cipher DES and has been already described in Chapter 2. The attack, which has later been generalised [Knu94b, BBS99, Knu98], takes advantage of non-uniform relations between plaintext and corresponding ciphertext pairs. Designing ciphers resistant to this attack and its generalisations has since then be of outstanding importance. While, to follow the structure of many symmetric cryptographic designs, the classical difference considered is the bit-wise addition modulo two, in [Ber92] Berson introduces the modular difference to study the MD/SHA family of hash functions. In [AS11], the authors try to use a similar method to cryptanalyse the block cipher PRESENT [BKL+07], featuring a bit-wise round-key addition. Even though this attempt has been unsuccessful, the idea of using alternative difference operations is for the first time taken into consideration and used in block ciphers with a bit-wise

key addition.

The aim of this work is to show that block ciphers may have different levels of resistance against differential attacks, depending on the additive law that is considered on the message space. Even if it is not essential, all the theory is developed on SPN-like ciphers.

Design principles

While bit-wise and modular differences have always been good natural candidates following the cipher's structure, in this work we investigate the possibility of using other differences. The aim being, then, to determine some operations which can weaken the security of well known ciphers. While given a message set of 2^n elements, the number of possible operations on this set is huge, in this work we propose a particular set of difference possibly threatening the security of key-alternating SPN block ciphers. Among the studied criteria to define suitable operations, we focus on operations which can be implemented relatively easily when defined over the basis elements. Such operations, defined from elementary abelian regular groups of translations, have been studied in the recent papers [CS17, BCS17]. In particular, to the best of our knowledge, the first time such operations were described and employed for cryptographic purposes was in [CS17], where the authors provided a description of a family of operations which are particularly suitable for implementation, and designed a toy cipher whose encryption functions are linear with respect to an alternative operation, which they called a *hidden sum*. In this part of the thesis instead, similar operations are studied in the differential context. Constraints coming from the combination of the bit-wise key addition with these operations are studied, and the notion of key distribution table is introduced. The differential uniformity, with respect to other operations, of some non-linear permutations such as the classical cubic function is studied. In particular we provide some conditions increasing the differential uniformity of this function. As a second contribution we propose an example of an SPN block cipher which is resistant against the classical differential attack, with XOR differences, but it is not resistant against a differential

attack which makes use of alternative differences coming from another operation defined on the message space. The designed experimental 15-bit block cipher possesses a structure similar to the one of PRINTcipher [KLPR10], with 3-bit S-boxes affine equivalent to the cubic function. The used linear layer is compatible with the chosen alternative difference and is described by a 15×15 matrix which presents some similarities with the permutation matrices of the block cipher PRINCE [BCG+12]. A general structure for potentially good diffusion layer is provided in this work, reducing the search for candidates round function which would weaken the security of SPNs when comparing our difference operation and the classical one.

Once showed how to define new operations $\circ$ on the message space, sufficient conditions for $\circ$-difference propagation during the encryption process are investigated. Based on the different component of SPN ciphers we search for operations satisfying the following properties.

parallel confusion layer Although the operation $\circ$ might be *a priori* defined on the whole message space V, studying differential properties of a confusion layer seen as a function with 2^n inputs may be impractical for standard-size ciphers. For this reason, in this work we choose to focus on operations which are applied in parallel to the different bricks, i.e. $\circ = (\circ^{(1)}, \circ^{(2)}, \ldots, \circ^{(b)})$, where for each $1 \leq j \leq b$, $\circ^{(j)}$ is an operation on $(\mathbb{F}_2)^s$. This allows us to independently study each S-box.

linear diffusion layer To limit the impact on the $\circ$-differential probability of a $\circ$-differential trail, we analyse only operations such that the diffusion layer is linear with respect to both $+$ and $\circ$. Indeed, if this is the case, the diffusion layer requires no probabilistic analysis.

However, as the chosen operation is different from the XOR, used to add the key at the different rounds of the cipher, differential probabilities have to be introduced when studying the interaction between $\circ$-differences and the key-addition layer. We show how to define a class of operations such that $\circ$-differences resulting from the key-addition layer do not depend on the state considered. In particular, we prove that $(x + k) \circ ((x \circ \Delta) + k)$ equals Δ for a subset of keys and does not depend on x for all keys. In this and in the

following chapter we explain how these requirements can be met.

The theory of alternative operations for cryptographic purposes has been developed in [CS17] and later in [BCS17], from which this work has drawn inspiration. Part of this description includes contents of [CS17, BCS17, CDVS06], sometimes stated and proved differently, according to the terminology and the notation of differential cryptanalysis.

4.2 Differential cryptanalysis revised

When developing new SPNs, designers provide hints on the immunity of the proposed cipher from standard statistical attacks, among which they certainly include differential cryptanalysis. Even if an exhaustive search for high-probability differentials cannot be performed, they usually provide an esteem of the probabilities of the best trails, assuming that these values can accurately measure the resistance to differential attacks. This is classically done with respect to the difference operator $\text{diff}\,(x, y) = x + y$, since this makes deterministic the output difference of the key addition layer. Then, once we are given an R-round SPN designed to be secure against differential attacks, we can assume that no difference propagates during the encryption process with a probability high enough to allow a distinguisher attack. In other words, we believe that for each r-round differential (Δ_I, Δ_O), when $r \sim R$, it holds $p_{\Delta_I}(\Delta_O) \sim 2^{-n}$. However, the fact that the key-addition is XOR-based does not force an attacker to use the XOR as the operation defining differentials. As a matter of fact, nothing guarantees that the differential probabilities computed with respect to the XOR are higher than those computed with respect to different operations, and hence that the security from XOR-based differential attacks implies the immunity from differential attacks induced by whatever difference operator. Our goal is to introduce another group operation, denoted with a *circle*, $\circ$ on the message space and to show that an SPN secure in the classical sense can be distinguished from a random permutation using the new operation considered.

Let $\circ$ be an additive group operation on V such that $x \circ x = 0$ for each

$x \in V$, different from the XOR. We aim at investigating whether it is possible to perform a distinguishing attack against the cipher, where all the chosen plaintext pairs are of the type $(x, \mathrm{diff}\,(x, \Delta))$ for a given $\Delta \in V$, and where the difference operator $\mathrm{diff}\,(x, y) = x \circ y$ is different from the one induced by the XOR. Since through all this chapter we will have to deal with differentials, differential probabilities, differential trails etc. which are computed with respect to different operations, we will denote the dependence on the operation by adding a prefix "+-" if the differential is induced by the XOR and "∘-" if the differential is induced by an operation circle.

In the remainder of this second part, we will explain in detail how to build new additive laws on V and how to study the interaction between the induced differentials and the layers of an SPN. We will provide a concrete example of cipher whose +-differential probabilities do not lead to a successful distinguishing attack, i.e. a cipher which is secure from differential cryptanalysis in the standard context, but against which a differential attack can be successfully performed by computing ∘-diffentials, where ∘ is an operation tailored to fit the structure of the chosen SPN.

4.3 New operations on the message space

Let us recall that we denote by T_+ the group of translations on V, i.e.

$$T_+ = \{\sigma_a \mid a \in V,\, x \mapsto x + a\}\,,$$

and let us stress again that the translation σ_k acts on a vector x in the same way the key addition layer acts on the message x, i.e. $x\sigma_k = x + k$. In order to represent the key addition by means of an action of the translation group on the message space, let us recall that T_+ is 2-elementary, abelian and regular (see Lemma 3.3.3). Moreover, the operation $+$ on V can be seen as the action of T_+ on V, i.e.

$$\forall a, b \in V \quad a + b = a\sigma_b.$$

Our goal is to define alternative operations on the vector space V by means of other 2-elementary abelian regular groups which can play the role of translation groups. Indeed, given any 2-elementary abelian regular subgroup $T < \mathrm{Sym}(V)$, we can represent $T = \{\tau_a \mid a \in V\}$, where for a given $a \in V$, τ_a is the unique element in T which maps 0 into a. Then, if we define

$$\forall a, b \in V \quad a \circ b \stackrel{\text{def}}{=} a\tau_b,$$

we obtain that $(V, \circ)$ is an additive group and $\circ$ induces a vector space structure on V, whose corresponding group of translations is $T_\circ = T$. The proof of this fact is straightforward. Indeed, given $a, b, c \in V$ the following conditions hold.

- $\circ$ is *abelian*: $a \circ b = a\tau_b = 0\tau_a\tau_b = 0\tau_b\tau_a = b\tau_a = b \circ a$;
- 0 is the *neutral element* with respect to $\circ$: $0 \circ a = 0\tau_a = a$;
- a is the *inverse* of a: $a \circ a = a\tau_a = 0(\tau_a)^2 = 0$;
- $\circ$ is *associative*:

$$(a \circ b) \circ c = a\tau_b\tau_c = 0\tau_a\tau_b\tau_c = 0\tau_b\tau_c\tau_a = (b \circ c) \circ a = a \circ (b \circ c).$$

Moreover, $(V, \circ)$ is a vector space over $\mathbb{F}_2$, which is isomorphic to $(V, +)$. However, this construction is too general and far from being practically usable. Indeed, even assuming that we are given a basis of $(V, \circ)$ and a procedure to compute the coefficient of each vector in the given basis, the computation of $a \circ b$ for each $a, b \in V$ requires at least the storage of the $n \times 2^n$ values of the translations defining the basis. For this reason, let us define a class of operations for which $a \circ b$ can be computed in polynomial time.

4.3.1 Efficiently-computable new operations

Alternative operations can be defined by means of elementary abelian regular subgroups of permutations. However, this hypothesis is too general to be useful in practice, since the computation of $a \circ b$ for each $a, b \in V$ requires at least the storage of n 2^n-valued functions. For this reason it becomes necessary to

individuate a smaller subgroup of $\mathrm{Sym}(V)$ which contains elementary abelian regular groups inducing operations that are efficiently computable. In this section we develop the procedure which led us to select some particular set of alternative operations. In particular, based on the result, firstly showed in [CS17], that operations defined from elementary abelian and regular subgroups $T_\circ < \mathrm{AGL}(V,+)$ can be easily computed, we focus on such groups and show a practical method to construct the corresponding operations.

Setting 1. *The operation $\circ$ is induced by a translation group $T_\circ$ which is elementary, abelian and regular, and such that $T_\circ < \mathrm{AGL}(V,+)$.*

Les us assume that $T_\circ < \mathrm{AGL}(V,+)$ is elementary, abelian and regular, and let $\circ$ be the corresponding operation induced. Then, given $a \in V$, the translation τ_a is an affine map with respect to $+$, which means that there exists a $+$-linear map (i.e. a matrix) M_a depending on a, and a $+$-translation σ_b for some $b \in V$ such that $\tau_a = M_a\sigma_b$. In addition, since $0\tau_a = a$, we obtain $b = a$ and so

$$\forall a \in V \ \exists M_a \in \mathrm{GL}(V,+) \quad \tau_a = M_a\sigma_a.$$

In the light of this, in what follows we will denote with M_a the matrix defining τ_a. The following result shows how $+$ interacts with an operation $\circ$ induced by an elementary abelian regular subgroup of the affine group. In particular, it shows that $\circ$ is not distributive over $+$.

Proposition 4.3.1. *Let $T_\circ < \mathrm{AGL}(V,+)$ an elementary abelian regular group. For each $a,b,c \in V$ it holds*

$$(a+b)\circ c = a\circ c + b\circ c + c.$$

Proof. Let $a,b,c \in V$. Then

$$\begin{aligned}(a+b)\circ c &= (a+b)M_c + c \\ &= (aM_c + c) + (bM_c + c) + c \\ &= a\circ c + b\circ c + c,\end{aligned}$$

hence the statement is proven. □

This result has, as a consequence, that $a\circ b$ can be computed in polynomial time.

Corollary 4.3.2. *For each $a, b \in V$, if $a = \sum_{i=1}^{n} \xi_i e_i$, with $\xi_i \in \mathbb{F}_2$, it holds*

$$a \circ b = \begin{cases} \sum_{\xi_i \neq 0} b \circ e_i & \text{if weight}(a) \text{ is odd,} \\ \left(\sum_{\xi_i \neq 0} b \circ e_i\right) + b & \text{if weight}(a) \text{ is even.} \end{cases}$$

Proof. Let $a, b, c, d \in V$. Using Proposition 4.3.1 we obtain

$$\begin{aligned} (a+b+c) \circ d &= a \circ d + (b+c) \circ d + d \\ &= a \circ d + b \circ d + c \circ d + d + d \\ &= a \circ b + b \circ d + c \circ d. \end{aligned}$$

Using this fact and the result of Proposition 4.3.1, the proof of the corollary is straightforward. □

Proposition 4.3.3. *Let $T_\circ$ be as in Setting 1. The map $a \mapsto M_a$ is an homorphism between $(V, \circ)$ and $\mathrm{GL}(V, +)$ equipped with the matrix multiplication.*

Proof. Let $a, b, x \in V$. It holds

$$\begin{aligned} x \circ (a \circ b) &= xM_{a \circ b} + (a \circ b) \\ &= xM_{a \circ b} + aM_b + b \end{aligned}$$

and

$$\begin{aligned} (x \circ a) \circ b &= (xM_a + a) \circ b \\ &= (xM_a + a)M_b + b \\ &= xM_aM_b + aM_b + b. \end{aligned}$$

Being $\circ$ associative, the desired result is proven. □

Corollary 4.3.4. *The set $\{M_a \mid a \in V\}$ is a commutative subgroup of $\mathrm{GL}(V, +)$ and every element has order 2.* □

Example 4.3.5. Let us show an operation which is built from a 2-elementary abelian and regular subgroup of $\mathrm{AGL}(V, +)$, which is different from the XOR.

Let us assume $n = 3$, hence $V = (\mathbb{F}_2)^3$. In order to describe the operation, let us show all the matrices M_a, with $a \in V$:

$$M_{0_x} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} M_{1_x} = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix} M_{2_x} = \begin{pmatrix} 0 & 1 & 1 \\ 0 & 1 & 0 \\ 1 & 1 & 0 \end{pmatrix} M_{3_x} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$$

$$M_{4_x} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} M_{5_x} = \begin{pmatrix} 0 & 1 & 1 \\ 0 & 1 & 0 \\ 1 & 1 & 0 \end{pmatrix} M_{6_x} = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix} M_{7_x} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

It is an easy (but tedious) task to verify that

$$\{1_x, 2_x, 6_x\} = \{(0,0,1), (0,1,0), (1,1,0)\}$$

is a basis for $(V, \circ)$. However, on the other hand

$$\begin{aligned} e_1 \circ e_2 &= (0,0,1) \circ (0,1,0) \\ &= (0,0,1) M_{2_x} + (0,1,0) \\ &= (0,0,1) \begin{pmatrix} 0 & 1 & 1 \\ 0 & 1 & 0 \\ 1 & 1 & 0 \end{pmatrix} + (0,1,0) \\ &= (1,1,0) + (0,1,0) \\ &= (1,0,0) \\ &= e_3, \end{aligned}$$

therefore the canonical basis is not a basis for $(V, \circ)$.

As we have just shown, we do not have a "canonical" basis for $(V, \circ)$ in general. This does not represent a major issue since, by means of Corollary 4.3.2, we can always compute $a \circ b$ by moving the problem in $(V, +)$, where we always know a basis. However, in the following sections, we will refine the hypotheses on $T_\circ$ in order to obtain operations which are suitable for performing a differential attack against an SPN, and simultaneously we will determine conditions which ensure that the canonical basis is always a basis for $(V, \circ)$.

4.4 Interaction with the key-addition layer

Let now $\circ$ be an operation as in Setting 1 and let us assume we want to use it for a differential attack. The classical differential attack exploits the property, as well as others, that each $+$-difference is maintained the same after the round key is XORed. This is never the case when considering $\circ$-differences. Indeed, for each pair of messages x and $x \circ \Delta$ having $\circ$-difference fixed to Δ, after the addition with the round key k we get

$$(x + k) \circ ((x \circ \Delta) + k). \tag{4.1}$$

It is easy to show that Eq. (4.1) $= \Delta$ for each $x, k \in V$ if and only if $+ = \circ$. However, it may be possible that for some *weak* key $k \in V$ it holds $x+k = x \circ k$ for each $x \in V$. Then, in that case, every occurrence of "$+k$" in Eq. (4.1) can be replaced by "$\circ k$", and hence the output difference to the key-addition layer becomes $(x \circ k) \circ ((x \circ \Delta) \circ k) = \Delta$, which is exactly what happens in the classical attack when differences pass through the key-addition layer. Let us give a formal definition of the (possibly empty) set of weak keys.

Definition 4.4.1. Let $\circ$ be any operation on V. A vector $k \in V$ is called a *weak key* if for each $x \in V$ it holds $x + k = x \circ k$. The set

$$W_\circ \stackrel{\text{def}}{=} \{k \mid k \in V,\ k \text{ is a weak key}\}$$

is called the *set of the weak keys*.

In this context, the following result is helpful. The proof may be found in [CDVS06].

Theorem 4.4.2. *Let $T_\circ$ be as in Setting 1. Then $T_\circ \cap T_+ \neq \emptyset$.* □

As a consequence of this theorem, if the translations defining $\circ$ are $+$-affinities, then weak keys exist. The proof for this fact relies on the following lemma.

Lemma 4.4.3. *Let $T_\circ$ be as in Setting 1. For each $a \in V$, $\sigma_a \in T_\circ$ if and only if $a \in W_\circ$.*

Proof. Let $a \in V$. If $\sigma_a \in T_\circ$, there exists $b \in V$ such that $\sigma_a = \tau_b$, and $a = 0\sigma_a = 0\tau_b = b$, hence $\sigma_a = \tau_a$. This proves $a \in W_\circ$. Conversely, if $a \in W_\circ$, then $\sigma_a = \tau_a \in T_\circ$. □

Theorem 4.4.4. *Let $T_\circ$ be as in Setting 1. Then $W_\circ$ is a non-trivial vector subspace of $(V, +)$ and $(V, \circ)$.*

Proof. Let $a \in V$. From Theorem 4.4.2 and from Lemma 4.4.3

$$\emptyset \neq \{a \mid a \in V,\ \sigma_a \in T_\circ \cap T_+\} = W_\circ,$$

which concludes the proof. □

The following result gives a bound on the dimension of the weak-key space, and it is due to Calderini [CS17].

Theorem 4.4.5. *Let $T_\circ$ be as in Setting 1 and let us assume $T_\circ \neq T_+$. Then $0 < \dim(W_\circ) \leq n - 2$.*

Proof. Let us assume by contradiction that $W_\circ = \mathrm{Span}\{a_1, a_2, \ldots, a_{n-1}\}$, and let a in $V \setminus W_\circ$. Let $b \in V$ and let us write $b = \sum_{i=1}^{n-1} \xi_i a_i + \xi_n a$, where $\xi_i \in \mathbb{F}_2$ for $1 \leq i \leq n$. Let us recall that, since for each $1 \leq i \leq n-1$ it holds $a_i + a = a_i \circ a = a_i M_a + a$, then $a_i = a_i M_a$. Furthermore $a \circ a = 0$, hence $aM_a = a$. Therefore

$$\begin{aligned} b \circ a &= \left(\sum_{i=1}^{n-1} \xi_i a_i + \xi_n a\right) \circ a \\ &= \left(\sum_{i=1}^{n-1} \xi_i a_i + \xi_n a\right) M_a + a \\ &= \sum_{i=1}^{n-1} \xi_i a_i M_a + \xi_n a M_a + a \\ &= \sum_{i=1}^{n-1} \xi_i a_i + \xi_n a + a \\ &= b + a. \end{aligned}$$

This proves that $a \in W_\circ$, which is a contradiction. □

Example 4.4.6. Let us recall the operation defined in Example 4.3.5. In the light of the previous result, for this operation the weak-key subspace is 1-dimensional. The non-trivial weak key for $\circ$ is $(1,1,1)$, as it can be noticed from $M_{7_x} = \mathbb{1}_3$, which means that $W_\circ = \{(0,0,0),(1,1,1)\}$.

The important role of $W_\circ$ is now disclosed: whenever k is a weak key, the key-addition layer σ_k behaves as a translation layer with respect to $\circ$-differences. However, in the general case, when k is not a weak key, the value of Eq. (4.1) is different from Δ, hence differential probabilities have to be introduced when studying the interaction between $\circ$-differences and the key-addition layer. With an eye on using $\circ$ to perform a differential attack, we can reasonably assert that the attack can succeed if we manage to weaken enough the non-linearity of the confusion layer. It is straightforward to notice that the larger the set of the weak keys, the more the operation $\circ$ is similar to $+$. For this reason we may be tempted to assume that a successful attack relies on considering an operation $\circ$ such that $\dim(W_\circ)$ is very little compared to n. On the other hand, for such an operation, the probabilities induced by the key-addition layer are lower that those induced by an operation with a larger weak-key set, since it happens more often that the output difference of the key-addition layer depends on the message and on the key, as well as on the input difference. The success of a differential attack using a different operation relies then, among other things, on finding the correct balance between n and $\dim(W_\circ)$. In order to understand that, we need a more practical way to represent the matrices described in Section 4.3.1, which will be explained in Section 4.4.2. Before doing this, let us define a further operation which simplifies computations involving $\circ$ and $+$.

4.4.1 Introducing a product

Let us recall that the operation $\circ$, as shown in Proposition 4.3.1, is not distributive over $+$. Since our goal is to determine the value of Eq. (4.1), let us introduce an operation on V which is induced by both $+$ and $\circ$ and helps in understanding their interaction. Although the following operation depends on $\circ$, we do not explicitly write its dependance to keep the notation lighter.

Definition 4.4.7. Let $\circ$ be an operation as in Setting 1. For each $a, b \in V$ let us define

$$a \cdot b \stackrel{\text{def}}{=} a + b + a \circ b.$$

The operation $\cdot$ is called the *dot product induced by* $\circ$.

Remark 4.4.8. The following facts are straightforward:

- the dot product $\cdot$ is abelian,
- for each $a \in V$, $a \cdot a = 0$,
- if $a \in W_\circ$ or $b \in W_\circ$, then $a \cdot b = 0$. In particular for each $a \in V$, $a \cdot 0 = 0$.

The importance of the dot product is established by the following result, whose straightforward proof which follows is intended to show in an easy way to the reader how all the previously defined operations combine together.

Theorem 4.4.9. *Let $\circ$ be an operation as in Setting 1 and let $\cdot$ be the dot product induced. Then $\cdot$ is distributive over $+$, i.e. $(V, +, \cdot)$ is an $\mathbb{F}_2$-algebra.*

Proof. Let $a, b, c \in V$. First of all $\cdot$ is associative. Indeed

$$\begin{aligned}
(a \cdot b) \cdot c &= (a + b + a \circ b) \cdot c \\
&= a + b + a \circ b + c + (a + b + a \circ b) \circ c \\
&= a + b + c + a \circ b + a \circ c + b \circ c + a \circ b \circ c \\
&= a + b + c + b \circ c + a \circ (b + c + b \circ c) \\
&= a \cdot (b + c + b \circ c) \\
&= a \cdot (b \cdot c).
\end{aligned}$$

Moreover,

$$\begin{aligned}
(a + b) \cdot c &= a + b + c + (a + b) \circ c \\
&= a + b + c + a \circ c + b \circ c + c \\
&= a + c + a \circ c + b + c + b \circ c \\
&= a \cdot c + b \cdot c,
\end{aligned}$$

hence $\cdot$ is distributive over $+$ and therefore $(V, +, \cdot)$ is an $\mathbb{F}_2$-algebra. $\square$

Now that the dot product has been introduced, we can rewrite Eq. (4.1) in a different way. This will lead to the definition of new hypotheses on $T_\circ$ which will make the operation $\circ$ more suitable for a differential attack.

Theorem 4.4.10. *Let $\circ$ be an operation as in Setting 1. Then for each $x, k, \Delta \in V$*

$$(x + k) \circ ((x \circ \Delta) + k) = \Delta + k \cdot \Delta + k \cdot \Delta \cdot x. \tag{4.2}$$

Proof. The proof directly follows using the distributivity of $\cdot$ over $+$. Indeed, let $x, k, \Delta \in V$. Then

$$\begin{aligned}
(x + k) \circ ((x \circ \Delta) + k) &= (x + k) \circ (x + \Delta + x \cdot \Delta + k) \\
&= x + k + x + \Delta + x \cdot \Delta + k \\
&\quad +(x + k) \cdot (x + \Delta + x \cdot \Delta + k) \\
&= \Delta + x \cdot \Delta + x \cdot x + x \cdot \Delta + x \cdot x \cdot \Delta \\
&\quad +x \cdot k + k \cdot x + k \cdot \Delta + k \cdot \Delta \cdot x + k \cdot k \\
&= \Delta + k \cdot \Delta + k \cdot \Delta \cdot x.
\end{aligned}$$

□

We have rewritten the output difference of the key-addition layer in a way it does not depend explicitly on $\circ$. However it still depends on x and k. What we gained is that we can derive an interpretation of the dot product that will allow to simplify Eq. (4.2). Before doing this by refining our assumptions, let us show a result which will be helpful in the following sections.

Lemma 4.4.11. *Let $\circ$ be an operation as in Setting 1. Then for each $a, b \in V$ it holds $\sigma_{a \cdot b} = \sigma_a \tau_b \sigma_a \tau_b$.*

Proof. Let $a, b \in V$. First of all let us notice that $\tau_b \sigma_a \tau_b = \tau_b^{-1} \sigma_a \tau_b \in T_+$ since $T_\circ < \mathrm{AGL}(V, +) = \mathrm{N}_{\mathrm{Sym}(V)}(T_+)$, and consequently $\sigma_a \tau_b \sigma_a \tau_b \in T_+$. From

$$0\, \sigma_a \tau_b \sigma_a \tau_b = a \circ b + a + b = a \cdot b$$

the desired holds. □

4.4.2 Assumptions on the weak keys

We showed that $W_\circ$ is a non-trivial vector space, i.e. every operation as in Setting 1 admits weak keys. For reasons that will be clearer later, we want to set the position of the weak keys in the block. For this reason, we assume that $W_\circ$ is generated by a given set of vectors. Even if this set might be chosen arbitrarily, we will assume from now on that $W_\circ$ is generated by the last d vector of the canonical basis, since this will force the matrices defining the operation to have a precise block form.

In what follows, we denote by d the dimension of the weak-key space, i.e. $d \stackrel{\text{def}}{=} \dim(W_\circ)$.

Setting 2. *The operation $\circ$ satisfies the hypotheses of Setting 1 and, if $d = \dim(W_\circ)$, then $W_\circ = \mathrm{Span}\{e_{n-d+1}, \dots, e_n\}$.*

Under this hypothesis, the matrices defining $\circ$ are into the block form showed in the following result.

Theorem 4.4.12. *Let $\circ$ be as in Setting 2. Then, for each $a \in V$ there exist $\Pi_a \in \mathrm{GL}\left((\mathbb{F}_2)^{n-d}, +\right)$ and $\Sigma_a \in (\mathbb{F}_2)^{(n-d)\times d}$ such that*

$$M_a = \begin{pmatrix} \Pi_a & \Sigma_a \\ \mathbb{0}_{d,n-d} & \mathbb{1}_d \end{pmatrix}.$$

Proof. Let $a \in V$, and let $i \in \{n-d+1, \dots, n\}$. Since $\circ$ is abelian and e_i is a weak key it holds

$$a + e_i = a \circ e_i = e_i \circ a = e_i M_a + a,$$

hence $e_i M_a = e_i$, which means that the i^{th} row of M_a is e_i. □

Understanding the dot product

Let us fix $a, b \in V$ and see in detail what the dot product represents. Since $V = W^\perp \oplus W$, we can write $a = (\bar{a}, \tilde{a})$, with $\bar{a} \in (\mathbb{F}_2)^{n-d}$ and $\tilde{a} \in (\mathbb{F}_2)^d$.

Then

$$\begin{aligned} a \cdot b &= aM_b + b + a + b \\ &= (\bar{a}, \tilde{a}) \begin{pmatrix} \Pi_b & \Sigma_b \\ \mathbb{0}_{d,n-d} & \mathbb{1}_d \end{pmatrix} + a \\ &= (\bar{a}\Pi_b, \bar{a}\Sigma_b + \tilde{a}) + a \\ &= (\bar{a}\Pi_b + \bar{a}, \bar{a}\Sigma_b)\,, \end{aligned} \tag{4.3}$$

which does not depend on $\tilde{a}$, the component of a in the space of weak keys.

4.4.3 Assumptions on $\circ$-affinities

Let us come back now to our main concern regarding the key-addition layer, i.e. the fact that its output difference depends on the message, on the key, and on the input difference, as displayed in Theorem 4.4.10. It is clear that we significantly reduce the impact of the key-addition layer on $\circ$-differential probabilities if, for example, we succeed in finding suitable hypotheses which make the value of Eq. (4.1) independent on the message x. This is obtainable, by virtue of Lemma 2.3.1 (*mutatis mutantis*), by forcing the XOR-translations to behave as "$\circ$-affinities". Indeed, writing

$$(x + k) \circ ((x \circ \Delta) + k) = (x\sigma_k) \circ ((x \circ \Delta)\sigma_k),$$

if $\sigma_k = f_k \tau_T$ for a "$\circ$-linear function" f_k and some $T \in V$, we obtain

$$\begin{aligned} (x + k) \circ ((x \circ \Delta) + k) &= (x\sigma_k) \circ ((x \circ \Delta)\sigma_k) \\ &= xf_k \circ T \circ ((x \circ \Delta)f_k \circ T) \\ &= xf_k \circ T \circ xf_k \circ \Delta f_k \circ T \\ &= \Delta f_k, \end{aligned}$$

which does not depend on x anymore. In order to do this, a precise definition of $\circ$-affinities is required. Let us recall that $\mathrm{AGL}(V, +)$, i.e. the group of all the $+$-affinities, can be seen as the normaliser in the symmetric group of the $+$-translation group, i.e. $\mathrm{AGL}(V, +) = \mathrm{N}_{\mathrm{Sym}(V)}(T_+)$. In a similar way, let us define

$$\mathrm{AGL}(V, \circ) \stackrel{\text{def}}{=} \mathrm{N}_{\mathrm{Sym}(V)}(T_\circ)$$

as the group of the $\circ$-affine functions. The stabiliser of 0 in $\mathrm{AGL}(V, \circ)$ represents the subgroup of all the $\circ$-linear functions, i.e.

$$\mathrm{GL}(V, \circ) \stackrel{\text{def}}{=} \mathrm{AGL}(V, \circ)_0.$$

The following important result [CS17] is a description of the matrices defining an operation $\circ$ such that the $+$-translations behave like $\circ$-affinities.

Theorem 4.4.13. *Let $T_\circ$ be as in Setting 2. If $T_+ < \mathrm{AGL}(V, \circ)$, then for each $a \in V$ there exists a matrix $\Sigma_a \in (\mathbb{F}_2)^{(n-d)\times d}$ such that*

$$M_a = \begin{pmatrix} \mathbb{1}_{n-d} & \Sigma_a \\ \mathbb{0}_{d,n-d} & \mathbb{1}_d \end{pmatrix}.$$

Proof. Let $a \in V$. Since $T_\circ$ satisfies the hypotheses of Setting 2, from Theorem 4.4.12 there exist $\Pi_a \in \mathrm{GL}\left((\mathbb{F}_2)^{n-d}, +\right)$ and $\Sigma_a \in (\mathbb{F}_2)^{(n-d)\times d}$ such that

$$M_a = \begin{pmatrix} \Pi_a & \Sigma_a \\ \mathbb{0}_{d,n-d} & \mathbb{1}_d \end{pmatrix}.$$

Moreover, from $T_+ < \mathrm{AGL}(V, \circ)$ we obtain the T_+ normalises $T_\circ$, which means that for each $b \in V$

$$\sigma_{a\cdot b} = \sigma_a^{-1}\tau_b^{-1}\sigma_a\tau_b \in T_\circ,$$

and, from Lemma 4.4.3, this is equivalent to saying that $a \cdot b \in W_\circ$. From this and from Eq. (4.3), $\left(\bar{b}\Pi_a + \bar{b}, \bar{b}\Sigma_a\right) \in W_\circ$, which implies $\bar{b}\Pi_a + \bar{b} = 0$. The conclusion yields from the generality of b. □

Setting 3. *The operation $\circ$ is as in Setting 2 and $T_+ < \mathrm{AGL}(V, \circ)$.*

Operations as in Setting 3 are also known as *effective hidden sums* [BCS17].

Remark 4.4.14. Notice that in Theorem 4.4.13 we also proved that, if the hypotheses of Setting 3 hold, for each $x, y \in V$, $x \cdot y$ is a weak key. In particular, from Remark 4.4.8 follows that every triple dot product is null, i.e. $a \cdot b \cdot c = 0$ for each $a, b, c \in V$. The opposite implication also applies, i.e. if $a \cdot b \cdot c = 0$ for each $c \in V$, then $a \cdot b \in W_\circ$.

In the previous hypothesis, the vectors $\{e_i\}_{i=1}^n$ are a basis for both $(V,+)$ and $(V,\circ)$, and this is basically granted from the fact that for each $a,b \in V$ it holds $a \cdot b \in W_\circ = \mathrm{Span}\{e_{n-d+1},\dots,e_n\}$. Indeed, let us consider $a \in V$ and let us decompose

$$a = \xi_1 e_1 + \xi_2 e_2 + \dots \xi_n e_n. \tag{4.4}$$

In order to determine the coefficients of the decomposition of a with respect to $\circ$ and to the canonical basis, we can proceed as follows: let us write the first two addends of Eq. (4.4) as $\xi_1 e_1 + \xi_2 e_2 = \xi_1 e_1 \circ \xi_2 e_2 + \xi_1 e_1 \cdot \xi_2 e_2$. Since $\xi_1 e_1 \cdot \xi_2 e_2 \in W_\circ$, we can equivalently write $\xi_1 e_1 + \xi_2 e_2 = \xi_1 e_1 \circ \xi_2 e_2 \circ \xi_1 e_1 \cdot \xi_2 e_2$, where $\xi_1 e_1 \cdot \xi_2 e_2 = \xi'_{w_i} e_{w_1}$ for some $n-d+1 \leq w_1 \leq n$ and $\xi'_{w_i} \in \mathbb{F}_2$. We can then rewrite Eq. (4.4) as

$$a = (\xi_1 e_1 \circ \xi_2 e_2 \circ \xi'_{w_1} e_{w_1}) + \xi_3 e_3 + \dots \xi_n e_n.$$

One can proceed in the same way until for all the non-weak vector of the canonical basis, every occurrence of $+$ is replaced by $\circ$, obtaining

$$a = \xi'_1 e_1 \circ \xi'_2 e_2 \circ \dots \xi'_n e_n. \tag{4.5}$$

Notice that, the coefficients ξ'_i in Eq. (4.5) satisfy the following:

- if $1 \leq i \leq n-d$, then $\xi'_i = \xi_i$;
- if $n-d+1 \leq i \leq n$, then $\xi'_i = \mu_i$, where μ_i's are the coefficient of $a \circ e_1 \circ e_2 \circ \dots \circ e_{n-d}$.

From this fact, an algorithm which computes the coefficients of a with respect to $\circ$ and the canonical basis can be easily derived.

Let us now show that, if the dimension of the space is sufficiently small, operations as in Setting 2 satisfy the hypotheses of Setting 3. This result still holds in a more general setting. A more general version can be found in [CS17].

Theorem 4.4.15. *Let $\circ$ be as in Setting 2. If $n \leq 6$, then $T_+ < \mathrm{AGL}(V,\circ)$.*

Proof. Let us assume $T_+ \not\leq \mathrm{AGL}(V, \circ)$. Then, by Remark 4.4.14 there exist $x, y, z \in V$ such that $x \cdot y \cdot z \neq 0$. Let us show that $x, y, z, x \cdot y, x \cdot z, y \cdot z, x \cdot y \cdot z$ are linearly independent. Let $\xi_i \in \mathbb{F}_2$ for $1 \leq i \leq 7$ such that

$$\xi_1 x + \xi_2 y + \xi_3 z + \xi_4 x \cdot y + \xi_5 x \cdot z + \xi_6 y \cdot y + \xi_7 x \cdot y \cdot z = 0.$$

By multiplying each member of the previous equation by $y \cdot z$ we obtain $\xi_1 x \cdot y \cdot z = 0$, which implies $\xi_1 = 0$. In the same way, by multiplying by $x \cdot z$ we prove $\xi_2 = 0$. Proceeding in this way one proves that $\xi_i = 0$ for each $1 \leq i \leq 7$. □

If Φ is an n-bit cipher this result may seems insignificant, since for sure $n >> 6$. However, operations $\circ$ can also be defined only on some bricks of the block, as we will show later. In this case, it does make sense to consider operations defined on smaller spaces, for example of the same size of the S-boxes. Consequently, Theorem 4.4.15 can be applied when considering e.g. ciphers having 3 or 4-bit S-boxes.

4.4.4 A more compact representation

In this section we will present a compact way to represent operations satisfying our last assumption, described also in [BCS17]. From this, some useful properties will be derived. As shown in Theorem 4.4.13, an operation as in Setting 3 is such that the matrices M_{e_i} for each $1 \leq i \leq n$ are in a precise block form. Let us denote by $\boldsymbol{b_{i,j}}$ the last d components of the j-th row of M_{e_i} in such a way that we can represent

$$M_{e_i} = \begin{pmatrix} \mathbb{1}_{n-d} & \Sigma_{e_i} \\ \mathbb{0}_{d,n-d} & \mathbb{1}_d \end{pmatrix} = \left(\begin{array}{c|c} & \boldsymbol{b_{i,1}} \\ \mathbb{1}_{n-d} & \vdots \\ & \boldsymbol{b_{i,n-d}} \\ \hline \mathbb{0}_{d,n-d} & \mathbb{1}_d \end{array}\right).$$

Notice that, by the assumption $W_\circ = \mathrm{Span}\{e_{n-d+1}, \ldots, e_n\}$ we have $\Sigma_{e_i} = 0$ for each $n-d+1 \leq i \leq n$. This incidentally implies that only $n-d$ matrices have to be stored in order to compute $\circ$. In the light of these considerations, an operation as in Setting 3 is defined when the matrices $\Sigma_{e_1}, \Sigma_{e_2}, \ldots, \Sigma_{e_{n-d}}$

are given. Interpreting each row of Σ_{e_i} as an element of the finite field $\mathbb{F}_{2^d}$, we can easily represent all the Σ_{e_i}s in a matrix in $(\mathbb{F}_{2^d})^{(n-d)\times(n-d)}$ as follows:

$$(\Sigma_{e_1}, \Sigma_{e_2}, \dots, \Sigma_{e_{n-d}}) = \begin{pmatrix} \boldsymbol{b_{1,1}} & \boldsymbol{b_{1,2}} & \cdots & \boldsymbol{b_{1,n-d}} \\ \boldsymbol{b_{2,1}} & \boldsymbol{b_{2,2}} & \cdots & \boldsymbol{b_{2,n-d}} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{b_{n-d,1}} & \boldsymbol{b_{n-d,2}} & \cdots & \boldsymbol{b_{n-d,n-d}} \end{pmatrix}. \tag{4.6}$$

We say that the matrix in Eq. (4.6) *defines* the operation $\circ$.

Remark 4.4.16. Let us highlight some properties of the elements $\boldsymbol{b_{i,j}}$:

- from the fact that $T_\circ$ is 2-elementary, for each $1 \leq i \leq n-d$ it holds $e_i \circ e_i = 0$, which means $\boldsymbol{b_{i,i}} = \mathbf{0}$;
- since for each $1 \leq i, j \leq n-d$ we have $e_i \circ e_j = e_j \circ e_i$, then $\boldsymbol{b_{i,j}} = \boldsymbol{b_{j,i}}$.

The following theorem is a characterisation of the operations satisfying the hypotheses of Setting 3.

Theorem 4.4.17. *Let $\circ$ be an operation as in Setting 1, and let $d = \dim(W_\circ) \leq n-2$. A matrix $\Theta \in \mathbb{F}_{2^d}^{(n-d)\times(n-d)}$ defines an operation $\circ$ such that $T_+ < \mathrm{AGL}(V, \circ)$ and $W_\circ = \mathrm{Span}\{e_{n-d+1}, \dots, e_n\}$ (i.e. an operation as in Setting 3) if and only if Θ is zero-diagonal, symmetric and no $\mathbb{F}_2$-linear combination of columns of Θ is the null vector.*

Proof. Part of this result has already been proven in Theorem 4.4.13. Indeed, if we assume that the hypotheses of Setting 3 hold, then for each $1 \leq i \leq n-d$ we can write

$$M_{e_i} = \begin{pmatrix} \mathbb{1}_{n-d} & \Sigma_{e_i} \\ \mathbb{0}_{d,n-d} & \mathbb{1}_d \end{pmatrix},$$

hence Θ is the matrix built as in Eq. (4.6), whose columns are filled with the rows of the matrices Σ_{e_i}'s. From Remark 4.4.16 it follows that Θ is zero-diagonal and symmetric. Moreover, let us assume that an $\mathbb{F}_2$-linear combination of columns of Θ is the null vector. Without loss of generality,

let us assume $\Sigma_{e_1} + \Sigma_{e_2} = 0$. From this it follows that

$$\begin{aligned} M_{e_1 \circ e_2} = M_{e_1} M_{e_2} &= \begin{pmatrix} \mathbb{1}_{n-d} & \Sigma_{e_1} \\ \mathbb{0}_{d,n-d} & \mathbb{1}_d \end{pmatrix} \begin{pmatrix} \mathbb{1}_{n-d} & \Sigma_{e_2} \\ \mathbb{0}_{d,n-d} & \mathbb{1}_d \end{pmatrix} \\ &= \begin{pmatrix} \mathbb{1}_{n-d} & \Sigma_{e_1} + \Sigma_{e_2} \\ \mathbb{0}_{d,n-d} & \mathbb{1}_d \end{pmatrix} \\ &= \mathbb{1}_n, \end{aligned}$$

which implies $e_1 \circ e_2 = w$, for some $w \in W_\circ$, i.e. $e_1 = e_2 \circ w = e_2 + w$. This proves that $e_1 + e_2 \in W_\circ$, which is a contradiction. □

In the light of the previous result, the following definition comes naturally.

Definition 4.4.18. A matrix $\Theta_\circ \in \mathbb{F}_{2^d}^{(n-d)\times(n-d)}$ is called the *defining matrix of an operation* $\circ$ as in Setting 3 if

$$\Theta_\circ = \begin{pmatrix} \mathbf{0} & \boldsymbol{b_{2,1}} & \cdots & \boldsymbol{b_{n-d,1}} \\ \boldsymbol{b_{2,1}} & \mathbf{0} & \cdots & \boldsymbol{b_{n-d,2}} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{b_{n-d,1}} & \boldsymbol{b_{n-d,2}} & \cdots & \mathbf{0} \end{pmatrix} \tag{4.7}$$

and no $\mathbb{F}_2$-linear combination of columns of $\Theta_\circ$ is the null vector. In this case, the operation $\circ$ is defined by letting $\Sigma_{e_i} = \Theta_\circ[\cdot, i]$ for each $1 \leq i \leq n - d$.

In the following result (see also [BCS17]) we prove a more precise lower bound for the dimension of the weak-key space, improving consequently what already proven in Theorem 4.4.2 and Theorem 4.4.5. Although the minimal hypotheses are those of Setting 1, we prove the result in the case that matter for our purposes, i.e. the one of Setting 3. The proof relies on the following lemma, whose proof may be found e.g. in [MMMM13].

Lemma 4.4.19. *There is no symmetric zero-diagonal invertible matrix of odd dimension over the field* $\mathbb{F}_2$. □

Theorem 4.4.20. *Let* $\circ$ *be an operation as in Setting 3, and let us assume* $T_\circ \neq T_+$. *Then*

$$2 - (n \mod 2) \leq \dim(W_\circ) \leq n - 2.$$

Proof. The upper bound has been proved in more general hypotheses in Theorem 4.4.5. Let us now assume n even and assume by contradiction that $d = 1$. Then, if Θ is the defining matrix of the operation, Θ is a matrix in $(\mathbb{F}_2)^{(n-1)\times(n-1)}$ which is symmetric and zero-diagonal. Notice also that the condition on the $\mathbb{F}_2$-linear combinations of columns of Θ given in Theorem 4.4.17, in the case $d = 1$ is equivalent to saying that Θ is invertible. The previous lemma leads to the desired contradiction. □

The cases of $n = 3$, $n = 4$, and other small values

The problem of counting the operations as in Setting 3 having a given size n and a given compatible weak-key space dimension d is equivalent to counting all the possible matrices like in Theorem 4.4.17, i.e. all the possible defining matrices. This task is demanding, and it has partially treated in the seminal works [CS17, BCS17]. However a complete description of these operations can be easily given if the space size is small. With an eye on considering operations having the size of classical small S-boxes, let us focus on the case $n \in \{3, 4\}$.

n=3

Example 4.4.21. Due to Theorem 4.4.20, $d = 1$ is the only value which is admissible with $n = 3$. In this case, a matrix Θ is the defining matrix of an operation with $n = 3$ and $d = 1$ if and only if $\Theta \in (\mathbb{F}_2)^{(2\times 2)}$ and it is symmetric, zero-diagonal, and non-singular. Only one matrix in $(\mathbb{F}_2)^{(2\times 2)}$ meets these requirements. Such a matrix is

$$\Theta = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

The submatrices Σ_{e_1} and Σ_{e_2} of the matrices M_{e_1} and M_{e_2} can be read in the first and in the second column of Θ, i.e. the operation is defined by the following matrices

$$M_{e_1} = \left(\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & 1 & 1 \\ \hline 0 & 0 & 1 \end{array}\right), \quad M_{e_2} = \left(\begin{array}{cc|c} 1 & 0 & 1 \\ 0 & 1 & 0 \\ \hline 0 & 0 & 1 \end{array}\right), \quad M_{e_3} = \mathbb{1}_3 \ .$$

Since it will be helpful in the remainder of the work, let us reserve the symbol $\diamond$ for the previously defined operation.

n=4

Example 4.4.22. In this case, again for Theorem 4.4.20, $d = 2$ is the only admissible value. A defining matrix is of the form

$$\begin{pmatrix} \mathbf{0} & \boldsymbol{\alpha} \\ \boldsymbol{\alpha} & \mathbf{0} \end{pmatrix},$$

where $\boldsymbol{\alpha}$ is non-null in the field $\mathbb{F}_{2^2}$. Since it can be chosen if three different ways, we obtain that three operations as in Setting 3 can be considered when $n = 4$. Those are defined by the three lists below:

$$\left[M_{e_1} = \left(\begin{array}{c|cc} \mathbb{1}_2 & 0 & 0 \\ & 0 & 1 \\ \hline \mathbb{0}_2 & \multicolumn{2}{c}{\mathbb{1}_2} \end{array}\right), \; M_{e_2} = \left(\begin{array}{c|cc} \mathbb{1}_2 & 0 & 1 \\ & 0 & 0 \\ \hline \mathbb{0}_2 & \multicolumn{2}{c}{\mathbb{1}_2} \end{array}\right), \; M_{e_3} = M_{e_4} = \mathbb{1}_4 \right],$$

$$\left[M_{e_1} = \left(\begin{array}{c|cc} \mathbb{1}_2 & 0 & 0 \\ & 1 & 0 \\ \hline \mathbb{0}_2 & \multicolumn{2}{c}{\mathbb{1}_2} \end{array}\right), \; M_{e_2} = \left(\begin{array}{c|cc} \mathbb{1}_2 & 1 & 0 \\ & 0 & 0 \\ \hline \mathbb{0}_2 & \multicolumn{2}{c}{\mathbb{1}_2} \end{array}\right), \; M_{e_3} = M_{e_4} = \mathbb{1}_4 \right],$$

$$\left[M_{e_1} = \left(\begin{array}{c|cc} \mathbb{1}_2 & 0 & 0 \\ & 1 & 1 \\ \hline \mathbb{0}_2 & \multicolumn{2}{c}{\mathbb{1}_2} \end{array}\right), \; M_{e_2} = \left(\begin{array}{c|cc} \mathbb{1}_2 & 1 & 1 \\ & 0 & 0 \\ \hline \mathbb{0}_2 & \multicolumn{2}{c}{\mathbb{1}_2} \end{array}\right), \; M_{e_3} = M_{e_4} = \mathbb{1}_4 \right].$$

More in general For $n \in \{5, 6, 7\}$ the number of the operations as in Setting 3 has been computed in [BCS17] using the software MAGMA. A summary of the results obtained is displayed in Tab. 4.1. For $n = 8$ only partial results are known.

4.4.5 Differential probabilities and key-addition layer

In this section we will study how differential probabilities may vary as a function of the operation which is used to compute differentials. In order to keep

n	d	number of operations as in Setting 3
3	1	1
4	2	3
5	1	28
	2	42
	3	7
6	2	3969
	3	462
	4	15
7	1	13888
	2	937440
	3	254968
	4	3990
	5	31
8	2	unknown
	3	unknown
	4	unknown
	5	32500
	6	63

Table 4.1: Number of operations for small values of n and d.

the notation light, from now on, by saying *an operation* we will be referring to an operation satisfying the hypotheses of Setting 3.

Let us recall that $a + b = a \circ b + a \cdot b = a \circ b \circ a \cdot b$. By definition of $\cdot$ we can say that $a \cdot b$ represents the error committed when confusing $a \circ b$ with $a + b$. This consideration justifies the following definition.

Definition 4.4.23. Let $\circ$ be an operation as in Setting 3. Let us define the *set of errors*

$$U_\circ \stackrel{\text{def}}{=} \{a \cdot b \mid \ a, b \in V\},$$

i.e. the set of all the possible dot products in V. Each element in $U_\circ$ is called an *error*.

Remark 4.4.24. As we already noticed in Remark 4.4.14, every error is a weak key. Hence $U_\circ \subseteq W_\circ$. Therefore, for each $x, y \in V$ there exists $u_{x,y} \in U_\circ$ such that

$$x + y = x \circ y + u_{x,y}, \tag{4.8}$$

with $u_{x,y} = x \cdot y = (0, \tilde{x}\Sigma_y)$. It is easy to notice that $U_\circ$ is composed of all the possible vectors $w \in W_\circ$ whose last d-components are all the possible $\mathbb{F}_2$-linear combinations of the rows of the matrices Σ_y for each $y \in V$, i.e.

$$U_\circ = \left\{ \left(\underbrace{0, \dots, 0}_{n-d}, \sum_{\xi_j \in \mathbb{F}_2} \left(\xi_j \left(\sum_{\mu_i \in \mathbb{F}_2} \mu_i \Sigma_{e_i} \right) [j, \cdot] \right) \right) \right\}.$$

Example 4.4.25. In Fig. 4.1, $+$ and the operation $\diamond$ defined in Example 4.4.21 are compared. Notice that $W_\diamond = \{0, e_3\} = \{0_\mathrm{x}, 1_\mathrm{x}\}$, and so the first two rows and columns in the tables are equal. Different entries are emphasised. Moreover, since $U_\diamond \subseteq W_\diamond$, it also holds $U_\diamond = W_\diamond$, and consequently, if $x + y \neq x \diamond y$, then from Eq. (4.8) it follows $x + y = x \diamond y + 1_\mathrm{x}$, as it can be seen in Fig. 4.1.

We can then restate Theorem 4.4.10 in the light of the newer hypotheses on $\circ$, showing that in the hypotheses of Setting 3, as expected, the output $\circ$-difference after the key-addition layer does not depend on the message x.

Theorem 4.4.26. *Let $\circ$ be an operation as in Setting 3. Then for each $x, k, \Delta \in V$*

$$(x + k) \circ ((x \circ \Delta) + k) = \Delta + k \cdot \Delta \in \Delta + U_\circ. \tag{4.9}$$

Proof. The proof comes directly from Theorem 4.4.10, where the triple product in Eq. (4.2) vanishes for Remark 4.4.14. □

Remark 4.4.27. It is worth noting here that the expected output difference after the key-addition layer, given in input a difference Δ, can be either Δ or Δ plus an error, which depends on Δ and on the key k used. Hence, the larger the number $\#U_\circ - 1$ of non-null errors, the less the effect of the key-addition layer can be controlled.

$+$	0_x	1_x	2_x	3_x	4_x	5_x	6_x	7_x
0_x	0_x	1_x	2_x	3_x	4_x	5_x	6_x	7_x
1_x	1_x	0_x	3_x	2_x	5_x	4_x	7_x	6_x
2_x	2_x	3_x	0_x	1_x	6_x	7_x	4_x	5_x
3_x	3_x	2_x	1_x	0_x	7_x	6_x	5_x	4_x
4_x	4_x	5_x	6_x	7_x	0_x	1_x	2_x	3_x
5_x	5_x	4_x	7_x	6_x	1_x	0_x	3_x	2_x
6_x	6_x	7_x	4_x	5_x	2_x	3_x	0_x	1_x
7_x	7_x	6_x	5_x	4_x	3_x	2_x	1_x	0_x

$\diamond$	0_x	1_x	2_x	3_x	4_x	5_x	6_x	7_x
0_x	0_x	1_x	2_x	3_x	4_x	5_x	6_x	7_x
1_x	1_x	0_x	3_x	2_x	5_x	4_x	7_x	6_x
2_x	2_x	3_x	0_x	1_x	7_x	6_x	5_x	4_x
3_x	3_x	2_x	1_x	0_x	6_x	7_x	4_x	5_x
4_x	4_x	5_x	7_x	6_x	0_x	1_x	3_x	2_x
5_x	5_x	4_x	6_x	7_x	1_x	0_x	2_x	3_x
6_x	6_x	7_x	5_x	4_x	3_x	2_x	0_x	1_x
7_x	7_x	6_x	4_x	5_x	2_x	3_x	1_x	0_x

Figure 4.1: Comparison between operation $+$ and $\diamond$

In order to keep track of the bias introduced by the key-addition layer with respect to $\circ$-differences, let us store in a table all the information which are required to predict how $\circ$-differences pass through the key-addition layer. While the DDT is a bi-dimensional table based on the confusion layer, gathering information on the number of messages following a differential, Theorem 4.4.26 states that we can construct an equivalent table, for operations in Setting 3, gathering information on the number of keys following a differential.

Definition 4.4.28. Let $\circ$ be an operation as in Setting 3. The *key distribution table* (KDT) *of* $\circ$ is the integer table $\in \mathbb{Z}^{n\times n}$ where

$$\mathrm{KDT}^{\circ}[\Delta_I, \Delta_O] \stackrel{\text{def}}{=} \#\{k \in (\mathbb{F}_2)^n \mid \Delta_I + k \cdot \Delta_I = \Delta_O\}.$$

The key distribution table can be read in the following way: whenever the input difference is in the weak-key space, i.e. $\Delta_I \in W_\circ$, then, no matter the key considered, the output difference after the key-addition layer is Δ_I with probability 1. This is because, if $\Delta_I \in W_\circ$, then

$$(x + k) \circ ((x \circ \Delta_I) + k) = \Delta_I + k \cdot \Delta_I = \Delta_I$$

for each k, since the error corresponding to k and Δ_I is always null. If $\Delta_I \notin W_\circ$, the output difference equals $\Delta_I + k \cdot \Delta_I$. The error may be zero, leading to the output difference Δ_I (this is always the case e.g. when $k \in W_\circ$), or may be different from zero, leading to $\Delta_I + u$ for some $u \in U_\circ$.

Example 4.4.29. We computed the key distribution table of the operation $\diamond$ defined in Example 4.4.21. Since, as shown in Example 4.4.25, $U_\diamond = W_\diamond = \{0_\mathrm{x}, 1_\mathrm{x}\}$, then considering e.g. the $\diamond$-difference $\Delta_I = 2_\mathrm{x}$, the $\diamond$-difference after key-addition layer may be either $\Delta_O = 2_\mathrm{x} = 2_\mathrm{x} + 0_\mathrm{x}$ or $\Delta_O = 3_\mathrm{x} = 2_\mathrm{x} + 1_\mathrm{x}$. Each event happens with probability $1/2$, as it can be noticed in Fig. 4.2.

Example 4.4.30. Let $n = 5$, $d = 2$, and let us consider the operation $\circ$ having the following defining matrix

$$\Theta = \begin{pmatrix} \mathbf{0} & \boldsymbol{a} & \boldsymbol{b} \\ \boldsymbol{a} & \mathbf{0} & \boldsymbol{a} \\ \boldsymbol{b} & \boldsymbol{a} & \mathbf{0} \end{pmatrix},$$

where $\boldsymbol{a} = (1, 1)$ and $\boldsymbol{b} = (1, 0)$. The table KDT° is displayed in Fig. 4.3.

	0_x	1_x	2_x	3_x	4_x	5_x	6_x	7_x
0_x	8	·	·	·	·	·	·	·
1_x	·	8	·	·	·	·	·	·
2_x	·	·	4	4	·	·	·	·
3_x	·	·	4	4	·	·	·	·
4_x	·	·	·	·	4	4	·	·
5_x	·	·	·	·	4	4	·	·
6_x	·	·	·	·	·	·	4	4
7_x	·	·	·	·	·	·	4	4

Figure 4.2: Key distribution table of $\diamond$

The tables displayed in the two previous examples are symmetric. This is a general rule, as showed in the following result.

Theorem 4.4.31. *For each operation $\circ$ as in Setting 3, the table* KDT° *is symmetric.*

Proof. Let us fix $\Delta_1, \Delta_2, k \in V$ and suppose that $\Delta_1 + k \cdot \Delta_1 = \Delta_2$. Then

$$\begin{aligned} \Delta_2 + k \cdot \Delta_2 &= \Delta_1 + k \cdot \Delta_1 + k \cdot (\Delta_1 + k \cdot \Delta_1) \\ &= \Delta_1 + k \cdot \Delta_1 + k \cdot \Delta_1 + k \cdot k \cdot \Delta_1 \\ &= \Delta_1, \end{aligned}$$

therefore $\mathrm{KDT}[\Delta_1, \Delta_2] = \mathrm{KDT}[\Delta_2, \Delta_1]$. □

Notice that the symmetry of the key-distribution table is based on the fact that triple products are always null, i.e. it crucially depends on the hypotheses of Setting 3.

It is worth noting that the number of non-zero entries in a key distribution table is an important feature in terms of understanding differential weaknesses of the cipher. Indeed, the less they are, the more the key-addition layer effect on $\circ$-differences can be controlled. We will consider in the next chapter a particular case which seems the most convenient for this purpose. In order to compute this number in general, the following result is helpful.

	0_x	1_x	2_x	3_x	4_x	5_x	6_x	7_x	8_x	9_x	A_x	B_x	C_x	D_x	E_x	F_x	10_x	11_x	12_x	13_x	14_x	15_x	16_x	17_x	18_x	19_x	$1A_x$	$1B_x$	$1C_x$	$1D_x$	$1E_x$	$1F_x$
0_x	32	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
1_x	·	32	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
2_x	·	·	32	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
3_x	·	·	·	32	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
4_x	·	·	·	·	8	8	8	8	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
5_x	·	·	·	·	8	8	8	8	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
6_x	·	·	·	·	8	8	8	8	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
7_x	·	·	·	·	8	8	8	8	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
8_x	·	·	·	·	·	·	·	·	16	·	·	16	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
9_x	·	·	·	·	·	·	·	·	·	16	16	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
A_x	·	·	·	·	·	·	·	·	·	16	16	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
B_x	·	·	·	·	·	·	·	·	16	·	·	16	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
C_x	·	·	·	·	·	·	·	·	·	·	·	·	8	8	8	8	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
D_x	·	·	·	·	·	·	·	·	·	·	·	·	8	8	8	8	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
E_x	·	·	·	·	·	·	·	·	·	·	·	·	8	8	8	8	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
F_x	·	·	·	·	·	·	·	·	·	·	·	·	8	8	8	8	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
10_x	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	8	8	8	8	·	·	·	·	·	·	·	·	·	·	·	·
11_x	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	8	8	8	8	·	·	·	·	·	·	·	·	·	·	·	·
12_x	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	8	8	8	8	·	·	·	·	·	·	·	·	·	·	·	·
13_x	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	8	8	8	8	·	·	·	·	·	·	·	·	·	·	·	·
14_x	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	16	·	16	·	·	·	·	·	·	·	·	·
15_x	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	16	·	16	·	·	·	·	·	·	·	·
16_x	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	16	·	16	·	·	·	·	·	·	·	·	·
17_x	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	16	·	16	·	·	·	·	·	·	·	·
18_x	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	8	8	8	8	·	·	·	·
19_x	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	8	8	8	8	·	·	·	·
$1A_x$	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	8	8	8	8	·	·	·	·
$1B_x$	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	8	8	8	8	·	·	·	·
$1C_x$	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	16	16	·	·
$1D_x$	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	16	16	·	·
$1E_x$	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	16	16
$1F_x$	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	16	16

Figure 4.3: Key distribution table for the 5-bit operation defined in Example 4.4.30

Theorem 4.4.32. *Let $\circ$ be an operation as in Setting 3. For each $a \in V$ it holds* $\mathrm{Rank}(\Sigma_a) \leq \min(n-d-1, d)$.

Proof. If $a \in W_\circ$ there is nothing to prove. If not, then $a = (\bar{a}, \tilde{a})$, with $\bar{a} \neq 0$. Since $0 = a \circ a = aM_a + a$, it follows $a(M_a + \mathbb{1}_n) = 0$. This implies, from the description of M_a given in Theorem 4.4.13, that

$$a \in \mathrm{Ker} \begin{pmatrix} 0 & \Sigma_a \\ 0 & 0 \end{pmatrix}.$$

Therefore $\bar{a} \in \mathrm{Ker}(\Sigma_a)$. From this it follows

$$\mathrm{Rank}(\Sigma_a) = \dim(\mathrm{Im}(\Sigma_a)) = n - d - \dim(\mathrm{Ker}(\Sigma_a)) \leq n - d - 1.$$

Now, if $n-d-1 \leq d$, then the result holds. If not, then $d < n-d-1 < n-d$, whence $\mathrm{Rank}(\Sigma_a) \leq d = \min(n-d-1, d)$. □

Theorem 4.4.33. *Let $\circ$ be an operation as in Setting 3. The number of non-zero entries in each row of the key distribution table* KDT° *is upper bounded by* $2^{\min(n-d-1,d)}$.

Proof. Given a fixed $\Delta \in V$, the number of non-zero entries in the row $\mathrm{KDT}[\Delta, \cdot]$ depends on the values of $k \cdot \Delta$, for each $k \in V$. Since $k \cdot \Delta = \bar{k}\Sigma_\Delta \in \mathrm{Im}(\Sigma_\Delta)$, and $\dim(\mathrm{Im}(\Sigma_\Delta)) = \mathrm{Rank}(\Sigma_\Delta) \leq \min(n-d-1, d)$ for Theorem 4.4.32, the desired holds. □

Remark 4.4.34. Notice that the upper bound reaches its minimum value, i.e. 2, if $d = 1$ or $d = n-2$.

Corollary 4.4.35. *Let $\circ$ be an operation as in Setting 3 and let $\Delta_I \in V$. Then, for each $\Delta_O \in V$ it holds*

$$\mathrm{KDT}^\circ[\Delta_I, \Delta_O] \in \left\{0, 2^{n-\mathrm{Rank}\left(\Sigma_{\Delta_I}\right)}\right\}.$$

Proof. Let $\Delta_O \in V$ such that $\mathrm{KDT}[\Delta_I, \Delta_O] \neq 0$. Two keys $k_1, k_2 \in V$ are such that $k_1 \cdot \Delta_I = k_2 \cdot \Delta_I = \Delta_O$ if $\bar{k_1}$ and $\bar{k_2}$ are in the same class modulo $\mathrm{Ker}(\Sigma_{\Delta_I})$. Recalling that the value of $k \cdot \Delta$ does not depend on the last d

bits of k, then $\mathrm{KDT}^{\circ}[\Delta_I, \Delta_O]$ is the number of elements contained in each class modulo $\mathrm{Ker}(\Sigma_{\Delta_I})$ multiplied by 2^d. Therefore

$$\begin{aligned}\mathrm{KDT}^{\circ}[\Delta_I, \Delta_O] &= 2^d\, 2^{\dim\left(\mathrm{Ker}\left(\Sigma_{\Delta_I}\right)\right)} \\ &= 2^d\, 2^{n-d-\mathrm{Rank}\left(\Sigma_{\Delta_I}\right)} \\ &= 2^{n-\mathrm{Rank}\left(\Sigma_{\Delta_I}\right)}.\end{aligned}$$

□

Example 4.4.36. Notice that, in general, the set of errors is not a vector space. Indeed, let us consider the following operation having $n = 8$ and $d = 4$ and defined by the defining matrix

$$\Theta = \begin{pmatrix} (0,0,0,0) & (0,0,0,1) & (1,0,0,1) & (1,0,1,1) \\ (0,0,0,1) & (0,0,0,0) & (1,1,1,1) & (0,1,1,1) \\ (1,0,0,1) & (1,1,1,1) & (0,0,0,0) & (1,0,0,1) \\ (1,0,1,1) & (0,1,1,1) & (1,0,0,1) & (0,0,0,0) \end{pmatrix}.$$

Computing all the possible dot products, one can notice that $\#U_{\circ} = 15$, hence $U_{\circ}$ is not a vector space. However, $n = 8$ is the first value for which this happens.

4.5 Interaction with the confusion layer

While in classical differential cryptanalysis differential probabilities are only induced by the confusion layer, in the previous section we illustrate that, with new operations, probabilities are also added by the key-addition layer. For the probability of a $\circ$-differential to be larger than the probability of a $+$-differential, we should either have trails with larger probabilities and / or more trails. The first goal can only be achieved if the values in the DDT of the S-box computed with respect to $\circ$ are larger that those in the classical DDT computed with respect to the XOR. In the following section we show through several examples that this is possible. Before doing so, let us notice that the difference distribution table of a vectorial Boolean function can be defined more in general, considering differences induced by whatever operation whose group of translations is elementary, abelian, and regular (XOR

included).

Let $\circ$ be an operation as in Setting 1. Proceeding in the same way of Section 2.4.1, the *derivative of a vectorial Boolean function* $f : (\mathbb{F}_2)^n \to (\mathbb{F}_2)^n$ *in the direction* $u \in (\mathbb{F}_2)^n$ *with respect to the operation* $\circ$, denoted by $\partial_u^\circ f$ is the function

$$\begin{array}{rccc} \partial_u^\circ f : & (\mathbb{F}_2)^n & \to & (\mathbb{F}_2)^n \\ & x & \mapsto & xf \circ (x \circ u)f. \end{array}$$

Then, the *difference distribution table* (DDT°) *of* f *with respect to the operation* $\circ$ is the integer table $\mathrm{DDT}_f^\circ \in \mathbb{Z}^{n \times n}$ defined for $u \in (\mathbb{F}_2)^n$ and $v \in (\mathbb{F}_2)^n$ as

$$\mathrm{DDT}_f^\circ[u, v] \stackrel{\text{def}}{=} \delta_f^\circ(u, v),$$

where

$$\delta_f^\circ(u, v) \stackrel{\text{def}}{=} \#\{x \in (\mathbb{F}_2)^n \mid x\, \partial_u^\circ f = v\}.$$

The *differential uniformity of* f *with respect to* $\circ$ is defined as

$$\delta^\circ(f) \stackrel{\text{def}}{=} \max_{\substack{u,v \\ u \neq 0}} \mathrm{DDT}_f^\circ[u, v],$$

and the function f is said to be δ*-differentially uniform with respect to* $\circ$ if $\delta = \delta^\circ(f)$.

In the following examples it is shown how the non-linearity of vectorial Boolean functions used as S-boxes for famous block ciphers may change when differentials are computed with respect to another operation.

Example 4.5.1. Let us consider the following S-box $\gamma' : (\mathbb{F}_2)^3 \to (\mathbb{F}_2)^3$ which is affinely equivalent to the power function $x \mapsto x^3$:

x	0_x	1_x	2_x	3_x	4_x	5_x	6_x	7_x
$x\,\gamma'$	0_x	6_x	2_x	1_x	5_x	7_x	4_x	3_x

.

We have computed the difference distribution table of γ' with respect to $+$ and to the operation $\diamond$ defined in Example 4.4.21. The given function is well-known to be APN with respect to $+$ [Nyb93]. However, this property does not hold when looking at $\diamond$-differences. As can be noticed in Fig. 4.4, where the difference distribution tables of γ' with respect to $+$ and $\diamond$ respectively are displayed, γ' is 8-differentially uniform with respect to $\diamond$.

$+$	0_x	1_x	2_x	3_x	4_x	5_x	6_x	7_x
0_x	8	·	·	·	·	·	·	·
1_x	·	·	2	2	·	·	2	2
2_x	·	2	2	·	2	·	·	2
3_x	·	2	·	2	2	·	2	·
4_x	·	2	2	·	·	2	2	·
5_x	·	2	·	2	·	2	·	2
6_x	·	·	·	·	2	2	2	2
7_x	·	·	2	2	2	2	·	·

$\diamond$	0_x	1_x	2_x	3_x	4_x	5_x	6_x	7_x
0_x	8	·	·	·	·	·	·	·
1_x	·	·	·	4	·	·	4	·
2_x	·	·	4	·	·	·	·	4
3_x	·	4	·	·	·	4	·	·
4_x	·	4	·	·	·	4	·	·
5_x	·	·	4	·	·	·	·	4
6_x	·	·	·	·	8	·	·	·
7_x	·	·	·	4	·	·	4	·

Figure 4.4: Difference distribution table of $\gamma' : x \mapsto x^3$ over $(\mathbb{F}_2)^3$ defined in Example 4.5.1 with respect to $+$ and $\diamond$

Example 4.5.2. The S-box `S2` of the cipher SERPENT [BAK98] is 4-differentially uniform with respect to $+$. However, if differentials are computed with respect to the operation $\circ$ defined in Example 4.4.22 obtained considering $\boldsymbol{\alpha} = (0, 1)$, we obtain that `S2` is a 10-differentially uniform function with respect to the given operation $\circ$. The complete DDTs of `S2` with respect to $+$ and $\circ$ are displayed in Fig. 4.5.

In general, the following bound holds.

Theorem 4.5.3. *Let $\circ$ an operation as in Setting 3. Let $\gamma : (\mathbb{F}_2)^n \to (\mathbb{F}_2)^n$ a vectorial Boolean function and let us assume that γ is δ^+-differentially uniform and δ°-differentially uniform with respect to $+$ and $\circ$, respectively. Then we have the following relation between δ^+ and δ°:*

$$\delta^\circ \leq \min\left(\delta^+ \left(\#U_\circ\right)^2, 2^n\right).$$

In particular for $\Delta_I \neq 0$, we have

$$\max_{\Delta_O} \mathrm{DDT}^\circ_\gamma[\Delta_I, \Delta_O] \leq \begin{cases} \min\left(\delta^+ \#U_\circ, 2^n\right) & \Delta_I \in W_\circ \\ \min\left(\delta^+ \left(\#U_\circ\right)^2, 2^n\right) & \Delta_I \notin W_\circ. \end{cases}$$

Proof. Let $\Delta_I, \Delta_O \in V$. We are interested in the number of solutions of the following equation:

$$x\gamma \circ (x \circ \Delta_I)\gamma = \Delta_O. \tag{4.10}$$

Rewriting Eq. (4.10) introducing the dot products, we obtain

$$x\gamma + (x + \Delta_I + u)\gamma = \Delta_O + w, \tag{4.11}$$

where $u = u_{x,\Delta_I} = x \cdot \Delta_I$, $w = w_{x,\Delta_I} = x\gamma \cdot (x \circ \Delta_I)\gamma$, and $u, w \in U_\circ$. For any possible choice of $u, w \in U_\circ$, Eq. (4.11) admits at most δ solutions, hence $\delta^\circ_\gamma(\Delta_I, \Delta_O) \leq \delta^+ \left(\#U_\circ\right)^2$ In particular, if $\Delta_I \in W_\circ$, then $u = 0$, and consequently we derive $\delta^\circ_\gamma(\Delta_I, \Delta_O) \leq \delta^+ \#U_\circ$. The proof is then concluded. □

The bound of the previous result can be quite loose though, especially when $\#U_\circ$ is large or when $\#U_\circ$ in not a subspace. However, in the case described in the following section, and more in general for small values of n

+	0_x	1_x	2_x	3_x	4_x	5_x	6_x	7_x	8_x	9_x	A_x	B_x	C_x	D_x	E_x	F_x
0_x	16	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
1_x	·	·	·	·	·	2	·	2	·	·	2	2	2	·	4	2
2_x	·	·	·	4	·	4	·	·	·	4	·	·	·	·	·	4
3_x	·	4	2	·	·	·	2	·	·	2	·	·	2	·	2	2
4_x	·	·	·	·	·	·	4	·	·	·	4	4	·	4	·	·
5_x	·	4	·	2	2	2	2	·	2	·	·	·	2	·	·	·
6_x	·	·	2	2	2	2	·	·	2	2	·	·	·	·	2	2
7_x	·	·	·	·	4	2	·	2	·	·	2	2	2	·	·	2
8_x	·	·	·	2	·	2	·	4	·	2	·	·	·	4	·	2
9_x	·	·	·	2	·	·	·	2	4	2	2	2	2	·	·	·
A_x	·	·	2	·	2	·	4	·	2	·	4	·	·	·	2	·
B_x	·	4	·	·	2	·	2	·	2	2	·	·	2	·	·	2
C_x	·	·	2	·	2	·	·	·	2	·	·	4	·	4	2	·
D_x	·	4	2	2	·	2	2	·	·	·	·	·	2	·	2	·
E_x	·	·	2	·	2	·	·	4	2	·	·	·	·	4	2	·
F_x	·	·	4	2	·	·	·	2	·	2	2	2	2	·	·	·

∘	0_x	1_x	2_x	3_x	4_x	5_x	6_x	7_x	8_x	9_x	A_x	B_x	C_x	D_x	E_x	F_x
0_x	16	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
1_x	·	·	·	·	2	·	·	2	·	·	·	4	2	·	·	6
2_x	·	·	·	4	·	4	·	·	·	4	·	·	·	·	4	·
3_x	·	4	2	·	·	·	·	2	2	·	·	·	2	·	·	4
4_x	·	4	·	·	·	·	·	4	·	·	·	4	4	·	·	·
5_x	·	·	·	2	·	4	2	·	·	2	4	·	·	2	·	·
6_x	·	·	2	·	10	·	·	·	2	·	·	·	·	·	·	2
7_x	·	·	·	2	·	·	2	·	·	2	4	·	·	2	4	·
8_x	·	·	·	·	2	·	·	2	6	·	·	2	4	·	·	·
9_x	·	·	·	4	·	·	4	·	·	2	2	·	·	2	2	·
A_x	·	4	2	·	·	·	·	2	4	·	·	2	·	·	·	2
B_x	·	·	·	·	·	4	4	·	·	2	2	·	·	2	2	·
C_x	·	4	4	·	·	·	·	·	2	·	·	2	2	·	·	2
D_x	·	·	·	2	·	4	2	·	·	·	2	·	·	4	2	·
E_x	·	·	6	·	2	·	·	4	·	·	·	2	2	·	·	·
F_x	·	·	·	2	·	·	2	·	·	4	2	·	·	4	2	·

Figure 4.5: Difference distribution table of the S-box S2 of SERPENT with respect to + and ∘

compatible with the dimension of a real-life S-box, it is tight in most of the cases.

Although we did not reach a deep knowledge in the interaction between a general confusion layer and any operation $\circ$, since it depends on the complex interaction between the two additive operations on the vector space $+$ and $\circ$ and the finite field addition and multiplication, we achieved some partial results which can be helpful in order to understand, for example, how we obtained the S-box of Example 4.5.1.

4.5.1 On the cubic function in odd dimension

Let us recall that, when interpreting $x \in \mathbb{F}_{2^n}$ as a vector in $(\mathbb{F}_2)^n$ and, vice versa, a vector $x \in (\mathbb{F}_2)^n$ as an element of the finite field $\mathbb{F}_{2^n}$, we are tacitly fixing a bijective correspondence $\Psi : (\mathbb{F}_2)^n \to \mathbb{F}_{2^n}$. For example, when considering $x \in (\mathbb{F}_2)^n$ and writing e.g. $\mathrm{Tr}(x)$, we mean $(\mathrm{Tr}(x\Psi))\,\Psi^{-1}$, where Tr denote the *trace* over $\mathbb{F}_2$. Analogously, if $y \in (\mathbb{F}_2)^n$ such that $y \neq 0$, then x/y denotes $\left(x\Psi\,(y\Psi)^{-1}\right)\Psi^{-1}$. The aim of this section is to illustrate that, depending on the way we identify $(\mathbb{F}_2)^n$ and $\mathbb{F}_{2^n}$, we can transform an APN permutation into a permutation with higher differential uniformity with respect to another operation. For this reason, let us assume n odd and let us study the power function $x \mapsto x^3$ on $\mathbb{F}_{2^n}$. Recall that such a map is a permutation when n is odd and it is APN with respect to the XOR [Nyb93].

Following the classical proof used to derive the differential properties of the cubic function, we aim at determining the number of solutions of the equation

$$x^3 \circ (x \circ \Delta_I)^3 = \Delta_O \tag{4.12}$$

for each $\Delta_I, \Delta_O \in (\mathbb{F}_2)^n$ and for each operation $\circ$. As already said, the interaction between all the operations defined is complex, hence let us focus on the case which matters more for our purposes (see Remark 4.4.34), i.e. the case of operations $\circ$ such that $d = n-2$. Recall that for such operations, by Remark 4.4.24, it holds $\#U_\circ = 2$. Let us only focus on the case where Δ_I is the non-null error. The next result shows that if $U_\circ$ is the 1-dimensional

space spanned by Δ_I and Eq. (4.12) admits at least four solution for some Δ_O (which at the same time means that $x \mapsto x^3$ is δ-differentially uniform with respect to $\circ$, with $\delta \geq 4$), then $\mathrm{Tr}(1/\Delta_I) = 0$. Also note that according to Theorem 4.5.3, $\max_{\Delta_O} \mathrm{DDT}^{\circ}[\Delta_I, \Delta_O] \leq 4$ in the case where $\Delta_I \in U_{\circ}$, $\Delta_I \neq 0$.

Proposition 4.5.4. *Let $\circ$ be an operation as in Setting 3 and such that $d = n - 2$. Let $\Delta_I \in V \setminus \{0\}$ such that $U_{\circ} = \mathrm{Span}\{\Delta_I\}$. If*

$$\max_{\Delta_O} \mathrm{DDT}^{\circ}[\Delta_I, \Delta_O] > 2,$$

then $\mathrm{Tr}(1/\Delta_I) = 0$. Equivalently, if $\mathrm{Tr}(1/\Delta_I) = 1$, then the equation $x^3 \circ (x \circ \Delta_I)^3 = \Delta_O$ admits at most 2 solutions for each Δ_O.

Proof. Since $U_{\circ} \subseteq W_{\circ}$, Δ_I is also a weak vector, hence Eq. (4.12) can be written as

$$x^3 + (x + \Delta_I)^3 = \Delta_O + w, \tag{4.13}$$

where $w = \varepsilon_{x^3,(x+\Delta_I)^3} \in W_{\circ}$, according to Remark 4.4.24. Moreover, notice that when Eq. (4.13) holds, then

$$\left(\frac{x}{\Delta_I}\right)^2 + \frac{x}{\Delta_I} = 1 + \frac{\Delta_O + w}{\Delta_I^3},$$

which, by means of a change of variable, can be written as

$$x^2 + x = 1 + \frac{\Delta_O + w}{\Delta_I^3}.$$

This implies, since n is odd, that

$$\mathrm{Tr}\left(\frac{\Delta_O + w}{\Delta_I^3}\right) = 1. \tag{4.14}$$

Let us now assume that Eq. (4.13) admits four solutions $x_1, x_2, x_1 + \Delta_I$ and $x_2 + \Delta_I$. This means that there exist $w_1, w_2 \in U_{\circ}$ such that $x_1^3 + (x_1 + \Delta_I)^3 = \Delta_O + w_1$ and $x_2^3 + (x_2 + \Delta_I)^3 = \Delta_O + w_2$. Notice that $w_1 \neq w_2$. Indeed, if we assume $w_1 = w_2$, then Eq. (4.13), and consequently Eq. (4.12), admits four solutions, running counter to the fact that $x \mapsto x^3$ is APN. Since $w_i \in U_{\circ}$,

we can assume without loss of generality that $w_1 = 0$ and $w_2 = \Delta_I$. From Eq. (4.14) we obtain

$$\mathrm{Tr}\left(\frac{\Delta_O + w_1}{\Delta_I^3}\right) = \mathrm{Tr}\left(\frac{\Delta_O + w_2}{\Delta_I^3}\right) = 1,$$

which implies

$$\mathrm{Tr}\left(\frac{\Delta_O}{\Delta_I^3}\right) = 1 \text{ and } \mathrm{Tr}\left(\frac{\Delta_I}{\Delta_I^3}\right) = \mathrm{Tr}\left(\frac{1}{\Delta_I^2}\right) = \mathrm{Tr}\left(\frac{1}{\Delta_I}\right) = 0.$$

□

We proved that, if we want Eq.(4.13) to admit at least four solutions, which would mean that $x \mapsto x^3$ is at least 4-differentially uniform with respect to an operation $\circ$ such that $d = n - 2$, then we have to choose Ψ in such a way that $\mathrm{Tr}\left((u\Psi)^{-1}\right) = 0$, where u is the non-null error in $U_\circ$. Notice that this argument can be generalised to other APN quadratic functions and illustrate that differential uniformity with respect to another operation is influenced by the chosen correspondence Ψ.

Remark 4.5.5. We can also show that in the case where $U_\circ = \{0, \alpha\}$ and $\Delta_I \notin W_\circ$, some necessary conditions for 8-differential uniformity are

$$\mathrm{Tr}\left(\frac{\alpha}{\Delta_I^3}\right) = \mathrm{Tr}\left(\frac{\alpha}{(\Delta_I + \alpha)^3}\right) = 0.$$

Indeed, proceeding as in Proposition 4.5.4, in this case we obtain that x is a solution of Eq.(4.12) if and only if

$$\left(\frac{x}{\Delta_I}\right)^2 + \frac{x}{\Delta_I} = 1 + \frac{\Delta_O + w}{(\Delta_I + u)^3},$$

where $u = u_{x,\Delta_I}$ and $w = w_{x^3,(x+\Delta_I)^3}$, which implies

$$\mathrm{Tr}\left(\frac{\Delta_O + w}{(\Delta_I + u)^3}\right) = 1. \tag{4.15}$$

Reasoning as in Proposition 4.5.4, if Eq.(4.12) admits eight solution of the kind $x_i, x_i + \Delta_I$ for $1 \leq i \leq 4$, then the corresponding $(u_i, w_i) \in \{0, \alpha\}^2$ must

be all different. The four cases are given in the following table:

Case	u	w	Condition
1	0	0	$\mathrm{Tr}\left(\frac{\Delta_O}{\Delta_I^3}\right) = 1$
2	0	α	$\mathrm{Tr}\left(\frac{\Delta_O + \alpha}{(\Delta_I + \alpha)^3}\right) = 1$
3	α	0	$\mathrm{Tr}\left(\frac{\Delta_O}{(\Delta_I + \alpha)^3}\right) = 1$
4	α	α	$\mathrm{Tr}\left(\frac{(\Delta_O + \alpha)}{(\Delta_I + \alpha)^3}\right) = 1$

If the conditions in the four cases of the table are fulfilled, combining together Case 1 and Case 2, and Case 3 and Case 4, we obtain some necessary conditions for $\max_{\Delta_O} \mathrm{DDT}^{\circ}[\Delta_I, \Delta_O] = 8$:

$$\mathrm{Tr}\left(\frac{\alpha}{\Delta_I^3}\right) = \mathrm{Tr}\left(\frac{\alpha}{(\Delta_I + \alpha)^3}\right) = 0.$$

The method described above has been used to build the S-box γ' of Example 4.5.1, where we choose Ψ with the property $\mathrm{Tr}\left((e_3\Psi)^{-1}\right) = 0$. Moreover, in this case, we computationally proved, using MAGMA, the following result on $x \mapsto x^3$ on $(\mathbb{F}_2)^3$.

Proposition 4.5.6. *The function $x \mapsto x^3$ on $(\mathbb{F}_2)^3$ is APN with respect to $\diamond$ if and only if Ψ is such that $\mathrm{Tr}\left((e_3\Psi)^{-1}\right) = 1$ and it is 8-differentially uniform with respect to $\diamond$ in the other case.* □

In this chapter we have defined the hypotheses which make operations coming from alternative group of translations suitable for a differential attack. As recalled in Section 1.2.1, three different layers define the round functions of a SPN block cipher. In this chapter, we provided properties selecting a subset of operations with a particular behavior with respect to the key-addition and the confusion layer, considered singularly. We have noticed that the magnitude of differential probabilities with respect to a different additive law on the message space may depend on a wise choice of the parameters n and d which define the operation. In particular, in Theorem 4.5.3 we proved that, if the size of $U_\circ$ is too small, we have less chances to significantly weaken the non-linearity of the Boolean function considered.

Even if the size $U_\circ$ depends on the relation between n and d, it is easy to prove that $U_\circ$ is small when $d \sim n-2$ or $d \sim 1$, i.e. when d is closer to its lower or upper bound (see Theorem 4.4.5). On the other hand, by virtue of Theorem 4.4.33, the cases $d \sim n-2$ or $d \sim 1$ look the most convenient since they allow $\circ$-differences to pass through the key-addition layer with the highest probabilities. It may seems that a choice of n and d which is good for the confusion layer is bad for the key-addition layer, and vice versa. This should not surprise anyone: when $n \sim d$, almost every vector is weak, i.e. the operation $+$ and $\circ$ are similar. In this case, since the confusion layer is usually designed to maximise the non-linearity with respect to the XOR, we cannot expect from an operation similar to $+$ to induce differential probabilities significantly different than the one induced by the XOR. However, all the keys which are weak pass unharmed through the key-addition layer. The same holds in the case $d \sim 1$. Indeed, even if few vectors are weak, the size of $U_\circ$ is small, hence no advantages in terms of non-linearity are obtained. In the next chapter we will show how to find the correct balance between n and d. Even if the previous considerations are correct, we are not taking into account yet the hypothesis that the confusion layer is a parallel map. Moreover, the choice of the parameter is also conditioned by another crucial point which has been so far completely neglected: the impact of the diffusion layer on $\circ$-differential probabilities. Indeed, to define an SPN with an *a priori* weaker resistance to $\circ$-differential attack than $+$-differential attack, we should also select an appropriate diffusion layer. The next chapter focuses on these aspects. In particular, in Section 5.1, we present some general properties of the diffusion layer.

5 Designing a cipher

Our goal in this section is to first design an operation $\circ$, and then a cipher that can be attacked using $\circ$-differentials. The attack can be successful if the operation interacts with the layers of the chosen cipher in such a way that $\circ$ induces differential probabilities which are high enough to allow a distinguishing attack. However, if we want to show that our attack is meaningful, we have also to show that such an attack fails it the attacker tries to use ordinary differentials. Before putting things together, the way the diffusion layer impacts on $\circ$-differential probabilities requires a careful study.

5.1 Interaction with the diffusion layer

In the previous chapter we refined our assumptions on operations $\circ$ in order to obtain a higher differential uniformity of the S-boxes with respect to $\circ$ and possibly high probabilities induced by the key-addition layer. However, we did not consider so far how $\circ$-differences propagate through the diffusion layer, which is, in our model, a $+$-linear map. Notice that the role of the diffusion layer, in the sense of keeping the cipher safe from differential attacks, is to spread the differences as fast and as far as possible in the block, i.e. to quickly activate as many S-boxes as possible. However, the diffusion layer does not have a direct role in terms of differential probability when differentials are computed with respect to the XOR, since it is a XOR-linear map, and consequently each $+$-differential is deterministic over the diffusion

layer. On the other hand though, in the case of $\circ$-differentials, a cryptanalyst willing to predict the output difference of the diffusion layer λ, given the input $\circ$-difference Δ, needs to determine the distribution of the elements of the kind of

$$x\lambda \circ (x \circ \Delta)\lambda. \tag{5.1}$$

knowing that, in general, λ is not linear with respect to $\circ$. The cryptanalyst has to take into consideration that the n-bit linear map is then a huge $\circ$-non-linear function with 2^n inputs, which will make further analysis non-trivial. Indeed, writing Eq. (5.1) in terms of the dot product, we obtain

$$x\lambda \circ (x \circ \Delta)\lambda = \Delta\lambda + (x \cdot \Delta)\lambda + x\lambda \cdot \Delta\lambda + x\lambda \cdot (x \cdot \Delta)\lambda,$$

which clearly depends on $x \in (\mathbb{F}_2)^n$. This should be enough to convince the reader that a successful attack with respect to an alternative operation $\circ$ may rely on the linearity of the diffusion layer also with respect to $\circ$.

5.1.1 Compatible diffusion layers

For the reason explained above, from now on we focus on operations $\circ$ and on ciphers whose diffusion layer is an invertible matrix which is also $\circ$-linear. Let us give a name to these functions.

Definition 5.1.1. Let $\circ$ be an operation as in Setting 3. We denote the group of the homomorphisms of $(V, +, \cdot)$ by $H_\circ \stackrel{\text{def}}{=} \mathrm{Hom}(V, +, \cdot)$. A matrix $\lambda \in (\mathbb{F}_2)^{n\times n}$ is said to be *compatible* with the operation $\circ$ if $\lambda \in H_\circ$. Equivalently, when λ is compatible with $\circ$, we also say that $\circ$ is compatible with λ.

Is not hard to prove that every map $\lambda \in H_\circ$ is linear with respect to $+$ and $\circ$, i.e. $H_\circ = \mathrm{GL}(V, +) \cap \mathrm{GL}(V, \circ)$.

Proposition 5.1.2. *Let $\circ$ be an operation as in Setting 3 and let $\lambda \in (\mathbb{F}_2)^{n\times n}$ be compatible with $\circ$. Then $\lambda \in \mathrm{GL}(V, \circ)$.*

Proof. Let $a, b \in V$. Then $(a \circ b)\lambda = a\lambda + b\lambda + (a \cdot b)\lambda$. Moreover, since λ is compatible with $\circ$, $(a \cdot b)\lambda = a\lambda \cdot b\lambda$, hence it follows

$$(a \circ b)\lambda = a\lambda + b\lambda + a\lambda \cdot b\lambda = a\lambda \circ b\lambda.$$

□

The following result is useful to give a description of $H_\circ$.

Theorem 5.1.3. *Let $\circ$ be an operation as in Setting 3, and let $W_\circ$ and $U_\circ$ be the weak-keys subspace and the set of errors respectively. Then, for each $\lambda \in H_\circ$ it holds $W_\circ\lambda = W_\circ$ and $U_\circ\lambda = U_\circ$.*

Proof. Let $\lambda \in H_\circ$ and let us prove that $W_\circ\lambda = W_\circ$. Let $a \in W_\circ$ and let $b \in V$. Then

$$a\lambda \circ b\lambda = (a \circ b)\lambda = (a + b)\lambda = a\lambda + b\lambda,$$

which proves, being λ invertible, that $a\lambda \in W_\circ$, and consequently $W_\circ\lambda = W_\circ$. On the other hand, if $a \in U_\circ$, then $a = b \cdot c$ for some $b, c \in V$. Then $a\lambda = (b \cdot c)\lambda = b\lambda \cdot c\lambda$, hence $a\lambda \in U_\circ$. □

The previous result is important since it gives a precise structure to the matrices in $H_\circ$.

Lemma 5.1.4. *Let $\circ$ be an operation. If $\lambda \in (\mathbb{F}_2)^{n\times n}$ is compatible with $\circ$, then for each $a \in V$ it holds $M_a\lambda = \lambda M_{a\lambda}$.*

Proof. Let $a, b \in V$. Then, by Proposition 5.1.2, $(b \circ a)\lambda = b\lambda \circ a\lambda$, which means that $(bM_a + a)\lambda = b\lambda M_{a\lambda} + a\lambda$, i.e. $bM_a\lambda = b\lambda M_{a\lambda}$. From the generality of b, the desired follows. □

Theorem 5.1.5. *Let $\circ$ be an operation. If $\lambda \in (\mathbb{F}_2)^{n\times n}$ is compatible with $\circ$, then*

$$\lambda = \begin{pmatrix} A & B \\ \mathbb{0}_{d,n-d} & D \end{pmatrix},$$

where $A \in \mathrm{GL}\left((\mathbb{F}_2)^{n-d}, +\right)$, $D \in \mathrm{GL}\left((\mathbb{F}_2)^{d}, +\right)$, $B \in (\mathbb{F}_2)^{(n-d)\times d}$, and for each $a \in V$ it holds $\Sigma_a D = A\Sigma_{a\lambda}$.

Proof. Let us write λ into the block form

$$\lambda = \begin{pmatrix} A & B \\ C & D \end{pmatrix}.$$

Since, from Theorem 5.1.3, $W_\circ\lambda = W_\circ$ and $W_\circ$ is spanned by the last d vectors of the canonical basis, it holds $C = \mathbb{0}_{d,n-d}$, and consequently A and D are

invertible. Furthermore, for Lemma 5.1.4, for $a \in V$ it holds $M_a\lambda = \lambda M_{a\lambda}$. This means that

$$\begin{pmatrix} \mathbb{1}_n & \Sigma_a \\ \mathbb{0}_{d,n-d} & \mathbb{1}_d \end{pmatrix} \begin{pmatrix} A & B \\ 0 & D \end{pmatrix} = \begin{pmatrix} A & B \\ 0 & D \end{pmatrix} \begin{pmatrix} \mathbb{1}_n & \Sigma_{a_\lambda} \\ \mathbb{0}_{d,n-d} & \mathbb{1}_d \end{pmatrix},$$

and consequently

$$\begin{pmatrix} A & B + \Sigma_a D \\ 0 & D \end{pmatrix} = \begin{pmatrix} A & A\Sigma_{a\lambda} + B \\ 0 & D \end{pmatrix}.$$

It follows that

$$\Sigma_a D = A\Sigma_{a\lambda}. \tag{5.2}$$

□

Remark 5.1.6. Notice that λ is compatible with $\circ$ if the condition of Eq. (5.2) is satisfied, regardless the choice of the matrix B. Indeed, when λ is invertible, being an homomorphism of $(V, +, \cdot)$ means that $(a \cdot b)\lambda = a\lambda \cdot b\lambda$ for each a and b, and such a condition relates only to weak vectors, which are not influenced by B.

Theorem 5.1.5 and the previous lemma have been first proved and used in [BCS17] to derive a polynomial-time algorithm which takes as input a matrix $\lambda \in (\mathbb{F}_2)^{n\times n}$ and returns an operation $\circ$ as in Setting 3 compatible with λ. As an application, the authors show an example of operation which is compatible with the diffusion layer of PRESENT. However, the output of this algorithm is a huge[1] class of operations which are too general for our purpose. The main reason why we do not use here this algorithm is that we experimentally noticed the best results in terms of the ratio between the $\circ$-differential uniformity and the $+$-differential uniformity of the confusion layer when n is small. For this reason, we claim that the best operations for a differential attack are obtained as a concatenation of smaller operations, having the size of the S-boxes. We refer to these by saying *parallel operations*. It is not clear if the output of the algorithm presented in [BCS17] contains parallel operations. Since our main scope is to weaken the non-linearity

[1] In the case of PRESENT, the algorithm computes 2^{2360} operations compatible with the diffusion layer.

of the confusion layer, we decide to prioritise the choice of the operation, and consequently to focus on the opposite problem, i.e. given a convenient parallel operation $\circ$, determine diffusion layers which are compatible with $\circ$. In the light of this, notice that Theorem 5.1.5 might impose a restriction on the size of the weak-key space. For example, when d is such that $n - d$ exceeds the size of the S-boxes, the zero block in λ may prevent the layer from having good diffusion properties. In the following section, we address this problem in a case that, as we will show, represents the right compromise between the dimension of the space and of the weak-key space, i.e. $d = n-2$.

5.2 The case $d = n - 2$

As we have seen in the previous chapter, unlike the standard differential attack where only the confusion layer induces differential probabilities, in the case of the attack with respect to different operations we also have to consider the issues coming from the key-addition layer. In light of this, the case where $\circ$-differential probabilities are the highest possible, in the sense of Theorem 4.4.33, deserves particular attention. Let $\circ$ be any operation such that $d = \dim(W_\circ) = n - 2$. As we have shown in Section 4.4.4, every operation with such hypotheses is individuated by a defining matrix of the kind

$$\Theta = \begin{pmatrix} \mathbf{0} & \boldsymbol{b} \\ \boldsymbol{b} & \mathbf{0} \end{pmatrix},$$

where $\boldsymbol{b}$ is any non-null vector in $(\mathbb{F}_2)^{n-2}$. Consequently it holds

$$M_{e_1} = \left(\begin{array}{c|c} \mathbb{1}_2 & \begin{matrix} \mathbf{0} \\ \boldsymbol{b} \end{matrix} \\ \hline \mathbb{0}_{n-2,2} & \mathbb{1}_{n-2} \end{array}\right), \quad M_{e_2} = \left(\begin{array}{c|c} \mathbb{1}_2 & \begin{matrix} \boldsymbol{b} \\ \mathbf{0} \end{matrix} \\ \hline \mathbb{0}_{n-2,2} & \mathbb{1}_{n-2} \end{array}\right),$$

and $M_{e_i} = \mathbb{1}_n$ for each $3 \leq i \leq n$. Notice that, in this case, $U_\circ$ is a subspace and from Remark 4.4.24 it holds $U_\circ = \mathrm{Span}\{u\}$, where $u = (0, 0, \boldsymbol{b}) \in (\mathbb{F}_2)^n$. Consequently, since $U_\circ$ contains only one non-null error, the output difference of the key-addition layer, when the input difference Δ is given, is Δ or $\Delta + u$, each one with probability $1/2$. As we showed, as far as $\circ$-differential passing through the key-addition layer and the corresponding probabilities

are concerned, this is the best result we can obtain.

Let us now provide a characterisation of the matrices compatible with an operation $\circ$ such that $d = n - 2$, i.e. a description of $H_\circ$. In the last part of the chapter, this result will be useful for selecting a diffusion layer for the illustration of our attack against a particular instance of SPN.

Theorem 5.2.1. *Let $\circ$ be an operation as in Setting 3 and such that $\dim(W_\circ) = n - 2$, and let $u = (0, 0, \boldsymbol{b}) \in (\mathbb{F}_2)^n$ be the generator of $U_\circ$. Let $\lambda \in (\mathbb{F}_2)^{n \times n}$. The following are equivalent:*

(i) λ is compatible with $\circ$;

(ii) there exist $A \in \mathrm{GL}\left((\mathbb{F}_2)^2, +\right)$, $D \in \mathrm{GL}\left((\mathbb{F}_2)^{n-2}, +\right)$, and $B \in (\mathbb{F}_2)^{2\times(n-2)}$, such that

$$\lambda = \begin{pmatrix} A & B \\ \mathbb{0}_{n-2,2} & D \end{pmatrix}$$

and $\boldsymbol{b}D = \boldsymbol{b}$.

Proof. Firstly, from Theorem 5.1.5, λ decomposes into the block form

$$\lambda = \begin{pmatrix} A & B \\ \mathbb{0}_{n-2,2} & D \end{pmatrix},$$

where $A \in \mathrm{GL}\left((\mathbb{F}_2)^2, +\right)$, $D \in \mathrm{GL}\left((\mathbb{F}_2)^{n-2}, +\right)$, $B \in (\mathbb{F}_2)^{2\times(n-d)}$. From Theorem 5.1.3, since $U_\circ = \{0, u\}$, one obtains $u\lambda = u$, and hence $\boldsymbol{b}D = \boldsymbol{D}$. Conversely, let us assume (ii) and prove that given $x, y \in V$ it holds $(x \cdot y)\lambda = x\lambda \cdot y\lambda$. If $x \in W_\circ$, then also $x\lambda \in W_\circ$, hence there is nothing to prove. For the same reason $(x \cdot y)\lambda = x\lambda \cdot y\lambda$ if and only if

$$\left((x_1, x_2, 0, \ldots, 0) \cdot (y_1, y_2, 0, \ldots, 0)\right)\lambda = (x_1, x_2, 0, \ldots, 0)\lambda \cdot (x_1, x_2, 0, \ldots, 0)\lambda,$$

thus it is sufficient to consider the case $x = e_1$ and $y = e_2$. It is easy to check that both the products $e_1 \cdot e_2$ and $e_1\lambda \cdot e_2\lambda$ equal u, hence from $u\lambda = u$ the desired holds. □

From Theorem 5.2.1 it also follows that the number of matrices compatible with an operation $\circ$ of size n and such that $d = n - 2$ can be counted.

Corollary 5.2.2. *Let $\circ$ be an operation as in Setting 3 and such that* $\dim(W_\circ) = n-2$. *Then*

$$\#H_\circ = 3 \cdot 2^{3(n-2)} \cdot \#\,\mathrm{GL}\left((\mathbb{F}_2)^{n-3}, +\right),$$

where $\#\,\mathrm{GL}\left((\mathbb{F}_2)^{n-3}, +\right) = \prod_{j=0}^{n-4}(2^{n-3} - 2^j)$. □

Notice that in the present case, as we could have noticed by looking at the table in Fig. 4.1, also the problem of counting the number of all the possible operations is trivial. Since the number of operations equals the number of all the possible defining matrices, it is straightforward that, when $d = n-2$, there exist $2^{n-2} - 1$ alternative operations as is Assumpion 3 different from the XOR.

5.3 Experiments on a small cipher

In this section we design a small cipher and we perform some experiments on it, in order to show that the differential attack with an alternative operation can be effective. Before doing this, it is worth spending some words on why we decided for a small-size cipher and not for a standard one. As we already pointed out in Section 4.5, having trails with larger probabilities with respect to the operation $\circ$ than with respect to $+$ represents a necessary condition for the success of the distinguishing attack. However, this is not sufficient. In fact, let us consider again Example 4.5.1. From one hand it is true that the entries in DDT° are significantly larger of those in DDT^+. The counterpart is that the number of non-null entries in DDT^+ is larger than those of DDT°, which means that, in the second case, for a given differential there may be many more $+$-differential trails, even if with a lower probability. This might result in a differential whose differential probability, obtained as the sum of the probabilities of all its differential trails, is higher in the case of $+$ than in the one of $\circ$. In this sense, the role of the diffusion layer is crucial, since the probability of any differential is the sum of the probabilities of all the trails which compose the differential, and the number of such trails depends on the good diffusion properties of the linear layer. It should be clear that, if we want to prove that a given cipher is secure in the standard setting and not in the one of $\circ$-differentials, a comparison between the best $+$-differential trail

and the best $\circ$-differential trail may be rather inconclusive. For this reason, we believe that the only possible approach is to perform an exhaustive search for all the differentials, and to compare the values of the best $+$-differential and of the best $\circ$-differential. The choice of a standard size for the block would make these computations unfeasible.

From now on let us assume $n = 15$ and let us write $V = \oplus_{i=1}^{5}(\mathbb{F}_2)^3$, i.e. let us assume V is decomposed as the sum of five 3-dimensional bricks. We will design shortly a 15-bit SPN having five identical 3-bit S-boxes. Before doing this, let us describe the operation $\hat{\circ}$ we will be using for the attack.

5.3.1 The operation $\hat{\circ}$

One particular operation between those discussed in Section 5.2 deserves attention, i.e. the one obtained when $\boldsymbol{b} = (1, 0, \ldots, 0) \in (\mathbb{F}_2)^{13}$, which is defined by the matrices

$$M_{e_1} = \left(\begin{array}{c|cccc} \mathbb{1}_2 & 0 & 0 & \ldots & 0 \\ & 1 & 0 & \ldots & 0 \\ \hline \mathbb{0}_{13,2} & \multicolumn{4}{c}{\mathbb{1}_{13}} \end{array}\right), \quad M_{e_2} = \left(\begin{array}{c|cccc} \mathbb{1}_2 & 1 & 0 & \ldots & 0 \\ & 0 & 0 & \ldots & 0 \\ \hline \mathbb{0}_{13,2} & \multicolumn{4}{c}{\mathbb{1}_{13}} \end{array}\right),$$

and $M_{e_i} = \mathbb{1}_{15}$ for $3 \leq i \leq 15$. The reason why we focus on this operation is that $\hat{\circ}$ is a parallel operation with respect to the message space decomposition $V = \oplus_{i=1}^{5}(\mathbb{F}_2)^3$. In particular, it acts as the operation $\diamond$ defined in Example 4.4.21 on the first brick, and as $+$ on the remaining ones, i.e. $\hat{\circ} = (\diamond, +, +, +, +)$. For example, if $x, y \in V$

$$\begin{aligned}(x_1, x_2, x_3, x_4, \ldots, x_{15}) \hat{\circ} (y_1, y_2, y_3, y_4, \ldots, y_{15}) \\ = ((x_1, x_2, x_3) \diamond (y_1, y_2, y_3), x_4 + y_4, \ldots, x_{15} + y_{15}).\end{aligned}$$

In the case of the operation $\hat{\circ}$, Theorem 5.2.1 becomes as follows.

Theorem 5.3.1. *Let $\lambda \in (\mathbb{F}_2)^{15\times 15}$. The following are equivalent:*

(i) $\lambda \in H_{\hat{\circ}}$;

$$\lambda = \left(\begin{array}{cccccc|cccccc|ccc}
1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\
0 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 \\
\hdashline
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\hline
0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 0 \\
0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\
\hline
0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 1 & 1 \\
0 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 1 \\
\hline
0 & 0 & 1 & 1 & 1 & 0 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 \\
0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 1 \\
\hline
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 1 \\
0 & 0 & 1 & 1 & 1 & 0 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 \\
0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0
\end{array}\right)$$

Figure 5.1: A diffusion layer compatible with $\hat{\circ}$

(ii) there exist $A \in \mathrm{GL}\left((\mathbb{F}_2)^2, +\right)$, $D' \in \mathrm{GL}\left((\mathbb{F}_2)^{12}, +\right)$, $B \in (\mathbb{F}_2)^{2\times 13}$ *and* $B' \in (\mathbb{F}_2)^{12\times 1}$ *such that*

$$\lambda = \left(\begin{array}{c|cc}
A & \multicolumn{2}{c}{B} \\
\hline
\mathbb{0}_{13,2} & 1 & \mathbb{0}_{1,12} \\
\hdashline
 & B' & D'
\end{array}\right).$$

□

Example 5.3.2. From Theorem 5.3.1, the 15×15 binary matrix λ displayed in Fig. 5.1 is compatible with $\hat{\circ}$. Notice that, when used with a parallel confusion layer featuring 3-bit S-boxes, the zero block $\mathbb{0}_{13,2}$ does not represent a weakness.

5.3.2 The target cipher

Let us consider the R-round SPN defined by classical round functions of the type $\varepsilon_{i,K} = \gamma\lambda\sigma_{k_i}$, where γ acts on every brick as the S-box γ' defined in

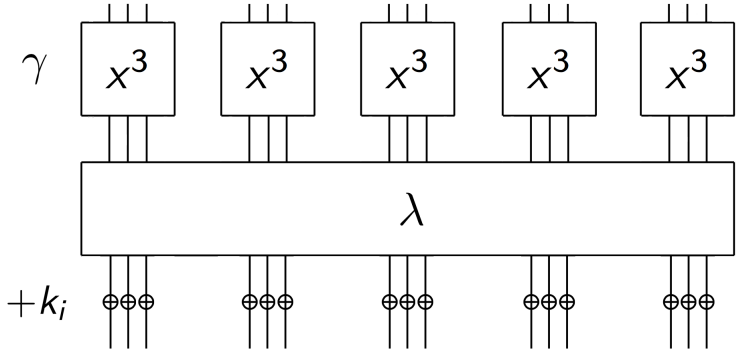

Figure 5.2: 1-round encryption of the 15-bit target cipher

Example 4.5.1 and λ is the matrix of Fig. 5.1. In Fig 5.2 a 1-round encryption of our target cipher is displayed. Recalling that the operation $\diamond$ is suitable for attacking the cubic S-box of Example 4.5.1, which we have chosen as S-box for our cipher, and that the operation $\hat{\circ}$ behaves as the XOR from the 4^{th} bit of the block onwards, we are basically performing an attack against the first S-box of the cipher, whose differential properties with respect to $\hat{\circ}$ are weaker, as already shown in Example 4.5.1. On the other hand, recall that the values in the $\text{DDT}^{\diamond}_{\gamma'}$ need to be rescaled, due to the effect of the further $\diamond$-differential probabilities induced by the key-addition layer. What we show next is that the magnitude difference between the entries of the two DDTs is sufficiently large for a distinguishing attack, despite the effect of the key-addition.

5.3.3 Results and conclusions

For the cipher of Section 5.3.2, we have performed experiments to study its resistance to differential cryptanalysis. Note that only the resistance to differential cryptanalysis is considered and we do not claim any other resistance criteria for the security of this small cipher.

Setting the attack We did not specify yet how we generate the key used for the attack. As it has been discussed in [DR07] and [BG10], the proba-

bility of a differential trail may depend on the choice of the master key used to encrypt the messages. In order to take this fact into account, we generate a key-schedule by selecting the round keys k_i's uniformly at random in V, for each master key, i.e. we are considering the cipher as a *long-key cipher*. Moreover, in our computation we considered 2^{11} possible key-schedule samples, in order to have a good estimate of the expected differential probability of the best differential on r rounds of the cipher. We computed by means of an exhaustive search all the possible differentials, for each possible key assignment, and furthermore considered the average of the obtained results. Let us now discuss the results obtained when $r = 5$.

A simulation of the attack on 5 rounds The experimental computations show that the best 5-round differential $(\Delta_{I_+}, \Delta_{O_+}) = (\texttt{0007}_\text{x}, \texttt{1301}_\text{x})$ occurs with probability $2^{-14.567}$ where the difference taken into consideration is the classical $+$-difference. Using the operation $\hat{\text{o}}$ of Section 5.3.1 instead, the best 5-round differential is $(\Delta_{I_{\hat{\text{o}}}}, \Delta_{O_{\hat{\text{o}}}}) = (\texttt{3000}_\text{x}, \texttt{019D}_\text{x})$ with probability $2^{-14.296}$. Let now `DIST` be an algorithm intended to distinguish the 5-round cipher from a random permutation, and let `ORACLE` be an encryption oracle. Consider the following games.

Procedure 1 Game 1

1: `DIST` computes $P \stackrel{\text{def}}{=} [(x, x + \Delta_{I_+})|\ x \in V]$
2: `ORACLE` select uniformly at random $\alpha \in \{0, 1\}$
3: **if** $\alpha = 1$ **then**
4: `ORACLE` picks a random key K and **returns**

$$C \stackrel{\text{def}}{=} [aE_K + bE_K|\ (a, b) \in P]$$

5: **else** `ORACLE` **returns** a list C of $\#P$ elements chosen uniformly at random in V
6: **if** $\Delta_{O_+} \in C$ **then**
7: `DIST` **returns** 1
8: **else** `DIST` **returns** 0

We can compute the maximal success probability of the distinguishing attack as the probability that at least one pair $(x, x + \Delta_{I_+})$ or $(x, x\hat{\text{o}}\Delta_{I_{\hat{\text{o}}}})$ follows

Procedure 2 Game 2

1: DIST computes $P \stackrel{\text{def}}{=} [(x, x \hat{\diamond} \Delta_{I_\diamond}) |\, x \in V]$
2: ORACLE select uniformly at random $\alpha \in \{0, 1\}$
3: **if** $\alpha = 1$ **then**
4: ORACLE picks a random key K and **returns**

$$C \stackrel{\text{def}}{=} [aE_K \,\hat{\diamond}\, bE_K |\, (a, b) \in P]$$

5: **else** ORACLE **returns** a list C of $\#P$ elements chosen uniformly at random in V
6: **if** $\Delta_{O_\diamond} \in C$ **then**
7: DIST **returns** 1
8: **else** DIST **returns** 0

the differential, assuming that, when using the full codebook, differentials are binomially distributed over the keys [DR07]. Hence, letting the *success probability* of the algorithm DIST with respect to Game 1 and Game 2 be defined respectively as

$$\mathcal{S}^i_{\texttt{DIST}} \stackrel{\text{def}}{=} \mathbb{P}(\texttt{DIST} = 1 |\, \alpha = 1), \quad i \in \{1, 2\},$$

we obtain that if $\mathcal{S}^1_{\texttt{DIST}} > 1/2$ or $\mathcal{S}^2_{\texttt{DIST}} > 1/2$, then DIST manages to distinguish the 5-round cipher from a random permutation using $+$-differences or $\hat{\diamond}$-differences, respectively. With the probabilities previously given, we find that in more than 50% of the cases the differential is not fulfilled for the $+$-difference and we can conclude that a basic distinguishing attack does not succeed. In the same setting using the $\hat{\diamond}$-differences, the differential appears at least once for about 56% of the keys.

Consequently this represents an example of a small cipher which looks like a classically secure SPN, and for which considering an operation different from the one used for the key-addition produces a successful distinguishing attack. This may also be rephrased in another way: a designer may be tempted to claim that a 5-round encryption is sufficient to grant security from differential cryptanalysis to the cipher. The designer would not be wrong since we proved that, considering only classical differences, the

distinguishing attack fails. However, a cryptanalyst using $\hat{\circ}$-differentials may succeed in breaking the 5-round cipher. Therefore, an R-round encryption with $R > 5$ is required to provide security with respect to the differential attack with a more general class of operations.

While at the time of writing the impact of the key addition is well understood, the question of searching large-scale diffusion layers and of understanding the impact of the new operations on the differential uniformity of classical S-boxes remains open.

In conclusion, we proved that a cipher which appears to be secure with respect to the classical differential attack may be actually weak with respect to a differential attack where the difference used comes from another group operation on the message space. We essentially showed that, depending on the operation considered, a cipher can have different levels of resistance against differential attacks. Considering the class of effectively computable operations introduced in [CS17], we studied the interaction between the latter and the layers of a SPN, and designed operations which made our differential attack possible. We finally provided an example of a 15-bit SPN which cannot be distinguished from a random permutation in the classical context, whereas a distinguishing attack succeeds when considering $\hat{\circ}$-differences, where $\hat{\circ}$ is an operation built *ad hoc* for the purpose.

Part III

On the design of wave ciphers

6 On wave functions

The last part of this thesis is devoted to the design and the study of the algebraic security of wave ciphers, a family of Feistel Networks which is described in this chapter.

6.1 Overview and motivation

Let us recall that we are considering two families of symmetric cryptosystems, namely Substitution-Permutation Networks and Feistel Networks, which are obtained as a composition of several round functions. Each round function is a key-dependent permutation of the plaintext space, designed in a way to provide both confusion and diffusion. Confusion is provided by means of a non-linear layer which applies vectorial Boolean functions, called S-boxes, whereas a linear map, called diffusion layer, provides diffusion. In order to perform decryption, invertible layers and the Feistel structure are used in SPN and FN, respectively. In the framework of SPNs decryption is performed applying in reverse order the inverse of each layer of the cipher. In the case of FNs, it is the Feistel structure itself to guarantee a fast decryption.

It is well-established that the non-linearity of the confusion layer is a crucial parameter for the security of the cipher. In particular, in order to prevent statistical attacks (e.g. differential [BS91a] and linear [Mat93] cryptanalysis), block ciphers' designers are interested in invertible S-boxes reaching the

best possible differential uniformity, which is two. Functions satisfying such property, i.e. almost-perfect non-linear functions (see Definition 2.4.7), are extensively studied. However, as recalled in Section 2.4.2, no APN permutation is known in the cases $s \in \{4, 8\}$, which for implementation needs represent the optimal sizes of an S-box. For this reason, in this part of the thesis we focus on defining a new framework for block ciphers, whose S-boxes are not bijective, and consequently can be APN functions with s inputs, s even. More precisely, we focus on injective confusion layers which enlarge the message, and on surjective diffusion layers which reduce the message to its original size. By appending a key addition to these, we obtain a generalised round function which we call a *wave function*. Consequently a *wave cipher* is a block cipher featuring wave functions in its structure. In order to guarantee an efficient decryption, we propose to use wave functions inside an FN-like framework. As far as the security of wave ciphers is concerned, we focus on a group-theoretical analysis, giving sufficient conditions for the primitivity of the group generated by the round functions (see Chapter 3). Recall that the cryptanalysts' interest into the imprimitivity of the group generated by the round functions of a block cipher arises from the study performed by Paterson [Pat99], who showed how the imprimitivity of the group can be exploited to construct a trapdoor that may be hard to detect. In particular, he gave an example of a DES-like cipher, which can be easily broken since its round functions generate an imprimitive group, but which is resistant to both linear and differential cryptanalysis. In this part, we show that ciphers having such a wave structure are provably secure, under some cryptographic assumptions, with respect to the imprimitivity attack described in Section 3.3. It is worth mentioning here that, in order to prove the security of a given wave cipher with respect to other classical statistical attacks (e.g. linear and differential cryptanalysis), it is needed to analyse the single instance under consideration.

6.2 Wave ciphers

The aim of this section is to define ciphers whose inner layers are not necessarily invertible, which allow to use APN vectorial Boolean functions as

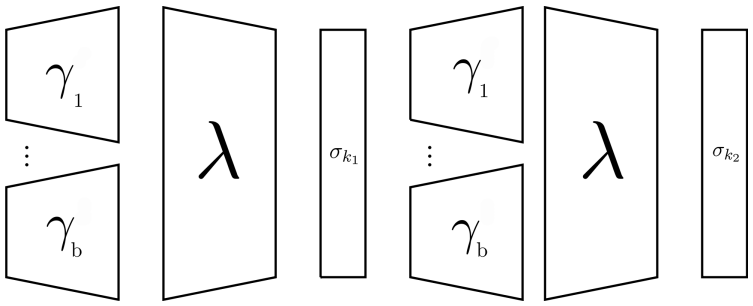

Figure 6.1: Wave functions

S-boxes, even when the S-box input size is four or eight. We focus on the case of wave-shaped round functions, which feature a first layer which enlarges the state, a second which reduces its size, and a key addition. These round functions are employed in the place of classical round functions for both SPNs and FNs. To do so, let us define an auxiliary space $W = (\mathbb{F}_2)^m$, with $n \leq m$ such that $\dim(W) = m = bt$ and $W = W_1 \oplus W_2 \oplus \ldots \oplus W_b$. The subspaces W_js are also called *bricks of* W.

What follows is a generalisation of the concept of classical round function given in Definition 1.3.1.

Definition 6.2.1. For each $k \in V$, a *wave function* induced by k is a map $\varepsilon_k : V \to V$ of the type $\varepsilon_k = \gamma\lambda\sigma_k$, where

- $\gamma : V \to W$ is an injective non-linear transformation (parallel S-box) which acts in a parallel way on each V_j, i.e.

$$(x_1, x_2, \ldots, x_n)\gamma = \big((x_1, \ldots, x_s)\gamma_1, \ldots, (x_{s(b-1)+1}, \ldots, x_n)\gamma_b\big)\,.$$

  The maps $\gamma_j : V_j \to W_j$ are called S-boxes;

- $\lambda : W \to V$ is a surjective linear map;

- $\sigma_k : V \to V, x \mapsto x + k$ is the round key addition.

Figure 6.1 depicts the composition of two consecutive wave functions.

Notice that, in general, we do not require that a wave function is invertible. However, if it is necessary, the following result gives a condition on the confusion and diffusion layers which ensures that a wave function is a permutation.

Lemma 6.2.2. *Let $\varepsilon_k = \gamma\lambda\sigma_k$ be a wave function. The following are equivalent:*

(i) $\{a + b \mid a, b \in \operatorname{Im}\gamma\} \cap \operatorname{Ker}\lambda = \{0\}$;

(ii) $\varepsilon_k \in \operatorname{Sym}(V)$.

Proof. Let us assume (i). Let $x_1, x_2 \in V$ such that $x_1\varepsilon_k = x_2\varepsilon_k$. Then $(x_1\gamma + x_2\gamma)\lambda = 0$, so $x_1\gamma + x_2\gamma \in \{a + b \mid a, b \in \operatorname{Im}\gamma\} \cap \operatorname{Ker}\lambda = \{0\}$, and hence $x_1\gamma = x_2\gamma$. Since γ is injective, it follows $x_1 = x_2$. Conversely, let $x \in \{a + b \mid a, b \in \operatorname{Im}\gamma\} \cap \operatorname{Ker}\lambda$. Then there exist $x_1, x_2 \in V$ such that $x = x_1\gamma + x_2\gamma$ and $x\lambda = 0$, that is $x_1\gamma\lambda = x_2\gamma\lambda$. Therefore $x_1\varepsilon_k = x_2\varepsilon_k$ and hence $x_1 = x_2$, which implies $x = 0$. □

Remark 6.2.3. Notice that it always holds $0 \in \{a+b \mid a, b \in \operatorname{Im}\gamma\} \cap \operatorname{Ker}\lambda$. Moreover, notice that if we assume that $0\gamma = 0$, then the first condition of the previous lemma implies that $\operatorname{Im}\gamma \cap \operatorname{Ker}\lambda = \{0\}$.

6.2.1 Feistel Networks with wave functions

Since our goal is to use the previously defined wave functions inside a cipher, we now define a wave cipher as an FN whose F-function is a wave function. Feistel Network's straightforward decryption encourages this choice.

Before defining wave ciphers, we generalise the security requirement of proper and strongly proper diffusion layers (see Definition 2.4.13 and Definition 2.4.14) to the case of surjective maps. Let us also recall that a *wall* of V (resp. W) is any non-trivial and proper sum of bricks of V (resp. W).

Definition 6.2.4. A surjective linear transformation $\lambda : W \longrightarrow V$ is a *proper diffusion layer* if for each wall $W' = \bigoplus_{j\in J} W_j$ of W and $V' = \bigoplus_{j\in J} V_j$ of V, where $\emptyset \neq J \subset \{1, \ldots, b\}$, it holds

$$V'\lambda^{-1} \not\subset W' + \operatorname{Ker}\lambda.$$

In other terms, if $\pi : W \longrightarrow W/\operatorname{Ker}\lambda$ is the *canonical projection of* W onto $W/\mathrm{Ker}(\lambda)$, and $\hat{\lambda} : W/\operatorname{Ker}\lambda \to V$ is such that $w + \operatorname{Ker}\lambda \mapsto w\lambda$, λ is proper in the sense of Definition 6.2.4 if there exists no wall $W' = \bigoplus_{j\in J} W_j$ of W and $V' = \bigoplus_{j\in J} V_j$ of V such that $W'\pi\hat{\lambda} = V'$, i.e. no wall of W is sent by $\pi\hat{\lambda}$ into a wall of V.

We are now ready to give the definition of a wave cipher, which involves the notion of Feistel operator already explained in Definition 1.2.4.

Definition 6.2.5. An R-round *wave cipher* Φ is a family of encryption functions $\{E_K \mid K \in \mathcal{K}\} \subseteq \mathrm{Sym}(V \times V)$ such that for each $K \in \mathcal{K}$ the map E_K is the composition of R functions. More precisely $E_K = \overline{\varepsilon_{1,K}}\,\overline{\varepsilon_{2,K}} \ldots \overline{\varepsilon_{R,K}}$, where $\varepsilon_{i,K} = \gamma\lambda\sigma_{k_i}$ is an n-bit wave function such that

- λ is a proper diffusion layer, in the sense of Definition 6.2.4,
- the key-schedule $\mathcal{K} \to V^R$, $K \mapsto (k_1, k_2, \ldots, k_R)$, is surjective with respect to any brick.

The function $\rho \stackrel{\text{def}}{=} \gamma\lambda$ is called the *generating function of the wave cipher*.

Let us notice that the ciphers previously introduced are FNs featuring a wave function as F-function. Indeed, given $(x_1, x_2) \in V \times V$ one has

$$(x_1, x_2)\overline{\varepsilon_{i,K}} = (x_1, x_2)\begin{pmatrix} \mathbb{0}_n & \mathbb{1}_n \\ \mathbb{1}_n & \varepsilon_{i,K} \end{pmatrix} = (x_2, x_1 + x_2\varepsilon_{i,K}),$$

where the operator $\overline{\varepsilon_{i,K}}$ induces the Feistel structure, as shown in Figure 6.2. Moreover $\overline{\varepsilon_{i,K}}$ is invertible with the following inverse

$$\overline{\varepsilon_{i,K}}^{-1} = \begin{pmatrix} \varepsilon_{i,K} & \mathbb{1}_n \\ \mathbb{1}_n & \mathbb{0}_n \end{pmatrix}.$$

Note that, as for any FN, the inverse $\overline{\varepsilon_{i,K}}^{-1}$ of the round function $\overline{\varepsilon_{i,K}}$ does not involve the inverse of the wave function $\varepsilon_{i,K}$.

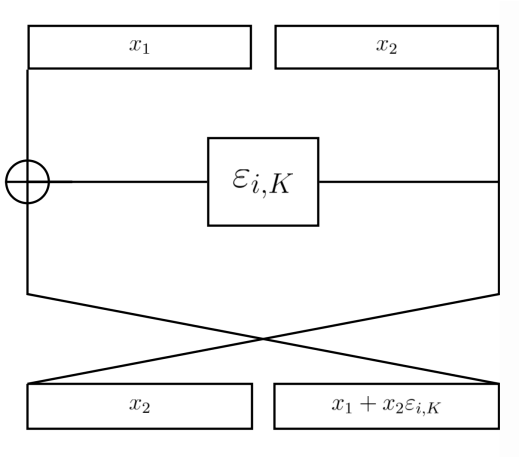

Figure 6.2: Feistel structure of wave ciphers

6.2.2 The group generated by the rounds of a wave cipher

Let $T_{(0,n)} \stackrel{\text{def}}{=} \{\sigma_{(0,k)} \mid (x_1, x_2) \mapsto (x_1, x_2 + k)\} < \mathrm{Sym}(V \times V)$. Let ρ be the generating function of a wave cipher Φ, and $\overline{\rho}$ the corresponding Feistel operator

$$\overline{\rho} = \begin{pmatrix} \mathbb{0}_n & \mathbb{1}_n \\ \mathbb{1}_n & \rho \end{pmatrix}.$$

Then $\overline{\varepsilon_{i,K}} = \overline{\rho}\,\sigma_{(0,k_i)}$, and so $\langle T_{(0,n)}, \overline{\rho} \rangle$ is the group generated by the round functions of the wave cipher Φ.

Remark 6.2.6. It is worth noting here that $\langle T_{(0,n)}, \overline{\rho} \rangle$ is well defined even if ρ is not a permutation. However, the strategy we adopt in the following chapter to prove the primitivity of the group under consideration requires the assumption that ρ is invertible.

The study of the group $\langle T_{(0,n)}, \overline{\rho} \rangle$ previously defined is the subject of the next chapter, where we determine conditions ensuring that such a group is primitive.

7 Group-theoretical study of wave ciphers

In this last chapter, we first show a group-theoretical result which, as consequence, links the primitivity for a Substitution-Permutation Network and the primitivity for a Feistel Network having respectively round functions and F-functions with the same structure. By exploiting this result we prove that the group generated by the round functions of a wave cipher is primitive under some reasonable cryptographic assumptions on the underlying wave functions.

7.1 Security reduction

Let us consider the group generated by the rounds of an FN which uses as F-functions the round functions of a primitive SPN. Here we prove a group-theoretical result which implies the primitivity of the group under consideration. In particular this result is used to show that the group generated by the round functions of a wave cipher with an invertible generating function is primitive if the group[1] generated by the round functions of an SPN-like cipher having as round functions the same wave functions is primitive, as depicted in Fig. 7.1.

[1]Note that the hypothesis that the wave functions are invertible allows to consider this group.

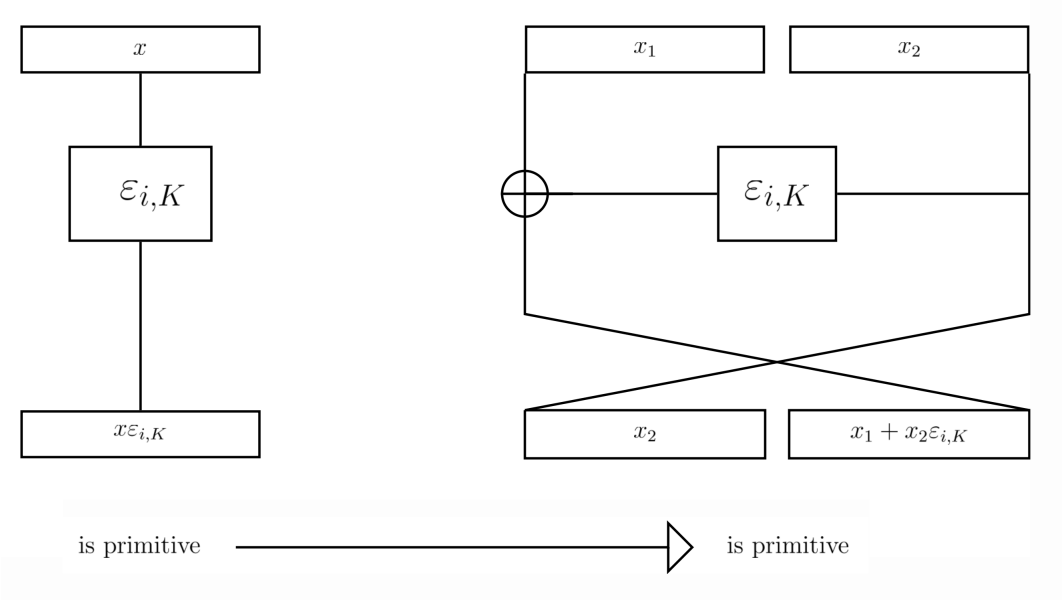

Figure 7.1: FN to SPN reduction

Let us recall that

- $T_n = \{\sigma_k \mid x \mapsto x + k\} < \mathrm{Sym}(V)$,
- $T_{(0,n)} = \{\sigma_{(0,k)} \mid (x_1, x_2) \mapsto (x_1, x_2 + k)\} < \mathrm{Sym}(V \times V)$,

and let us define

- $T_{(n,0)} \stackrel{\text{def}}{=} \{\sigma_{(k,0)} \mid (x_1, x_2) \mapsto (x_1 + k, x_2)\} < \mathrm{Sym}(V \times V)$,
- $T_{(n,n)} \stackrel{\text{def}}{=} \{\sigma_{(k_1,k_2)} \mid (x_1, x_2) \mapsto (x_1 + k_1, x_2 + k_2)\} < \mathrm{Sym}(V \times V)$.

Notice that $T_n \cong T_{(0,n)} \cong T_{(n,0)} < T_{(n,n)}$.

Let ρ be any element in $\mathrm{Sym}(V)$, $\overline{\rho}$ be the corresponding Feistel operator, and let $\Gamma \stackrel{\text{def}}{=} \langle T_{(0,n)}, \overline{\rho} \rangle$. Since we aim at characterising imprimitivity blocks for Γ using Lemma 3.3.2 and Lemma 3.3.3, we need to individuate a transitive subgroup of Γ. For this reason, the following alternative presentation of Γ is useful.

Lemma 7.1.1. $\Gamma = \langle T_{(n,n)}, \overline{\rho} \rangle$.

Proof. Obviously $\Gamma = \langle T_{(0,n)}, \overline{\rho} \rangle < \langle T_{(n,n)}, \overline{\rho} \rangle$. On the other hand, given $x_1, x_2, k \in V$ one has

$$\begin{aligned}(x_1, x_2)\overline{\rho}\sigma_{(0,k)} =& (x_1, x_2) \begin{pmatrix} \mathbb{0}_n & \mathbb{1}_n \\ \mathbb{1}_n & \rho \end{pmatrix} \sigma_{(0,k)} \\ =& (x_2, x_1 + x_2\rho + k) \\ =& (x_1 + k, x_2) \begin{pmatrix} \mathbb{0}_n & \mathbb{1}_n \\ \mathbb{1}_n & \rho \end{pmatrix} \\ =& (x_1, x_2)\sigma_{(k,0)}\overline{\rho}.\end{aligned}$$

Hence for each $k \in V$ it holds $\overline{\rho}\sigma_{(0,k)} = \sigma_{(k,0)}\overline{\rho}$, and consequently $\sigma_{(k,0)} \in \Gamma$. Therefore for each $k_1, k_2 \in V$, $\sigma_{(k_1,k_2)} = \sigma_{(k_1,0)}\sigma_{(0,k_2)} \in \Gamma$. □

Being $T_{(n,n)}$ a transitive subgroup of Γ and noticing that the subgroups of $T_{(n,n)}$ are of the form $\{\sigma_u : u \in U\}$, where U is a subgroup of $V \times V$, we obtain the following.

Lemma 7.1.2. *If Γ is imprimivitive in its action on $V \times V$, then a block system is made of the cosets of a subgroup of $V \times V$, i.e. it is*

$$\{U + v \mid v \in V \times V\},$$

where U is a non-trivial and proper subgroup of $V \times V$.

Proof. See Lemma 3.3.2 and Lemma 3.3.3. □

According to Lemma 7.1.2, in order to prove that Γ is primitive it is sufficient to prove that no subgroup of $V \times V$ is a block. The following theorem, known as Goursat's Lemma [Gou89], characterises the subgroups of the direct product of two groups in terms of suitable sections of the direct factors (see also [Pet09]). We apply this result to the additive group $V \times V$.

Theorem 7.1.3. *Let G_1 and G_2 be two groups. There exists a bijection between*

1. *the set of all subgroups of the direct product $G_1 \times G_2$, and*

2. the set of all triples $(A/B, C/D, \psi)$*, where*

- A *is a subgroup of* G_1*,*
- C *is a subgroup of* G_2*,*
- B *is a normal subgroup of* A*,*
- D *is a normal subgroup of* C*, and*
- $\psi : A/B \to C/D$ *is a group isomorphism.*

In this bijection, each subgroup of $G_1 \times G_2$ *can be uniquely written as*

$$U_\psi = \{(a, c) \in A \times C : (a + B)\psi = c + D\}.$$

□

Note that the isomorphism $\psi : A/B \to C/D$ is induced by a homomorphism $\varphi : A \to C$ such that $(a + B)\psi = a\varphi + D$ for any $a \in A$, and $B\varphi \leq D$. Such homomorphism is not unique.

Lemma 7.1.4. *In the above notation, given any homomorphism* φ *inducing* ψ*, we have*

$$U_\psi = \{(a, a\varphi + d) : a \in A, d \in D\}. \tag{7.1}$$

Proof. Note first that the right-hand side of Eq. (7.1) is contained in U_ψ, since for $a \in A$ and $d \in D$ we have $(a + B)\psi = a\varphi + D = a\varphi + d + D$, that is, $(a, a\varphi + d) \in U_\psi$. Moreover U_ψ is contained in the right-hand side of Eq. (7.1). Indeed, if $(a, c) \in U_\psi$ we have $a\varphi + D = (a + B)\psi = c + D$, so that $c = a\varphi + d$ for some $d \in D$. □

This is our main result of this section.

Theorem 7.1.5. *Let* $\rho \in \mathrm{Sym}(V) \setminus \mathrm{AGL}(V)$*, let* $\overline{\rho}$ *be the corresponding Feistel operator, and denote by* $\Gamma = \langle T_n, \rho \rangle$ *and by* $\overline{\Gamma} = \langle T_{(0,n)}, \overline{\rho} \rangle$*. If* Γ *is primitive on* V*, then* $\overline{\Gamma}$ *is primitive on* $V \times V$*.*

Before proving Theorem 7.1.5, we show how this group-theoretical result can be helpful to us. Let $\Phi = \{E_K \mid K \in \mathcal{K}\} \subseteq \mathrm{Sym}(V \times V)$ be an R-round wave cipher. Denoting by $\rho \stackrel{\text{def}}{=} \gamma\lambda$ its generating function, one has, as shown in

Section 6.2.2, that $\Gamma_\infty(\Phi) = \langle T_{(0,n)}, \overline{\rho} \rangle$ is the group generated by the round functions of the wave cipher Φ. Moreover, $\langle T_n, \rho \rangle$ is the group generated by the wave-shaped round functions of an SPN-like cipher whose round functions are $\varepsilon_{i,K} = \rho\,\sigma_{k_i}$. Therefore, from Theorem 7.1.5, next result directly follows.

Corollary 7.1.6. *Let Φ be a wave cipher, let $\rho \in \mathrm{Sym}(V)$ its generating function and $\overline{\rho}$ the Feistel operator induced by ρ. If $\langle T_n, \rho \rangle$ is primitive on V, then $\Gamma_\infty(\Phi) = \langle T_{(0,n)}, \overline{\rho} \rangle$ is primitive on $V \times V$.* □

Proof of Theorem 7.1.5. Let us suppose that $\overline{\Gamma} = \langle T_{(0,n)}, \overline{\rho} \rangle = \langle T_{(n,n)}, \overline{\rho} \rangle$ is imprimitive, so there exists a non-trivial and proper subgroup U of $V \times V = (\mathbb{F}_2)^n \times (\mathbb{F}_2)^n$ such that $\{U + (v_1, v_2) \mid (v_1, v_2) \in V \times V\}$ is a block system. In particular,

$$U\overline{\rho} = U + (v_1, v_2) \tag{7.2}$$

for some $(v_1, v_2) \in V \times V$. Since $(0,0)\overline{\rho} = (0, 0\rho)$, we can assume $v_1 = 0$ and $v_2 = 0\rho$. With reference to Lemma 7.1.4 and its notation, we have $U = \{(a, a\varphi + d) \mid a \in A, d \in D\}$, and by Eq. (7.2), for any $a \in A$ and $d \in D$ there exist $x \in A$ and $y \in D$ such that

$$(a, a\varphi + d) \begin{pmatrix} \mathbb{0}_n & \mathbb{1}_n \\ \mathbb{1}_n & \rho \end{pmatrix} = (x, x\varphi + y + 0\rho),$$

that is

$$(a\varphi + d, a + (a\varphi + d)\rho) = (x, x\varphi + y + 0\rho).$$

Hence, it holds $x = a\varphi + d$, and considering $a = 0$, we obtain $D \leq A$. Otherwise, considering $d = 0$, we obtain $A\varphi \leq A$. Similarly, we have

$$U\overline{\rho}^{-1} = U + (v_1', v_2') \tag{7.3}$$

for some $(v_1', v_2') \in V \times V$. Since $\overline{\rho}^{-1} = \begin{pmatrix} \rho & \mathbb{1}_n \\ \mathbb{1}_n & \mathbb{0}_n \end{pmatrix}$, we can consider $v_1' = 0\rho$ and $v_2' = 0$. In this case, for any $a \in A$ and $d \in D$ there exist $x \in A$ and $y \in D$ such that

$$(a\rho + a\varphi + d, a) = (x + 0\rho, x\varphi + y).$$

Hence we have $x = a\rho + a\varphi + d + 0\rho$. Substituting $x = a\varphi + d$ in $x\varphi + y$ and being φ a homomorphism, it holds $y = a + a\rho\varphi + a\varphi^2 + d\varphi + 0\rho\varphi$. Then,

considering $a = 0$, we obtain $y = d\varphi$, and thus $D\varphi \leq D$. Now, in the general case, letting $(v_1, v_2) \in V \times V$ it holds

$$(U + (v_1, v_2))\overline{\rho} = U + (v_1', v_2') \tag{7.4}$$

for some $(v_1', v_2') \in V \times V$. By definition of $\overline{\rho}$, we can take $v_1' = v_2$ and $v_2' = v_1 + v_2\rho$. By Lemma 7.1.4 and by Eq. (7.4), for any $a \in A$ and $d \in D$ there exist $x \in A$ and $y \in D$ such that

$$(a + v_1, a\varphi + d + v_2) \begin{pmatrix} \mathbb{0}_n & \mathbb{1}_n \\ \mathbb{1}_n & \rho \end{pmatrix} = (x + v_2, x\varphi + y + v_1 + v_2\rho),$$

that is,

$$(a\varphi + d + v_2, a + v_1 + (a\varphi + d + v_2)\rho) = (x + v_2, x\varphi + y + v_1 + v_2\rho),$$

hence we have $x = a\varphi + d$. Substituting $x = a\varphi + d$ in $x\varphi + y + v_1 + v_2\rho$,

$$a + v_1 + (a\varphi + d + v_2)\rho + a\varphi^2 + v_1 + v_2\rho = y + d\varphi.$$

Then, considering $a = 0$, we obtain $(d + v_2)\rho = y + d\varphi + v_2\rho$. Since $D\varphi \leq D$, then $y + d\varphi \in D$ and so

$$(D + v_2)\rho = D + v_2\rho.$$

Note that we obtain the equality since ρ is a permutation. If $D \neq \{0\}, (\mathbb{F}_2)^n$, then we proved that the imprimitivity of $\overline{\Gamma}$ implies the imprimitivity of Γ. To complete the proof, it remains to consider the cases $D = (\mathbb{F}_2)^n$ and $D = \{0\}$.
[$\mathbf{D = (\mathbb{F}_2)^n}$] We proved that $D \leq A$, and from the hypotheses holds that $D \leq C$ and ψ is an isomorphism between A/B and C/D. Since $D = (\mathbb{F}_2)^n$, we have $D = C = A = B = (\mathbb{F}_2)^n$, which contradicts that U is a proper subgroup of $V \times V$.
[$\mathbf{D = \{0\}}$] First, note that in this case $B\varphi = \{0\}$. Moreover, by Lemma 7.1.4,

$$U = \{(a, a\varphi) \mid a \in A\},$$

and by Eq. (7.4) for any $a \in A$ there exists $x \in A$ such that

$$(a\varphi + v_2, a + v_1 + (a\varphi + v_2)\rho) = (x + v_2, x\varphi + v_1 + v_2\rho).$$

Proceedings as before, it holds

$$a + a\varphi^2 = (a\varphi + v_2)\rho + v_2\rho. \tag{7.5}$$

Note that for any $a \in B \leq A$, $a\varphi = 0$ and so we obtain $a + v_2\rho = v_2\rho$ for any $a \in B$, that is, $B = \{0\}$. Therefore, if $D = \{0\}$, also $B = \{0\}$ and so $\varphi = \psi$ is an isomorphism between A and C. Moreover, since $A\varphi$ is contained in both A and C, then $A = C$ and φ is an automorphism of A. If $A = \{0\}$, then $A = C = D = B = \{0\}$, which contradicts that U is non-trivial. If A is a proper subgroup of $(\mathbb{F}_2)^n$, then by Eq. (7.5) and since both $a + a\varphi^2$ and $a\varphi$ belong to A we have

$$(A + v_2)\rho = A + v_2\rho,$$

and so Γ is imprimitive. If $A = (\mathbb{F}_2)^n$, in Eq. (7.5) we can consider $v_2 = 0$ since $a\varphi + v_2$ is an element of $A = (\mathbb{F}_2)^n$, so we have

$$(a\varphi)\rho = a + a\varphi^2 + 0\rho.$$

Since the function $x + x\varphi^2$ is linear, we proved that $\rho \in \mathrm{AGL}(V)$, which is a contradiction. □

7.2 Conditions on SPN-like wave ciphers

In the light of Theorem 7.1.5, given a wave cipher Φ whose generating function ρ is invertible, we obtain that the group $\Gamma_\infty(\Phi)$ is primitive if we manage to prove that the group $\langle T_n, \rho \rangle$ is primitive. The latter represents the group generated by the rounds of an SPN-like cipher featuring wave functions in the place of classical round functions. Although for such a cipher it may be difficult to compute the computational inverse of the encryption functions, since it has an SPN structure with non-invertible layers, we can still study its theoretical properties. In this section we underline which properties of the generating function ρ guarantee that $\langle T_n, \rho \rangle$ is primitive. From now on let us assume that $\rho \in \mathrm{Sym}(V)$.

Let $\rho = \gamma\lambda$ be the generating function of a wave cipher. From now on we also assume that γ maps 0 to 0, since it is always possible to add 0γ to the

round key of the previous round. Then, since λ is linear, it holds $0\rho = 0$.

In the following, we give a generalisation of Definition 2.4.10, which is a condition in our second main theorem. Let us recall that, as in Section 6.2, $V = V_1 \oplus V_2 \oplus \ldots \oplus V_b$ and $W = W_1 \oplus W_2 \oplus \ldots \oplus W_b$, with $V_j = (\mathbb{F}_2)^s$ and $W_j = (\mathbb{F}_2)^t$ for each $1 \leq j \leq b$.

Definition 7.2.1. Let $1 \leq j \leq b$, $\gamma_j : V_j \to W_j$ be an S-box such that $0\gamma_j = 0$, and $\lambda : W \to V$ be a surjective linear map. Given $0 \leq \delta < s$, γ_j is *δ-non-invariant with respect to λ* if for any proper subspaces $V' < V_j$ and $W' < W_j$ such that $V'\gamma_j + \operatorname{Ker}\lambda \cap W_j = W'$, then $\dim(W') < s - \delta$.

Notice that if $0 \leq \delta < \delta' < s$ and γ_j is δ'-non-invariant with respect to λ, then it is also δ-non-invariant with respect to λ.

Lemma 7.2.2. *Let $\rho = \gamma\lambda \in \mathrm{Sym}(V)$ be the generating function of a wave cipher. Then $\langle T_n, \rho \rangle$ is imprimitive if and only if there exists a proper and non-trivial subgroup U of V such that $(u+v)\gamma + v\gamma \in U\lambda^{-1}$, for any $u \in U$ and $v \in V$. In this case, $\{U + v \mid v \in V\}$ is a block system for $\langle T_n, \rho \rangle$.*

Proof. Since $T_n \leq \langle T_n, \rho \rangle$, if $\langle T_n, \rho \rangle$ is imprimitive, then $\{U + v \mid v \in V\}$ is a block system, for some proper and non-trivial subgroup U of V. Let $v \in V$, then $(U+v)\rho = U + v\rho = U + v\gamma\lambda$. Therefore for any $u \in U$ and $v \in V$ it holds $(u+v)\gamma\lambda + v\gamma\lambda \in U$ and, since λ is linear, $(u+v)\gamma + v\gamma \in U\lambda^{-1}$. □

The following is the main result of this section.

Theorem 7.2.3. *Let $\rho = \gamma\lambda \in \mathrm{Sym}(V)$ be the generating function of a wave cipher Φ. If*

(i) there exists $1 \leq \delta < s$ such that for each $1 \leq j \leq b$ the S-box γ_j is

- *2^δ-differentially uniform,*
- *δ-non-invariant with respect to λ,*

(ii) $\operatorname{Ker}\lambda = \bigoplus_{j=1}^{b} \operatorname{Ker}\lambda \cap W_j$ (λ has a parallel kernel*),*

then $\langle T_n, \rho \rangle$ is primitive (and so it is $\Gamma_\infty(\Phi)$).

Proof. Suppose that $\langle T_n, \rho \rangle$ is imprimitive. For Lemma 7.2.2, a block system is of the form $\{U + v \mid v \in V\}$, for any proper and non-trivial subgroup U of V. Since U is an imprimitivity block and $\rho \in \langle T_n, \rho \rangle$, $U\rho = U + v$ for some $v \in V$. Moreover, since $0\rho = 0$, we obtain $U + v = U$, and consequently $U\rho = U\gamma\lambda = U$. Moreover

$$U\gamma + \operatorname{Ker} \lambda = U\lambda^{-1} \subseteq W, \tag{7.6}$$

and so $U\gamma + \operatorname{Ker} \lambda$ is a subspace of W. For $1 \leq j \leq b$, let $\pi_j : V \longrightarrow V_j$ be the j-th projection with respect to the decomposition $V = V_1 \oplus \ldots \oplus V_b$, and $I \stackrel{\text{def}}{=} \{\, j \mid j \in \{1, \ldots, b\}, U\pi_j \neq \{0\}\}$. Then two cases are possible: either $U \cap V_j = V_j$ for each $j \in I$, or there exists $j \in I$ such that $U \cap V_j \neq V_j$.

In the first case $U = \bigoplus_{j \in I} V_j$ is a wall. From Eq. (7.6) it holds

$$\left(\bigoplus_{j \in I} V_j \right) \gamma + \operatorname{Ker} \lambda = \left(\bigoplus_{j \in I} V_j \right) \lambda^{-1}. \tag{7.7}$$

Since γ is a parallel transformation, we have

$$\left(\bigoplus_{j \in I} V_j \right) \gamma \subset \bigoplus_{j \in I} W_j. \tag{7.8}$$

Thus, from Eq. (7.7) and Eq. (7.8) it follows that

$$\left(\bigoplus_{j \in I} V_j \right) \lambda^{-1} \subset \bigoplus_{j \in I} W_j + \operatorname{Ker} \lambda,$$

which is a contradiction since λ is proper, in the sense of Definition 6.2.4.

In the second case, let us assume there exists $j \in I$ such that $U \cap V_j \neq V_j$. From Eq. (7.6) we have

$$(U\gamma + \operatorname{Ker} \lambda) \cap W_j = U\lambda^{-1} \cap W_j, \tag{7.9}$$

where, since both γ and $\operatorname{Ker}(\lambda)$ are parallel by definition of γ and for (ii),

$$(U\gamma + \operatorname{Ker} \lambda) \cap W_j = U\gamma \cap W_j + \operatorname{Ker} \lambda \cap W_j = (U \cap V_j)\gamma_j + \operatorname{Ker} \lambda \cap W_j. \tag{7.10}$$

Indeed, let $u = (u_1\gamma_1, u_2\gamma_2, \ldots, u_b\gamma_b) \in U\gamma$, $v = (v_1, v_2, \ldots, v_b) \in \operatorname{Ker}\lambda$, and let us assume that $w \stackrel{\text{def}}{=} u\gamma + v \in (U\gamma + \operatorname{Ker}\lambda) \cap W_j$, hence $w = (0, \ldots, 0, w_j, 0, \ldots, 0)$. For $l \neq j$ we obtain $u_l\gamma_l = v_l$, hence $v_l \in \operatorname{Im}\gamma_l \cap (\operatorname{Ker}\lambda \cap W_l)$. From Remark 6.2.3 and since $\operatorname{Ker}\lambda$ is parallel, we have $\operatorname{Im}\gamma_l \cap (\operatorname{Ker}\lambda \cap W_l) = \{0\}$, therefore $v_l = u_l = 0$. Thus, Eq. (7.9) and Eq. (7.10) imply that

$$(U \cap V_j)\gamma_j + \operatorname{Ker}\lambda \cap W_j = U\lambda^{-1} \cap W_j,$$

and, since γ_j is δ-non-invariant with respect to λ, then

$$\dim(U\lambda^{-1} \cap W_j) < s - \delta. \tag{7.11}$$

Furthermore, let $u \in U$ such that $u_j \stackrel{\text{def}}{=} u\pi_j \neq 0$ and $v_j \in V_j$. Since $\langle T_n, \rho \rangle$ is imprimitive, by Lemma 7.2.2 it follows that $(u+v_j)\gamma+v_j\gamma \in U\lambda^{-1}$. Moreover $u\gamma \in U\gamma \subset U\lambda^{-1}$, and so $u\gamma + (u + v_j)\gamma + v_j\gamma \in U\lambda^{-1}$, whose components are null, except possibly for those of the j-th brick, i.e.

$$u_j\gamma_j + (u_j + v_j)\gamma_j + v_j\gamma_j \in U\lambda^{-1} \cap W_j, \tag{7.12}$$

which implies that $\operatorname{Im}(\hat{\gamma}_{j_{u_j}}) + u_j\gamma_j \subset U\lambda^{-1} \cap W_j$. Being γ_j 2^δ-differentially uniform, it is also weakly 2^δ-differentially uniform, and since $u_j \neq 0$ we obtain

$$2^{s-\delta-1} < |\operatorname{Im}(\hat{\gamma}_{j_{u_j}})| \leq |U\lambda^{-1} \cap W_j|,$$

therefore $\dim(U\lambda^{-1} \cap W_j) \geq s - d$, which contradicts Eq. (7.11). □

Notice that in the proof of Theorem 7.2.3 we actually exploited that every S-box is weakly 2^δ-differentially uniform. Hence, we also proved the more general following result.

Theorem 7.2.4. *Let $\rho = \gamma\lambda \in \operatorname{Sym}(V)$ be the generating function of a wave cipher Φ. If*

(i) there exists $1 \leq \delta < s$ such that for each $1 \leq j \leq b$ the S-box γ_j is

- *weakly 2^δ-differentially uniform,*
- *δ-non-invariant with respect to λ,*

(ii) $\operatorname{Ker}\lambda = \bigoplus_{j=1}^{b} \operatorname{Ker}\lambda \cap W_j$ (λ has a parallel kernel*),*

then $\langle T_n, \rho \rangle$ *is primitive (and so it is* $\Gamma_\infty(\Phi)$*).* □

The hypothesis of each S-box being δ-non-invariant with respect to λ in Theorem 7.2.3 can be weakened by adding a reasonable requirement on the diffusion layer. However, for this result does not exist an alternative version using the weak differential uniformity.

Theorem 7.2.5. *Let* $\rho = \gamma\lambda \in \mathrm{Sym}(V)$ *be the generating function of a wave cipher* Φ*. If*

(i) there exists $1 \leq \delta < s$ *such that for each* $1 \leq j \leq b$ *the S-box* γ_j *is*

- 2^δ*-differentially uniform,*
- $(\delta - 1)$*-non-invariant with respect to* λ*,*

(ii) $\operatorname{Ker}\lambda = \bigoplus_{j=1}^{b} \operatorname{Ker}\lambda \cap W_j$*,*

(iii) for each $1 \leq j \leq b$ $\dim(\operatorname{Ker}\lambda \cap W_j) < s - \delta$*,*

then $\langle T_n, \rho \rangle$ *is primitive (and so it is* $\Gamma_\infty(\Phi)$*).*

Proof. The proof proceeds exactly as that of Theorem 7.2.3. In this slightly different setting induced from a further requirement on λ, we can conclude that $U \cap V_j \neq \{0\}$. Indeed, being

$$(U \cap V_j)\gamma_j + \operatorname{Ker}\lambda \cap W_j = U\lambda^{-1} \cap W_j,$$

and having $\dim(U\lambda^{-1} \cap W_j) \geq s - \delta$ and $\dim(\operatorname{Ker}\lambda \cap W_j) < s - \delta$, there must be a non-zero element in $(U \cap V_j)\gamma_j$, and consequently a non-zero element $z \in U \cap V_j$. Then, reasoning as before, using Lemma 7.2.2 one can prove that $\operatorname{Im}(\hat{\gamma}_{j_z}) \subset U\lambda^{-1} \cap W_j$ and $|\operatorname{Im}(\hat{\gamma}_{j_z})| \geq 2^{s-\delta}$. Moreover, $0 \notin \operatorname{Im}(\hat{\gamma}_{j_z})$, since $z \neq 0$ and γ_j is injective. Hence

$$|U\lambda^{-1} \cap W_j| \geq 2^{s-\delta} + 1,$$

and therefore $\dim(U\lambda^{-1} \cap W_j) \geq s - \delta + 1$. The hypothesis of $(\delta - 1)$-non-invariance of γ_j leads to a contradiction, hence the desired holds. □

x	0_x	1_x	2_x	3_x	4_x	5_x	6_x	7_x	8_x	9_x	A_x	B_x	C_x	D_x	E_x	F_x
$x\gamma_1$	0_x	$1F_x$	9_x	C_x	F_x	$1C_x$	12_x	14_x	2_x	1_x	19_x	11_x	A_x	4_x	7_x	$1A_x$

Figure 7.2: A 4x5 APN S-box

7.2.1 A wave cipher with a 4x5 APN S-box

Let us now assume $n = 16$, $m = 20$, $s = 4$, $t = 5$ and $b = 4$. The function $\gamma_1 : (\mathbb{F}_2)^4 \to (\mathbb{F}_2)^5$ displayed in Figure 7.2 represents an example of a 4x5 injective function, which is APN, as it can be noted looking at its DDT displayed in Table 7.3. With an eye on using this function as an S-box for a wave function, and on using Theorem 7.2.5 to prove the primitivity of the corresponding group, one has to verify that there exists a diffusion layer satisfying the hypothesis of Lemma 6.2.2. It holds $\mathrm{Im}(\gamma_1) \subset (\mathbb{F}_2)^5$; moreover it is easy to check that $|\{a + b \mid a, b \in \mathrm{Im}(\gamma_1)\}| = 31$, and the missing vector in $(\mathbb{F}_2)^5$ is $\xi \stackrel{\text{def}}{=} 17_x$. Assuming that we want to design a 16-bit generating function for a wave cipher whose confusion layer γ applies 4 copies of the S-box γ_1 and whose diffusion layer features a parallel kernel, it is sufficient to determine a proper diffusion layer λ such that $\mathrm{Ker}\,\lambda = \mathrm{Span}_{\mathbb{F}_2}\{(\xi,0,0,0),(0,\xi,0,0),(0,0,\xi,0),(0,0,0,\xi)\}$, where 0 here denotes the zero vector in $(\mathbb{F}_2)^5$. The matrix displayed in Figure 7.4 is an example of such a layer. The hypothesis (i) of Lemma 6.2.2 is satisfied, hence all the produced wave functions are bijective, and the given diffusion layer features a parallel kernel, i.e.

$$\mathrm{Ker}\,\lambda = \bigoplus_{j=1}^{b} \mathrm{Ker}\,\lambda \cap W_j.$$

Moreover, we checked using MAGMA that γ_1 is 0-non-invariant with respect to $\mathrm{Ker}\,\lambda$. Consequently, the hypotheses of Theorem 7.2.5 are satisfied with $\delta = 1$, and hence the obtained generating function $\rho = \gamma\lambda$ is such that the group $\langle T_n, \rho\rangle$ is primitive. Then Theorem 7.1.5 implies that the group $\Gamma_\infty(\Phi)$ generated by the rounds of a wave cipher having $\gamma\lambda$ as generating function is primitive.

	0_x	1_x	2_x	3_x	4_x	5_x	6_x	7_x	8_x	9_x	A_x	B_x	C_x	D_x	E_x	F_x	10_x	11_x	12_x	13_x	14_x	15_x	16_x	17_x	18_x	19_x	$1A_x$	$1B_x$	$1C_x$	$1D_x$	$1E_x$	$1F_x$
0_x	16	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
1_x	·	·	·	2	·	2	2	·	2	·	·	·	·	·	2	·	·	·	·	2	·	·	·	·	·	·	·	·	·	2	·	2
2_x	·	·	·	·	·	·	·	·	2	2	·	·	·	2	·	·	2	·	·	2	·	·	·	·	·	·	·	2	·	2	2	·
3_x	·	·	·	2	·	·	·	·	·	·	·	·	2	·	2	·	2	·	·	2	·	·	2	·	2	·	·	2	·	·	·	·
4_x	·	·	·	2	·	2	·	·	2	·	·	2	·	·	·	2	·	·	·	·	·	·	·	·	2	·	·	2	·	·	2	·
5_x	·	·	·	2	·	·	2	·	·	·	·	2	·	·	·	·	2	·	·	·	·	·	2	·	·	·	·	·	2	2	2	·
6_x	·	·	·	·	·	2	2	·	·	·	·	2	·	·	·	·	2	·	2	2	·	2	·	·	·	·	·	2	·	·	·	·
7_x	·	·	·	2	·	·	2	·	·	·	·	·	·	2	·	·	·	·	·	·	2	2	·	·	2	·	·	2	·	2	·	·
8_x	·	·	2	·	·	2	·	·	·	·	·	·	·	·	·	·	2	·	·	·	·	2	·	·	2	·	·	·	·	2	2	·
9_x	·	2	·	·	·	·	·	·	2	·	·	2	·	·	·	·	·	·	·	2	·	2	2	·	2	·	·	·	·	2	·	·
A_x	·	·	·	·	·	·	2	·	2	·	·	2	·	2	2	·	2	·	·	·	·	·	·	·	2	2	·	·	·	·	·	·
B_x	·	·	·	·	·	·	2	·	2	·	·	·	·	·	2	·	·	2	·	·	·	2	2	·	·	·	·	2	·	·	2	·
C_x	·	·	·	·	·	2	·	·	·	·	2	2	·	2	2	·	·	·	·	·	·	·	·	2	·	·	·	2	·	2	·	·
D_x	·	·	·	2	2	·	·	·	·	·	·	2	·	2	2	·	·	·	·	2	·	2	·	·	·	·	·	·	·	·	2	·
E_x	·	·	·	2	·	2	·	2	2	·	·	·	·	2	·	·	2	·	·	·	·	2	2	·	·	·	·	·	·	·	·	·
F_x	·	·	·	·	·	2	2	·	·	·	·	·	·	2	·	·	·	·	·	2	·	·	2	·	2	·	2	·	·	·	2	·

Figure 7.3: Difference distribution table of the S-box γ_1 defined in Section 7.2.1

$$\lambda \stackrel{\text{def}}{=} \begin{pmatrix}
0 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 1 & 1 & 1 & 0 & 1 & 0 \\
0 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\
1 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 1 & 1 & 1 \\
0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 1 & 1 & 1 & 1 \\
1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\
0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\
0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\
1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 1 \\
0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 1 \\
0 & 0 & 1 & 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 \\
1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 \\
0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\
1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 \\
0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\
1 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 1 & 0 & 1 \\
0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 1 & 0
\end{pmatrix}$$

Figure 7.4: An example of 20×16 proper diffusion layer with parallel kernel

7.3 Conclusions and open problems

In this part of the thesis we proposed a new family of ciphers, called wave ciphers, whose round functions are the composition of layers not all invertible. Round functions of a wave cipher are wave functions, bijective functions obtained as the composition of injective non-linear confusion layers enlarging the message, surjective linear diffusion layers reducing the message size, and a key addition. Relaxing the requirement that the S-boxes are permutations allowed to consider APN functions to build confusion layers. In particular we gave an example of 4×5 APN function which can be used as S-box in a wave cipher. We proposed to use wave functions as F-functions of Feistel Networks, where computing inverse functions is not required in order to perform decryption. With regard to their security we showed that, under the assumption that the generating function is invertible, and under suitable non-linearity properties of the Boolean functions involved, the group generated by the round functions of a wave ciphers acts primitively.

Several problems arise from this new construction, such as determining conditions on the wave functions to ensure that the group generated by the round functions of a wave cipher is the alternating or the symmetric group, or studying the resistance of instances of wave ciphers with respect to other statistical attacks, for example studying the impact of differential and linear cryptanalysis on the wave-shaped structure. Moreover, to the best of our knowledge, $s \times t$ APN functions with $s < t$ are not very much investigated in literature. Finally note that, in order to prove that $\Gamma_\infty(\Phi) = \langle\, T_{(0,n)}, \overline{\rho}\,\rangle$ is primitive, we adopted the strategy of considering an SPN-like cipher having as round functions the same wave functions of Φ, and we used Theorem 7.1.5 to deduce the primitivity of $\Gamma_\infty(\Phi)$ from the primitivity of $\langle T_n, \rho\rangle$. This forced us to suppose $\rho \in \mathrm{Sym}(V)$. However, the bijectivity of ρ is not required to define a wave cipher. More importantly, computer simulations lead us to think that non-invertible generating functions provide better levels of non-linearity and consequently better resistance to differential attacks. For this reason, one of our interests is to prove the same result in more general hypotheses on ρ.

List of Figures

List of Tables

Bibliography

[ACDVS14] R. Aragona, A. Caranti, F. Dalla Volta, and M. Sala. On the group generated by the round functions of translation based ciphers over arbitrary finite fields. *Finite Fields and Their Applications*, 25:293–305, 2014.

[ACS17] R. Aragona, A. Caranti, and M. Sala. The group generated by the round functions of a GOST-like cipher. *Annali di Matematica Pura ed Applicata (1923-)*, 196(1):1–17, 2017.

[ACTT16] R. Aragona, M. Calderini, A. Tortora, and M. Tota. Primitivity of PRESENT and other lightweight ciphers. *Journal of Algebra and Its Applications*, page 1850115, 2016.

[AIK+00] K. Aoki, T. Ichikawa, M. Kanda, M. Matsui, S. Moriai, J. Nakajima, and T. Tokita. Camellia: A 128-bit block cipher suitable for multiple platforms-design and analysis. In *Selected Areas in Cryptography*, volume 2012, pages 39–56. Springer, 2000.

[AS11] F. Abazari and B. Sadeghian. Cryptanalysis with ternary difference: Applied to block cipher PRESENT. Cryptology ePrint Archive, Report 2011/022, 2011.

[BAB93] I. Ben-Aroya and E. Biham. Differential cryptanalysis of Lucifer. In *Advances in Cryptology CRYPTO*, volume 93, pages 187–199. Springer, 1993.

[BAK98] E. Biham, R. Anderson, and L. Knudsen. Serpent: A new block cipher proposal. In *Fast Software Encryption*, pages 222–238. Springer, 1998.

[BBS99] E. Biham, A. Biryukov, and A. Shamir. Cryptanalysis of Skipjack reduced to 31 rounds using impossible differentials. In *International Conference on the Theory and Applications of Cryptographic Techniques*, pages 12–23. Springer, 1999.

[BCG+12] J. Borghoff, A. Canteaut, T. Güneysu, E. B. Kavun, M. Knezevic, L. R. Knudsen, G. Leander, V. Nikov, C. Paar, C. Rechberger, et al. PRINCE–a low-latency block cipher for pervasive computing applications. In *International Conference on the Theory and Application of Cryptology and Information Security*, pages 208–225. Springer, 2012.

[BCS17] C. Brunetta, M. Calderini, and M. Sala. Algorithms and bounds for hidden sums in cryptographic trapdoors. *arXiv:1702.08384*, 2017.

[BD93] T. Beth and C. Ding. On almost perfect nonlinear permutations. In *Workshop on the Theory and Application of of Cryptographic Techniques*, pages 65–76. Springer, 1993.

[BDMW10] K. A. Browning, J. F. Dillon, M. T. McQuistan, and A. J. Wolfe. An APN permutation in dimension six. *Finite Fields: theory and applications*, 518:33–42, 2010.

[Ber92] T. A. Berson. Differential cryptanalysis mod 2ˆ32 with applications to MD5. In *Eurocrypt*, volume 658, pages 71–80. Springer, 1992.

[BF17] A. Bannier and E. Filiol. Partition-based trapdoor ciphers. In *Partition-based Trapdoor Ciphers*. InTech, 2017.

[BG10] C. Blondeau and B. Gérard. Links between theoretical and effective differential probabilities: Experiments on PRESENT. *IACR Cryptology ePrint Archive*, 2010:261, 2010.

[BKL+07] A. Bogdanov, L. R. Knudsen, G. Leander, C. Paar, A. Poschmann, M. JB Robshaw, Y. Seurin, and C. Vikkelsoe. PRESENT: An ultra-lightweight block cipher. In *CHES '07*, pages 450–466. Springer, 2007.

[BL08] M. Brinkmann and G. Leander. On the classification of APN functions up to dimension five. *Designs, Codes and Cryptography*, 49(1-3):273–288, 2008.

[Blo11] C. Blondeau. *La cryptanalyse différentielle et ses généralisations*. PhD thesis, Université Pierre et Marie Curie-Paris VI, 2011.

[BS91a] E. Biham and A. Shamir. Differential cryptanalysis of DES-like cryptosystems. *Journal of Cryptology*, 4(1):3–72, 1991.

[BS91b] E. Biham and A. Shamir. Differential cryptanalysis of Feal and N-hash. In *Advances in Cryptology EUROCRYPT 91*, pages 1–16. Springer, 1991.

[BS91c] E. Biham and A. Shamir. Differential cryptanalysis of Snefru, Khafre, REDOC-II, Loki and Lucifer. In *Annual International Cryptology Conference*, pages 156–171. Springer, 1991.

[Cam99] P. J. Cameron. *Permutation groups*, volume 45. Cambridge University Press, 1999.

[Car10] C. Carlet. Vectorial Boolean functions for cryptography. *Boolean models and methods in mathematics, computer science, and engineering*, 134:398–469, 2010.

[CCD00] A. Canteaut, P. Charpin, and H. Dobbertin. Binary m-sequences with three-valued crosscorrelation: a proof of Welch's conjecture. *IEEE Transactions on Information Theory*, 46(1):4–8, 2000.

[CDVS06] A. Caranti, F. Dalla Volta, and M. Sala. Abelian regular subgroups of the affine group and radical rings. *Publ. Math. Debrecen*, 69(3):297–308, 2006.

[CDVS09a] A. Caranti, F. Dalla Volta, and M. Sala. An application of the O'Nan-Scott theorem to the group generated by the round functions of an AES-like cipher. *Designs, Codes and Cryptography*, 52(3):293–301, 2009.

[CDVS09b] A. Caranti, F. Dalla Volta, and M. Sala. On some block ciphers and imprimitive groups. *Applicable algebra in engineering, communication and computing*, 20(5):339–350, 2009.

[CG75] D. Coppersmith and E. Grossman. Generators for certain alternating groups with applications to cryptography. *SIAM Journal on Applied Mathematics*, 29(4):624–627, 1975.

[CS17] M. Calderini and M. Sala. Elementary abelian regular subgroups as hidden sums for cryptographic trapdoors. *arXiv:1702.00581*, 2017.

[CSV17] M. Calderini, M. Sala, and I. Villa. A note on APN permutations in even dimension. *Finite Fields and Their Applications*, 46:1–16, 2017.

[Dob99a] H. Dobbertin. Almost perfect nonlinear power functions on GF (2^n): the Niho case. *Information and Computation*, 151(1-2):57–72, 1999.

[Dob99b] H. Dobbertin. Almost perfect nonlinear power functions on GF(2^n): the Welch case. *IEEE Transactions on Information Theory*, 45(4):1271–1275, 1999.

[Dol10] V. Dolmatov. GOST 28147-89: Encryption, decryption, and message authentication code (MAC) algorithms. Technical report, 2010.

[DR07] J. Daemen and V. Rijmen. Probability distributions of correlation and differentials in block ciphers. *Journal of Mathematical Cryptology JMC*, 1(3):221–242, 2007.

[DR13] J. Daemen and V. Rijmen. *The design of Rijndael: AES-the advanced encryption standard*. Springer Science & Business Media, 2013.

[Gol68] R. Gold. Maximal recursive sequences with 3-valued recursive cross-correlation functions (corresp.). *IEEE transactions on Information Theory*, 14(1):154–156, 1968.

[Gou89] E. Goursat. Sur les substitutions orthogonales et les divisions régulières de l'espace. *Ann. sci. ecole norm.*, (3):9–102, 1889.

[HX01] K. D. L. Hollmann and Q. Xiang. A proof of the Welch and Niho conjectures on cross-correlations of binary m-sequences. *Finite Fields and Their Applications*, 7(2):253–286, 2001.

[Kas71] T. Kasami. The weight enumerators for several classes of subcodes of the 2nd order binary Reed-Muller codes. *Information and Control*, 18(4):369–394, 1971.

[Ker83] A. Kerckhoffs. La cryptographic militaire. *Journal des sciences militaires*, pages 5–38, 1883.

[KLPR10] L. R. Knudsen, G. Leander, A. Poschmann, and M. J.B. Robshaw. PRINTcipher: A block cipher for IC-printing. In *CHES*, volume 6225, pages 16–32. Springer, 2010.

[Knu94a] L. R. Knudsen. Block ciphers: analysis, design and applications. *DAIMI Report Series*, 23(485), 1994.

[Knu94b] L. R. Knudsen. Truncated and higher order differentials. In *International Workshop on Fast Software Encryption*, pages 196–211. Springer, 1994.

[Knu98] L. Knudsen. DEAL - a 128-bit block cipher. In *NIST AES Proposal*, 1998.

[KR11] L. R. Knudsen and M. Robshaw. *The block cipher companion*. Springer Science & Business Media, 2011.

[KRS03] B. S. Kaliski, R. L. Rivest, and A. T. Sherman. Is the Data Encryption Standard a group? In *Advances in Cryptology EUROCRYPT*, volume 85, pages 81–95, 2003.

[LMM91] X. Lai, J. Massey, and S. Murphy. Markov ciphers and differential cryptanalysis. In *Advances in Cryptology EUROCRYPT 91*, pages 17–38. Springer, 1991.

[Mas88] J. L. Massey. An introduction to contemporary cryptology. *Proceedings of the IEEE*, 76(5):533–549, 1988.

[Mat93] M. Matsui. Linear cryptanalysis method for DES cipher. In *Workshop on the Theory and Application of of Cryptographic Techniques*, pages 386–397. Springer, 1993.

[Mil82] F. Miller. *Telegraphic code to insure privacy and secrecy in the transmission of telegrams*. CM Cornwell, 1882.

[MMMM13] D. S. Mackey, N. Mackey, C. Mehl, and V. Mehrmann. Skew-symmetric matrix polynomials and their Smith forms. *Linear Algebra and Its Applications*, 438(12):4625–4653, 2013.

[Nyb93] K. Nyberg. Differentially uniform mappings for cryptography. In *Workshop on the Theory and Application of of Cryptographic Techniques*, pages 55–64. Springer, 1993.

[Pat99] K. G. Paterson. Imprimitive permutation groups and trapdoors in iterated block ciphers. In *FSE*, volume 99, pages 201–214. Springer, 1999.

[Pet09] J. Petrillo. Student research projects Goursat's other theorem. *The College Mathematics Journal*, 40(2):119–124, 2009.

[Pub77] Federal Information Processing Standards Publication. Data Encryption Standard and others. *National Bureau of Standards, US Department of Commerce*, 1977.

[Rij97] V. Rijmen. *Cryptanalysis and design of iterated block ciphers*. PhD thesis, Doctoral Dissertation, October 1997, KU Leuven, 1997.

[RP94] V. Rijmen and B. Preneel. Cryptanalysis of McGuffin. In *International Workshop on Fast Software Encryption*, pages 353–358. Springer, 1994.

[Sha49] C. E. Shannon. Communication theory of secrecy systems. *Bell Labs Technical Journal*, 28(4):656–715, 1949.

[Spe07] KASUMI Specification. General report on the design, specification and evaluation of 3GPP standard confidentiality and integrity algorithms. *Version*, 1:8–17, 2007.

[SW08] R. Sparr and R. Wernsdorf. Group theoretic properties of Rijndael-like ciphers. *Discrete Applied Mathematics*, 156(16):3139–3149, 2008.

[SW15] R. Sparr and R. Wernsdorf. The round functions of KASUMI generate the alternating group. *Journal of Mathematical Cryptology*, 9(1):23–32, 2015.

[Wer92] R. Wernsdorf. The one-round functions of the DES generate the alternating group. In *Eurocrypt*, pages 99–112. Springer, 1992.

[Wer02] R. Wernsdorf. The round functions of Rijndael generate the alternating group. In *FSE*, pages 143–148. Springer, 2002.

[Wer10] R. Wernsdorf. The round functions of SERPENT generate the alternating group. Technical report, available at http://csrc.nist.gov/archive/aes/round2/comments/20000512-rwernsdorf.pdf, 2010.

[ZBL+15] W. Zhang, Z. Bao, D. Lin, V. Rijmen, B. Yang, and I. Verbauwhede. RECTANGLE: a bit-slice lightweight block cipher suitable for multiple platforms. *Science China Information Sciences*, 2015.