Weakness of Block Ciphers Using Highly Nonlinear Confusion Functions

Anne Canteaut, Marion Videau

INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Weakness of block ciphers using highly nonlinear confusion functions

Anne Canteaut — Marion Videau

N° 4367

Février 2002

THÈME 2

ISSN 0249-6399 ISRN INRIA/RR--4367--FR+ENG

INRIA
ROCQUENCOURT

Weakness of block ciphers using highly nonlinear confusion functions

Anne Canteaut* , Marion Videau*

Thème 2 — Génie logiciel
et calcul symbolique
Projet Codes

Rapport de recherche n° 4367 — Février 2002 — 29 pages

Abstract: The level of security of an iterated block cipher is mainly quantified in terms of resistance against known attacks. In particular the resistance against the two main generic attacks, the linear and the differential ones has been formalized in terms of "provable security" which lies on some properties of the confusion functions used in the system. It demands especially that these functions have a high nonlinearity. However such a property implies in the case of almost bent functions, that the Walsh spectrum is divisible by a high power of 2. We show how this provides a new upper bound for the degree of the product of Boolean components of an almost bent function. This result leads to a higher order differential attack on any 5-round Feistel cipher using an almost bent function as a round function. We also show that it is precisely the origin of the weakness of MISTY1 allowing a 7-th order differential attack.

Key-words: iterated block cipher, higher order differential cryptanalysis, almost bent function, Boolean function, Walsh spectrum, MISTY1

* INRIA Rocquencourt

Unité de recherche INRIA Rocquencourt
Domaine de Voluceau, Rocquencourt, BP 105, 78153 Le Chesnay Cedex (France)
Téléphone : +33 1 39 63 55 11 — Télécopie : +33 1 39 63 53 30

Faiblesse des chiffrements par blocs utilisant des fonctions hautement non-linéaires

Résumé : La sécurité des chiffrements itératifs par blocs s'exprime essentiellement en termes de résistance contre des cryptanalyses connues. D'où le concept de "sécurité démontrable" dérivé de la formalisation des propriétés que doit vérifier un système résistant aux deux principales attaques génériques que sont la cryptanalyse différentielle et la cryptanalyse linéaire. Dans ce cadre, les fonctions de confusion utilisées dans le chiffrement doivent être hautement non linéaires donc au mieux presque courbes. Or, ces fonctions sont caractérisées par un spectre de Walsh divisible par une grande puissance de 2. Cette propriété permet de trouver une borne supérieure, pour le degré de la fonction de chiffrement globale, qui croît plus lentement que le calcul a priori le laisserait supposer. Cette particularité utilisée dans une cryptanalyse différentielle d'ordre supérieur permet d'attaquer tout chiffrement de Feistel à cinq tours. Elle est également à l'origine de la faiblesse de MISTY1 autorisant une attaque différentielle d'ordre 7.

Mots-clés : chiffrement itératif par blocs, cryptanalyse différentielle d'ordre supérieur, fonction presque courbe, fonction booléenne, spectre de Walsh, MISTY1

1 Introduction

This paper focuses on a large class of symmetric block ciphers called *iterated block ciphers*. In such a cipher the ciphertext is obtained by iteratively applying a keyed function, called the *round function*, to the plaintext. The underlying idea of this construction is that many iterations of a cryptographically weak round function is expected to lead to a cryptographically strong encryption function. The development of cryptanalysis in the last ten years has led to the definition of some design criteria for iterated block ciphers. These criteria correspond to some mathematical properties of the round function. Especially, the use of a highly nonlinear round function ensures a high resistance to linear attacks [24, 25]. The functions with maximal nonlinearity are called almost bent. They only exist for an odd number of variables, but they also guarantee the best resistance to differential cryptanalysis [8]. Such functions are used for instance in the block cipher MISTY [26]. Here, we show that these optimal functions present some particular properties which introduce other weaknesses in the cipher. This vulnerability comes from the fact that all values occurring in the Walsh spectrum of an almost bent function are divisible by a high power of 2. Most highly nonlinear functions of an even number of variables present a similar structure, except the inverse function. Such a spectral property for a round function F leads to an upper bound on the degree of the function $F \circ F$ which grows much slower than $deg(F)^2$. Therefore, any iterated cipher using an almost bent function may be vulnerable to a higher order differential attack [20, 18], even if the round function has a high degree. This weakness leads to a new design criterion for iterated block ciphers: the Walsh spectrum of the round function should contain at least one value which is not divisible by a higher power of 2. The S-box used in AES is the only known highly nonlinear function which fulfills this requirement.

The paper is organized as follows. Section 2 recalls the basic structure of an iterated block cipher, the principle of a last round attack completed by the cases of the differential, linear and higher order differential attacks. Section 3 deals with the properties derived from the divisibility of the Walsh spectrum. Section 4 describes a higher order differential attack on any 5-round Feistel cipher. The last section is a generalization of an attack on MISTY1.

2 Cryptanalysis of iterated block ciphers

To define an iterated block cipher more formally, we consider a family $(F_k)_{k \in \mathcal{K}}$ of permutations of the set of n-bit words, $\mathbf{F}_2^n$, indexed by a value $k \in \mathcal{K}$ where $\mathcal{K}$ is called the round key space. The encryption function of the iterated block cipher with block size n, with r rounds and with round function F_k is the keyed permutation of $\mathbf{F}_2^n$ defined by

$$x_i = F_{k_i}(x_{i-1}) \quad \text{for } 1 \leq i \leq r \ ,$$

where x_0 is the plaintext and x_r is the ciphertext. The vector $(k_1, \ldots, k_r)$ is called the key and its components are the *round keys*. The round keys may be derived from a unique master key which is shorter than the concatenation of all round keys.

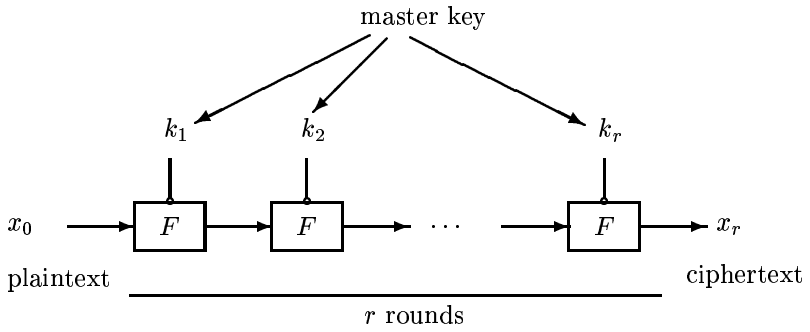

Figure 1: Iterated block cipher

2.1 Last-round attacks

Most attacks on iterated block ciphers are divide-and-conquer techniques which recover the last round key k_r from the knowledge of some pairs of plaintexts and ciphertexts. The other round keys $(k_1, \ldots, k_{r-1})$ (or the entire master key) may then be recovered either directly from k_r (e.g. by exhaustive search) or one after another by successively applying the last round attack on the cipher obtained by removing the last round. In a last round attack, we consider the *reduced cipher*, i.e., the cipher obtained by removing the final round of the original cipher. The reduced cipher corresponds to the function

$$G_{(k_1,\ldots,k_{r-1})} = F_{k_{r-1}} \circ \ldots \circ F_{k_1}.$$

The key point in a last-round attack is to be able to distinguish the reduced cipher from a random permutation for all possible values of the first $(r-1)$ round keys $k_1, \ldots, k_{r-1}$. Some information on k_r can be recovered by applying a discriminator to all functions

$$H_k : x_0 \mapsto F_k^{-1}(x_r) = F_k^{-1}\left(F_{k_r} \circ G_{(k_1,\ldots,k_{r-1})}(x_0)\right) \quad , \quad k \in \mathcal{K}$$

(k describes here the set of all possible values of k_r). If the guess k matches the actual last round key k_r, then F_k^{-1} inverts the last encryption round and H_k corresponds to the reduced cipher. On the contrary, when k is a wrong guess, we get

$$H_k = F_k^{-1} \circ F_{k_r} \circ F_{k_{r-1}} \circ \ldots \circ F_{k_1} \ .$$

Since it essentially corresponds to the reduced cipher followed by two more encryption rounds, this function is supposed to act like a random permutation. This assumption is called the *hypothesis of wrong-key randomization* [13, 21].

Now, we give a more formal description. We refer to [13, 17] for a detailed presentation of last-round attacks.

Definition 1 *Let $\mathcal{P}_n$ denotes the set of all permutations of $\mathbf{F}_2^n$ and let $\mathcal{F}$ be a subset of $\mathcal{P}_n$. A* discriminator *for $\mathcal{F}$ with respect to a subset $(x_1, \ldots, x_N)$ of $\mathbf{F}_2^n$ is a function*

$$\begin{array}{rccc} \mathcal{D}: & (\mathbf{F}_2^n)^N & \rightarrow & \mathbf{F}_2 \\ & (y_1, \ldots, y_N) & \mapsto & \mathcal{D}(y_1, \ldots, y_N) \end{array}$$

for which there exists $\varepsilon > 0$ *such that*

$$\mid Pr_{f \in \mathcal{F}} [\mathcal{D}(f(x_1), \ldots, f(x_N)) = 1] - Pr_{\pi \in_R \mathcal{P}_n} [\mathcal{D}(\pi(x_1), \ldots, \pi(x_N)) = 1] \mid > \varepsilon.$$

$\pi \in_R \mathcal{P}_n$ means that π is a randomly chosen permutation of $\mathbf{F}_2^n$. The formula above shows that a discriminator is a function which allows to distinguish the subset of the permutations corresponding to the reduced cipher from a set of randomly chosen permutations.

Now, the existence of a discriminator $\mathcal{D}$ for the family of reduced ciphers,

$$\mathcal{G} = \{G_{\mathbf{k}},\ \mathbf{k} = (k_1, \ldots, k_{r-1}) \in \mathcal{K}^{r-1}\}$$

with respect to a set $(x_1, \ldots, x_N)$ leads to a last-round attack. The discriminator $\mathcal{D}$ should satisfy the *hypothesis of fixed-key equivalence*, i.e., it should return the same value for almost all reduced ciphers in $\mathcal{G}$ [13, 14, 21]. This hypothesis obviously holds when the round key is introduced by addition, i.e., $F_k(x) = F(x + k)$. This situation occurs in many ciphers, like DES, AES... The last-round attack derived from $\mathcal{D}$ is as follows:

Algorithm 1
INPUT: $(c_1, \ldots, c_N)$, *the* N *ciphertexts corresponding to the plaintexts* $(x_1, \ldots, x_N)$.
OUTPUT: *A set of candidates for the last-round key* k_r.

For all $k \in \mathcal{K}$
  For i *from 1 to* N *do* $y_i \leftarrow F_k^{-1}(c_i)$
  If $\mathcal{D}(y_1, \ldots, y_N) = 1$ *then return* k.

The attack requires the knowledge of N pairs of plaintexts-ciphertexts. Its average cost is $\#\mathcal{K} \times (NT_{F^{-1}} + T_{\mathcal{D}})$, where $\#\mathcal{K}$ is the size of the round key space, $T_{\mathcal{D}}$ is the average cost of the discriminator and $T_{F^{-1}}$ is the average number of operations required for evaluating F_k^{-1}. Notice that the costs of the most commonly used discriminators are proportional to N. If the attack returns several round keys, it can be repeated with another discriminator.

2.2 Basic properties of Boolean functions

Several specific properties of the reduced cipher may yield a discriminator. Now, we define some basic notions related to Boolean functions, which appear in the most commonly used last-round attacks.

A *Boolean function* f *of* n *variables* is a function from $\mathbf{F}_2^n$ into $\mathbf{F}_2$. It can be expressed as a polynomial in $x_1, \ldots, x_n$, called its *algebraic normal form*:

$$f(x_1, \ldots, x_n) = \sum_{u \in \mathbf{F}_2^n} a_u \left(\prod_{i=1}^{n} x_i^{u_i} \right).$$

The *degree* of f, denoted by $deg(f)$, is the degree of its algebraic normal form, i.e.

$$\deg f = \max_{u \in \mathbf{F}_2^n,\ a_u \neq 0} wt(u).$$

where $wt(u)$ is the Hamming weight of u.

Differential and higher order differential attacks involve the derivatives of the reduced cipher.

Definition 2 *[23] Let F be a function from $\mathbf{F}_2^n$ into $\mathbf{F}_2^n$. For any $a \in \mathbf{F}_2^n$, the* derivative of F with respect to a *is the function*

$$D_a F(x) = F(x+a) + F(x) \ .$$

For any t-dimensional subspace V of $\mathbf{F}_2^n$, the t-th derivative of F with respect to V *is the function*

$$D_V F = D_{a_1} D_{a_2} \dots D_{a_t} F \ ,$$

where $(a_1, \dots, a_t)$ is any basis of V.

Linear cryptanalysis has concern with the Walsh spectrum of the reduced cipher. In the following, the usual dot product between two vectors x and y is denoted by $x \cdot y$. For any $\alpha \in \mathbf{F}_2^n$, φ_α is the linear function of n variables: $x \mapsto \alpha \cdot x$.

For any Boolean function f of n variables, we denote by $\mathcal{F}(f)$ the following value related to the Walsh (or Fourier) transform of f:

$$\mathcal{F}(f) = \sum_{x \in \mathbf{F}_2^n} (-1)^{f(x)} = 2^n - 2wt(f) \ ,$$

where $wt(f)$ is the Hamming weight of f, i.e., the number of $x \in \mathbf{F}_2^n$ such that $f(x) = 1$.

Definition 3 *The* Walsh spectrum *of a Boolean function f of n variables is the multiset*

$$\{\mathcal{F}(f + \varphi_\alpha), \alpha \in \mathbf{F}_2^n\} \ .$$

The Walsh spectrum *of a vectorial function F from $\mathbf{F}_2^n$ into $\mathbf{F}_2^n$ consists of the Walsh spectra of all Boolean functions $\varphi_\alpha \circ F : x \mapsto \alpha \cdot F(x)$, $\alpha \neq 0$. Therefore, it corresponds to the multiset*

$$\{\mathcal{F}(\varphi_\alpha \circ F + \varphi_\beta),\ \ \alpha \in \mathbf{F}_2^n \setminus \{0\}, \beta \in \mathbf{F}_2^n\} \ .$$

Definition 4 *The* nonlinearity *of a function F from $\mathbf{F}_2^n$ into $\mathbf{F}_2^n$ is the Hamming distance between all $\varphi_\alpha \circ F, \alpha \in \mathbf{F}_2^n$, $\alpha \neq 0$, and the set of affine functions. It is given by*

$$2^{n-1} - \frac{1}{2}\mathcal{L}(F) \quad \textit{where} \quad \mathcal{L}(F) = \max_{\alpha \in \mathbf{F}_2^n \setminus \{0\}} \max_{\beta \in \mathbf{F}_2^n} |\mathcal{F}(\varphi_\alpha \circ F + \varphi_\beta)| \ .$$

In the following, we focus on three classes of last-round attacks: differential cryptanalysis, linear cryptanalysis and higher-order differential cryptanalysis. There are some other attacks on iterated block cipher. For example, a last-round attack can be performed when the reduced cipher, seen as a univariate polynomial in $\mathbf{F}_{2^n}[X]$, is close to a low-degree polynomial [16]. But, the mathematical nature of the property exploited by the latter attack is different.

2.3 Differential cryptanalysis

Differential cryptanalysis was introduced by Biham and Shamir [2]. It can be applied when the reduced cipher has a derivative which is not uniformly distributed. More precisely, assume that there exist two nonzero elements a and b in $\mathbf{F}_2^n$ such that for any $\mathbf{k} = (k_1, \ldots, k_{r-1})$,

$$\#\{x \in \mathbf{F}_2^n,\ D_a G_{\mathbf{k}}(x) = b\} \simeq A\ ,$$

for a large integer A. This property leads to a discriminator $\mathcal{D}$ for $\mathcal{G}$ with respect to any subset of $(x_{2i}, x_{2i} + a)_{0 \le i < N/2}$ where $(x_{2i})_{0 \le i < N/2}$ is a set of $N/2$ randomly chosen elements in $\mathbf{F}_2^n$ and $N \ge \frac{2^n}{A-1}$.

$$\begin{aligned} \mathcal{D}((y_0, y_1, \ldots, y_{2i}, y_{2i+1}, \ldots)) &= 1 \text{ if } \#\{i,\ 1 \le i \le N/2,\ y_{2i} + y_{2i+1} = b\} \simeq \frac{AN}{2^{n+1}}, \\ &= 0 \text{ if } \#\{i,\ 1 \le i \le N/2,\ y_{2i} + y_{2i+1} = b\} \simeq \frac{N}{2^{n+1}}\ . \end{aligned}$$

It follows that a cipher is resistant to differential cryptanalysis if the reduced cipher is such that, for any $\mathbf{k} \in \mathcal{K}^{r-1}$ and for any nonzero a in $\mathbf{F}_2^n$, the output distribution of $x \mapsto D_a G_{\mathbf{k}}(x)$ is close to the uniform distribution. A necessary security condition is that the round function satisfies this property; it may be a sufficient condition for some ciphers, e.g. for Feistel ciphers [31]. Therefore, the round function F_k of an iterated cipher should satisfy the following requirement: for any $k \in \mathcal{K}$,

$$\delta_{F_k} = \max_{a, b \neq 0} \#\{x \in \mathbf{F}_2^n,\ F_k(x + a) + F_k(x) = b\}$$

should be small. As the number of solutions $x \in \mathbf{F}_2^n$ of $D_a F_k(x) = b$ is even (because x_0 is a solution if and only if $x_0 + a$ is a solution), we can deduce:

Proposition 1 *[31] For any function F from $\mathbf{F}_2^n$ into $\mathbf{F}_2^n$, we have*

$$\delta_F \ge 2\ .$$

In case of equality, F is said to be almost perfect nonlinear (APN).

Note that the terminology APN comes from the general bound

$$\delta_F \ge 2^{m-n}$$

for a function from $\mathbf{F}_2^m$ into $\mathbf{F}_2^n$, where the functions achieving this bound are called *perfect nonlinear functions* [28]. Such functions only exist when m is even and $m \ge 2n$ [29].

The definition of APN functions can be expressed in terms of second derivatives:

Proposition 2 *A function F from $\mathbf{F}_2^n$ into $\mathbf{F}_2^n$ is APN if and only if, for any nonzero elements a and b in $\mathbf{F}_2^n$, with $a \neq b$, we have*

$$D_a D_b F(x) \neq 0 \text{ for all } x \in \mathbf{F}_2^n\ .$$

All known APN functions are functions of an odd number of variables. Actually, it is conjectured that, for any function F from $\mathbf{F}_2^n$ into $\mathbf{F}_2^n$ with n even, we have

$$\delta_F \geq 4 \ .$$

This statement is proved for some particular cases, most notably for power functions [3, 9].

2.4 Linear cryptanalysis

Linear cryptanalysis exploits the existence of a linear combination of the n output bits of the reduced cipher which is close to an affine function [24, 25]. Let us assume that there exists two nonzero elements a and b in $\mathbf{F}_2^n$ such that for any $\mathbf{k} = (k_1, \ldots, k_{r-1})$

$$|\mathcal{F}(\varphi_a \circ G_{\mathbf{k}} + \varphi_b)| \simeq A \ ,$$

for a large integer A. This property leads to a discriminator $\mathcal{D}$ for $\mathcal{G}$ with respect to any subset$(x_1, \ldots, x_N)$ of randomly chosen elements:

$$\begin{aligned} \mathcal{D}(y_1, \ldots, y_N) &= 1 \text{ if } |\sum_{i=1}^{N} (-1)^{a \cdot y_i + b \cdot x_i}| \simeq \frac{AN}{2^n}, \\ &= 0 \text{ if } |\sum_{i=1}^{N} (-1)^{a \cdot y_i + b \cdot x_i}| \simeq 0 \ . \end{aligned}$$

The security criterion corresponding to linear cryptanalysis is that all functions $\varphi_a \circ G_{\mathbf{k}}$, $a \neq 0$ should be far away from all affine functions. Therefore, a necessary condition is that all F_k, $k \in \mathcal{K}$, have a high nonlinearity, i.e. a high value for $2^{n-1} - \frac{1}{2}\mathcal{L}(F)$.

Proposition 3 [33, 8] *For any function $F : \mathbf{F}_2^n \to \mathbf{F}_2^n$,*

$$\mathcal{L}(F) \geq 2^{\frac{n+1}{2}} \ .$$

In case of equality F is called almost bent (AB).

For a function F from $\mathbf{F}_2^n$ into $\mathbf{F}_2^n$, we have

$$\mathcal{L}(F) \geq 2^{\frac{n}{2}}$$

where the functions achieving this bound were called *bent* functions in [29], as a generalization of the famous Boolean bent functions.

The minimum value of $\mathcal{L}(F)$ where F is a function from $\mathbf{F}_2^n$ into $\mathbf{F}_2^n$ can only be achieved when n is odd. For even n, some functions with $\mathcal{L}(F) = 2^{\frac{n}{2}+1}$ are known and it is conjectured that this value is the minimum [11, 32].

A particular property of almost bent functions is that their Walsh spectrum is unique.

Proposition 4 *[8] The Walsh spectrum of an almost bent function F from $\mathbf{F}_2^n$ into $\mathbf{F}_2^n$ takes the values 0 and $\pm 2^{\frac{n+1}{2}}$ only.*

This property implies that any almost bent function is *almost perfect nonlinear*, but the converse is false.

2.5 Higher-order differential cryptanalysis

Higher-order differential cryptanalysis was introduced by Knudsen [20]. It exploits the fact that the reduced cipher $G_{\mathbf{k}}$ has a constant t-th derivative. Assume that there exists a t-dimensional subspace $V \subset \mathbf{F}_2^n$ such that for any $\mathbf{k} = (k_1, \ldots, k_{r-1})$ we have

$$D_V G_{\mathbf{k}} = c \tag{1}$$

where c is a constant which does not depend on $\mathbf{k}$. In accordance with definition 2, we have for any subspace V [23]

$$D_V F(x) = \sum_{x \in V} F(x) \text{ for all } x \in \mathbf{F}_2^n \ ,$$

where the above sum is an addition over $\mathbf{F}_2^n$. Then, we derive the following discriminator for $\mathcal{G}$ with respect to the set $(x_0, \ldots, x_{2^t-1})$ of elements of any coset of V, $x_0 + V$ with $x_0 \in \mathbf{F}_2^n$:

$$\mathcal{D}(y_1, \ldots, y_{2^t}) = 1 \text{ if and only if } \sum_{i=1}^{2^t} y_i = c \ .$$

Then a basic higher order differential attack can be described as follows:

1. Select a random plaintext $x_0 \in \mathbf{F}_2^n$ and get the ciphertexts c_v corresponding to all plaintexts $x_0 + v$, $v \in V$.
2. Compute c with the reduced cipher using any round keys (e.g. $k_1, \ldots, k_{r-1} = 0$).
3. For each candidate round key $\widehat{k}$, compute

$$\sigma(\widehat{k}) = \sum_{x \in V} F_{\widehat{k}}^{-1}(c_v) \ .$$

Any key $\widehat{k}$ for which $\sigma(\widehat{k}) = c$ is a candidate for the last round key. If the attack returns several round keys, it could be repeated for different values of x_0. The running-time of the attack corresponds to $\#\mathcal{K} \cdot 2^t$ evaluations of F^{-1} where $\#\mathcal{K}$ is the size of the round key space and t is the dimension of V. It requires the knowledge of 2^t chosen plaintexts.

The main problem in this attack is then to find a subspace V satisfying (1) and having the lowest possible dimension. A natural candidate for V arises when the degree of the reduced cipher is known.

Definition 5 *The* degree *of a function F from $\mathbf{F}_2^n$ into $\mathbf{F}_2^n$ is the maximum degree of its Boolean components:*

$$deg(F) = \max_{1 \le i \le n} deg(\varphi_{e_i} \circ F)$$

where $(e_1, \ldots, e_n)$ denotes the canonical basis of $\mathbf{F}_2^n$.

Actually, we have

Proposition 5 *[23] Let F be a function from $\mathbf{F}_2^n$ into $\mathbf{F}_2^n$ of degree d. Then, for any $(d+1)$-dimensional subspace $V \subset \mathbf{F}_2^n$, we have*

$$D_V F(x) = 0 \quad \textit{for all } x \in \mathbf{F}_2^n \ .$$

Note that the dimension of the smallest subspace V satisfying $D_V F = 0$ may be smaller than $\deg(F)+1$. Since

$$\max_{\mathbf{k}\in\mathcal{K}^{r-1}} \deg(G_{\mathbf{k}}) \leq \left(\max_{k\in\mathcal{K}} \deg(F_k)\right)^{r-1},$$

it follows that a cipher is vulnerable to higher-order differential cryptanalysis when its round function has a low degree. This property was used by Jakobsen and Knudsen [18] for breaking a cipher example proposed in [31], whose round function is a quadratic AB permutation. However, this condition is not sufficient and a stronger requirement on the round function will be exhibited in the following.

All three properties involved in differential, linear and higher-order differential attacks are invariant under both right and left composition by a linear permutation of $\mathbf{F}_2^n$ [30]. Then, they only concern the *confusion part* of the round function.

3 Divisibility of the Walsh spectrum and degree of a composed function

In this section, we focus on the degree of a function $F' \circ F$ where F and F' are two mappings from $\mathbf{F}_2^n$ into $\mathbf{F}_2^n$. We show that the trivial bound

$$deg(F' \circ F) \leq deg(F') \cdot deg(F)$$

can be improved when the values occurring in the Walsh spectrum of F are divisible by a high power of 2. This situation especially occurs when F is an almost bent function (see Prop 4).

Definition 6 *The Walsh spectrum of a function F from $\mathbf{F}_2^n$ into $\mathbf{F}_2^m$ is said to be* 2^ℓ-divisible *if all its values are divisible by 2^ℓ. Moreover, it is said* exactly 2^ℓ-divisible *if, additionally, it contains at least one value which is not divisible by $2^{\ell+1}$.*

The divisibility of the values occurring in the Walsh spectrum of a function F provides an upper bound on its degree. As a direct consequence of [6, Lemma 3], we obtain the following bound:

Proposition 6 *Let F be a function from $\mathbf{F}_2^n$ into $\mathbf{F}_2^m$. If the Walsh spectrum of F is 2^ℓ-divisible, then $deg(F) \leq n - \ell + 1$.*

The i-th Boolean component of $F' \circ F$ can be expressed as $f'(F_1(x), \ldots, F_n(x))$, where f' is the i-th Boolean component of F' and $(F_1, \ldots, F_n)$ denote the Boolean components of F. Using the algebraic normal form of f', we can write this function as $\sum_J \prod_{j \in J} F_j(x)$ where each product involves at most $deg(f')$ Boolean components of F. We deduce that the degree of $F' \circ F$ cannot exceed the degree of a product of $deg(F')$ Boolean components of F.

Now, we focus on the Walsh spectrum of the product of some Boolean functions. We use the following lemma.

Lemma 1 *Let $f_1, \ldots, f_k$ be k Boolean functions of n variables, with $k > 0$. We have*

$$\mathcal{F}(\sum_{i=1}^{k} f_i) = 2^{n-1}\left[(-1)^k + 1\right] + \sum_{I \subset \{1,\ldots,k\}} (-2)^{|I|-1} \mathcal{F}(\prod_{i \in I} f_i) \ . \tag{2}$$

Moreover, for any nonzero α in $\mathbf{F}_2^n$, we have

$$\mathcal{F}(\sum_{i=1}^{k} f_i + \varphi_\alpha) = \sum_{I \subset \{1,\ldots,k\}} (-2)^{|I|-1} \mathcal{F}(\prod_{i \in I} f_i + \varphi_\alpha) \ .$$

Proof.

- We first prove Relation (2) by induction on k. It obviously holds for $k = 1$. For $k = 2$, we have

  $$wt(f_1 + f_2) = wt(f_1) + wt(f_2) - 2wt(f_1 f_2) \ ,$$

  leading to (2). Now, we suppose that Relation (2) is satisfied for all $i \leq k$, and we want to show that it holds for $(k+1)$ functions. We have

  $$\mathcal{F}(\sum_{i=1}^{k+1} f_i) = \mathcal{F}(\sum_{i=1}^{k} f_i) + \mathcal{F}(f_{k+1}) - 2\mathcal{F}((\sum_{i=1}^{k} f_i) f_{k+1}) + 2^n \ .$$

  By applying the induction hypothesis to $f_1, \ldots f_k$ and to $f_1 f_{k+1}, \ldots f_k f_{k+1}$, we obtain

  $$\begin{aligned} \mathcal{F}(\sum_{i=1}^{k+1} f_i) &= 2^{n-1}\left[(-1)^k + 1\right] + \sum_{I \subset \{1,\ldots,k\}} (-2)^{|I|-1} \mathcal{F}(\prod_{i \in I} f_i) + \mathcal{F}(f_{k+1}) \\ &\quad -2^n\left[(-1)^k + 1\right] + \sum_{I \subset \{1,\ldots,k\}} (-2)^{|I|} \mathcal{F}(\prod_{i \in I} f_i f_{k+1}) + 2^n \\ &= 2^{n-1}[(-1)^{k+1} + 1] + \sum_{I \subset \{1,\ldots,k+1\}} (-2)^{|I|-1} \mathcal{F}(\prod_{i \in I} f_i) \ . \end{aligned}$$

- Now, we consider a nonzero element α in $\mathbf{F}_2^n$. For any Boolean function f, we have from (2):

$$\mathcal{F}(f+\varphi_\alpha) = \mathcal{F}(f) + \mathcal{F}(\varphi_\alpha) - 2\mathcal{F}(f\varphi_\alpha) + 2^n = \mathcal{F}(f) - 2\mathcal{F}(f\varphi_\alpha) + 2^n \ .$$

Therefore, we have

$$\mathcal{F}(\sum_{i=1}^{k} f_i + \varphi_\alpha) = \mathcal{F}(\sum_{i=1}^{k} f_i) - 2\mathcal{F}(\sum_{i=1}^{k} f_i\varphi_\alpha) + 2^n \ .$$

Now, Relation (2) applied to $f_1, \ldots, f_k, \varphi_\alpha$ leads to

$$\begin{aligned}
\mathcal{F}(f+\varphi_\alpha) &= 2^{n-1}\left[(-1)^{k+1}+1\right] + \mathcal{F}(\varphi_\alpha) + \sum_{I\subset\{1,\ldots,k\}} (-2)^{|I|-1}\mathcal{F}(\prod_{i\in I} f_i) \\
&\quad + \sum_{I\subset\{1,\ldots,k\}} (-2)^{|I|}\mathcal{F}(\prod_{i\in I} f_i\varphi_\alpha) \\
&= 2^{n-1}\left[(-1)^{k+1}+1\right] + \sum_{I\subset\{1,\ldots,k\}} (-2)^{|I|-1}\left[\mathcal{F}(\prod_{i\in I} f_i) - 2\mathcal{F}(\prod_{i\in I} f_i\varphi_\alpha)\right] \\
&= 2^{n-1}\left[(-1)^{k+1}+1\right] + \sum_{I\subset\{1,\ldots,k\}} (-2)^{|I|-1}\left[\mathcal{F}(\prod_{i\in I} f_i + \varphi_\alpha) - 2^n\right] \\
&= 2^{n-1}\left[(-1)^{k+1}+1\right] + 2^{n-1}\left[\sum_{i=1}^{k}(-2)^i\binom{k}{i}\right] \\
&\quad + \sum_{I\subset\{1,\ldots,k\}} (-2)^{|I|-1}\mathcal{F}(\prod_{i\in I} f_i + \varphi_\alpha) \\
&= \sum_{I\subset\{1,\ldots,k\}} (-2)^{|I|-1}\mathcal{F}(\prod_{i\in I} f_i + \varphi_\alpha) \ .
\end{aligned}$$

◇

Using the previous relation between the Walsh coefficients of the sum of k Boolean functions and the Walsh coefficients of their product, we obtain:

Theorem 1 *Let $f_1, \ldots, f_k$ be k Boolean functions of n variables, with $k > 0$. Suppose that for any subset I of $\{1, \ldots, k\}$ we have*

$$\forall \alpha \in \mathbf{F}_2^n, \ \ \mathcal{F}(\sum_{i\in I} f_i + \varphi_\alpha) \equiv 0 \bmod 2^\ell \ .$$

Then, for any $I \subset \{1, \ldots, k\}$ of size at most ℓ, we have

$$\forall \alpha \in \mathbf{F}_2^n, \ \ \mathcal{F}(\prod_{i\in I} f_i + \varphi_\alpha) \equiv 0 \bmod 2^{\ell+1-|I|} \ . \tag{3}$$

Therefore,

$$deg(\prod_{i\in I} f_i) \leq n - \ell + |I| \ .$$

Proof.
We prove Relation (3) by induction on the size of I. The result obviously holds for $|I| = 1$. We now assume that (3) holds for any I with $|I| \leq w$ and we consider a subset $I \subset \{1, \ldots, k\}$ of size $w+1$. From Lemma 1, we have for any $\alpha \in \mathbf{F}_2^n$

$$(-2)^w \mathcal{F}(\prod_{i\in I} f_i + \varphi_\alpha) \equiv \mathcal{F}(\sum_{i\in I} f_i + \varphi_\alpha) - \sum_{J\subset I, J\neq I} (-2)^{|J|-1} \mathcal{F}(\prod_{j\in J} f_j + \varphi_\alpha) \bmod 2^n \ .$$

From induction hypothesis, we derive that

$$(-2)^w \mathcal{F}(\prod_{i\in I} f_i + \varphi_\alpha) \equiv \mathcal{F}(\sum_{i\in I} f_i + \varphi_\alpha) \bmod 2^\ell \ .$$

Therefore, we have

$$\mathcal{F}(\prod_{i\in I} f_i + \varphi_\alpha) \equiv 0 \bmod 2^{\ell-w} \ .$$

The upper bound on the degree is a direct consequence of (3) and Proposition 6. ◇

By applying the previous theorem to the Boolean components of a mapping F from $\mathbf{F}_2^n$ into $\mathbf{F}_2^n$, we derive the following corollary.

Corollary 1 *Let F be a function from $\mathbf{F}_2^n$ into $\mathbf{F}_2^n$ such that its Walsh spectrum is 2^ℓ-divisible. Then, the degree of the product of any t Boolean components of F is at most $n-\ell+t$.*

Therefore, for any function F' from $\mathbf{F}_2^n$ into $\mathbf{F}_2^n$, we have

$$deg(F' \circ F) \leq n - \ell + deg(F') \ .$$

As a consequence, when F is an almost bent function, we obtain

$$deg(F' \circ F) \leq \frac{n-1}{2} + deg(F') \ .$$

The result presented in Corollary 1 was already proved for the particular case of *power functions*. Here, we identify $\mathbf{F}_2^n$ with the finite field with 2^n elements, $\mathbf{F}_{2^n}$. In this context, any function F from $\mathbf{F}_2^n$ into $\mathbf{F}_2^n$ can be expressed as a unique univariate polynomial in $\mathbf{F}_{2^n}[X]$, $F(X) = \sum_{u=0}^{2^n-1} a_u X^u$. The degree of F (in the sense of Definition 5) is given by $deg(F) = \max_{u, a_u \neq 0} w_2(u)$, where $w_2(u)$ denotes the number of ones in the 2-adic expansion of u,

$u = \sum_{i=0}^{n-1} u_i 2^i$. The case of power functions is of great interest since all known highly nonlinear mappings are equivalent to some power functions $x \mapsto x^s$ over $\mathbf{F}_{2^n}$. Now, if we write F' as a univariate polynomial $F'(X) = \sum_{u=0}^{2^n-1} a_u X^u$, we obtain for $F : x \mapsto x^s$ that $F' \circ F(x) = \sum_{u=0}^{2^n-1} a_u X^{us \bmod (2^n-1)}$. Therefore, $deg(F' \circ F) \leq \max_{u, a_u \neq 0} w_2(us \bmod (2^n-1))$. This bound is related to the divisibility of the Walsh spectrum of F by the following proposition [4, Coro. 2]. The result is directly derived from McEliece's theorem which provides the weight divisibility of a cyclic code [27]. We refer to [7, 4] for the link between cyclic codes and power functions.

Proposition 7 *Let $F : x \mapsto x^s$ be a power function over $\mathbf{F}_{2^n}$. Then, the Walsh spectrum of F is 2^ℓ-divisible if and only if, for any integer u, $1 \leq u \leq 2^n - 1$, we have*

$$w_2(us \bmod (2^n-1)) \leq n - \ell + w_2(u) \ .$$

4 Cryptanalysis of 5-round Feistel ciphers using highly nonlinear functions

We now focus on 5-round Feistel ciphers. In a Feistel cipher with block size $2n$, the round function is defined by

$$\begin{array}{rccc} F_k: & \mathbf{F}_2^n \times \mathbf{F}_2^n & \rightarrow & \mathbf{F}_2^n \times \mathbf{F}_2^n \\ & (L, R) & \mapsto & (R, L + S_k(R)) \end{array}$$

where S_K is a function from $\mathbf{F}_2^n$ into $\mathbf{F}_2^n$ called the confusion function. In the following, L_i (resp. R_i) denotes the left part (resp. right part) of the output of the i-th round.

In a 5-round Feistel cipher, the right part of the output of the third round, R_3, can be derived from the ciphertext (L_5, R_5) and the last-round key:

$$R_3 = R_5 + S_{k_5}(L_5) \ .$$

Moreover, when we consider any plaintext (x, c_0) whose right part is a given constant c_0, R_3 can be computed from x by only two iterations of the confusion function :

$$R_3(x) = x + c_1 + S_{k_3}(c_0 + S_{k_2}(x + c_1))$$

where x stands for the left half of the plaintext and c_0, c_1 are some constants.

When the Walsh spectra of all functions S_k are 2^ℓ-divisible, we can apply Corollary 1. Then, we obtain the following upper bound for the degree of R_3:

$$deg(R_3) \leq n - \ell + \max_{k \in \mathcal{K}} \deg(S_k).$$

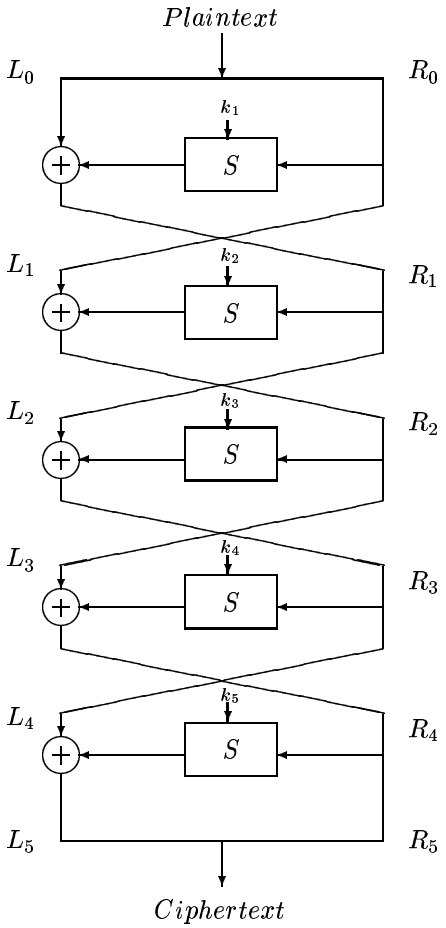

Figure 2: A 5-round Feistel cipher

Let $\delta = \min(\max_{k \in \mathcal{K}} deg(S_k)^2 + 1, n - \ell + 1 + \max_{k \in \mathcal{K}} deg(S_k))$, if we consider the attack described in Section 2.5, we have exhibited a new attack on the last round key with average running-time of $\#\mathcal{K} \cdot 2^{\delta}$, where $\#\mathcal{K}$ is the size of the round key space. This attack is feasible as soon as $\delta \leq n$.

Algorithm 2
INPUT: $(L_i, R_i)_{1\leq i\leq 2^\delta}$: *the ciphertexts corresponding to the plaintexts* $(x_i, c_0)_{1\leq i\leq 2^\delta}$, *where* c_0 *is any fixed element of* $\mathbf{F}_2^n$ *and* $(x_i)_{1\leq i\leq 2^\delta}$ *is a* δ*-dimensional subspace of* $\mathbf{F}_2^n$
OUTPUT: *A set of candidates for the last-round key* k_5.

For all $k \in \mathcal{K}$
  For i *from 1 to* 2^δ *do* $y_i \leftarrow R_i + S_k(L_i)$
  If $\sum_{i=1}^{2^\delta} y_i = 0$ *then return* k.

For example, if all S_k are almost bent, a higher order differential attack can be performed except when $\max_{k\in\mathcal{K}} deg(S_k) = \dfrac{n+1}{2}$, i.e., when S_k is an almost bent function of maximum degree. A similar situation occurs when S is a function of an even number of variables which has the highest known nonlinearity, $\mathcal{L}(S_k) = 2^{\frac{n}{2}+1}$. All known functions satisfying this property are equivalent to one of the power functions given in Table 1 (or to one of their inverses) [11].

exponents s	condition on n	divisibility	
$2^{n-1}-1$	$n \equiv 0 \bmod 2$	2^2	[22]
2^k+1, with $\gcd(k,n)=2$ and $k<\frac{n}{2}$	$n \equiv 2 \bmod 4$	$2^{\frac{n}{2}+1}$	[12, 30]
$2^{2k}-2^k+1$, with $\gcd(k,n)=2, k<\frac{n}{2}$	$n \equiv 2 \bmod 4$	$2^{\frac{n}{2}+1}$	[19]
$2^{\frac{n}{2}}+2^{\frac{n+2}{4}}+1$	$n \equiv 2 \bmod 4$	$2^{\frac{n}{2}+1}$	[10]
$2^{\frac{n}{2}}+2^{\frac{n}{2}-1}+1$	$n \equiv 2 \bmod 4$	$2^{\frac{n}{2}+1}$	[10]
$\sum_{i=0}^{n/2} 2^{ik}$, with $\gcd(k,n)=1, k<\frac{n}{2}$	$n \equiv 0 \bmod 4$	$2^{\frac{n}{2}}$	[11]
$2^{\frac{n}{2}}+2^{\frac{n}{4}}+1$	$n \equiv 4 \bmod 8$	$2^{\frac{n}{2}}$	[11]

Table 1: Known power permutations x^s on $\mathbf{F}_{2^n}$, n even, with the highest nonlinearity and exact divisibility of their Walsh spectra

All optimal functions for n even are such that their Walsh spectra are divisible either by $2^{\frac{n}{2}}$ or by $2^{\frac{n}{2}+1}$, except the inverse function whose Walsh spectrum is exactly 4-divisible. If the Walsh spectrum of the confusion function S_k is $2^{\frac{n}{2}+1}$-divisible, then $deg(S_k) \leq n/2$. Therefore, the attack is always feasible. When the Walsh spectrum of S_k is $2^{\frac{n}{2}}$-divisible, the attack can be performed except if $deg(S_k) \in \{n/2, n/2+1\}$. These results are summed up in Table 2 (general case).

It is also possible to improve this attack when the round key in the Feistel cipher is inserted by addition, i.e., $S_k(x) = S(x+k)$. In that case, we obtain the following expression for R_3:

$$R_3(x) = x + c_1 + S(c_0 + k_3 + S(x + c_1 + k_2)).$$

function S_k	divisibility	General case	
		differential order	feasibility
n odd $\mathcal{L}(S_k) = 2^{\frac{n+1}{2}}$	$2^{\frac{n+1}{2}}$	$\frac{n+1}{2} + \max_{k \in \mathcal{K}} \deg(S_k)$	except for $\deg(S_k) = \frac{n+1}{2}$
n even $\mathcal{L}(S_k) = 2^{\frac{n}{2}+1}$	$2^{\frac{n}{2}+1}$	$\frac{n}{2} + \max_{k \in \mathcal{K}} \deg(S_k)$	always feasible
	$2^{\frac{n}{2}}$	$\frac{n}{2} + 1 + \max_{k \in \mathcal{K}} \deg(S_k)$	except for $\deg(S_k) \in \{\frac{n}{2}, \frac{n}{2}+1\}$

Table 2: Higher order differential attack on a 5-round Feistel cipher using a highly nonlinear confusion function S_k : general case

Let G be the function defined by $G : x \mapsto S(k_3 + c_0 + S(x + c_1 + k_2))$ and let G' be defined by $G' : x \mapsto S(k_3 + c_0 + S(x))$. Then, we know that $deg(G') \leq n - \ell + deg(S)$. The expression of G' shows that the terms containing the constants c_0 or k_3 are the result of the product of at most $(deg(S) - 1)$ Boolean components of S. Thus, their degree is at most $n - \ell + deg(S) - 1$. We then deduce that the terms of maximal degree in G' are independent of the constants. In particular we have for any subspace V of dimension $(n - \ell + deg(S))$:

$$\forall a \in \mathbf{F}_2^n, \quad D_V G'(a) = \sum_{v \in V} G'(a + v) = c$$

where c is independent of any kind of constants. We can see that G is obtained by translating G', so we have:

$$\forall a \in \mathbf{F}_2^n, \quad \sum_{v \in V} G(a + v) = \sum_{v \in V} G'(a + v + c_1 + k_2) = D_V G'(a + c_1 + k_2) = c.$$

The constant c can be computed, for example, with the null value for all the subkeys. The above attack requires $2^{n-\ell+deg(S)}$ pairs of plaintexts-ciphertexts and $2^{2n-\ell+deg(S)}$ evaluations for the function S. It can be performed for any almost bent function S (see Table 3).

The inverse function is very specific in this context (see Table 1). Its Walsh coefficients $\mathcal{F}(\varphi_a \circ F + \varphi_b)$ are all the values which are divisible by 4, such that

$$|\mathcal{F}(\varphi_a \circ F + \varphi_b)| \leq 2^{\frac{n}{2}+1} \quad [22].$$

Therefore its Walsh spectrum has the smallest possible divisibility for a permutation of $\mathbf{F}_2^n$, n even. Moreover, it is proved that $x \mapsto x^{2^{n-1}-1}$ is the only power permutation of $\mathbf{F}_{2^n}$ (up to equivalence) whose Walsh spectrum is exactly 4-divisible [15]. Thus, the inverse function

function S	divisibility	$S_k(x) = S(x+k)$	
		differential order	feasibility
n odd	$2^{\frac{n+1}{2}}$	$\frac{n-1}{2} + deg(S)$	always feasible
n even	$2^{\frac{n}{2}+1}$	$\frac{n}{2} - 1 + deg(S)$	always feasible
$\mathcal{L}(S) = 2^{\frac{n}{2}+1}$	$2^{\frac{n}{2}}$	$\frac{n}{2} + deg(S)$	except for $deg(S) = \frac{n}{2} + 1$

Table 3: Higher order differential attack on a 5-round Feistel cipher using a highly nonlinear confusion function S: case where the round keys are inserted by addition

is the only confusion function which is optimal with respect to all resistance criteria; it opposes the best resistance to differential, linear and higher order differential attacks. This function is used in the new block cipher standard AES.

5 Higher order differential cryptanalysis on a generalization of MISTY1

MISTY is a model of block ciphers designed by Matsui in 1997 [26]. It was proposed in two variants, MISTY1 and MISTY2, the former being the object of this study. More precisely, M'1, the version of MISTY1 reduced to 5 rounds and without FL functions (linear roud functions) is "provably secure" against both differential and linear crytanalysis, therefore the background of the attack is this simplified version of the algorithm. Indeed, in [34] it is shown that M'1 can be attacked with a 7-th order differential. Moreover in [1] the attack is extended to the case where any almost bent power function of degree 3 on $\mathbf{F}_2^7$ is used for the S_7-box. In this section we extend the use of this higher order differential attack to a generalization of the algorithm M'1 where the block size becomes variable and whose value is $16m$ bits, instead of the original one of 64 bits. In this generalization, we show that the weakness of M'1 is due to the use of an almost bent function as a round function.

5.1 The M'1 algorithm

The cipher M'1 is depicted in Figure 3. In the following x_0 and x_1 are the left and right halves of the plaintext. Similarly, (x_{i+1}, x_i) denotes the intermediate value after i rounds.

Notation 1 *Let u be a $16m$ bit word. We denote by u^L, u^R, u^{L_k}, u^{R_k}, respectively the left and right halves of u and the k left and right most bits. The $\|$ symbol stands for the concatenation of two binary words.*

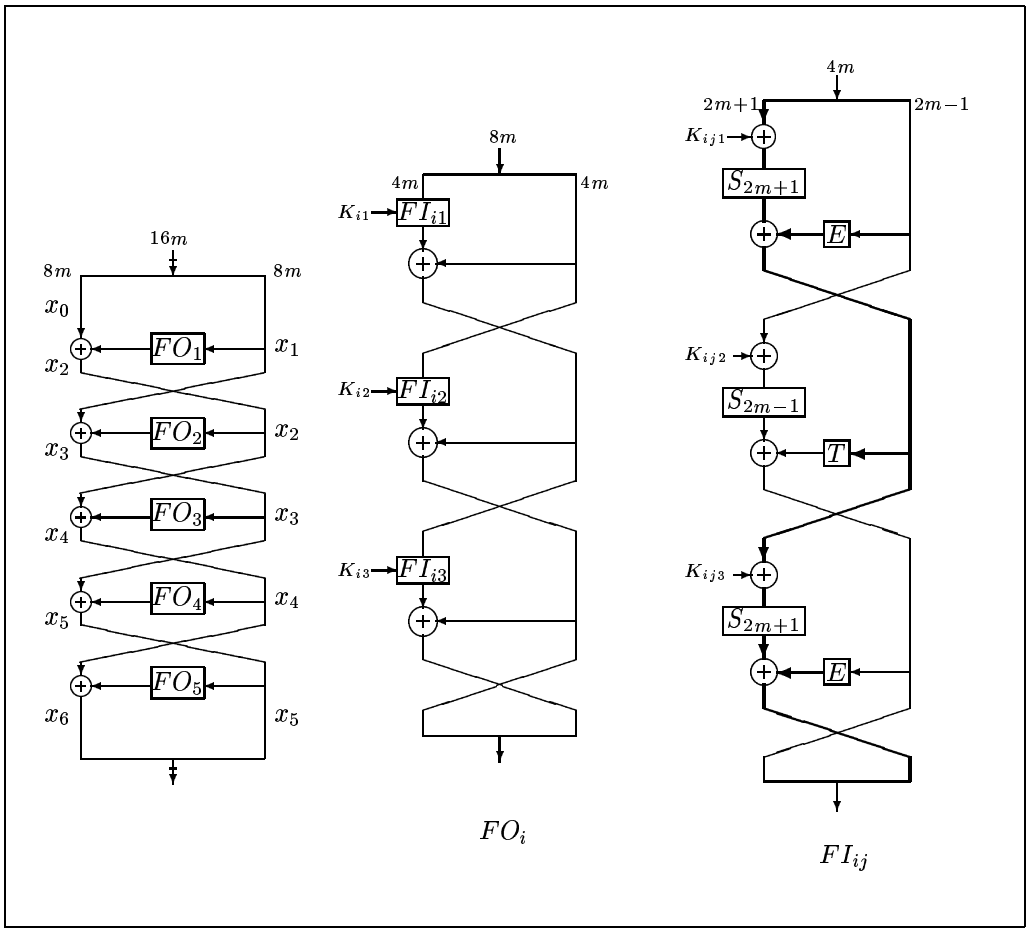

Figure 3: The 5-round Feistel cipher M'1 with equivalent key schedule

The cipher uses both following functions. The "zero-extend" function is defined by:

$$\begin{array}{rlcl} E: & \mathbf{F}_2^{2m-1} & \to & \mathbf{F}_2^{2m+1} \\ & (u_1, \cdots, u_{2m-1}) & \mapsto & (u_1, \cdots, u_{2m-1}, 0, 0) \end{array}$$

and the "truncate" function by:

$$\begin{array}{rlcl} T: & \mathbf{F}_2^{2m+1} & \to & \mathbf{F}_2^{2m-1} \\ & (u_1, \cdots, u_{2m+1}) & \mapsto & (u_1, \cdots, u_{2m-1}). \end{array}$$

The nonlinear part of the cipher consists of two permutations, S_{2m-1} and S_{2m+1} respectively defined over $\mathbf{F}_{2^{2m-1}}$ and $\mathbf{F}_{2^{2m+1}}$. In the original cipher, we have $S_7(x) = A(x^{81})$ over

$\mathbf{F}_{2^7}$ where A is an affine permutation and and S_9 is a quadratic almost bent permutation over $\mathbf{F}_{2^9}$.

5.2 The general principal of the attack

Let V be the $(2m-1)$-dimensional subspace of plaintexts of $16m$ bits whose form is $(0_{6m+1} \parallel x \parallel 0_{8m})$ where x is in $\mathbf{F}_2^{2m-1}$. W denotes the subspace such that $V \times W = \mathbf{F}_2^{16m}$. We are interested in ciphering plaintexts whose form is:

$$\underbrace{(0_{6m+1} \parallel x \parallel 0_{8m})}_{P} + \underbrace{(w_0 \parallel 0_{2m-1} \parallel w_1)}_{w}$$

where $P \in V$ and w is a fixed constant in W.

We now consider the function G_K defined as follows:

$$\begin{array}{rccc} G_K: & \mathbf{F}_2^{2m-1} & \rightarrow & \mathbf{F}_2^{2m-1} \\ & x & \mapsto & x_4^{L_{2m-1}}. \end{array}$$

To sum up the higher order differential attack proposed in [34], with $m = 4$ and the original S_7 and S_9 boxes, we can say that the 7-th order derivative of G_K with respect to V is a constant independent from the secret key K:

$$\forall w \in W, \quad \sum_{x \in V} G_K(x+w) = c. \tag{4}$$

Here, we show and explain how this property can be generalized to different block sizes.

First, we give the exact expression of $x_4^{L_{2m-1}}$.

5.3 The detailed structure of $x_4^{L_{2m-1}}$

5.3.1 Expressions of FO_i and FI_{ij} outputs

fAs shown in Figure 4, for $FO_i(x_i)$ we have:

$$\begin{array}{rcl} [FO_i(x_i)]^L & = & FI_{i2}(x_i^R, K_{i2}) + FI_{i1}(x_i^L, K_{i1}) + x_i^R \\ [FO_i(x_i)]^R & = & FI_{i3}(FI_{i1}(x_i^L, K_{i1}) + x_i^R, K_{i3}) + [FO_i(x_i)]^L \end{array}$$

and $FI_{ij}(h)$:

$$\begin{array}{ll} 1: & S_{2m+1}(h^{L_{2m+1}} + K_{ij1}) \\ 2: & S_{2m-1}(h^{R_{2m-1}} + K_{ij2}) \\ 3: & S_{2m+1}\left(S_{2m+1}(h^{L_{2m+1}} + K_{ij1}) + E(h_{2m-1}^R) + K_{ij3}\right) \end{array}$$

At the end we obtain:

$$\begin{array}{rcl} [FI_{ij}(h)]^{L_{2m-1}} & = & S_{2m-1}(h^{R_{2m-1}} + K_{ij2}) + T \circ S_{2m+1}(h^{L_{2m+1}} + K_{ij1}) + h^{R_{2m-1}} \\ [FI_{ij}(h)]^{R_{2m+1}} & = & S_{2m+1}\left(S_{2m+1}(h^{L_{2m+1}} + K_{ij1}) + E(h^{R_{2m-1}}) + K_{ij3}\right) \\ & & + E\left([FI_{ij}(h)]^{L_{2m-1}}\right). \end{array}$$

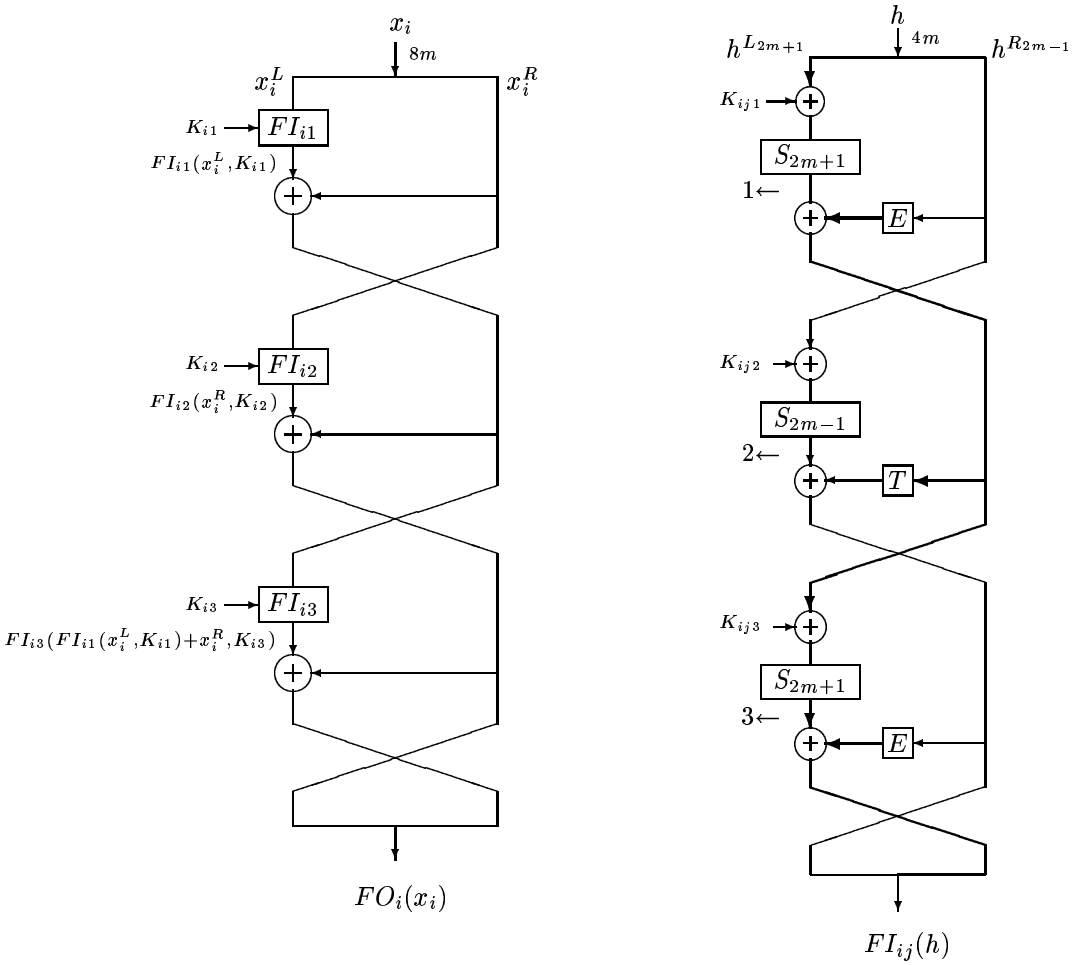

Figure 4: Transformations in details

The following paragraph is devoted to the calculus of $x_4^{L_{2m-1}}$. c_i will denote any kind of constants.

5.3.2 The first round

$$x_2 = FO_1(x_1, K_1) + x_0 = (\underbrace{c_2}_{4m} \parallel \underbrace{c_1}_{2m+1} \parallel \underbrace{x + c_0}_{2m-1}).$$

5.3.3 The second round

$$x_3 = FO_2(x_2, K_2) + x_1 = (\mu \parallel \lambda) + x_1,$$

with:

$$\begin{aligned}\mu \;=\; [FO_2(x_2,K_2)]^L &= FI_{22}(x_2^R,K_{22}) + FI_{21}(x_2^L,K_{21}) + x_2^R \\ &= FI_{22}(c_1 \parallel x+c_0) + \underbrace{cst + (c_1 \parallel x+c_0)}_{c_3 \parallel x+c_4}\end{aligned}$$

where $FI_{21}(x_2^L,K_{21}) + x_2^R = cst + (c_1 \parallel x+c_0) = (c_3 \parallel x+c_4)$.

$$\lambda = [FO_2(x_2,K_2)]^R = FI_{23}(c_3 \parallel x+c_4) + FI_{22}(c_1 \parallel x+c_0) + (c_3 \parallel x+c_4).$$

On the other hand, we have:

$$\begin{aligned}[FI_{22}(c_1 \parallel x+c_0)]^{L_{2m-1}} &= S_{2m-1}(x+c_0+K_{222}) + T \circ S_{2m+1}(c_1+K_{221}) + x + c_0 \\ &= S_{2m-1}(x+c_5) + x + c_6,\end{aligned}$$

hence:

$$\mu^{L_{2m-1}} = S_{2m-1}(x+c_5) + x + c_9\ ;$$

By the same way we obtain:

$$\begin{aligned}[FI_{23}(c_3 \parallel x+c_4)]^{L_{2m-1}} &= S_{2m-1}(x+c_4+K_{232}) + T \circ S_{2m+1}(c_3+K_{231}) + x + c_4 \\ &= S_{2m-1}(x+c_7) + x + c_8,\end{aligned}$$

hence:

$$\lambda^{L_{2m-1}} = S_{2m-1}(x+c_7) + S_{2m-1}(x+c_5) + c_{10}.$$

We also have:

$$\begin{aligned}[FI_{22}(c_1 \parallel x+c_0)]^{R_{2m+1}} &= S_{2m+1}\left(S_{2m+1}(c_1+K_{221}) + E(x+c_0) + K_{223}\right) \\ &\quad + E \circ S_{2m-1}(x+c_5) + E(x) + E(c_6) \\ &= S_{2m+1}(E(x)+c_{11}) + E \circ S_{2m-1}(x+c_5) + E(x) + c_{12} \\ [FI_{23}(c_3 \parallel x+c_4)]^{R_{2m+1}} &= S_{2m+1}\left(S_{2m+1}(c_3+K_{221}) + E(x+c_4) + K_{223}\right) \\ &\quad + E \circ S_{2m-1}(x+c_7) + E(x) + E(c_8) \\ &= S_{2m+1}(E(x)+c_{13}) + E \circ S_{2m-1}(x+c_7) + E(x) + c_{14},\end{aligned}$$

which leads to the following expressions for $\mu^{R_{2m+1}}$ and $\lambda^{R_{2m+1}}$:

$$\begin{aligned}\mu^{R_{2m+1}} &= S_{2m+1}\left(E(x)+c_{11}\right) + E \circ S_{2m-1}(x+c_5) + c_{15} \\ \lambda^{R_{2m+1}} &= S_{2m+1}\left(E(x)+c_{13}\right) + E \circ S_{2m-1}(x+c_7) + E(x) + c_{14} \\ &\quad + S_{2m+1}\left(E(x)+c_{11}\right) + E \circ S_{2m-1}(x+c_5) + E(x) + c_{12} + E(x) + E(c_4) \\ &= S_{2m+1}\left(E(x)+c_{13}\right) + E \circ S_{2m-1}(x+c_7) \\ &\quad + S_{2m+1}\left(E(x)+c_{11}\right) + E \circ S_{2m-1}(x+c_5) + E(x) + c_{16}.\end{aligned}$$

So the final result is:

$$\begin{aligned}\mu &= (S_{2m-1}(x+c_5) + x + c_9 \parallel S_{2m+1}\left(E(x)+c_{11}\right) + E \circ S_{2m-1}(x+c_5) + c_{15}) \\ \lambda &= (S_{2m-1}(x+c_7) + S_{2m-1}(x+c_5) + c_{10} \parallel S_{2m+1}\left(E(x)+c_{13}\right) + E \circ S_{2m-1}(x+c_7) \\ &\quad + S_{2m+1}\left(E(x)+c_{11}\right) + E \circ S_{2m-1}(x+c_5) + E(x) + c_{16})\,.\end{aligned}$$

As $x_3 = (\mu \parallel \lambda) + x_1$, we have

$$x_3 = (\mu + c_{17} \parallel \lambda + c_{18}).$$

5.3.4 The third round

We are interested in $x_4^{L_{2m-1}}$, the $(2m-1)$ most left bits of x_4. If we note γ the left half of FO_3 output, we want to determine $\gamma^{L_{2m-1}}$.

$$\gamma = [FO_3(x_3)]^L = FI_{32}(\lambda + c_{18}) + FI_{31}(\mu + c_{17}) + \lambda + c_{18}.$$

and

$$\begin{aligned}
[FI_{31}(\mu + c_{17})]^{L_{2m-1}} &= S_{2m-1}(\mu^{R_{2m-1}} + c_{17}^{R_{2m-1}} + K_{312}) \\
&\quad + T \circ S_{2m+1}(\mu^{L_{2m+1}} + c_{17}^{L_{2m+1}} + K_{311}) + \mu^{R_{2m-1}} + c_{17}^{R_{2m-1}} \\
[FI_{32}(\lambda + c_{18})]^{L_{2m-1}} &= S_{2m-1}(\lambda^{R_{2m-1}} + c_{18}^{R_{2m-1}} + K_{322}) \\
&\quad + T \circ S_{2m+1}(\lambda^{L_{2m+1}} + c_{18}^{L_{2m+1}} + K_{321}) + \lambda^{R_{2m-1}} + c_{18}^{R_{2m-1}},
\end{aligned}$$

so we have:

$$\begin{aligned}
\gamma^{L_{2m-1}} &= \mu^{R_{2m-1}} + \lambda^{R_{2m-1}} + \lambda^{L_{2m-1}} + c_{19} + T \circ S_{2m+1}(\mu^{L_{2m+1}} + c_{20}) \\
&\quad + T \circ S_{2m+1}(\lambda^{L_{2m+1}} + c_{21}) + S_{2m-1}(\mu^{R_{2m-1}} + c_{22}) + S_{2m-1}(\lambda^{R_{2m-1}} + c_{23}).
\end{aligned}$$

As $x_4^{L_{2m-1}} = \gamma^{L_{2m-1}} + \underbrace{x_2^{L_{2m-1}}}_{=cst}$, we obtain

$$\begin{aligned}
x_4^{L_{2m-1}} = {} & \mu^{R_{2m-1}} + \lambda^{R_{2m-1}} + \lambda^{L_{2m-1}} + c_{24} + T \circ S_{2m+1}(\mu^{L_{2m+1}} + c_{20}) \\
& + T \circ S_{2m+1}(\lambda^{L_{2m+1}} + c_{21}) + S_{2m-1}(\mu^{R_{2m-1}} + c_{22}) + S_{2m-1}(\lambda^{R_{2m-1}} + c_{23}).
\end{aligned} \tag{5}$$

where

$$\begin{aligned}
\mu^{L_{2m-1}} &= S_{2m-1}(x + c_5) + x + c_9 \\
\lambda^{L_{2m-1}} &= S_{2m-1}(x + c_7) + S_{2m-1}(x + c_5) + c_{10} \\
\mu^{R_{2m+1}} &= S_{2m+1}\left(E(x) + c_{11}\right) + E \circ S_{2m-1}(x + c_5) + c_{15} \\
\lambda^{R_{2m+1}} &= S_{2m+1}\left(E(x) + c_{13}\right) + E \circ S_{2m-1}(x + c_7) \\
&\quad + S_{2m+1}\left(E(x) + c_{11}\right) + E \circ S_{2m-1}(x + c_5) + E(x) + c_{16}.
\end{aligned}$$

where all c_i are some constants depending on the round keys.

5.4 The analysis of the degree of $x_4^{L_{2m-1}}$

The aim of our study is to determine the degree of the Boolean components of $x_4^{L_{2m-1}}$, and to show that it is not as high as we can think at first.

We restrict our study to the case where S_{2m+1} is a quadratic function, as in the original cipher. We assume that the almost bent permutation S_{2m-1} can be written as $S_{2m-1}(x) = L(x^e)$ where L is a linear permutation. We denote by d the degree of S_{2m-1} and we assume that $2d < 2m-1$, i.e., that the degree of S_{2m-1} differs from the highest possible degree for an almost bent function over $\mathbf{F}_2^{2m-1}$. These conditions obviously imply that we can neglect the terms $T \circ S_{2m+1}(\mu^{L_{2m+1}} + c_{20}) + T \circ S_{2m+1}(\lambda^{L_{2m+1}} + c_{21})$ in (5) for a $(2m-1)$-th order differential.

Notation 2 *We denote by $[F]_d$ the terms in the algebraic normal form of F whose degree are at least d.*

It appears that the terms of degree $2m-1$ in $x_4^{L_{2m-1}}$ correspond to

$$\left[x_4^{L_{2m-1}}\right]_{2m-1} = \left[S_{2m-1}(\mu^{R_{2m-1}} + c_{22})\right]_{2m-1} + \left[S_{2m-1}(\lambda^{R_{2m-1}} + c_{23})\right]_{2m-1} .$$

5.4.1 Terms of highest degree in $S_{2m-1}(\lambda^{R_{2m-1}} + c_{23})$

We first consider the terms of highest degree in $S_{2m-1}(\lambda^{R_{2m-1}} + c_{23})$. We make a change of variable, since we consider all $x \in \mathbf{F}_2^{2m-1}$. Then,

$$[S_{2m-1}(\lambda^{R_{2m-1}} + c_{23})]_{2m-1} = [S_{2m-1}(g(x))]_{2m-1}$$

with

$$\begin{aligned} g(x) &= S_{2m-1}(x) + S_{2m-1}(x + c_{28}) + T \circ S_{2m+1}(E(x) + c_{29}) + T \circ S_{2m+1}(E(x) + c_{30}) \\ &\quad + x + c_{31} \\ &= D_{c_{28}} S_{2m-1}(x) + A(x, c_{29}, c_{30}, c_{31}) , \end{aligned}$$

where all terms of A have degree at most 1. Therefore, all terms of $S_{2m-1}(g(x))$ correspond to the product of β_1 components of $D_{c_{28}} S_{2m-1}$ and of β_2 components of $A(x, c_{29}, c_{30}, c_{31})$ where $\beta_1 + \beta_2 = d$. The degree of such a term is then lower than $\beta_1(d-1) + (d - \beta_1)$ as $deg(D_{c_{28}} S_{2m-1}) \leq d - 1$. When $\beta_1 = d$ (and then $\beta_2 = 0$), this term corresponds to a product of derivatives with respect to c_{28}. Hence it has the same value on x and $x + c_{28}$ for all $x \in \mathbf{F}_2^{2m-1}$ and it cannot have degree $2m-1$. Therefore, the degree $2m-1$ can only be obtain for $\beta_1 \leq d - 1$. In such cases, the degree admits the upper bound $(d-1)^2 + 1$. It follows that $S_{2m-1}(g(x))$ have degree at most $(2m-2)$ if

$$d < 1 + \sqrt{2m-2} .$$

Note that this condition is satisfied by the original parameters ($m = 4$ and $d = 3$).

5.4.2 Terms of highest degree in $S_{2m-1}(\mu^{R_{2m-1}} + c_{22})$

Now, we apply a similar treatment to $S_{2m-1}(\mu^{R_{2m-1}} + c_{22})$, where

$$\mu^{R_{2m+1}} = S_{2m+1}\left(E(x) + c_{11}\right) + E \circ S_{2m-1}(x + c_5) + c_{15}.$$

We also make a change of variable. Then,

$$[S_{2m-1}(\mu^{R_{2m-1}} + c_{22})]_{2m-1} = [S_{2m-1}(t(x))]_{2m-1}$$

with

$$t(x) = S_{2m-1}(x) + T \circ S_{2m+1}(E(x) + c_{25}) + c_{26}.$$

Moreover, the explicit writing of the almost bent power function $S_{2m-1}(x) = L(x^e)$ leads to:

$$L^{-1}(t(x)) = x^e + Q(x) + A(x, c_{25}, c_{26})$$

where Q contains quadratic terms only and A affine or constant terms (since c_{25} and c_{26} only appear in linear or constant terms).

In [1], Babbage and Frisch give the following explanation for the 7-th order differential attack on the original cipher: the only way to obtain a term of degree 7 in $S_{2m-1}(t(x))$ with $d = 3$ is to multiply at least two terms of degree 3 of $L^{-1}(t(x))$ and another term. But, the terms of degree 3 in $L^{-1}(t(x))$ come from the almost bent function S_7, and they observe that the product of any two Boolean components of S_7 has degree at most 5 [1, Fact 2]. This observation is a direct consequence of Corollary 1. Thus, the maximum degree that we can obtain is at most 7.

More generally, all terms in $[S_{2m-1}(t(x))]_{2m-1}$ are the result of the product of β_1 terms from x^e, β_2 terms from $Q(x)$ and β_3 terms from $A(x, c_{25}, c_{26})$, with $\beta_1 + \beta_2 + \beta_3 = d$. In other terms, we can write them as: $x^{e\lambda_1} \cdot x^{\lambda_2} \cdot x^{\lambda_3} \cdot c$ where λ_1, λ_2 and λ_3 are integers lower than 2^{2m-1} and verifying $w_2(\lambda_1) = \beta_1$, $w_2(\lambda_3) \leq \beta_3$ and $w_2(\lambda_2) \leq 2\beta_2$ as λ_2 is the sum of β_2 integers whose 2-weights equal 2. Such a term depends on a constant only if $\beta_3 \neq 0$. Its degree is then:

$$w_2\left((e\lambda_1 + \lambda_2 + \lambda_3) \bmod \left(2^{2m-1} - 1\right)\right)$$

and the attack could be done as soon as $w_2\left((e\lambda_1 + \lambda_2 + \lambda_3) \bmod \left(2^{2m-1} - 1\right)\right) < 2m - 1$. Now, we derive from Proposition 7

$$\begin{aligned} w_2((e\lambda_1 + \lambda_2 + \lambda_3) \bmod (2^{2m-1} - 1)) &\leq w_2(e\lambda_1 \bmod (2^{2m-1} - 1)) + w_2(\lambda_2) + w_2(\lambda_3) \\ &\leq (m-1) + \beta_1 + 2\beta_2 + \beta_3 \quad (6) \\ &\leq (m-1) + d + \beta_2 \quad (7) \end{aligned}$$

as $\beta_1 + \beta_2 + \beta_3 = d$.

Such a term depends on the constants only if $\beta_3 \geq 1$. We then have $\beta_2 \leq d - 1$. But the terms including a high value for β_2 ($\beta_2 \geq d-2$) correspond to one of the following particular cases:

- Case $\beta_1 = 0$. Then, we have $\beta_2 + \beta_3 = d$. We deduce that

$$\begin{aligned} w_2((e\lambda_1 + \lambda_2 + \lambda_3) \bmod (2^{2m-1} - 1)) &= w_2(\lambda_2 + \lambda_3 \bmod (2^{2m-1} - 1)) \\ &\leq 2\beta_2 + \beta_3 \leq 2d - \beta_3 \leq 2m - 3 \end{aligned}$$

  since $\beta_3 \geq 1$. Note that this completely solves the case $\beta_2 = d - 1$.

- Case $\beta_1 = 1$ and $\beta_2 = d-2$. As $w_2(\lambda_1) = w_2(\lambda_3) = 1$, we have $\lambda_1 = 2^i$ and $\lambda_3 = 2^j$. Therefore,

$$\begin{aligned} w_2((e\lambda_1 + \lambda_2 + \lambda_3) \bmod (2^{2m-1}-1)) &= w_2(2^i e + \lambda_2 + 2^j \bmod (2^{2m-1}-1)) \\ &= w_2(e + \lambda_2' + 2^k \bmod (2^{2m-1}-1)) \\ &\leq w_2(e) + w_2(\lambda_2') + 1 \\ &\leq d + (d-2) + 1 \leq 2m-3 \ . \end{aligned}$$

Both previous situations include the case $\beta_2 \geq d-2$. Now, for any $\beta_2 \leq d-3$, we derive from (7) that

$$w_2((e\lambda_1 + \lambda_2 + \lambda_3) \bmod (2^{2m-1}-1)) \leq m-1+2d-3 \ .$$

This upper bound cannot exceed $(2m-2)$ as soon as

$$d < \frac{m+3}{2} \ .$$

This study emphasizes that for any block size $16m$, with a S_{2m+1} box of degree 2, the cipher is vulnerable to a higher order cryptanalysis of degree $2m-1$ as soon as the degree d of the almost bent function S_{2m-1} satisfies

$$d < \min(1+\sqrt{2m-2}, \ \frac{m+3}{2}) \ .$$

The condition required by the first bound is clearly the most restrictive one, since it does not exploit the almost bent property. For any S_{2m-1} of degree 3, the cipher is vulnerable when $m \geq 4$ and for S_{2m-1} of degree 4 when $m \geq 6$. The attackable degrees are classified in the following table.

m	block size	attackable degrees
3	48	$d \leq 2$
4	64	$d \leq 3$ (original parameters)
5	80	$d \leq 3$
6	96	$d \leq 4$
10	160	$d \leq 5$

Then, our study points out that the property of high divisibility of the Walsh spectrum of the confusion function is at the origin of the vulnerability of such a cipher. This property leads to the following new design criterion: the Walsh spectrum of the confusion function should contain at least one value which is not divisible by a higher power of 2.

References

[1] S. Babbage and L. Frisch. On MISTY1 Higher Order Differential Cryptanalysis. In *Proceedings of ICISC 2000*, number 2015 in Lecture Notes in Computer Science, pages 22–36. Springer-Verlag, 2000.

[2] E. Biham and A. Shamir. Differential cryptanalysis of DES-like cryptosystems. *Journal of Cryptology*, 4(1):3–72, 1991.

[3] A. Canteaut. Differential cryptanalysis of Feistel ciphers and differentially uniform mappings. In *Selected Areas on Cryptography, SAC'97*, pages 172–184, Ottawa, Canada, 1997.

[4] A. Canteaut, P. Charpin, and H. Dobbertin. A new characterization of almost bent functions. In *Fast Software Encryption 99*, number 1636 in Lecture Notes in Computer Science, pages 186–200. Springer-Verlag, 1999.

[5] A. Canteaut, P. Charpin, and H. Dobbertin. Weight divisibility of cyclic codes, highly nonlinear functions of $GF(2^m)$, and crosscorrelation of maximum-length sequences. *SIAM Journal of Discrete Mathematics*, 13(1):105–138, 2000.

[6] C. Carlet. Two new classes of bent functions. In *Advances in Cryptology - EUROCRYPT'93*, number 765 in Lecture Notes in Computer Science, pages 77–101. Springer-Verlag, 1994.

[7] C. Carlet, P. Charpin, and V. Zinoviev. Codes, Bent Functions and Permutations Suitable for DES-like Cryptosystems. *Designs, Codes and Cryptography*, 15(2):125–156, 1998.

[8] F. Chabaud and S. Vaudenay. Links between differential and linear cryptanalysis. In *Advances in Cryptology - EUROCRYPT'94*, number 950 in Lecture Notes in Computer Science, pages 356–365. Springer-Verlag, 1995.

[9] P. Charpin, A. Tietäväinen, and V. Zinoviev. On binary cyclic codes with minimum distance $d = 3$. *Problems of Information Transmission*, 33(4):287–296, 1997.

[10] T. Cusick and H. Dobbertin. Some new 3-valued crosscorrelation functions of binary m-sequences. *IEEE Transactions on Information Theory*, 42:1238–1240, 1996.

[11] H. Dobbertin. One-to-one highly nonlinear power functions on $GF(2^n)$. *Appl. Algebra Engrg. Comm. Comput.*, 9(2):139–152, 1998.

[12] R. Gold. Maximal recursive sequences with 3-valued recursive crosscorrelation functions. *IEEE Transactions on Information Theory*, 14:154–156, 1968.

[13] C. Harpes. *Cryptanalysis of iterated block ciphers*, volume 7 of *ETH Series in Information Processing*. Hartung-Gorre Verlag, Konstanz, 1996.

[14] C. Harpes, G. Kramer, and J. L. Massey. A generalization of linear cryptanalysis and the applicability of Matsui's piling-up lemma. In *Advances in Cryptology - EUROCRYPT'95*, number 921 in Lecture Notes in Computer Science, pages 24–38. Springer-Verlag, 1995.

[15] T. Helleseth. Some results about the cross-correlation function between two maximal linear sequences. *Discrete Mathematics*, 16:209–232, 1976.

[16] T. Jakobsen. Cryptanalysis of block ciphers with probabilistic non-linear relations of low degree. In *Advances in Cryptology - CRYPTO'98*, number 1462 in Lecture Notes in Computer Science, pages 212–222. Springer-Verlag, 1998.

[17] T. Jakobsen. *Higher-order cryptanalysis of block ciphers.* PhD thesis, Technical University of Denmark, 1999.

[18] T. Jakobsen and L.R. Knudsen. The interpolation attack on block ciphers. In *Fast Software Encryption 97*, number 1267 in Lecture Notes in Computer Science. Springer-Verlag, 1997.

[19] T. Kasami. The weight enumerators for several classes of subcodes of the second order binary Reed-Muller codes. *Information and Control*, 18:369–394, 1971.

[20] L. R. Knudsen. Truncated and higher order differentials. In *Fast Software Encryption - Second International Workshop*, number 1008 in Lecture Notes in Computer Science, pages 196–211. Springer-Verlag, 1995.

[21] Z. Kukorelly. *On the validity of certian hypotheses used in linear cryptanalysis*, volume 13 of *ETH Series in Information Processing.* Hartung-Gorre Verlag, Konstanz, 1999.

[22] G. Lachaud and J. Wolfmann. The weights of the orthogonal of the extended quadratic binary Goppa codes. *IEEE Transactions on Information Theory*, 36(3):686–692, 1990.

[23] X. Lai. Higher order derivatives and differential cryptanalysis. In *Proc. "Symposium on Communication, Coding and Cryptography", in honor of J. L. Massey on the occasion of his 60'th birthday*, 1994.

[24] M. Matsui. Linear cryptanalysis method for DES cipher. In *Advances in Cryptology - EUROCRYPT'93*, number 765 in Lecture Notes in Computer Science. Springer-Verlag, 1994.

[25] M. Matsui. The first experimental cryptanalysis of the Data Encryption Standard. In *Advances in Cryptology - CRYPTO'94*, number 839 in Lecture Notes in Computer Science. Springer-Verlag, 1995.

[26] M. Matsui. New Block Encryption Algorithm MISTY. In *Proceedings of the Fourth International Workshop of Fast Software Encryption*, number 1267 in Lecture Notes in Computer Science, pages 54–68. Springer-Verlag, 1997.

[27] R.J. McEliece. Weight congruence for p-ary cyclic codes. *Discrete Mathematics*, 3:177–192, 1972.

[28] W. Meier and O. Staffelbach. Nonlinearity criteria for cryptographic functions. In *Advances in Cryptology - EUROCRYPT'89*, number 434 in Lecture Notes in Computer Science, pages 549–562. Springer-Verlag, 1990.

[29] K. Nyberg. Perfect nonlinear S-boxes. In *Advances in Cryptology - EUROCRYPT'91*, number 547 in Lecture Notes in Computer Science, pages 378–385. Springer-Verlag, 1991.

[30] K. Nyberg. Differentially uniform mappings for cryptography. In *Advances in Cryptology - EUROCRYPT'93*, number 765 in Lecture Notes in Computer Science, pages 55–64. Springer-Verlag, 1993.

[31] K. Nyberg and L.R. Knudsen. Provable security against differential cryptanalysis. In *Advances in Cryptology - CRYPTO'92*, number 740 in Lecture Notes in Computer Science, pages 566–574. Springer-Verlag, 1993.

[32] D.V. Sarwate and M.B. Pursley. Crosscorrelation properties of pseudorandom and related sequences. *Proceedings of the IEEE*, 68(5):593–619, 1980.

[33] V.M. Sidelnikov. On mutual correlation of sequences. *Soviet Math. Dokl.*, 12:197–201, 1971.

[34] H. Tanaka, K. Hisamatsu, and T. Kaneko. Strength of MISTY1 without FL function for Higher Order Differential Attack. In *Applied Algebra, Algebraic Algorithms and Error-Correcting Codes*, number 1719 in Lecture Notes in Computer Science, pages 221–230. Springer-Verlag, 1999.

Unité de recherche INRIA Rocquencourt
Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex (France)

Unité de recherche INRIA Lorraine : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex (France)
Unité de recherche INRIA Rennes : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)
Unité de recherche INRIA Rhône-Alpes : 655, avenue de l'Europe - 38330 Montbonnot-St-Martin (France)
Unité de recherche INRIA Sophia Antipolis : 2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex (France)

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399