Degree of Composition of Highly Nonlinear Functions and Applications to Higher Order Differential Cryptanalysis

Anne Canteaut and Marion Videau

INRIA - projet CODES
B.P. 105 - 78153 Le Chesnay Cedex, France
{Anne.Canteaut, Marion.Videau}@inria.fr

Abstract. To improve the security of iterated block ciphers, the resistance against linear cryptanalysis has been formulated in terms of provable security which suggests the use of highly nonlinear functions as round functions. Here, we show that some properties of such functions enable to find a new upper bound for the degree of the product of its Boolean components. Such an improvement holds when all values occurring in the Walsh spectrum of the round function are divisible by a high power of 2. This result leads to a higher order differential attack on any 5-round Feistel ciphers using an almost bent substitution function. We also show that the use of such a function is precisely the origin of the weakness of a reduced version of MISTY1 reported in [23, 1].

Keywords. Block ciphers, higher order differential cryptanalysis, Boolean functions, nonlinearity.

1 Introduction

The development of cryptanalysis in the last ten years has led to the definition of some design criteria for block ciphers. These criteria correspond to some mathematical properties of the round function which is used in an iterated block cipher. In particular, the use of a highly nonlinear round function ensures a high resistance to linear attacks [16, 17]. 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 [6]. Such functions are used for instance in the block cipher MISTY [18]. 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 [12, 10], 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 main spectral properties of the round function which are involved in differential and linear cryptanalysis. The general principle of a higher order differential attack is described in Section 3. Section 4 then investigates the link between the divisibility of the Walsh coefficients of a function and the degree of the product of its Boolean components. This result leads to a higher order differential attack on any 5-round Feistel cipher using an almost bent substitution function. Finally, we point out that the attack of a reduced version of MISTY1 presented in [23, 1] is a direct consequence of the use of almost bent S-boxes. We show that a similar attack can be performed for different block sizes and almost bent S-boxes.

2 Spectral Properties of a Round Function

In an iterated block cipher, the ciphertext is obtained by iteratively applying a round function F to the plaintext. In an r-round iterated cipher, we have

$$x_i = F(x_{i-1}, K_i)$$

where x_0 is the plaintext, x_r is the ciphertext and the r-round keys $(K_1, \ldots, K_r)$ are usually derived from a unique secret key. For any fixed round key K, the round function $F_K : x \mapsto F(x, K)$ is a permutation of the set of n-bit vectors, $\mathbb{F}_2^n$, where n is the block size. The resistance of such cipher to some particular attacks can be quantified by some properties of the round function.

A *Boolean function f of n variables* is a function from $\mathbb{F}_2^n$ into $\mathbb{F}_2$. It can be expressed as a polynomial, called its *algebraic normal form*. The *degree* of f, denoted by $\deg(f)$, is the degree of its algebraic normal form. The following notation will be extensively used in the paper. The usual dot product between two vectors x and y is denoted by $x \cdot y$. For any $\alpha \in \mathbb{F}_2^n$, 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:

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

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

Definition 1. *The* Walsh spectrum *of a Boolean function f of n variables is the multiset*

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

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

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

A linear attack against a cipher with round function F exploits the existence of a pair (α, β) with $\alpha \neq 0$ such that, for almost all round keys K, the function $x \mapsto \varphi_\alpha \circ F_K(x) + \varphi_\beta(x)$ takes the same value for most values of $x \in \mathbb{F}_2^n$. Therefore, all functions $\varphi_\alpha \circ F_K$ should be far from all affine functions. This requirement is related to the nonlinearity of the functions F_K.

Definition 2. *[21] The* nonlinearity *of a function F from $\mathbb{F}_2^n$ into $\mathbb{F}_2^n$ is the Hamming distance between all $\varphi_\alpha \circ F, \alpha \in \mathbb{F}_2^n$, $\alpha \neq 0$, and the set of affine functions. It is given by*

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

Proposition 1. *[6] For any function $F : \mathbb{F}_2^n \to \mathbb{F}_2^n$,*

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

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

This minimum value for $\mathcal{L}(F)$ can only be achieved if 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. Note that the Walsh spectrum of a function is invariant under both right and left compositions by a linear permutation of $\mathbb{F}_2^n$.

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

Proposition 2. *[6] The Walsh spectrum of an almost bent function F from $\mathbb{F}_2^n$ into $\mathbb{F}_2^n$ takes the values 0 and $\pm 2^{\frac{n+1}{2}}$ only.*

This property implies that any almost bent function is *almost perfect nonlinear* [22], i.e., that it ensures the best resistance to differential cryptanalysis. Therefore, the use of an almost bent function as round function (or as substitution function) provides a high resistance to both linear and differential attacks. These functions are used in MISTY [18]. Similarly, AES uses a function of an even number of variables which has the highest known nonlinearity.

3 Higher Order Differential Attacks

Higher order differential cryptanalysis was introduced by Knudsen [12]. As a generalization of differential cryptanalysis, it relies on some properties of higher order derivatives of a vectorial function. In the following, we denote by $\oplus$ the bitwise exclusive-or.

Definition 3. *[14] Let F be a function from $\mathbb{F}_2^n$ into $\mathbb{F}_2^m$. For any $a \in \mathbb{F}_2^n$, the* derivative of F with respect to a *is the function*

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

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

$$D_V F = D_{a_1} D_{a_2} \ldots D_{a_k} F \ ,$$

where $(a_1, \ldots, a_k)$ is any basis of V. Moreover, we have for any $x \in \mathbb{F}_2^n$

$$D_V F(x) = \bigoplus_{v \in V} F(x \oplus v) \ .$$

We now consider an r-round iterated cipher with block size n and round function F. We call *reduced cipher*, the cipher obtained by removing the final round of the original cipher. The reduced cipher corresponds to the function $G = F_{K_{r-1}} \circ \ldots \circ F_{K_1}$.

Suppose that there exists a k-dimensional subspace $V \subset \mathbb{F}_2^n$ such that

$$D_V G(x) = c \text{ for all } x \in \mathbb{F}_2^n$$

where c is a constant in $\mathbb{F}_2^n$ which does not depend on the round keys $K_1, \ldots, K_{r-1}$. Then, for any round keys the reduced cipher G satisfies

$$\forall x \in \mathbb{F}_2^n, \ \bigoplus_{v \in V} G(x \oplus v) = c \ . \tag{1}$$

This property leads to the following chosen plaintext attack.

1. Select a random plaintext $x_0 \in \mathbb{F}_2^n$ and get the ciphertexts c_v corresponding to all plaintexts $x_0 \oplus v$, $v \in V$.
2. Compute c by applying (1) to the reduced cipher with any round keys (e.g. $K_1, \ldots, K_{r-1} = 0$).
3. For each candidate round key k_r, compute

$$\sigma(k_r) = \bigoplus_{x \in V} F_{k_r}^{-1}(c_v) \ .$$

The key k_r for which $\sigma(k_r) = c$ is the correct last-round key with a high probability. 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 2^{m+k} evaluations of F^{-1} where m is the size of the round key and k is the dimension of V. It requires the knowledge of 2^k 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 4. *The* degree *of a function F from $\mathbb{F}_2^n$ into $\mathbb{F}_2^n$ is the maximum degree of its Boolean components:* $\deg(F) = \max_{1 \le i \le n} \deg(\varphi_{e_i} \circ F)$ *where $(e_i)_{1 \le i \le n}$ denotes the canonical basis of $\mathbb{F}_2^n$.*

For any F of degree d, we obviously have $D_V F = 0$ for any $(d+1)$-dimensional subspace $V \subset \mathbb{F}_2^n$. Therefore, if the reduced cipher G has degree at most d for all round keys, it is possible to perform a differential attack of order $(d+1)$.

The degree of the round function F provides a trivial upper bound on the degree of the reduced cipher:

$$\deg(G) \leq (\deg(F))^{r-1} \ .$$

This bound was directly used by Jakobsen and Knudsen [10] for breaking a cipher example proposed in [22], whose round function is an almost bent quadratic permutation. Unfortunately, this method can only be used when the degree of the round function is very low. It clearly appears that another approach has to be used when the degree of the round function is strictly greater than $\sqrt{n}$ since $(\deg(F))^{r-1} > n$ for any $r \geq 3$.

4 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 $\mathbb{F}_2^n$ into $\mathbb{F}_2^n$. We show that the trivial bound

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

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 Proposition 2).

Definition 5. *The Walsh spectrum of a function F from $\mathbb{F}_2^n$ into $\mathbb{F}_2^m$ 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 [15, Page 447]. The following proposition is a direct consequence of [4, Lemma 3].

Proposition 3. *Let F be a function from $\mathbb{F}_2^n$ into $\mathbb{F}_2^m$. 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. Its proof can be found in [3].

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

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

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

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

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*

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

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

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

Therefore,

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

Proof. We prove Relation (2) by induction on the size of I. The result obviously holds for $|I| = 1$. We now assume that (2) 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 $\alpha \in \mathbb{F}_2^n$

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

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 (2) and Proposition 3.

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

Corollary 1. *Let F be a function from $\mathbb{F}_2^n$ into $\mathbb{F}_2^n$ 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 $\mathbb{F}_2^n$ into $\mathbb{F}_2^n$, we have

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

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 $\mathbb{F}_2^n$ with the finite field with 2^n elements, $\mathbb{F}_{2^n}$. In this context, any function F from $\mathbb{F}_2^n$ into $\mathbb{F}_2^n$ can be expressed as a unique univariate polynomial in $\mathbb{F}_{2^n}[X]$, $F(X) = \sum_{u=0}^{2^n-1} a_u X^u$. The degree of F (in the sense of Definition 4) 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 (up to a linear permutation of $\mathbb{F}_2^n$) to some power functions $x \mapsto x^s$ over $\mathbb{F}_{2^n}$. 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 [2, Coro. 2]. The result is directly derived from McEliece's theorem which provides the weight divisibility of a cyclic code [19]. We refer to [5, 2] for the link between cyclic codes and power functions.

Proposition 4. *Let $F : x \mapsto x^s$ be a power function over $\mathbb{F}_{2^n}$. 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*

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

5 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

$$\begin{aligned} F_K \colon \mathbb{F}_2^n \times \mathbb{F}_2^n &\to \mathbb{F}_2^n \times \mathbb{F}_2^n \\ (L, R) &\mapsto (R, L \oplus S_K(R)) \end{aligned}$$

where S_K is a function from $\mathbb{F}_2^n$ into $\mathbb{F}_2^n$ called the substitution 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:

$$R_3 = R_5 \oplus S_{K_5}(L_5) \ .$$

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 substitution function :

$$R_3(x) = x \oplus c_1 \oplus S_{K_3}(c_0 \oplus S_{K_2}(x \oplus c_1))$$

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

When the Walsh spectrum of the substitution function S_K is 2^ℓ-divisible for all values of K, we can apply Corollary 1. Then, we obtain the following upper bound for the degree of R_3:

$$\deg(R_3) \leq n - \ell + \deg(S) \ .$$

Thus, if we consider the attack described in Section 3, we have exhibited a new attack on the last round key with average running-time of $2^{m+\delta}$, where m is the size of the round key and $\delta = \min(\deg(S)^2 + 1, n - \ell + \deg(S) + 1)$. This attack is feasible as soon as $\delta \leq n$. For example, if S is almost bent, a higher order differential attack can be performed except when $\deg(S) = (n+1)/2$, i.e., when S 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, $\mathcal{L}(S) = 2^{\frac{n}{2}+1}$. All known functions satisfying this property are equivalent (up to a linear permutation of $\mathbb{F}_2^n$) to one of the power functions given in Table 1 (or to one of their inverses) [8]. 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. Note that the Walsh spectrum of the inverse function has the smallest possible divisibility for a function whose nonlinearity is even. If the Walsh spectrum of the substitution function S is $2^{\frac{n}{2}+1}$-divisible, then $\deg(S) \leq n/2$. Therefore, the attack is always feasible. When the Walsh spectrum of S is $2^{\frac{n}{2}}$-divisible, the attack can be performed except if $\deg(S) \in \{n/2, n/2+1\}$. 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., $S_K(x) = S(x \oplus K)$. In that case, we obtain the following expression for R_3:

$$R_3(x) = x \oplus c_1 \oplus S(c_0 \oplus K_3 \oplus S(x \oplus c_1 \oplus K_2)) \ .$$

Let G be the function defined by $G : x \mapsto S(K_3 \oplus c_0 \oplus S(x \oplus c_1 \oplus K_2))$ and let G' be defined by $G' : x \mapsto S(K_3 \oplus c_0 \oplus S(x))$. 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 $n - \ell + \deg(S) - 1$.

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

exponents s	condition on n	divisibility	
$2^{n-1}-1$	$n \equiv 0 \bmod 2$	2^2	[13]
2^k+1, with $\gcd(k,n)=2$ and $k<\frac{n}{2}$	$n \equiv 2 \bmod 4$	$2^{\frac{n}{2}+1}$	[9, 20]
$2^{2k}-2^k+1$, with $\gcd(k,n)=2, k<\frac{n}{2}$	$n \equiv 2 \bmod 4$	$2^{\frac{n}{2}+1}$	[11]
$2^{\frac{n}{2}}+2^{\frac{n+2}{4}}+1$	$n \equiv 2 \bmod 4$	$2^{\frac{n}{2}+1}$	[7]
$2^{\frac{n}{2}}+2^{\frac{n}{2}-1}+1$	$n \equiv 2 \bmod 4$	$2^{\frac{n}{2}+1}$	[7]
$\sum_{i=0}^{n/2} 2^{ik}$, with $\gcd(k,n)=1, k<\frac{n}{2}$	$n \equiv 0 \bmod 4$	$2^{\frac{n}{2}}$	[8]
$2^{\frac{n}{2}}+2^{\frac{n}{4}}+1$	$n \equiv 4 \bmod 8$	$2^{\frac{n}{2}}$	[8]

Table 2. Higher order differential attack on a 5-round Feistel cipher using a highly nonlinear substitution function S

function S $\mathcal{L}(S)$	div.	General case		$S_K(x)=S(x\oplus K)$	
		differential order	feasibility	differential order	feasibility
$2^{\frac{n+1}{2}}$ n odd	$2^{\frac{n+1}{2}}$	$\deg(S)+\frac{n+1}{2}$	except for $\deg(S)=\frac{n+1}{2}$	$\deg(S)+\frac{n-1}{2}$	always feasible
$2^{\frac{n}{2}+1}$ n even	$2^{\frac{n}{2}+1}$	$\deg(S)+\frac{n}{2}$	always feasible	$\deg(S)+\frac{n}{2}-1$	always feasible
	$2^{\frac{n}{2}}$	$\deg(S)+\frac{n}{2}+1$	except for $\deg(S)\in\{\frac{n}{2},\frac{n}{2}+1\}$	$\deg(S)+\frac{n}{2}$	except for $\deg(S)=\frac{n}{2}+1$

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 $(n-\ell+\deg(S))$:

$$\forall a \in \mathbb{F}_2^n,\quad D_V G'(a)=\bigoplus_{v\in V} G'(a\oplus v)=c$$

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

$$\forall a \in \mathbb{F}_2^n,\quad \bigoplus_{v\in V} G(a\oplus v)=\bigoplus_{v\in V} G'(a\oplus v\oplus c_1\oplus K_2)=D_V G'(a\oplus c_1\oplus K_2)=c.$$

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 2).

6 Higher Order Differential Cryptanalysis on a Generalization of MISTY1

MISTY is a model of block ciphers proposed by Matsui [18] and presented under two forms MISTY1 and MISTY2. MISTY1 is the object of this study. M'1, the version of MISTY1 reduced to 5 rounds without FL functions is provably secure against both differential and linear cryptanalysis. Therefore, the background of the attack is this simplified algorithm. In [23] it is shown that M'1 can be attacked with a 7-th order differential. In [1], the attack is extended to the case where any almost bent power function of degree 3 on $\mathbb{F}_2^7$ is used for the S_7-box.

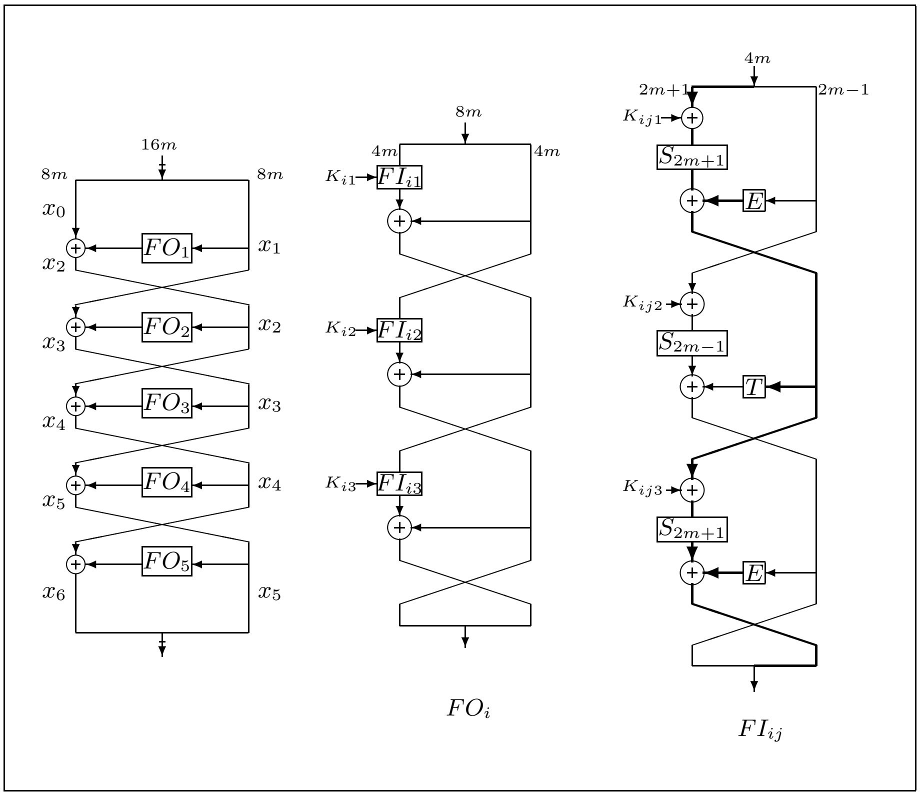

Fig. 1. The 5-round Feistel cipher M'1 with equivalent key schedule

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 $16m$ bits (see Fig. 1). The original value is 64 bits. In this generalization, we show that the weakness of M'1 is due to the use of an almost bent substitution function.

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.*

The cipher uses the "zero-extend" function, E, and the 'truncate" function, T, which are respectively defined by:

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

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

The nonlinear part of the cipher consists of two permutations, S_{2m-1} and S_{2m+1}, respectively defined over $\mathbb{F}_{2^{2m-1}}$ and $\mathbb{F}_{2^{2m+1}}$. In the original cipher, we have $S_7(x) = L(x^{81})$ over $\mathbb{F}_{2^7}$ where L is a linear permutation and S_9 is a quadratic almost bent permutation of $\mathbb{F}_{2^9}$.

Let V be the $(2m-1)$-dimensional subspace of plaintexts of $16m$ bits whose form is $(0_{6m+1} \| x \| 0_{8m})$ where x is in $\mathbb{F}_2^{2m-1}$. Let W denote the subspace $\{(w_0 \| 0_{2m-1} \| w_1), w_0 \in \mathbb{F}_2^{6m+1}, w_1 \in \mathbb{F}_2^{8m}\}$. We are interested in ciphering plaintexts $P \oplus w$ where $P \in V$ and $w = (w_0 \| w_1)$ is a fixed constant in W. We now consider the function G_K defined as follows:

$$G_K : x \mapsto x_4^{L_{2m-1}} \ .$$

To sum up the higher order differential attack proposed in [23], 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:

$$\forall w \in W, \quad \bigoplus_{x \in V} G_K(x \oplus w) = c. \tag{3}$$

Here, we show that this property can be generalized to different block sizes.

We need the exact expression of $x_4^{L_{2m-1}}$. The details of this computation are given in [3]. We obtain:

$$\begin{aligned} x_4^{L_{2m-1}} = \ & \mu^{R_{2m-1}} \oplus \lambda^{R_{2m-1}} \oplus \lambda^{L_{2m-1}} \oplus c_{24} \oplus T \circ S_{2m+1}(\mu^{L_{2m+1}} \oplus c_{20}) \\ & \oplus T \circ S_{2m+1}(\lambda^{L_{2m+1}} \oplus c_{21}) \oplus S_{2m-1}(\mu^{R_{2m-1}} \oplus c_{22}) \\ & \oplus S_{2m-1}(\lambda^{R_{2m-1}} \oplus c_{23}) \ , \end{aligned} \tag{4}$$

where

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

and all c_i are some constants depending on the round keys. The aim of our study is to determine the degree of the Boolean components of $x_4^{L_{2m-1}}$.

We restrict our study to the case where S_{2m+1} is a quadratic function, as in the original cipher. We suppose 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 $\mathbb{F}_2^{2m-1}$. These conditions obviously imply that we can neglect the terms $T \circ S_{2m+1}(\mu^{L_{2m+1}} \oplus c_{20}) \oplus T \circ S_{2m+1}(\lambda^{L_{2m+1}} \oplus c_{21})$ in (4) for a $(2m-1)$-th order differential.

We denote by $[F]_d$ the terms in the algebraic normal form of F whose degree are at least d. It clearly appears that the terms of degree $2m-1$ in $x_4^{L_{2m-1}}$ correspond to

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

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

We first consider the terms of highest degree in $S_{2m-1}(\lambda^{R_{2m-1}} \oplus c_{23})$. We make a change of variable, since we consider all $x \in \mathbb{F}_2^{2m-1}$. Then, $[S_{2m-1}(\lambda^{R_{2m-1}} \oplus c_{23})]_{2m-1} = [S_{2m-1}(g(x))]_{2m-1}$ with

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

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 $x \oplus c_{28}$ for all $x \in \mathbb{F}_2^{2m-1}$ 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 $d = 3$).

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

Now, we apply a similar treatment to $S_{2m-1}(\mu^{R_{2m-1}} \oplus c_{22})$, where $\mu^{R_{2m+1}} = S_{2m+1}(E(x) \oplus c_{11}) \oplus E \circ S_{2m-1}(x \oplus c_5) \oplus c_{15}$. We also make a change of variable. Then, $[S_{2m-1}(\mu^{R_{2m-1}} \oplus c_{22})]_{2m-1} = [S_{2m-1}(t(x))]_{2m-1}$ with

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

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

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

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 4

$$\begin{aligned} w_2 &((e\lambda_1 + \lambda_2 + \lambda_3) \bmod (2^{2m-1} - 1)) \\ &\leq w_2(e\lambda_1 \bmod (2^{2m-1} - 1)) + w_2(\lambda_2) + w_2(\lambda_3) \\ &\leq (m - 1) + \beta_1 + 2\beta_2 + \beta_3 \leq (m - 1) + d + \beta_2 \end{aligned} \tag{5}$$

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 case 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 (5) 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 substitution 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 substitution 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. 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.
3. A. Canteaut and M. Videau. Weakness of block ciphers using highly nonlinear confusion functions. Research Report 4367, INRIA, February 2002. Available on `http://www.inria.fr/rrrt/rr-4367.html`.
4. 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.

5. C. Carlet, P. Charpin, and V. Zinoviev. Codes, bent functions and permutations suitable for DES-like cryptosystems. *Designs, Codes and Cryptography*, 15:125–156, 1998.
6. F. Chabaud and S. Vaudenay. Links between differential and linear cryptanalysis. In A. De Santis, editor, *Advances in Cryptology - EUROCRYPT'94*, number 950 in Lecture Notes in Computer Science, pages 356–365. Springer-Verlag, 1995.
7. T. Cusick and H. Dobbertin. Some new 3-valued crosscorrelation functions of binary m-sequences. *IEEE Transactions on Information Theory*, 42:1238–1240, 1996.
8. H. Dobbertin. One-to-one highly nonlinear power functions on $GF(2^n)$. *Appl. Algebra Engrg. Comm. Comput.*, 9(2):139–152, 1998.
9. R. Gold. Maximal recursive sequences with 3-valued recursive crosscorrelation functions. *IEEE Transactions on Information Theory*, 14:154–156, 1968.
10. 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, pages 28–40. Springer-Verlag, 1997.
11. 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.
12. 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.
13. 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.
14. 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.
15. F.J. MacWilliams and N.J.A. Sloane. *The Theory of Error-Correcting Codes.* North-Holland, 1977.
16. M. Matsui. Linear cryptanalysis method for DES cipher. In *Advances in Cryptology - EUROCRYPT'93*, number 765 in Lecture Notes in Computer Science, pages 386–397. Springer-Verlag, 1993.
17. 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.
18. M. Matsui. New Block Encryption Algorithm MISTY. In *Fast Software Encryption 97*, number 1267 in Lecture Notes in Computer Science, pages 54–68. Springer-Verlag, 1997.
19. R.J. McEliece. Weight congruence for p-ary cyclic codes. *Discrete Mathematics*, 3:177–192, 1972.
20. 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.
21. K. Nyberg. On the construction of highly nonlinear permutations,. In *Advances in Cryptology - EUROCRYPT'92*, number 658 in Lecture Notes in Computer Science, pages 92–98. Springer-Verlag, 1993.
22. 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.

23. 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.