Cryptanalysis of block ciphers and weight divisibility of some binary codes

Anne Canteaut, Pascale Charpin, Marion Videau

Cryptanalysis of block ciphers and weight divisibility of some binary codes

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

Abstract

The resistance of an iterated block cipher to most classical attacks can be quantified by some properties of its round function. The involved parameters (nonlinearity, degrees of the derivatives...) for a function F from $\mathbf{F}_2^m$ into $\mathbf{F}_2^m$ are related to the weight distribution of a binary linear code $\mathcal{C}_F$ of length $2^m - 1$ and dimension $2m$. In particular, the weight divisibility of $\mathcal{C}_F$ appears as an important criterion in the context of linear cryptanalysis and of higher-order differential attacks. When the round function F is a power permutation over $\mathbf{F}_{2^m}$, the associated code $\mathcal{C}_F$ is the dual of a primitive cyclic code with two zeroes. Therefore, McEliece's theorem provides a powerful tool for evaluating the resistance of some block ciphers to linear and higher-order differential attacks.

Keywords: block ciphers, cryptanalysis, Boolean functions, almost bent functions, cyclic codes.

1 Introduction

This paper focuses on a large class of symmetric block ciphers called *iterated block ciphers*. In such systems 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 are expected to lead to a cryptographically strong encryption function.

A round function can in general be split up into three steps: a key dependant function (usually a bitwise xor), a non linear part called the *confusion function* and a linear permutation.

The design of such ciphers relies on the development of their cryptanalysis. It is particularly true since the publication of two generic attacks: the *differential* and the *linear* cryptanalysis. They are at the origin of a new approach to define the security of a cipher by some mathematical properties of the round function. This leads to the concept of *provable security* [34]. However those properties imply strong requirements on the algebraic structures of the round function and it appears that it can weaken the cipher.

In this paper we show how optimal functions in regard to differential and linear cryptanalysis suffer from weaknesses against *higher-order differential* attacks. The explanation has to be found the high *weight divisibility* of the code associated to the round function. From this property we have derived a new upper bound for the degree of some composed function. Then, it has been used to mount a generic attack on any 5-round *Feistel cipher* using an *almost bent* function as a round function and to explain the origin of a weakness in a reduced version of MISTY1 presented in [38, 1]. Finally we sum up the requirements, known up to now, a confusion function must verify to ensure the security of an iterated block cipher.

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 m-bit words, $\mathbf{F}_2^m$, indexed by a value $k \in \mathcal{K}$. In a cryptographic context, the round function F is public; F depends of a parameter k and each value of k provides a permutation F_k. The set $\mathcal{K}$ is called the *round key space*.

The encryption function of the iterated block cipher with block size m, with r rounds and with round function F is the keyed permutation of $\mathbf{F}_2^m$

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.

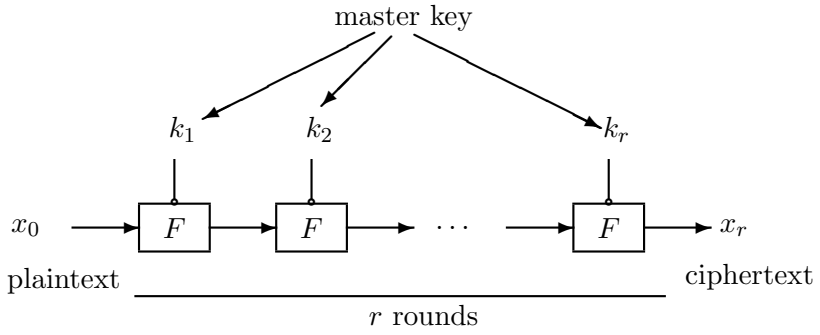

Figure 1: Iterated block cipher

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

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

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

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

(k describes here the set of 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* [14, 24].

Now, we give a more formal description. We refer to [14, 19] for a detailed presentation of last-round attacks.

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

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

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

$$\begin{array}{rcl} \varepsilon & < & \mid Pr_{f \in \mathcal{F}}\left[\mathcal{D}(f(x_1), \ldots, f(x_N)) = 1\right] \\ & & - \ Pr_{\pi \in_R \mathcal{P}_m}\left[\mathcal{D}(\pi(x_1), \ldots, \pi(x_N)) = 1\right] \mid . \end{array}$$

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 $\mathcal{G}$ [14, 15, 24]. 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:

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 m variables* is a function from $\mathbf{F}_2^m$ into $\mathbf{F}_2$. It can be expressed as a polynomial in $x_1, \ldots, x_m$, called its *algebraic normal form*. The *degree* of f, denoted by $deg(f)$, is the degree of its algebraic normal form.

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

Definition 2 [26] *Let F be a function from $\mathbf{F}_2^m$ into $\mathbf{F}_2^m$. For any $a \in \mathbf{F}_2^m$, 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 $\mathbf{F}_2^m$, the t-th derivative of F with respect to V *is the function*

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

where $(a_1, \ldots, a_t)$ 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 $\alpha \in \mathbf{F}_2^m$, φ_α denotes the linear function of m variables: $x \mapsto \alpha \cdot x$. For any Boolean function f of m variables, we denote by $\mathcal{F}(f)$ the following value related to the Walsh (or Fourier) transform of f:

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

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

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

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

The Walsh spectrum *of a vectorial function F from $\mathbf{F}_2^m$ into $\mathbf{F}_2^m$ 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*

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

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

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

In the following, we focus on three classes of last-round attacks: differential cryptanalysis, linear cryptanalysis and higher-order differential cryptanalysis. There exist 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 $\mathbf{F}_{2^m}[X]$, is close to a low-degree polynomial [18]. But, the mathematical nature of the property exploited by the latter attack is quite 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 $\mathbf{F}_2^m$ such that for any $\mathbf{k} = (k_1, \ldots, k_{r-1})$ the reduced cipher $G_\mathbf{k}$ (as defined by 1) satisfies:

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

for a large integer A. This property leads to a discriminator $\mathcal{D}$ for the family $\mathcal{G}$ of reduced ciphers with respect to any subset $(x_{2i}, x_{2i} + a \ ; \ 0 \leq i < N/2)$, where the x_{2i} form a set of $N/2$ randomly chosen elements in $\mathbf{F}_2^m$ and $N \geq \frac{2^{m+1}}{A-1}$.

Set $J = \mathcal{D}(y_0, y_1, \ldots, y_{2i}, y_{2i+1}, \ldots)$ with i in the range $[0, N/2[$, where y_{2i} and y_{2i+1} correspond respectively to the unknown values $G_\mathbf{k}(x_{2i})$ and

$G_{\mathbf{k}}(x_{2i}+a)$ for a correct guess of the last-round key. Then

$$\begin{aligned} J &= 1 \text{ if } \#\{i,\ 0 \le i < N/2,\ y_{2i}+y_{2i+1}=b\} \simeq \frac{AN}{2^{m+1}}, \\ &= 0 \text{ if } \#\{i,\ 0 \le i < N/2,\ y_{2i}+y_{2i+1}=b\} \simeq \frac{N}{2^{m+1}}\ . \end{aligned}$$

It follows that a cipher is resistant to differential cryptanalysis if each $G_{\mathbf{k}}$, $\mathbf{k} \in \mathcal{K}^{r-1}$, is such that for any nonzero $a \in \mathbf{F}_2^m$, 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 [34]. Therefore, the round function F of an iterated cipher should satisfy the following requirement: for any $k \in \mathcal{K}$,

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

should be small. As the number of solutions $x \in \mathbf{F}_2^m$ 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 [34] *For any function F from $\mathbf{F}_2^m$ into $\mathbf{F}_2^m$, 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

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

for a function from $\mathbf{F}_2^n$ into $\mathbf{F}_2^m$, where the functions achieving this bound are called *perfect nonlinear functions* [31]. Such functions only exist when n is even and $n \ge 2m$ [32].

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

$$\delta_F \ge 4\ .$$

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 m output bits of the reduced cipher which is close to an affine function [27, 28]. Let us assume that there exists two nonzero elements a and b in $\mathbf{F}_2^m$ 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:

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

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 $2^{m-1} - \frac{1}{2}\mathcal{L}(F)$.

Proposition 2 [37, 8] *For any function* $F : \mathbf{F}_2^m \to \mathbf{F}_2^m$,

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

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

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

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

where the functions achieving this bound were called *bent* functions in [32], as a generalization of the famous Boolean bent functions.

The minimum value of $\mathcal{L}(F)$ where F is a function from $\mathbf{F}_2^m$ into $\mathbf{F}_2^m$ can only be achieved when m is odd. For even m, some functions with $\mathcal{L}(F) = 2^{\frac{m}{2}+1}$ are known and it is conjectured that this value is the minimum [12, 36].

2.5 Higher-order differential cryptanalysis

Higher-order differential cryptanalysis was introduced by Knudsen in [23]. It exploits the fact that the reduced cipher $G_{\mathbf{k}}$, defined by (1), has a constant

t-th derivative. Assume that there exists a t-dimensional subspace $V \subset \mathbf{F}_2^m$ such that for any $\mathbf{k} = (k_1, \ldots, k_{r-1})$ we have

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

where $c \in \mathbf{F}_2^m$ is a constant which does not depend on $\mathbf{k}$. In accordance with Definition 2, we have:

$$D_V G_{\mathbf{k}}(x) = \sum_{v \in V} G_{\mathbf{k}}(x + v) \text{ for all } x \in \mathbf{F}_2^m \ ,$$

where the sum is an addition over $\mathbf{F}_2^m$. So we derive the following discriminator for $\mathcal{G}$ with respect to the set $(x_1, \ldots, x_{2^t})$ of elements of any coset of V, $a + V$ with $a \in \mathbf{F}_2^m$ (here y_i corresponds to $G_{\mathbf{k}}(x_i)$ for a correct guess of the last-round key):

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

A natural candidate for V arises when the degree of the reduced cipher is known; this comes from the next proposition whose proof can be found in [26].

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

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

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

Actually, we have

Proposition 3 *Let F be a function from $\mathbf{F}_2^m$ into $\mathbf{F}_2^m$ of degree d. Then, for any $(d+1)$-dimensional subspace $V \subset \mathbf{F}_2^m$, we have*

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

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

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

it obviously 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 [20] for breaking a cipher example proposed in [34], whose round function is a quadratic permutation. However, this condition is not sufficient and a stronger requirement on the round function will be exhibited in Section 4.2.

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 $\mathbf{F}_2^m$ [33]. Then, they only concern the *confusion part* of the round function.

3 Optimal cryptographic functions over $\mathbf{F}_2^m$ and weight distributions of some $[2^m - 1, 2m]$ binary codes

Carlet, Charpin and Zinoviev have pointed out that both APN and AB properties can be expressed in terms of error-correcting codes [7]. In the following, F denotes a function from $\mathbf{F}_2^m$ into $\mathbf{F}_2^m$. Since both APN and AB properties are invariant under translation, we here only consider the functions F such that $F(0, \ldots, 0) = 0$. We consider $\mathbf{F}_2^m$ as the ordered set

$$\{\ 0,\ \alpha_1,\ \ \ldots,\ \alpha_{2^m-1}\ \}\ .$$

Now, the linear binary code $\mathcal{C}_F$ of length $(2^m - 1)$ and dimension $2m$ is defined by its generator matrix:

$$G_F = \begin{pmatrix} \alpha_1 & \alpha_2 & \alpha_3 & \ldots & \alpha_{2^m-1} \\ F(\alpha_1) & F(\alpha_2) & F(\alpha_3) & \ldots & F(\alpha_{2^m-1}) \end{pmatrix}\ , \tag{2}$$

where each entry in $\mathbf{F}_2^m$ is viewed as a binary column vector of length m. It clearly appears that any codeword in $\mathcal{C}_F$ corresponds to a vector $(a \cdot \alpha_i + b \cdot F(\alpha_i), 1 \leq i < 2^m)$. Therefore, its Hamming weight is given by

$$\#\{i, 1 \leq i < 2^m,\ a \cdot \alpha_i + b \cdot F(\alpha_i) = 1\} = 2^{m-1} - \frac{1}{2}\mathcal{F}(\varphi_b \circ F + \varphi_a)\ .$$

Moreover, a vector $(c_1, \ldots, c_{2^m-1})$ belongs to the dual code $\mathcal{C}_F^{\perp}$ if and only if it satisfies:

$$\sum_{i=1}^{2^m-1} c_i\alpha_i = 0 \quad \text{and} \quad \sum_{i=1}^{2^m-1} c_i F(\alpha_i) = 0\ .$$

Then, we have that the minimum distance of $\mathcal{C}_F^\perp$ is at least 3. Moreover, there exist three different indexes i_1, i_2, i_3 such that

$$F(\alpha_{i_1}) + F(\alpha_{i_2}) + F(\alpha_{i_3}) + F(\alpha_{i_1} + \alpha_{i_2} + \alpha_{i_3}) = 0$$

if and only if $\mathcal{C}_F^\perp$ contains a codeword of Hamming weight 4 (or 3 if $\alpha_{i_1} + \alpha_{i_2} + \alpha_{i_3} = 0$). Therefore, we obtain the following correspondence:

Theorem 1 [7] *Let F be a function from $\mathbf{F}_2^m$ into $\mathbf{F}_2^m$ with $F(0) = 0$. Let $\mathcal{C}_F$ be the linear binary code of length $2^m - 1$ with generator matrix G_F described by (2). Then,*

(i) $\mathcal{L}(F) = \max_{c \in \mathcal{C}_F, c \neq 0} |2^m - 2wt(c)|$.

In particular, for odd m, F is AB if and only if for any non-zero codeword $c \in \mathcal{C}_F$,

$$2^{m-1} - 2^{\frac{m-1}{2}} \leq wt(c) \leq 2^{m-1} + 2^{\frac{m-1}{2}} \ .$$

(ii) *F is APN if and only if the code $\mathcal{C}_F^\perp$ has minimum distance 5.*

When the vector space $\mathbf{F}_2^m$ is identified with the finite field $\mathbf{F}_{2^m}$, the function F can be expressed as a unique polynomial of $\mathbf{F}_{2^m}[X]$. Now, we focus on power functions F, i.e., $F(x) = x^s$ over $\mathbf{F}_{2^m}$. Let α be a primitive element of $\mathbf{F}_{2^m}$. Any cyclic code $\mathcal{C}$ of length $(2^m - 1)$ can be defined by its generator polynomial whose roots are called the zeros of the code. The *defining set* of $\mathcal{C}$ is then the set

$$I(\mathcal{C}) = \{i \in \{0, \cdots, 2^m - 2\} |\ \alpha^i \text{ is a zero of } \mathcal{C}\} \ .$$

Since $\mathcal{C}$ is a binary code, its defining set is a union of 2-cyclotomic cosets modulo $(2^m - 1)$, $Cl(a)$, where $Cl(a) = \{2^j a \bmod (2^m - 1)\}$. From now on the defining set of a binary cyclic code of length $(2^m - 1)$ is identified with the representatives of the corresponding 2-cyclotomic cosets modulo $(2^m - 1)$. Then, when F is a power function $x \mapsto x^s$ over $\mathbf{F}_{2^m}$, the code $\mathcal{C}_F^\perp$ is the cyclic code of length $(2^m - 1)$ with defining set $\{1, s\}$ [7].

In the following, we investigate the weight divisibility of the code $\mathcal{C}_F$ associated to a function F from $\mathbf{F}_2^m$ to $\mathbf{F}_2^m$.

Definition 6 *A binary code $\mathcal{C}$ is said* 2^ℓ-divisible *if the weight of any of its codewords is divisible by 2^ℓ. Moreover $\mathcal{C}$ is said* exactly 2^ℓ-divisible *if, additionally, it contains at least one codeword whose weight is not divisible by $2^{\ell+1}$.*

The following theorem due to McEliece reduces the determination of the exact weight divisibility of binary cyclic codes to a combinatorial problem:

Theorem 2 [30] *A binary cyclic code is exactly 2^ℓ-divisible if and only if ℓ is the smallest number such that $(\ell+1)$ nonzeros of $\mathcal{C}$ (with repetitions allowed) have product 1.*

For a cyclic code with two nonzeros, McEliece's theorem can be reformulated as follows [5].

Corollary 1 *The dual of the cyclic code of length (2^m-1) with defining set $\{1,s\}$ is exactly 2^ℓ-divisible if and only if for all u such that $0 \leq u \leq 2^m-1$,*

$$w_2(A(u)) \leq w_2(u) + m - 1 - \ell$$

where $A(u) = us \bmod (2^m-1)$ and $w_2(u)$ corresponds to the number of 1s in the 2-adic expansion of u.

4 Optimal cryptographic functions and weight divisibility of some $[2^m-1, 2m]$ binary codes

We have pointed out that the cryptographic properties of a round function over $\mathbf{F}_2^m$ which guarantee a high resistance to linear and differential attacks have concern with the weight distributions of some $[2^m-1, 2m]$ linear binary codes. Now, we show that the weight divisibility of the associated code plays a major role in the context of linear cryptanalysis and of higher order differential attacks.

4.1 Characterization of almost bent functions

First, we give some general results on the weight distributions of linear binary codes with parameters $[2^m-1, 2m]$. These properties are derived from Pless power moment identities [35] and from some ideas due to Kasami [21, th. 13]. The proofs of the results given in this section can be found in [5, 7].

Theorem 3 *Let $\mathcal{C}$ be a $[2^m-1, 2m]$ linear code which does not contain the all-one vector $\mathbf{1} = (1, \cdots, 1)$. Assume that the minimum distance of the dual code $\mathcal{C}^\perp$ is at least 3. Let $A = (A_0, \cdots, A_{2^m-1})$ (resp. $B = (B_0, \cdots, B_{2^m-1})$) be the weight enumerator of $\mathcal{C}$ (resp. $\mathcal{C}^\perp$). Then, we have*

(i) *If w_0 is such that for all $0 < w < w_0$*

$$A_w = A_{2^m - w} = 0,$$

then

$$6(B_3 + B_4) \leq (2^m - 1)\left[(2^{m-1} - w_0)^2 - 2^{m-1}\right]$$

where equality holds if and only if

$$A_w = 0 \quad \textit{for all} \quad w \notin \{0, w_0, 2^{m-1}, 2^m - w_0\} \ .$$

(ii) *If w_1 is such that for all $w_1 < w < 2^{m-1}$*

$$A_w = A_{2^m - w} = 0,$$

then

$$6(B_3 + B_4) \geq (2^m - 1)\left[(2^{m-1} - w_1)^2 - 2^{m-1}\right]$$

where equality holds if and only if

$$A_w = 0 \quad \textit{for all} \quad w \notin \{0, w_1, 2^{m-1}, 2^m - w_1\} \ .$$

This theorem gives an upper bound on the value of w_0 for which all nonzero weights of $\mathcal{C}$ lie between w_0 and $2^m - w_0$:

Corollary 2 *Let $\mathcal{C}$ be a $[2^m - 1, 2m]$ linear code which does not contain the all-one vector $\mathbf{1} = (1, \cdots, 1)$. Assume that the minimum distance of the dual code $\mathcal{C}^\perp$ is at least 3. Let w_0 be the smallest w such that $0 < w < 2^{m-1}$ and*

$$A_w + A_{2^m - w} \neq 0$$

Then

$$w_0 \leq 2^{m-1} - 2^{\frac{m-1}{2}}$$

and equality holds if and only if the weight of every codeword in $\mathcal{C}$ belongs to $\{0, 2^{m-1}, 2^{m-1} \pm 2^{\frac{m-1}{2}}\}$. In this case the weight distribution of $\mathcal{C}$ is the same as the weight distribution of the dual of the 2-error-correcting BCH code, i.e.

Weight: w	*Number of words: A_w*
0	1
$2^{m-1} - 2^{(m-1)/2}$	$(2^m - 1)(2^{m-2} + 2^{(m-3)/2})$
2^{m-1}	$(2^m - 1)(2^{m-1} + 1)$
$2^{m-1} + 2^{(m-1)/2}$	$(2^m - 1)(2^{m-2} - 2^{(m-3)/2})$

Then we can deduce a fundamental result for characterizing AB functions by using divisibility properties. Assume that m is odd and consider a code $\mathcal{C}$ which satisfies the hypothesis of Theorem 3. Suppose that the minimum distance of the dual code $\mathcal{C}^\perp$ is 5 and $\mathcal{C}$ is $2^{\frac{m-1}{2}}$-divisible. Then, with notation of Theorem 3, we have $B_3 = B_4 = 0$ and $A_w = A_{2^m-w} = 0$ for all w in the range $[2^{m-1} - 2^{\frac{m-1}{2}}, 2^{m-1}]$. It follows from **(ii)** of Theorem 3 that the nonzero weights of $\mathcal{C}$ are 2^{m-1} and $2^{m-1} \pm 2^{\frac{m-1}{2}}$ only. According to the previous corollary, we obtain:

Corollary 3 *The equality $w_0 = 2^{m-1} - 2^{\frac{m-1}{2}}$ holds in Corollary 2 if and only if $\mathcal{C}$ is $2^{\frac{m-1}{2}}$-divisible and its dual has minimum distance 5.*

Corollary 4 *Let m be odd. Let F be a function from $\mathbf{F}_2^m$ into $\mathbf{F}_2^m$ and let $\mathcal{C}_F$ be the code defined in Theorem 1. Assume that $\mathcal{C}_F$ does not contain the all-one vector.*

Then F is AB if and only if F is APN and the code $\mathcal{C}_F$ is $2^{\frac{m-1}{2}}$-divisible.

Note that the divisibility condition on $\mathcal{C}_F$ equivalently means that all Walsh coefficients of F are divisible by $2^{\frac{m+1}{2}}$. Moreover, $\mathcal{C}_F$ does not contain the all-one vector means $\mathcal{L}(F) \neq 2^m$.

When F is a power function, $F : x \mapsto x^s$, the corresponding code C_F is the dual of a binary cyclic code of length (2^m-1) with defining set $\{1, s\}$. Its weight divisibility can therefore be obtained by applying McEliece's theorem, as expressed in Corollary 1. This leads to the following characterization of AB power functions:

Corollary 5 *Let $m = 2t + 1$. Assume that the power function $F : x \mapsto x^s$ on $\mathbf{F}_{2^m}$ satisfies $\mathcal{L}(F) \neq 2^m$. Then F is AB on $\mathbf{F}_{2^m}$ if and only if F is APN on $\mathbf{F}_{2^m}$ and*

$$\forall u, \;\; 1 \leq u \leq 2^m - 1, \;\; w_2(A(u)) \leq t + w_2(u) \tag{3}$$

where $A(u) = us \bmod (2^m - 1)$.

This result provides a fast algorithm for checking whether an APN power function is AB, and then for finding all AB power functions on $\mathbf{F}_{2^m}$. There are roughly $2^{m-1}/m$ cyclotomic representatives u such that $w_2(u) \leq (m - 1)/2$ and each test requires one modular multiplication on m-bit integers and two weight computations. Condition (3) can then be checked with around 2^m elementary operations and at no memory cost.

Moreover, the determination of the values of s such that $x \mapsto x^s$ is almost bent on $\mathbf{F}_{2^m}$ is now reduced to a combinatorial problem, and this technique was directly used to prove that some power functions are AB [4, 17]. Moreover, it leads to a very efficient method for proving that a given power function is not AB. For example, the APN power function $x \mapsto x^s$ over $\mathbf{F}_{2^{5g}}$ with $s = 2^{4g} + 2^{3g} + 2^{2g} + 2^g - 1$ does not satisfy the condition of Corollary 5 [5].

These recent results lead to the following list (up to equivalence) of known AB permutations (Table 1).

Table 1: Known AB power permutations x^s on $\mathbf{F}_{2^m}$, m odd

exponents s	m	
$2^j + 1$, $\gcd(j, m) = 1$, $1 \leq j \leq \frac{m-1}{2}$		[13, 33]
$2^{2j} - 2^j + 1$, $\gcd(j, m) = 1$, $2 \leq j \leq \frac{m-1}{2}$		[22]
$2^{\frac{m-1}{2}} + 3$		[4]
$2^{\frac{m-1}{2}} + 2^{\frac{m-1}{4}} - 1$	$m \equiv 1 \bmod 4$	[17]
$2^{\frac{m-1}{2}} + 2^{\frac{3m-1}{4}} - 1$	$m \equiv 3 \bmod 4$	[17]

4.2 Weight divisibility and resistance to higher-order differential attacks

Now, we show that the weight divisibility of the codes $\mathcal{C}_{F_k}$ associated to the round function F of an iterated block cipher has concern with its resistance to higher-order differential attacks. Recall that a higher-order differential attack can be performed when the reduced cipher has a low degree. The degree of the round function (*i.e.,* of the functions F_k) provides a trivial bound on the degree of the reduced cipher

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

but this bound can be exploited only when the degree of the round function is very low. Indeed another approach has to be used when the degree of the round function exceeds $\sqrt{m}$, since $(deg(F))^{r-1} > m$ for any $r \geq 3$.

Here, we focus on the degree of the function obtained by composing two permutations F and F'. We show that the trivial bound

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

can be improved when the weight divisibility of the code $\mathcal{C}_F$ associated with F is high.

First, we assume that F is a power function $F : x \mapsto x^s$ over $\mathbf{F}_{2^m}$. Now, any function F' from $\mathbf{F}_2^m$ into $\mathbf{F}_2^m$ can be written as a univariate polynomial in $\mathbf{F}_{2^m}[X]$:

$$F'(X) = \sum_{u=0}^{2^m-1} a_u X^u \ ,$$

where the degree of F' is: $\max\{w_2(u),\ a_u \neq 0\}$. Thus, we deduce that $F' \circ F(x) = \sum_{u=0}^{2^m-1} a_u X^{us \bmod (2^m-1)}$. Therefore,

$$deg(F' \circ F) \leq \max_{u, a_u \neq 0} w_2(us \bmod (2^m - 1)) \ ,$$

and McEliece's theorem directly provides a new bound on the degree of $F' \circ F$(see Corollary 1).

Theorem 4 *Let $F : x \mapsto x^s$ be a power function over $\mathbf{F}_{2^m}$ such that the associated code $\mathcal{C}_F$ defined in Theorem 1 is 2^ℓ-divisible. Then, for any function F' from $\mathbf{F}_2^m$ into $\mathbf{F}_2^m$, we have*

$$deg(F' \circ F) \leq m - 1 - \ell + deg(F') \ .$$

In particular, when F is an almost bent power function over $\mathbf{F}_{2^m}$, we obtain

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

The previous theorem leads to a general explanation for the weakness in the cipher MISTY1 [29] reported in [38] and [1]. The proposed attack relies on the fact that for any quadratic function Q from $\mathbf{F}_2^7$ into $\mathbf{F}_2^7$ we have $\deg(Q \circ S_7) \leq 5$, where S_7 is the almost bent power permutation $x \mapsto x^{81}$ over $\mathbf{F}_{2^7}$. Theorem 4 provides a generalization of this property to any AB power function over $\mathbf{F}_{2^7}$.

We can prove that Theorem 4 is not specific to power functions: it is also valid for any function F from $\mathbf{F}_2^m$ into $\mathbf{F}_2^m$. The proof of the next theorem, very technical, has to be found in [6].

Theorem 5 *Let F be a function from $\mathbf{F}_2^m$ into $\mathbf{F}_2^m$ such that the associated code $\mathcal{C}_F$ is 2^ℓ-divisible. Then, for any function F' from $\mathbf{F}_2^m$ into $\mathbf{F}_2^m$, we have*

$$deg(F' \circ F) \leq m - 1 - \ell + deg(F') \ .$$

Now, we show how the weight divisibility of the code associated with the round function can be exploited in a practical situation. As an example, we consider a 5-round Feistel cipher. In a Feistel cipher with block size $2m$, the round function is defined by

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

where S_k is a function from $\mathbf{F}_2^m$ into $\mathbf{F}_2^m$ 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, c_0 and c_1 being some constants.

When the confusion function S_k is such that all codes $\mathcal{C}_{S_k}$ are 2^ℓ-divisible, we can apply Theorem 5. Then, we obtain the following upper bound for the degree of R_3:

$$deg(R_3) \leq m - 1 - \ell + \max_{k \in \mathcal{K}} \deg(S_k).$$

Let $\delta = \min(\max_{k \in \mathcal{K}} \deg(S_k)^2 + 1, m - \ell + \max_{k \in \mathcal{K}} \deg(S_k))$. We have exhibited a higher-order differential attack on any 5-round Feistel cipher using a confusion function S_k which satisfies $\delta \leq m$:

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 $\mathbf{F}_2^m$ and $(x_i)_{1 \leq i \leq 2^\delta}$ is a δ-dimensional subspace of $\mathbf{F}_2^m$:

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, the previous higher order differential attack can be performed except when $\max_k \deg(S_k) = (m+1)/2$, i.e., when S_k is an almost bent function of maximum degree.

4.3 Highly nonlinear round functions having a low divisibility

The previous results point out that the use of an almost bent round function (or confusion function) is not suitable, even if the corresponding iterated block cipher is provably-secure against differential and linear cryptanalysis. An almost bent function may make the cipher vulnerable to higher-order differential cryptanalysis because of the high weight-divisibility of the associated code.

When m is even, the smallest known value of $\mathcal{L}(F)$ for a function F from $\mathbf{F}_2^m$ into $\mathbf{F}_2^m$ is $\mathcal{L}(F) = 2^{m/2+1}$. The only known functions (up to equivalence) achieving this bound are power functions. Since power permutations cannot be APN when m is even, that the security criteria corresponding to differential cryptanalysis and to linear cryptanalysis are not so strongly related. Moreover, the divisibility of the codes associated with these highly nonlinear functions varies. In particular, the degree of such a function is not upper-bounded since there is no requirement on the weight divisibility. All known functions satisfying this property are equivalent to one of the power functions given in Table 2 (or to one of their inverses) [12].

Table 2: Known power permutations x^s on $\mathbf{F}_{2^m}$, m even, with the highest nonlinearity and exact weight divisibility of the associated code

exponents s	m	divisibility	
$2^{m-1}-1$	$m \equiv 0 \bmod 2$	2	[25]
2^j+1, $\gcd(j,m)=2$, $j<\frac{m}{2}$	$m \equiv 2 \bmod 4$	$2^{\frac{m}{2}}$	[13]
$2^{2j}-2^j+1$, $\gcd(j,m)=2, j<\frac{m}{2}$	$m \equiv 2 \bmod 4$	$2^{\frac{m}{2}}$	[22]
$2^{\frac{m}{2}}+2^{\frac{m+2}{4}}+1$	$m \equiv 2 \bmod 4$	$2^{\frac{m}{2}}$	[10]
$2^{\frac{m}{2}}+2^{\frac{m}{2}-1}+1$	$m \equiv 2 \bmod 4$	$2^{\frac{m}{2}}$	[10]
$\sum_{i=0}^{m/2} 2^{ij}$, $\gcd(j,m)=1, j<\frac{m}{2}$	$m \equiv 0 \bmod 4$	$2^{\frac{m}{2}-1}$	[12]
$2^{\frac{m}{2}}+2^{\frac{m}{4}}+1$	$m \equiv 4 \bmod 8$	$2^{\frac{m}{2}-1}$	[12]

It appears that all known optimal functions for m even are such that the associated codes are either $2^{\frac{m}{2}-1}$-divisible or $2^{\frac{m}{2}}$-divisible, except the

inverse function. The inverse function is very specific in this context. The corresponding code $\mathcal{C}_F$ is the dual of the Melas code. Its weights are all even integers w such that $|w - \frac{2^m-1}{2}| \leq 2^{\frac{m}{2}}$ [25]. Therefore, it has the smallest possible weight divisibility when F is a permutation.

Moreover, $x \mapsto x^{2^{m-1}-1}$ is the only power permutation F of $\mathbf{F}_{2^m}$ (up to equivalence) which corresponds to an exactly 2-divisible code.

Proposition 4 [16] *Let m and s be two positive integers such that* $\gcd(s, 2^m - 1) = 1$. *Let $\mathcal{C}_s$ be the dual of binary cyclic code of length $(2^m - 1)$ with defining set $\{1, s\}$. Then, $\mathcal{C}_s$ is exactly 2-divisible if and only if $w_2(s) = m - 1$.*

Thus, the inverse function 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 [11].

The search for other confusion functions which are "almost optimal" with respect to these criteria leads to the following open problem.

Open problem 1 *Find all permutations F of $\mathbf{F}_2^m$ such that the associated code C_F defined in Theorem 1 satisfies:*

(i) $\max_{c \in \mathcal{C}_F, c \neq 0} |2^m - 2wt(c)|$ *is close to* $2^{\frac{m-1}{2}}$*;*

(ii) *The number of codewords of weight 3 and 4 in the dual code $\mathcal{C}_F^{\perp}$ is small;*

(iii) *The exact weight divisibility of $\mathcal{C}_F$ is low.*

For power functions, McEliece theorem provides a useful tool for finding such functions. For instance, the exponents s such that the code associated with $x \mapsto x^s$ is exactly 4-divisible are known.

Proposition 5 [5, Prop. 5.3] *Let m and s be two positive integers such that* $\gcd(s, 2^m - 1) = 1$. *Let $\mathcal{C}_s$ be the dual of the binary cyclic code of length $(2^m - 1)$ with defining set $\{1, s\}$. Then, $\mathcal{C}_s$ is exactly 4-divisible if and only if either $w_2(s) = m - 2$ or $w_2(s^{-1}) = m - 2$ where s^{-1} is the only integer in $\{0, \ldots, 2^m - 1\}$ such that $s^{-1}s \equiv 1 \bmod (2^m - 1)$.*

A necessary condition on s for obtaining an exactly 8-divisible code can also be found in [5, Prop. 5.4].

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. Binary m-sequences with three-valued crosscorrelation: A proof of Welch conjecture. *IEEE Trans. Inform. Theory*, 46(1):4–8, 2000.

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

[6] A. Canteaut and M. Videau. Weakness of block ciphers using highly nonlinear confusion functions. research report, INRIA, 2001. to appear.

[7] 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.

[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] J. Daemen and V. Rijmen. AES proposal: the Rijndael block cipher. Available at `http://csrc.nist.gov/encryption/aes/rijndael/`, 1999.

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

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

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

[15] 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.

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

[17] H. Hollman and Q. Xiang. A proof of the Welch and Niho conjectures on crosscorrelations of binary m-sequences. *Finite Fields and Their Applications*, 7(2):253–286, 2001.

[18] 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.

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

[20] 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.

[21] T. Kasami. Weight distributions of Bose-Chaudhuri-Hocquenghem codes. In *Proceedings of the conference on combinatorial mathematics and its applications*, pages 335–357. The Univ. of North Carolina Press, 1968.

[22] 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.

[23] 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.

[24] 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.

[25] 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.

[26] 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.

[27] 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.

[28] 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.

[29] 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.

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

[31] 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.

[32] 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.

[33] 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.

[34] 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.

[35] V. Pless. Power moment identities on weight distributions in error-correcting codes. *Info. and Control*, 3:147–152, 1963.

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

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

[38] 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.