Another View of the Division Property

Christina Boura, Anne Canteaut

Another view of the division property*

Christina Boura and Anne Canteaut

[1] University of Versailles, France, Christina.Boura@uvsq.fr
[2] Inria, Paris, France, Anne.Canteaut@inria.fr

Abstract. A new distinguishing property against block ciphers, called the division property, was introduced by Todo at Eurocrypt 2015. Our work gives a new approach to it by the introduction of the notion of parity sets. First of all, this new notion permits us to formulate and characterize in a simple way the division property of any order. At a second step, we are interested in the way of building distinguishers on a block cipher by considering some further properties of parity sets, generalising the division property. We detail in particular this approach for substitution-permutation networks. To illustrate our method, we provide low-data distinguishers against reduced-round PRESENT. These distinguishers reach a much higher number of rounds than generic distinguishers based on the division property and demonstrate, amongst others, how the distinguishers can be improved when the properties of the linear and the Sbox layer are taken into account. At last, this work provides an analysis of the resistance of Sboxes against this type of attacks, demonstrates links with the algebraic normal form of an Sbox as well as its inverse Sbox and exhibit design criteria for Sboxes to resist such attacks.

Keywords: division property, integral attacks, Sboxes, PRESENT.

1 Introduction

A new distinguishing property against block ciphers, called the division property, was recently introduced by Todo [25]. This property, that can be seen as a generalization of integral [16, 9] and higher-order differential [17, 15] distinguishers, was used to present new generic distinguishers against both the SPN and the Feistel constructions. Later, this attack was used by the same author to present the first cryptanalysis of the full block cipher MISTY [24].

If $u = (u_1, \ldots, u_n)$ is a vector of $\mathbf{F}_2^n$, we denote by x^u the coordinate product $x = (x_1, \ldots, x_n) \mapsto \prod_{i=1}^n x_i^{u_i}$. The division property, as introduced by Todo [25], is interested in the sum of this quantity taken over all vectors of X. More precisely, we say that a set $X \subseteq \mathbf{F}_2^n$ has the division property $\mathcal{D}_k^n$, for some $1 \leq k \leq n$, if the sum over all vectors x in X of the product x^u equals 0, for all vectors u that have a Hamming weight strictly less than k, i.e.

$$\bigoplus_{x \in X} x^u = 0 \text{ for all } u \in \mathbf{F}_2^n \text{ such that } wt(u) < k.$$

* Partially supported by the French Agence Nationale de la Recherche through the BRUTUS project under Contract ANR-14-CE28-0015.

The division property then generalizes integral attacks in the sense that $\mathcal{D}_2^n$ means that the set X is balanced, while $\mathcal{D}_n^n$ means that it is saturated. But the novelty is that it introduces intermediate properties, $\mathcal{D}_k^n$ for $3 \leq k \leq n-1$, which do not appear in classical integral attacks. Even if these intermediate properties do not have a simple interpretation like $\mathcal{D}_2^n$ and $\mathcal{D}_n^n$, they allow to easily propagate the property through the successive rounds of a cipher by capturing some information resulting from the algebraic degree of the round function. In a nutshell, the distinguishers described by Todo in [25, 24] are classical higher-order differential distinguishers, but they are exhibited by exploiting the classical properties used in integral attacks together with some algebraic properties related to the degree of several iterations of a nonlinear function like in [8, 6].

Our contribution. This work aims at providing new insights into the division property, presenting a new approach to it. This new approach enables us to provide a simpler formulation and interpretation of the division property of any order. It also improves the strength of the distinguishers that exploit this type of properties. For this, we introduce a new notion, that we call the *parity set*. The parity set of a set $X \subseteq \mathbf{F}_2^n$ is nothing more than the set of all exponents $u \in \mathbf{F}_2^n$ such that $\bigoplus_{x\in X} x^u = 1$. The main advantage of this new notion is that it completely characterises a set X in the sense that there is a one-to-one correspondence between sets and their parity sets. It also provides a very simple formulation of the division property. In particular, we show that the division property of any order can be expressed in an elegant way by using the theory of Reed-Muller codes. One of the first questions we investigate in this work is what does it mean for a set X to have the division property $\mathcal{D}_k^n$ for some special values of k. As previously explained, this question was treated for $k \in \{1, 2, n\}$ in previous works [25, 22, 14]. However, our approach, and especially the link with the Reed-Muller codes, permits us to recover in a much simpler way these previous results and to characterize the property for some other values of k.

We investigate next the question of how to build distinguishers for keyed permutations by means of parity sets. For this, we start by analyzing the distinguishers built by Todo in [25] and formulate them in terms of parity sets. These distinguishers were of a generic nature as they only exploited the classical integral properties and the propagation of the algebraic degree through the successive non-linear layers. It is thus natural to believe that these distinguishers can be improved if additional information besides the degree is taken into account, in the same spirit as in cube distinguishers [12, 1]. We investigate this issue here and provide a way to exploit these more precise properties. We then further show how to find distinguishers on iterated block ciphers, especially on substitution-permutation networks, by propagating some information on the parity set of the output set through the successive rounds of the cipher. For this, we provide a detailed analysis of the evolution of the parity set through the basic operations of the round function of an SPN cipher.

We illustrate the above technique by constructing low-data distinguishers on the PRESENT block cipher. With this example we aim at particularly showing how the generic distinguishers can be improved when the properties of the linear and the Sbox layer are taken into account. We manage to provide a distinguisher on 6 rounds of PRESENT with data complexity 2^{12}, while the generic distinguisher from [25], reaches only 3 rounds for the same quantity of data. Finally, we analyze the resistance of Sboxes against this type of attack, show a link with the algebraic normal forms of the Sbox and of its inverse, and we give a criterion that an Sbox must satisfy in order to resist this kind of attacks.

Organization of the paper. The rest of the paper is organized as follows. Section 2 introduces the notion of the parity set of a set and shows how it is related to Reed-Muller codes. Section 3 presents the link between this new notion and the division property, and it characterizes the division property of any order. It also focuses on the division property of low and high orders. Section 4 explains how to build distinguishers by means of parity sets and Section 5 analyzes the special case of SPN ciphers. In Section 6, low-data distinguishers for the block cipher PRESENT are presented. Finally, Section 7 discusses the properties that an Sbox must exhibit in order to resist the above attacks.

2 Parity set of a set

2.1 Preliminaries

A Boolean function of n variables can be alternatively represented as a multivariate polynomial, named the Algebraic Normal Form (aka ANF) of the function, or as a 2^n-bit vector (named the value vector) corresponding to all $f(x), x \in \mathbf{F}_2^n$.

Polynomial representation. We use the following notation for the monomials of n variables where u is an element of $\mathbf{F}_2^n$:

$$x^u = \prod_{i=1}^{n} x_i^{u_i} \ .$$

The following well-known lemma will be extensively used for evaluating a monomial at a given point.

Lemma 1. *Let x and u be two n-bit words. Then $x^u = 1$ if and only if $u \preceq x$, i.e., $u_i \leq x_i$ for all $1 \leq i \leq n$.*

The previous relation between n-bit words is a partial order. It equivalently means that the support of u is included in the support of x. In the whole paper, we will use the following notation for the set of all words less than (resp. greater than) a given word with respect to this partial order.

Notation 1 *Let $u \in \mathbf{F}_2^n$. Then, we define*

$$\begin{aligned}\mathsf{Prec}(u) &= \{x \in \mathbf{F}_2^n \ : x \preceq u\} \\ \mathsf{Succ}(u) &= \{x \in \mathbf{F}_2^n \ : u \preceq x\} \ .\end{aligned}$$

It is worth noticing that $\mathsf{Prec}(u)$ is a linear subspace of dimension $wt(u)$, while $\mathsf{Succ}(u)$ is an affine subspace of dimension $(n - wt(u))$.

Value vector. When a Boolean function is represented by its value vector, it is often convenient to use the terminology coming from coding theory since Boolean functions have been widely studied for error-correction. In this context, the value vector of a function is seen as a codeword from a Reed-Muller code [21, 20].

Definition 1 (Reed-Muller codes). *Let n be a positive integer and r an integer such $0 \leq r \leq n$. The* r-th order binary Reed-Muller code of length 2^n*, denoted by $\mathcal{R}(r, n)$, is the set formed by the value vectors of all Boolean functions of n variables with degree at most r:*

$$\mathcal{R}(r, n) = \{(f(x), x \in \mathbf{F}_2^n), f : \mathbf{F}_2^n \to \mathbf{F}_2 \textit{ with } \deg f \leq r\} .$$

2.2 Parity set of a set

We now define a new notion named *parity set*. We will show that any set is characterized by its parity set. Any property of a set can then also be expressed in terms of its parity set, and we will show that the division property has a very simple expression by means of parity sets.

Definition 2. *Let X be a set of elements in $\mathbf{F}_2^n$. The* parity set of X*, denoted by $\mathcal{U}(X)$, is the subset of $\mathbf{F}_2^n$ defined by*

$$\mathcal{U}(X) = \{u \in \mathbf{F}_2^n : \bigoplus_{x \in X} x^u = 1\} .$$

The parity set provides a complete characterization of a set, as shown by the following results.

Lemma 2. *Let G be the $2^n \times 2^n$ binary matrix whose entries are indexed by n-bit vectors and defined by*

$$G_{u,a} = a^u, \ a, u \in \mathbf{F}_2^n .$$

For any subset X of $\mathbf{F}_2^n$, the incidence vector of $\mathcal{U}(X)$ is equal to the product of G by the incidence vector of X.

Proof. The incidence vector of a set X, v_X, is the 2^n-bit vector having a one at position $x \in \mathbf{F}_2^n$ if and only if $x \in X$. Then, Gv_X is equal to the sum of all columns of G indexed by the elements in the support of v_X, i.e., indexed by the elements in X:

$$(Gv_X)_u = \bigoplus_{x \in X} x^u .$$

By definition, the support of $v_{\mathcal{U}(X)}$ is then the set of all positions u such that $(Gv_X)_u = 1$. □

We can now deduce that there is a one-to-one correspondence between sets and their parity sets.

Theorem 1. *Let G be the $2^n \times 2^n$ binary matrix defined by*

$$G_{u,a} = a^u, \ a, u \in \mathbf{F}_2^n \ .$$

Then, G is non-singular and $G^{-1} = G$. Therefore, for any subset U of $\mathbf{F}_2^n$, there exists a unique set $X \subset \mathbf{F}_2^n$ such that $\mathcal{U}(X) = U$.

Proof. The fact that G is non-singular can be deduced by using that it is a generator matrix of the Reed-Muller code of length 2^n and order n. This code has dimension 2^n [19, Page 376], i.e., G is invertible. Its inverse is equal to G itself. Indeed, for any $u, w \in \mathbf{F}_2^n$, we have

$$\begin{aligned}(G \times G)_{u,w} &= \bigoplus_{v \in \mathbf{F}_2^n} G_{u,v} G_{v,w} = \bigoplus_{v \in \mathbf{F}_2^n} v^u w^v \\ &= |\{v \in \mathbf{F}_2^n \ : u \preceq v \text{ and } v \preceq w\}| \bmod 2 \\ &= \begin{cases} 2^{wt(w)-wt(u)} \bmod 2 \text{ if } u \preceq w \\ 0 \text{ otherwise.} \end{cases}\end{aligned}$$

We then deduce that $(G \times G)_{u,w} = 1$ if and only if $u = w$, i.e., $G \times G = \mathsf{Id}$. As a direct consequence, we get that the mapping $v_X \mapsto v_{\mathcal{U}(X)}$ is an isomorphism of the set of 2^n-bit vectors. □

The fact that G is involutive provides a simple way to find the set X corresponding to a given parity set U. Indeed, X corresponds to the parity set of U. Some useful examples are described in the following corollary.

Corollary 1. *Let X be a subset of $\mathbf{F}_2^n$. Then,*

- *$\mathcal{U}(X)$ is empty if and only if X is empty.*
- *$\mathcal{U}(X) = \mathsf{Prec}(x)$ if and only if $X = \{x\}$.*
- *$\mathcal{U}(X) = \{u\}$ if and only if $X = \mathsf{Prec}(u)$.*
- *$\mathcal{U}(X) = \{\underline{1}\}$ if and only if $X = \mathbf{F}_2^n$,*

where $\underline{1}$ denotes the all-one vector in $\mathbf{F}_2^n$.

3 New insights into the division property

3.1 The division property by means of parity sets

The division property introduced by Todo in [25] is a distinguishing property of the set $E_k(X)$ for a given choice of the input set X, where E_k is (typically) a keyed permutation. This property must be independent from the choice of the secret key. We now reformulate the division property of order k, $\mathcal{D}_k^n$, on a set X by a simple property of $\mathcal{U}(X)$. Indeed, $\mathcal{D}_k^n$ corresponds to a lower bound on the weights of all elements in $\mathcal{U}(X)$.

Definition 3. *A set X of elements in $\mathbf{F}_2^n$ is said to fulfill the division property of order k, $\mathcal{D}_k^n$, if all elements in $\mathcal{U}(X)$ have weight at least k, i.e.,*

$$\mathcal{U}(X) \subseteq \{u \in \mathbf{F}_2^n : wt(u) \geq k\} \ .$$

It is worth noticing that in [25], the division property is defined for a multiset, i.e., the elements in X may appear with some multiplicity. However, the original division property for a multiset X equivalently corresponds to the division property for the set composed of all elements in X having an odd multiplicity. Therefore, we will only focus on sets, instead of multisets.

As a direct consequence of the matrix relationship exhibited in Lemma 2, we deduce the following two characterizations of the incidence vectors of the sets satisfying the division property of order k.

Proposition 1. *Let X be a set of elements in $\mathbf{F}_2^n$ and k be an integer $1 \leq k \leq n$. Then, the following assertions are equivalent:*

(i) *X fulfills the division property of order k, $\mathcal{D}_k^n$.*
(ii) *The incidence vector of X belongs to the Reed-Muller code of length 2^n and order $(n-k)$.*
(iii) *The incidence vector of X belongs to the dual of the Reed-Muller code of length 2^n and order $(k-1)$.*

Proof. Assertion (ii) equivalently means that the incidence vector of $\mathcal{U}(X)$ vanishes at all positions u with $wt(u) \leq k-1$. This means that, if G' denotes the restriction of G to the rows of index u with $wt(u) < k$, then $G'v_X$ is the all-zero vector. But G' is a generator matrix of the Reed-Muller code of length 2^n and order $(k-1)$. The set of all v_X such that $G'v_X = 0$ is therefore the dual (i.e., the orthogonal) of $\mathcal{R}(k-1, n)$. It is well-known (see e.g. [19, Page 375]) that, for any r, the dual of $\mathcal{R}(r, n)$ is the Reed-Muller code $\mathcal{R}(n-r-1, n)$. We deduce that $G'v_X = 0$ if and only if $v_X \in \mathcal{R}(n-k, n)$. □

The first one of the previous characterization, (ii), has been independently exhibited by Khovratovich [14], while the equivalent formulation (iii) is new.

Using the minimum distance of the Reed-Muller codes, we recover very easily a result from [22] on the minimal size of a set satisfying $\mathcal{D}_k^n$. More importantly, we are able to characterize the sets of minimal size satisfying $\mathcal{D}_k^n$.

Proposition 2. *Let X be a non-empty set of elements in $\mathbf{F}_2^n$ satisfying $\mathcal{D}_k^n$. Then*

$$|X| \geq 2^k \ .$$

Moreover, a set X of size 2^k satisfies $\mathcal{D}_k^n$ if and only if X is an affine subspace[1] *of dimension k.*

Proof. We here use that X satisfies $\mathcal{D}_k^n$ if and only if v_X belongs to $\mathcal{R}(n-k, n)$. It is well-known that the minimum distance of $\mathcal{R}(n-k, n)$ is 2^k [19, Page 375]. Using that $|X| = wt(v_X)$, we deduce that $|X| \geq 2^k$.

Moreover, it is known that the minimum-weight codewords in $\mathcal{R}(n-k, n)$ are the incidence vectors of the affine subspaces of dimension k [19, Page 380]. It follows that a set of size 2^k satisfies $\mathcal{D}_k^n$ if and only if it is an affine subspace of dimension k. □

[1] In the whole paper, the terminology *affine subspace* includes any linear subspace or any coset of a linear subspace.

3.2 Division property of low order

Since the codewords of Reed-Muller codes of low order have a very simple form, a simple characterization of the division properties of low order directly follows. The Reed-Muller code $\mathcal{R}(0, n)$ consists of the all-zero and all-one words, $\mathcal{R}(0, n) = \{\underline{0}, \underline{1}\}$, and $\mathcal{R}(1, n) \setminus \mathcal{R}(0, n)$ is composed of all incidence vectors of affine hyperplanes. Then, we easily recover the characterization of the division properties of order 1 and 2 exhibited in [25]:

- X fulfills $\mathcal{D}_1^n$ if and only if its cardinality is even.
- X fulfills $\mathcal{D}_2^n$ if and only if its cardinality is even and it has the *Balance property* [16], i.e., $\oplus_{x \in X} x = 0$.

Then, the division property of order $k > 2$ generalizes the balance property used in integral attacks [16] in the following sense.

Proposition 3. *Let X be a set of elements in $\mathbf{F}_2^n$. Then, the following assertions are equivalent:*

(i) *X fulfills the division property of order k, $\mathcal{D}_k^n$.*
(ii) *For any set of coordinates $\{i_1, \dots, i_t\} \subseteq \{1, \dots, n\}$ of size $t < k$ and any constant $\alpha \in \mathbf{F}_2^t$, the number of elements in X such that $x_{i_j} = \alpha_j$ for all $1 \leq j \leq t$ is even.*
(iii) *For any set of coordinates $\{i_1, \dots, i_t\} \subseteq \{1, \dots, n\}$ of size $t < k$, the number of elements in X such that $x_{i_j} = 0$ for all $1 \leq j \leq t$ is even.*

Proof. **(i) ⇒ (ii)** Let $I = \{i_1, \dots, i_t\}$ be a set of coordinates of size $t < k$, and u be the vector in $\mathbf{F}_2^n$ having support I. Then,

$$\{x \in X : x_{i_j} = \alpha_j, 1 \leq j \leq t\} = \{x \in X : (x \oplus \beta) \succeq u\}$$

where β is the n-bit vector such that $\beta_{i_j} = \alpha_j \oplus 1$ for $1 \leq j \leq t$, and $\beta_i = 0$ if $i \notin I$. It follows that

$$\begin{aligned}|\{x \in X : x_{i_j} = \alpha_j, 1 \leq j \leq t\}| \bmod 2 &= \bigoplus_{x \in X} (x \oplus \beta)^u = \bigoplus_{x \in X} \bigoplus_{v \preceq u} x^v \beta^{u \oplus v} \\ &= \bigoplus_{v \preceq u} \beta^{u \oplus v} \left(\bigoplus_{x \in X} x^v \right) = 0\end{aligned}$$

since the division property of order k implies that all $\bigoplus_{x \in X} x^v$ vanish when $wt(v) \leq wt(u) < k$.

(ii) ⇒ (iii) Trivial.

(iii) ⇒ (i) Let $u \in \mathbf{F}_2^n$ with $wt(u) < k$. We have

$$\begin{aligned}\bigoplus_{x \in X} x^u &= \bigoplus_{x \in X} ((x \oplus u) \oplus u)^u = \bigoplus_{v \preceq u} \bigoplus_{x \in X} (x \oplus u)^v u^{u \oplus v} \\ &= \bigoplus_{v \preceq u} \bigoplus_{x \in X} (x \oplus u)^v = \bigoplus_{v \preceq u} |\{x \in X : x_i = 0, \forall i \in \mathsf{Supp}(v)\}| \bmod 2 .\end{aligned}$$

From (iii), all sets involved in the previous sum have even size because $wt(v) < k$. We then deduce that $\bigoplus_{x \in X} x^u = 0$, i.e. X fulfills the division property of order k.

□

It is worth noticing that, more generally, the division property $\mathcal{D}_k^n$ implies that the number of elements in $X \cap A$ is even for any affine subspace A of dimension $t > n - k$ (see e.g. [7, Prop III.1]).

As a direct corollary, we can for instance characterize the division property of order 3.

Corollary 2. *Let X be a set of elements in $\mathbf{F}_2^n$. Then, X fulfills the division property of order* 3*, $\mathcal{D}_3^n$, if and only if X and all the n subsets*

$$\{x \in X \text{ with } x_i = 0\},\ 1 \leq i \leq n,$$

satisfy the balance property.

Example 1. The following set $X \in \mathbf{F}_2^5$ composed of 12 elements satisfy the division property of order 3:

$$\begin{array}{c|cccccccccccc}
x_1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
x_2 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 0 \\
x_3 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 0 \\
x_4 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 \\
x_5 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 1
\end{array}$$

3.3 Division property of high order

The division property of maximal order, i.e., $\mathcal{D}_n^n$, obviously corresponds to the fact that X is either empty, or equal to the whole set $\mathbf{F}_2^n$ (see the last item in Corollary 1). But, we are also able to characterize all sets satisfying $\mathcal{D}_{n-1}^n$.

Proposition 4. *Let X be a set of elements in $\mathbf{F}_2^n$. Then X fulfills $\mathcal{D}_{n-1}^n$ and not $\mathcal{D}_n^n$ if and only if X is an (affine) hyperplane of $\mathbf{F}_2^n$.*

Proof. Let v denote the incidence vector of X. From the previous proposition, we have X satisfies $\mathcal{D}_{n-1}^n$ and not $\mathcal{D}_n^n$ if and only if $v \in \mathcal{R}(1, n) \setminus \mathcal{R}(0, n)$. This set consists of the incidence vectors of all (affine) hyperplanes of $\mathbf{F}_2$. Then, this equivalently means that X is an (affine) hyperplane. □

For instance, it can be easily checked that the multiset of elements of $\mathbf{F}_2^4$ defined in [25, Page 293]

$$\{\texttt{0x0}, \texttt{0x3}, \texttt{0x3}, \texttt{0x3}, \texttt{0x5}, \texttt{0x6}, \texttt{0x8}, \texttt{0xb}, \texttt{0xd}, \texttt{0xe}\},$$

satisfies $\mathcal{D}_3^4$ because the corresponding set composed of all elements with an odd multiplicity

$$\{\texttt{0x0}, \texttt{0x3}, \texttt{0x5}, \texttt{0x6}, \texttt{0x8}, \texttt{0xb}, \texttt{0xd}, \texttt{0xe}\}$$

is a linear subspace of dimension 3 spanned by $\{\texttt{0x3}, \texttt{0x5}, \texttt{0x8}\}$.

4 Distinguishers based on parity sets

We now investigate how we can build some distinguishers for a given keyed permutation E_K by means of parity sets. The basic idea consists in choosing an input set X such that the parity set of the corresponding output set $E_K(X)$ has some specific property for any choice of the key.

Since the size of X determines the data complexity of the distinguisher, it has to be as small as possible. For this reason, X is always chosen to be an affine subspace since subspaces are the smallest possible sets satisfying the division property of a given order (see Prop. 2).

4.1 Todo's distinguishers

The strategy proposed by Todo to build a distinguisher is to exhibit an affine subspace $a + V$ such that the corresponding output set $E_K(a + V)$ satisfies the division property of order 2, i.e., such that $E_K(a+V)$ is balanced. This property can be easily interpreted in terms of higher-order derivatives in the sense of the following definition.

Definition 4. *[17] Let F be a function from $\mathbf{F}_2^n$ into $\mathbf{F}_2^m$. Let $a \in \mathbf{F}_2^n$. The derivative of F with respect to a is the function from $\mathbf{F}_2^n$ into $\mathbf{F}_2^m$ defined by*

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

For any k-dimensional subspace V of $\mathbf{F}_2^n$ and for any basis of V, $\{a_1, \dots, a_k\}$, the k-th order derivative of F with respect to V is the function defined by

$$D_V F(x) = D_{a_1} D_{a_2} \dots D_{a_k} F(x) = \bigoplus_{v \in V} F(x + v)$$

We introduce now the following notation.

Notation 2 *Let P be a permutation of $\mathbf{F}_2^n$ and $P_1, P_2, \dots, P_n$ be the n coordinates of P. If $x = (x_1, \dots, x_n)$ and $u = (u_1, \dots, u_n)$ are vectors of $\mathbf{F}_2^n$, we denote by $P^u(x)$ the coordinate product $\prod_{i=1}^n P_i(x)^{u_i}$.*

Proposition 5. *Let P be a permutation of $\mathbf{F}_2^n$. Let V be the linear subspace of $\mathbf{F}_2^n$ and $a \in \mathbf{F}_2^n$. Then, an element u belongs to $\mathcal{U}(P(a + V))$ if and only if the derivative of P^u with respect to V satisfies*

$$D_V P^u(a) = 1 \ .$$

In the particular case where $V = \mathsf{Prec}(v)$ for some $v \in \mathbf{F}_2^n$, the following formulation are equivalent:

(i) *For all $a \in \mathbf{F}_2^n$, $u \notin \mathcal{U}(P(a + V))$*

(ii) *The algebraic normal form of the Boolean function $x \mapsto P^u(x)$ contains no monomial multiple of x^v*

(iii) *The superpoly of v in $x \mapsto P^u(x)$ vanishes.*

Proof. The fact that $D_V P^u(a) = 1$ if and only if $u \in \mathcal{U}(P(a+V))$ is directly deduced from the definition of the derivative with respect to V using that

$$D_V P^u(a) = \bigoplus_{x \in a+V} P^u(x) \ .$$

The superpoly of v in P^u is defined [12] as the Boolean function p_v such that

$$P^u(x) = x^v p_v(x) + q(x)$$

where q does not contain any monomial multiple of x^v. Moreover, it has been proved in [12, Th. 1] that

$$p_v(a) = \bigoplus_{x \preceq v} P^u(a \oplus x) = D_V P^u(a) \ .$$

It then follows that, for $V = \mathsf{Prec}(v)$, $u \notin \mathcal{U}(P(a+V))$ for all a if and only if the superpoly of v in P^u vanishes, which equivalently means that P^u does not contain any monomial multiple of x^v. □

4.2 Improving Todo's distinguishers

The distinguishers presented in [25] correspond to the existence of a word v such that $E_K(a + \mathsf{Prec}(v))$ satisfies the division property of order 2 for all a, which equivalently means that the monomial x^v does not appear in any coordinate of E_K. Since these distinguishers are constructed by propagating some information on the smallest Hamming weight of the elements in the parity set, they are based on the fact that the weight of v exceeds the degree of E_K, where the degree of the coordinates of the cipher after several iterations can be upper bounded by exploiting the techniques introduced in [6, 5]. However, it clearly appears that this type of distinguishers can be improved in the following two directions:

- it may happen that a given monomial x^u does not appear in the coordinates of E_K even if $wt(u) \leq \deg P$. This type of property, derived from the sparsity of some coordinates of the cipher, has been extensively used in cube attacks, e.g. [1, 13, 11].
- it may happen that a given monomial x^u appears in one coordinate of E_K but not in all functions $x \mapsto E_K^v(x)$. Then, $E_K(a + \mathsf{Prec}(u))$ does not fulfill the division property of order 2; instead we obtain a weaker distinguisher based on the fact that a given v does not belong to the parity set of $E_K(a+\mathsf{Prec}(u))$.

These two ways of exploiting some additional information besides the degree of the function are illustrated in the following toy example. In the rest of the paper, binary words are represented in hexadecimal notation where the least significant bit corresponds to the rightmost bit in the binary word.

Example 2. Let us consider the 4-bit Sbox S used in PRESENT [2]. This Sbox has degree 3 which is the maximal degree for a permutation of $\mathbf{F}_2^4$. The rows of Table 1 describes all sets

$$V_s(u) = \{v \in \mathbf{F}_2^n : S^v(x) \text{ contains } x^u\} .$$

In others words, the entry at Row u and Column v in this table is an x if and only if x^u appears in the ANF of $x \mapsto S^v(x)$. Equivalently, the column of index v in this table corresponds the list of all monomials in the ANF of $x \mapsto S^v(x)$.

	$V_S(u)$															
	0	1	2	4	8	3	5	9	6	a	c	7	b	d	e	f
0	x			x	x						x					
1		x			x		x				x					
2			x		x				x		x					
4		x		x				x			x					
8		x	x	x	x	x					x					
3				x		x	x	x	x	x	x		x			
5							x	x			x					
9				x		x	x		x	x					x	
6		x			x			x	x	x	x					
a			x	x			x	x		x		x	x	x	x	x
c			x			x		x			x					
7			x		x	x		x	x				x	x		
b			x	x	x	x			x	x	x	x		x		x
d			x	x	x			x		x		x			x	
e							x					x	x	x	x	x
f																x

Table 1. Sets $V_S(u)$ for all $u \in \mathbf{F}_2^4$ for the PRESENT Sbox. All 4-bit words are represented in hexadecimal notation, and the rightmost bit of the word corresponds to the least significant bit.

Clearly, we cannot exhibit any u such that $S(a + \mathsf{Prec}(u))$ fulfills the division property of order 2 for all a (i.e. such that no coordinate of S contains x^u) if we exploit the degree of the Sbox only. Indeed, using that $\deg(S) = 3$, we deduce that the all-one vector $u =$ `0xf` is the only value satisfying this property. This does not provide any distinguisher since this holds for all permutations.

However, distinguishers can be found by using that not all $u \in \mathbf{F}_2^4$ of weight less than or equal to 3 appear in the ANF of the coordinates of S. Indeed $u =$ `0xe` does not appear in any of the coordinates of S, i.e., for $u =$ `0xe`, $S(a + \mathsf{Prec}(u))$ fulfills the division property of order 2 for all a. It is worth noticing that this is the only value of u which provides such a distinguisher.

Instead of searching for u such that the parity set of $S(a + \mathsf{Prec}(u))$ does not contain any word of weight 1, we can exhibit a few values v which belong to none of the sets $S(a + \mathsf{Prec}(u))$, $a \in \mathbf{F}_2^4$. For instance, we can observe that neither S^1 nor S^{e} contains a multiple of $x^3 = x_1x_2$. This means that $v = \mathtt{0x1}$ and $v = \mathtt{0xe}$ do not belong to a set $S(a + \mathsf{Prec}(\mathtt{0x3}))$. Similarly, the sets $S(a + \mathsf{Prec}(\mathtt{0x9}))$ do not contain $v = \mathtt{0x1}$ and $v = \mathtt{0xb}$. The sets $S(a + \mathsf{Prec}(\mathtt{0xc}))$ do not contain $v = \mathtt{0x1}$ and $v = \mathtt{0x6}$.

Clearly, these two ideas allow us to decrease the dimension of the input subspace $X = a + \mathsf{Prec}(u)$ involved in the distinguisher. But the distinguisher based on the division property of order 2 is obviously stronger than the second one. A precise evaluation of the advantages of these distinguishers is provided in Appendix.

5 Exhibiting distinguishers on SPN by means of parity sets

We now show how to find some distinguishers on iterated block ciphers, especially on substitution-permutation networks, by propagating some information on the parity set of the output set through the successive rounds of the cipher. As previously explained, we choose as input set an affine subspace.

5.1 Propagation through key addition

One of the difficulties for finding a distinguisher for a block cipher is that the distinguishing property must hold for any value of the secret key. For this reason, we need to exploit a property which can be easily propagated through the operation inserting the round key, which is usually an XOR. This is the case of differential properties, or of the algebraic degree. We can show that the parity set can also be easily propagated.

Proposition 6. *Let X be a subset of $\mathbf{F}_2^n$ with parity set $\mathcal{U}(X)$. Then, for any $k \in \mathbf{F}_2^n$, the parity set of $(k + X)$ satisfies*

$$\mathcal{U}(k + X) \subseteq \bigcup_{u \in \mathcal{U}(X)} \mathsf{Succ}(u) .$$

Proof. We use that

$$(x \oplus k)^v = \bigoplus_{u \preceq v} x^v k^{v \oplus u} .$$

It follows that

$$\bigoplus_{x \in X} (x \oplus k)^v = \bigoplus_{x \in X} \bigoplus_{u \preceq v} x^u k^{v \oplus u} = \bigoplus_{u \preceq v} k^{v \oplus u} \left(\bigoplus_{x \in X} x^u \right) .$$

Then, this sum equals zero for all $k \in \mathbf{F}_2^n$ if $\bigoplus_{x \in X} x^u = 0$ for all u such that $u \preceq v$. In other words, if the sum equals one, then there exists u in $\mathcal{U}(X)$ such that $u \preceq v$, i.e., v satisfies $v \succeq u$ for at least one $u \in \mathcal{U}(X)$. □

It is worth noticing that there is no general improvement of the previous result which holds without any further assumption on k or on X. Indeed, it is easy to check that, when $X = \{u\}$, we have that any $v \in \mathsf{Succ}(u)$ belongs to $\mathcal{U}(k + X)$ for some k (e.g. $k = (u+v)$ satisfies this property). Thus, $\mathsf{Succ}(u)$ is the smallest set which contains the parity sets of all cosets $(k + X)$ in this case.

5.2 Propagation through an Sbox

We now investigate how a parity set propagates through a permutation, for instance through an Sbox or through a linear permutation.

Proposition 7. *Let S be a permutation of $\mathbf{F}_2^n$. For any $v \in \mathbf{F}_2^n$, we define*

$$V_S(u) = \{v \in \mathbf{F}_2^n : S^v(x) \textit{ contains } x^u\}$$

Then, for any set X of elements of $\mathbf{F}_2^n$,

$$\mathcal{U}(S(X)) \subseteq \bigcup_{u \in \mathcal{U}(X)} V_S(u) \; .$$

Proof. By definition, the vectors v which may be in $\mathcal{U}(S(X))$ are those such that $S(x)^v$ contains a monomial x^u with $u \in \mathcal{U}(X)$. Otherwise, we have that $\bigoplus_{x \in X} S(x)^v = 0$. The result then directly follows. □

We will discuss in more details in Section 7 the properties of an Sbox which make it resistant or not to this attack. It is worth noticing that the previous proposition applies to any permutation S, including the case where S corresponds to the linear layer of the cipher.

Another case of interest is the case where the Sbox can be seen as the concatenation of several independent Sboxes, like in a typical Sbox layer.

Proposition 8. *Let X be a set of elements in $\mathbf{F}_2^{mt}$ and let S be an Sbox over $\mathbf{F}_2^{mt}$ which consists of the parallel application of t Sboxes $S_1, \ldots, S_t$ over $\mathbf{F}_2^m$: $S(x_1, \ldots, x_t) = (S_1(x_1), \ldots, S_t(x_t))$. Then,*

$$\mathcal{U}(S(X)) \subseteq \bigcup_{(u_1, \ldots, u_t) \in \mathcal{U}(X)} V_{S_1}(u_1) \times \ldots \times V_{S_t}(u_t)$$

where $V_{S_i}(u) = \{v \in \mathbf{F}_2^m : S_i^v(x) \textit{ contains } x^u\}$.

Proof. From Prop. 7, we know that

$$\mathcal{U}(S(X)) \subseteq \bigcup_{(u_1, \ldots, u_t) \in \mathcal{U}(X)} V_S(u) \; .$$

We then have to determine all $v = (v_1, \ldots, v_t) \in (\mathbf{F}_2^m)^t$ such that $S^v(x)$ contains $u = (u_1, \ldots, u_t)$. We use that

$$S^v(x) = S_1^{v_1}(x_1) S_2^{v_2}(x_2) \ldots S_t^{v_t}(x_t) \; .$$

Since only $S_i^{v_i}(x_i)$ may contain $x_i^{u_i}$, we deduce that $v \in V_S(u)$ if and only if $v_i \in V_{S_i}(u_i)$ for each $1 \leq i \leq n$. Therefore, $V_S(u)$ is the Cartesian product of all $V_{S_i}(u_i)$. □

5.3 Propagation through one round

We now consider an SPN where the round key is inserted by addition at the end of the round. This implies that each Sbox layer comes after a round-key addition. Thus, if $\mathcal{U}(X)$ denotes the parity set of the input set X before the key addition, then the parity set after the key addition is included in a union of sets of the form $\mathsf{Succ}(u)$, for some $u \in \mathbf{F}_2^n$. It follows that the parity set after the Sbox layer satisfies

$$\mathcal{U}(S(X+k)) \subseteq \bigcup_{u \in \mathcal{U}(X)} \left(\bigcup_{v \in \mathsf{Succ}(u)} V_S(v) \right) .$$

Therefore, propagating the information from $\mathcal{U}(X)$ to $\mathcal{U}(S(X+k))$ involves the sets

$$\mathcal{V}_S(u) = \bigcup_{v \in \mathsf{Succ}(u)} V_S(v)$$

which depend on the Sbox only.

These sets $\mathcal{V}_S(u)$ are then the relevant quantities involved in the propagation through the Sbox, instead of the sets $V_S(u), u \in \mathbf{F}_2^n$. For instance, Table 2 provides all sets $\mathcal{V}_S(u)$ for the Present Sbox.

	$\mathcal{V}_S(u)$															
	0	1	2	4	8	3	5	9	6	a	c	7	b	d	e	f
0	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
1		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
2		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
4		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
8		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
3			x	x	x	x	x	x	x	x	x	x	x	x		x
5			x	x	x	x	x	x	x	x	x	x	x	x	x	x
9			x	x	x	x	x	x	x	x	x	x		x	x	x
6		x	x		x	x	x	x	x	x	x	x	x	x	x	x
a			x	x	x	x	x	x	x	x	x	x	x	x	x	x
c			x	x	x	x	x	x		x	x	x	x	x	x	x
7			x		x	x		x	x				x	x		x
b			x	x	x	x			x	x	x	x		x		x
d			x	x	x			x		x		x			x	x
e							x					x	x	x	x	x
f																x

Table 2. Sets $\mathcal{V}_S(u)$ for all $u \in \mathbf{F}_2^4$ for the Present Sbox. All 4-bit words are represented in hexadecimal notation, and the rightmost bit of the word corresponds to the least significant bit.

The table representing all $\mathcal{V}_S(u)$ has a few generic properties which hold for any bijective Sbox. The first obvious remark is that the all-zero vector does not belong to any $\mathcal{V}_S(u)$ except when $u = \underline{0}$. Indeed S^0 is the all-one function and then does not contain any monomial except x^0. The following property is much more interesting.

Proposition 9. *Let S be any permutation of $\mathbf{F}_2^n$. Then,*

$$\mathcal{V}_S(\underline{1}) = \{\underline{1}\} ,$$

where $\underline{1}$ denotes the all-one vector in $\mathbf{F}_2^n$.

Proof. Since $\mathsf{Succ}(\underline{1}) = \underline{1}$, we have

$$\mathcal{V}_S(\underline{1}) = V_S(\underline{1}) = \{v \in \mathbf{F}_2^n : S^v(x) \text{ contains } x^{\underline{1}}\}$$

or equivalently $\mathcal{V}_S(\underline{1})$ is the set of all v such that $x \mapsto S^v(x)$ has degree n. It is known [6, Prop. 1] that $\deg(S^v) = n$ if and only if $v = \underline{1}$. □

Some further properties of Table 2, specific to the PRESENT Sbox will be studied in Section 7.

6 Low-data distinguishers on a few rounds of PRESENT

6.1 Distinguisher on 3 rounds

In [25] Todo presents generic distinguishers for ciphers based on the SPN construction. In particular, it is shown in Table 4 of this same paper that 3 rounds of an SPN construction whose nonlinear layer is composed of 16 Sboxes over $\mathbf{F}_2^4$ of degree 3 can be distinguished from a random permutation with data complexity 2^{12}. These results are therefore valid for PRESENT. This distinguisher improves upon the distinguishers exploiting the algebraic degree since the best upper bound on the degree of three rounds of such ciphers is $3^3 = 27$, leading to a distinguisher with data complexity 2^{28}.

Todo's distinguisher can be easily explained in the following way (see also [14]). Suppose that the input space X is composed of vectors taking all possible 2^4 values on the first three nibbles and where the last 13 nibbles are fixed to a same constant value for all vectors. In this case $|X| = 2^{12}$ and X can be seen now as a coset of V, i.e. $X = a + V$, where

$$V = \mathsf{Prec}(\mathtt{0x0000000000000fff}) .$$

After the application of the round-key addition and the Sbox layer to X, the output space Y satisfies the same integral property as X, i.e. $Y = b + V$. Denote now by $F = \mathcal{R}^2 \circ P$, where $\mathcal{R}$ stands for the round function and where P denotes the linear layer. One can easily see that as F contains two non-linear layers, of degree 3 each, $\deg(F) \leq 9$. Therefore, as $\dim Y = 12 > 9$, we have that

$$\bigoplus_{x \in a+V} E_K(X) = \bigoplus_{y \in b+V} F(Y) = \bigoplus_{y \in V} F(b \oplus y) = D_V F(b) = 0.$$

Equivalently, this generic distinguisher on 3 rounds uses the feature that none of the coordinates of E_K contains a multiple of x^u for $u = \texttt{0x0000000000000fff}$ (see Prop. 5). This distinguisher can therefore be very easily explained in terms of parity sets. Since $X = a + V$, where $V = \mathsf{Prec}(\texttt{0x0000000000000fff})$, we have that $\mathcal{U}(V) = \{\texttt{0x0000000000000fff}\}$ (from Corollary 1), implying that

$$\mathcal{U}(X) \subseteq \mathsf{Succ}(\texttt{0x0000000000000fff}).$$

Since the Sbox S is a permutation, for each Sbox $\mathcal{V}_S(\texttt{0xf}) = \{\texttt{0xf}\}$ (Prop. 9) meaning that after the first Sbox layer the parity set U of the resulting set is again included in $\mathsf{Succ}(\texttt{0x0000000000000fff})$. By defining the function F as before, we have that

$$\mathcal{U}(E_K(X)) \subseteq \bigcup_{u \in U} V_F(u).$$

But, $V_F(u) = \{v : F^v(x) \text{ contains } x^u\}$ contains no vector v with $wt(v) \leq 1$ when $wt(u) \geq 12$ since $\deg(F) \leq 9$. Therefore,

$$\mathcal{U}(E_K(X)) \subseteq \{v : wt(v) \geq 2\},$$

meaning that the output of the cipher restricted to 3 rounds has the balanced property (i.e., satisfies the division property of order 2).

6.2 Distinguisher on 4 rounds

As explained in [25] in the case of AES-like ciphers, such generic distinguishers can be improved by exploiting the structure of the linear layer. In the particular case of Present, the linear layer (see Figure 1) is a bit permutation. Moreover, because of its structure, two rounds of Present (without the last permutation layer) can be seen as the concatenation of four independent Superboxes operating on $\mathbf{F}_2^{16}$. With this structure, it is clear that any coordinate of the output at round $(r + 1)$ of the cipher only contains monomials involving inputs from the same Superbox at round r.

Therefore, by exploiting the linear layer of Present one can extend the previous distinguisher to one more round as follows. Suppose now that the input set X has the form $X = a + V$, with

$$V = \mathsf{Prec}(\texttt{0x000000000000fff0}).$$

In this case $\mathcal{U}(X) \subseteq \mathsf{Succ}(\texttt{0x000000000000fff0})$. The parity set remains unchanged after the application of the first nonlinear layer. By applying now the permutation layer we see that the parity set of the resulting set is included in

$$\mathsf{Succ}(\texttt{0x000e000e000e000e}),$$

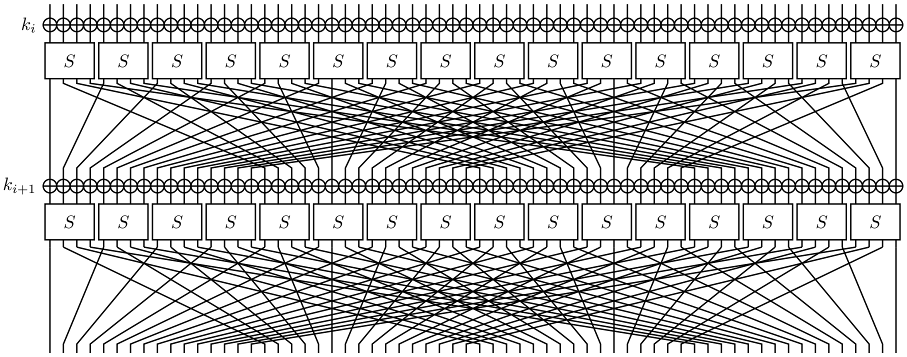

Fig. 1. Round function of the block cipher Present

leading to four active Superboxes. As previously explained, after the application of the Superboxes, so after the non-linear layer of the third round, any output coordinate only contains monomials coming from the same Superbox. Therefore, the resulting parity set is included in

$$\{u : wt(u) \geq 4\}.$$

By applying now the linear layer we have that the parity set is included in

$$\{u \text{ with } \geq 2 \text{ active nibbles}\} \cup \{\texttt{0x00}\ldots\texttt{0f}, \ldots, \texttt{0xf0}\ldots\texttt{0}\}.$$

This parity set is invariant under the application of the fourth nonlinear layer, meaning that

$$\mathcal{U}(E_K(X)) \subseteq \{v : wt(v) \geq 2\}.$$

We see therefore that the output set has the balanced property after 4 rounds.

However, it can be shown, that by only exploiting the properties of the linear layer of Present, this distinguisher cannot always be extended to five rounds. The following table shows a possible propagation of values in the parity sets, where some output coordinate of the 12th Sbox after 5 rounds may contain the monomial x^u, with $u =$ `0x000000000000fff0`. By looking at rows 2 and 3 of this table, we can see that this propagation can be realised, among others, if the Sbox makes the propagations `0xe` → `0x2` and `0xe` → `0x1` possible. All the elements of this table should be interpreted as hexadecimal values.

input	0 0 0 0	0 0 0 0	0 0 0 0	f f f 0
output S-layer 1st round	0 0 0 0	0 0 0 0	0 0 0 0	f f f 0
output P-layer 1st round	0 0 0 e	0 0 0 e	0 0 0 e	0 0 0 e
output S-layer 2nd round	0 0 0 2	0 0 0 1	0 0 0 1	0 0 0 1
output P-layer 2nd round	0 0 0 0	0 0 0 0	1 0 0 0	0 1 1 1
output S-layer 3rd round	0 0 0 0	0 0 0 0	1 0 0 0	0 1 1 1
output P-layer 3rd round	0 0 0 0	0 0 0 0	0 0 0 0	0 0 8 7
output S-layer 4th round	0 0 0 0	0 0 0 0	0 0 0 0	0 0 2 8
output P-layer 4th round	0 0 0 3	0 0 0 0	0 0 0 0	0 0 0 0
output S-layer 5th round	0 0 0 1	0 0 0 0	0 0 0 0	0 0 0 0

However, by looking closer at Table 1, one can see that these two propagations are not possible for the PRESENT Sbox, as 0xe can only be propagated to values of Hamming weight of at least 2.

6.3 Distinguishers on 5 and 6 rounds

The previous 4-round distinguisher on PRESENT exploited the algebraic degree of the Sbox together with the structure of the linear layer. We show here, that by further taking into account the particular form of the PRESENT Sbox, the previous distinguisher can be extended to one more round. For doing so, we consider the same input set X as for the 4-round distinguisher, i.e. $X = a + V$, with $V = \mathsf{Prec}(\texttt{0x000000000000fff0})$. As previously shown (see also the second row of the table above), after the application of the first round transformations, the parity set is included in $\mathsf{Succ}(\texttt{0x000e000e000e000e})$. However, the value 0xe presents, in the case of PRESENT, a particular interest. Indeed, notice that the row indexed by 0xe in Table 2 contains only a single element of weight 2 and no element of weight 1. This means that compared to other values u, $\mathcal{V}_S(u)$ for $u = \texttt{0xe}$ contains exceptionally few elements, making more than half of the transitions impossible. In particular, the transitions, 0xe → 0x2 and 0xe → 0x1 in the above 5-round path are not possible.

We checked by computer programming that in this setting, there is no u of weight 1 in $\mathcal{U}(E_k(X))$ if E_k is 5-round PRESENT. Therefore, the output set after 5 rounds has the balanced property and can then be distinguished from a random permutation.

However, it is not possible to extend the distinguisher in its actual form to 6 rounds. Indeed, we checked that after 6 rounds, many elements of weight 1 can be found in $\mathcal{U}(E_k(X))$. Nevertheless, it is still possible to exhibit a weaker distinguisher for 6 rounds of PRESENT, by exploiting the fact that the column corresponding to the element 0x1 in Table 2 is very sparse, meaning that more than half of the transitions u → 0x1 are not possible. We were able to check that in fact, among all the elements of weight 1 present in $\mathcal{U}(E_k(X))$, only the nibble values 0x2, 0x4 and 0x8 were possible. Therefore, we have exhibited 16 values which do not belong to $\mathcal{U}(E_k(X))$ after six rounds of PRESENT, leading to a distinguisher with data complexity 2^{12} and advantage $(1 - 2^{-16})$ (see Appendix for the evaluation of the advantage).

6.4 Distinguishers using more data

We provide in Table 3 a summary of all the distinguishers obtained for reduced-round versions of Present by using different sizes of the input space. These distinguishers were obtained by implementing the propagation of the parity set of the input subspace in a compact way. We mention here only strong distinguishers in the sense that we give only the results where the output set has the balanced property.

Input set	$\log_2(\#\text{texts})$	Rounds
`0x000000000000000f`	4	4
`0x000000000000fff0`	12	5
`0x00000000ffffffff`	32	6
`0xfffffffffffff000`	52	7
`0xfffffffffffffffe`	63	8

Table 3. Input sets leading to the division property of order 2 for reduced-round Present.

6.5 Changing the Sbox

We discuss in this section the strength of the above distinguishers in the case where the Sbox of Present is replaced by some other permutation of degree 3. For this, we consider exactly the same cipher but we change the nonlinear permutation. For instance, we consider the Sbox used in the block cipher Prince [3] and we study the propagation of the parity set of the resulting cipher. Table 4 provides the sets $\mathcal{V}_S(u)$ for all $u \in \mathbf{F}_2^4$ for this Sbox. However, it has to be noted that we do not propose in any case to replace the Sbox of Present for obtaining a more robust design in general. This change is done here only for being able to run the same experiments by keeping exactly the same other parameters. Applying for example the same attack to the Prince cipher and compare the results would not make any sense as the results could be different because of the different linear layers for example, making the impact of the Sbox unclear. This paragraph aims only at demonstrating that Present with some other Sbox can better resist against this type of distinguishers, but we do not argue that this would necessarily be the case for other type of attacks.

The following table is the equivalent of Table 2 for Present. As one can see, this table, is much less sparse than Table 2. In particular, even the rows the more sparse (rows corresponding to `0xb` and `0xd`), make at least half of the transitions possible. One can further notice that all rows and columns contain elements of weight 1 and 2. All the above indicate that whatever the input u of a particular Sbox, the set $\mathcal{V}_S(u)$ contains a high number of values, making only very few transitions impossible.

	$\mathcal{V}_S(u)$															
	0	1	2	4	8	3	5	9	6	a	c	7	b	d	e	f
0	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
1		x	x	x	x	x	x	x	x	x	x	x		x	x	x
2		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
4		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
8		x	x	x	x	x	x	x	x	x	x		x	x	x	x
3		x	x	x	x	x	x	x	x	x	x	x		x	x	x
5		x	x		x	x	x	x	x	x	x	x		x	x	x
9		x		x	x	x	x		x		x			x	x	x
6		x	x	x	x	x	x	x	x	x		x	x	x	x	x
a			x	x	x		x	x	x	x	x		x	x	x	x
c		x	x	x	x	x	x	x		x	x		x	x	x	x
7		x	x		x	x	x	x	x	x		x		x	x	x
b				x	x				x		x			x	x	x
d					x	x	x				x			x		x
e			x	x			x	x		x			x	x	x	x
f																x

Table 4. Sets $\mathcal{V}_S(u)$ for all $u \in \mathbf{F}_2^4$ for the PRINCE Sbox. All 4-bit words are represented in hexadecimal notation.

We were able to verify, that if we start with the same input space as before, the output set after five rounds does not satisfy the division property of order 2, while this was the case for the PRESENT Sbox. Indeed, Table 5 provides a path that is satisfied when this Sbox is used. As one can see, we end up with a vector of weight 1 in $\mathcal{U}(E_k)$. This was not the case with the original PRESENT Sbox, proving that the PRINCE Sbox is more resistant against this kind of property. A more detailed study of why this is so is provided in the following section.

input	0 0 0 0	0 0 0 0	0 0 0 0	f f f 0
output S-layer 1st round	0 0 0 0	0 0 0 0	0 0 0 0	f f f 0
output P-layer 1st round	0 0 0 e	0 0 0 e	0 0 0 e	0 0 0 e
output S-layer 2nd round	0 0 0 4	0 0 0 2	0 0 0 2	0 0 0 2
output P-layer 2nd round	0 0 0 0	1 0 0 0	0 1 1 1	0 0 0 0
output S-layer 3rd round	0 0 0 0	1 0 0 0	0 1 1 1	0 0 0 0
output P-layer 3rd round	0 0 0 0	0 0 0 0	0 0 0 0	0 8 7 0
output S-layer 4th round	0 0 0 0	0 0 0 0	0 0 0 0	0 1 1 0
output P-layer 4th round	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 6
output S-layer 5th round	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 1

Table 5. Example of a 5-round path that is satisfied when the PRINCE Sbox is plugged into the PRESENT block cipher. All entries should be interpreted as hexadecimal values.

7 Related security criterion for the Sbox

As illustrated by the previous attack, one of the main ingredients of the distinguisher is the particular form of the sets $\mathcal{V}_S(u)$ for the PRESENT Sbox, i.e., the particular form of Table 2. Two properties of this table are exploited:

1. Column of index `0x1` is very sparse and in particular it does not contain any element of weight 3 and it contains a single element of weight 2. This property is exploited at the end of the attack on 6 rounds, where we use that `0x1` only belongs to a very few sets $\mathcal{V}_S(u)$, i.e., only a few transitions $u \rightarrow$ `0x1` are possible.
2. Row of index `0xe` is very sparse and in particular it does not contain any element of weight 1 and it contains a single element of weight 2. This property is exploited when we use that $\mathcal{V}_S($`0xe`$)$ contains a few elements only, implying that many transitions `0xe` $\rightarrow v$ are impossible.

We now show where these unsuitable properties of the Sbox come from, and how they can be avoided. The first property has an obvious algebraic interpretation since each column in the table is derived from the ANF of a Boolean function $x \mapsto S^v(x)$. In particular, the columns having an index of weight 1 are derived from the ANF of the coordinates of the Sbox. The ANF of the PRESENT Sbox is

$$\begin{aligned}
S_1(x_1,x_2,x_3,x_4) &= x_1 + x_3 + x_4 + x_2x_3\\
S_2(x_1,x_2,x_3,x_4) &= x_2 + x_4 + x_2x_4 + x_3x_4 + x_1x_2x_3 + x_1x_2x_4 + x_1x_3x_4\\
S_3(x_1,x_2,x_3,x_4) &= 1 + x_3 + x_4 + x_1x_2 + x_1x_4 + x_2x_4 + x_1x_2x_4 + x_1x_3x_4\\
S_4(x_1,x_2,x_3,x_4) &= 1 + x_1 + x_2 + x_4 + x_2x_3 + x_1x_2x_3 + x_1x_2x_4 + x_1x_3x_4 \ .
\end{aligned}$$

The first weakness of the PRESENT Sbox then comes from the fact that its first coordinate has degree 2 only and contains a single quadratic term.

The second weakness is related to the fact that the monomial $x^{\mathsf{e}} = x_2x_3x_4$ appears neither in the coordinates of S, nor in most functions S_iS_j. This second property can be deduced in a much simpler way from the ANF of the inverse Sbox, as shown by the following proposition.

Lemma 3. *Let S be a permutation of $\mathbf{F}_2^n$. Then, for any $u, v \in \mathbf{F}_2^n$, the ANF of $x \mapsto S(x)^v$ contains x^u if and only if the ANF of $x \mapsto S^*(x)^{\overline{u}}$ contains $x^{\overline{v}}$, where $\overline{u}$ denotes the vector $u \oplus \underline{1}$ and S^* is the permutation $x \mapsto \overline{S^{-1}(\overline{x})}$.*

Proof. Let a_u denote the coefficient of x^u in the ANF of $x \mapsto S(x)^v$. Then

$$\begin{aligned}
a_u &= \bigoplus_{x \preceq u} S(x)^v\\
&= |\{x \in \mathbf{F}_2^n : x_i = 0, i \in \mathsf{Supp}(\overline{u}) \text{ and } S(x)_j = 1, j \in \mathsf{Supp}(v)\}| \bmod 2\\
&= |\{y \in \mathbf{F}_2^n : S^{-1}(y)_i = 0, i \in \mathsf{Supp}(\overline{u}) \text{ and } y_j = 1, j \in \mathsf{Supp}(v)\}| \bmod 2
\end{aligned}$$

where the last equality comes from the fact that S is a permutation, implying that there is a one-to-one correspondence between x and $y = S(x)$. We now replace y by $z = \overline{y}$ and use that $S^{-1}(y) = \overline{S^*(\overline{y})}$. Then,

$$\begin{aligned} a_u &= |\{z \in \mathbf{F}_2^n : S^{-1}(\overline{z})_i = 0, i \in \mathsf{Supp}(\overline{u}) \text{ and } z_j = 0, j \in \mathsf{Supp}(v)\}| \bmod 2 \\ &= |\{z \in \mathbf{F}_2^n : S^*(z)_i = 1, i \in \mathsf{Supp}(\overline{u}) \text{ and } z_j = 0, j \in \mathsf{Supp}(v)\}| \bmod 2 \\ &= \bigoplus_{z \preceq \overline{v}} (S^*(z))^{\overline{u}} \end{aligned}$$

which means that a_u is the coefficient of $x^{\overline{v}}$ in the ANF of $x \mapsto S^*(x)^{\overline{u}}$. □

The function S^* corresponding to the Present Sbox has the following ANF:

$$\begin{aligned} S_1^*(x_1, x_2, x_3, x_4) &= 1 + x_1 + x_2 + x_3 + x_4 + x_2x_4 \\ S_2^*(x_1, x_2, x_3, x_4) &= x_1 + x_4 + x_1x_3 + x_2x_3 + x_1x_2x_3 + x_1x_2x_4 + x_1x_3x_4 \\ S_3^*(x_1, x_2, x_3, x_4) &= 1 + x_2 + x_4 + x_1x_2 + x_1x_3 + x_1x_4 + x_3x_4 + x_1x_2x_3 \\ &\quad +x_1x_2x_4 + x_1x_3x_4 \\ S_4^*(x_1, x_2, x_3, x_4) &= x_2 + x_3 + x_2x_3 + x_1x_4 + x_3x_4 + x_1x_2x_3 + x_1x_3x_4 \end{aligned}$$

These ANF correspond to the rows with an index of weight 3 in Table 2. In particular, the form of the row of index `0xe` (defining $\mathcal{V}_S(\texttt{0xe})$) comes from the fact that the first coordinate of S^* has degree 2 only, and that its ANF contains a single monomial of degree 2.

Conversely, we can easily guarantee that, when an n-bit bijective Sbox is considered, all sets $\mathcal{V}_S(u)$ with $wt(u) < n$ contain an element of weight 1.

Proposition 10. *Let S be a permutation of $\mathbf{F}_2^n$. Then, all coordinates of S^{-1} have degree $(n-1)$ if and only if all sets $\mathcal{V}_S(u)$, with $u \neq \underline{1}$, contain at least one element of weight 1.*

Proof. Since the sets $\mathcal{V}_S(u)$ include all sets $\mathcal{V}_S(u')$ for $u' \succeq u$, we only have to prove the result for the sets $\mathcal{V}_S(u)$ with $wt(u) = n - 1$. When $wt(u) = n - 1$, we have

$$\mathcal{V}_S(u) = V_S(u) \cup V_S(\underline{1}) = V_S(u) \cup \{\underline{1}\}$$

from Prop. 9. Therefore, all sets $\mathcal{V}_S(u)$ with $wt(u) < n$ contain an element of weight 1 if and only if all sets $V_S(u)$ with $wt(u) = n - 1$ contain an element of weight 1. This equivalently means that all monomials of degree $(n - 1)$ appear in the ANF of the coordinates of S, or from Lemma 3 that all coordinates of S^* have degree $(n - 1)$. But, this last condition equivalently means that all coordinates of S^{-1} have degree $(n-1)$. Indeed, the monomials of highest degree in the ANF of a Boolean function f and in the ANF of $f^* : x \mapsto f(\overline{x}) + 1$ are the same. Then the i-th coordinates of S^* and S^{-1} have the same degree. □

It is worth noticing that the fact that all coordinates of S^{-1} have maximal degree is not equivalent to the fact that the same property holds for S, as shown in the following example.

Example 3. We consider the permutation of $\mathbf{F}_2^4$ corresponding to the inverse of the function G_{10} defined in the classification in [18, Table 6]. This Sbox has optimal cryptographic properties in the sense that it has the smallest possible differential uniformity and highest nonlinearity. Its coordinates are given by

$$\begin{aligned}
S_1(x_1,x_2,x_3,x_4) &= x_1 + x_1x_2 + x_2x_3 + x_3x_4 + x_1x_3x_4 + x_2x_3x_4\\
S_2(x_1,x_2,x_3,x_4) &= x_2 + x_1x_3 + x_2x_4 + x_3x_4 + x_1x_2x_4 + x_1x_3x_4 + x_2x_3x_4\\
S_3(x_1,x_2,x_3,x_4) &= x_3 + x_1x_2 + x_1x_3 + x_1x_4 + x_2x_4 + x_1x_2x_4 + x_1x_3x_4\\
S_4(x_1,x_2,x_3,x_4) &= x_4 + x_1x_2 + x_1x_3 + x_1x_2x_4\ .
\end{aligned}$$

It can then be checked that all coordinates of S have degree 3 while the monomial $x_1x_2x_3$ appears in none of the coordinates, implying that $\mathcal{V}_S(\texttt{0x7})$ contains no element of weight 1. Indeed,

$$\mathcal{V}_S(\texttt{0x7}) = \{\texttt{0x9}, \texttt{0xc}, \texttt{0x7}, \texttt{0xb}, \texttt{0xe}, \texttt{0xf}\}\ .$$

Actually, S does not satisfy the hypotheses of Prop. 10 since the first coordinate of S^{-1} has degree 2 only.

It is also worth noticing that the condition on the degree of the coordinates of S^{-1} is not invariant under composition (to the right or left) by an affine permutation.

A simple way to guarantee that all coordinates of the inverse Sbox have maximal degree consists in choosing for S an Sbox such that any linear combination of its coordinates (i.e. any of its components) has maximal degree. In this case, we obtain the following stronger result on the number of elements of weight 1 in $\mathcal{V}_S(u)$.

Corollary 3. *Let S be a permutation of $\mathbf{F}_2^n$ such that the Boolean functions $x \mapsto \lambda \cdot S(x)$ have degree $(n-1)$ for all nonzero $\lambda \in \mathbf{F}_2^n$. Then, for any $u \in \mathbf{F}_2^n$, $\mathcal{V}_S(u)$ contains at least $(n - wt(u))$ elements of weight* 1.

Proof. We will first prove the result for all u with $wt(u) = n-1$ by showing that all coordinates of S^{-1} have degree $(n-1)$. Let A denote the $n \times n$ binary matrix such that $a_{i,j}$ is the coefficient of the monomial of degree $(n-1)$ $x^{\overline{e_j}}$ in the ANF of the i-th coordinate of S where e_j denotes the n-bit word having a single one at Position j. A component of S, $x \mapsto \lambda \cdot S(x)$, $\lambda \neq 0$, has degree less than $(n-1)$ if and only if the corresponding linear combination of the rows of A vanishes, *i.e.*, $\lambda A = 0$. It follows that the number of non-trivial components of S with degree $(n-1)$ is equal to $2^{\mathrm{rk}(A)} - 1$. From Lemma 3, the coefficients of the monomials of degree $(n-1)$ in the coordinates of S^{-1} are defined by the transpose of A. Then, the number of non-trivial components of S^{-1} with degree $(n-1)$ is equal to $2^{\mathrm{rk}(A^T)} - 1$. We deduce from Prop. 10 that, if all components of S have degree $(n-1)$ (i.e., if $\mathrm{rk}(A) = \mathrm{rk}(A^T) = n$), all $\mathcal{V}_S(u)$ for $u \neq \underline{1}$ contain at least one element of weight 1.

Let us now consider any $u \in \mathbf{F}_2^n$. By definition, $\mathcal{V}_S(u)$ contains all sets $\mathcal{V}_S(\overline{e_i})$ with $i \notin \mathsf{Supp}(u)$ since the $k = (n - wt(u))$ words $\overline{e_i}$ of weight $(n-1)$ belong to $\mathsf{Succ}(u)$. As Matrix A has full rank, the k columns of A corresponding to the monomials $x^{\overline{e_i}}$ with $i \notin \mathsf{Supp}(u)$ have rank k, implying that this $k \times n$-submatrix has at least k nonzero rows. These rows correspond to the words of weight 1 which belong to $\mathcal{V}_S(u)$, implying that this set contains at least k vectors of weight 1. □

Example 4. The 4-bit Sbox used in PRINCE [3] (as well as all Sboxes with similar properties recommended for any cipher within the PRINCE-family [4, Appendix B]) has been chosen in such a way that all its nontrivial components have degree 3. Then, we can guarantee that any set $\mathcal{V}_S(u)$ contains at least $(4 - wt(u))$ elements of weight 1. This can be checked on Table 4.

It is worth noticing that the condition exhibited in Corollary 3 offers a similar guarantee for the inverse Sbox. Indeed, making the decryption function also immune to this type of attacks may be relevant, even if mounting the attack on the decryption function is much more difficult in practice because it requires the knowledge of plaintext-ciphertext pairs corresponding to chosen ciphertexts.

Application to the MISTY *Sboxes.* So far, the most important application of the division property is the cryptanalysis of the full MISTY1 [24]. It is then relevant to study the MISTY Sboxes in the light of the previous criteria. MISTY1 is an unbalanced Feistel network. It then uses two different Sboxes, which are both linearly equivalent to a power permutation. More precisely, S_7 is a permutation of degree 3 of $\mathbf{F}_2^7$ and S_9 is a permutation of degree 2 of $\mathbf{F}_2^9$.

We have then computed all sets $\mathcal{V}_S(u)$ for each of these Sboxes. Most notably, Tables 6 and 7 give the values of the minimal Hamming weight of an element in $\mathcal{V}_S(u)$ depending on the Hamming weight of u for the 7-bit Sbox S_7 and for the 9-bit Sbox S_9 respectively. These tables then recover the results on the propagation of the division property described in [24, Page 420]. The fact that, for both Sboxes, many sets $\mathcal{V}_S(u)$ have a large minimal weight show that the two MISTY Sboxes are weak regarding the division property. An interesting new observation is that, for S_7, for some vectors u of weight 3, for instance $u =$ `0x0b`, $\mathcal{V}_{S_7}(u)$ does not contain any vector of weight 1. This equivalently means that some monomials of degree 3 do not appear in any of the coordinates of S_7. This property of S_7, which is more precise that the propagation of the division property studied in [24], may then be exploited in an attack.

$wt(u)$	1	2	3	4	5	6	7
$\min\{wt(v) : v \in \mathcal{V}_{S_7}(u)\}$	1	1	1 or 2	2	2	4	7

Table 6. Minimum Hamming weight of $\mathcal{V}_{S_7}(u)$ depending on the Hamming weight of u for the 7-bit Sbox in MISTY1.

$wt(u)$	1	2	3	4	5	6	7	8	9
$\min\{wt(v) : v \in \mathcal{V}_{S_9}(u)\}$	1	1	2	2	3	3	4	4	9

Table 7. Minimum Hamming weight of $\mathcal{V}_{S_9}(u)$ depending on the Hamming weight of u for the 9-bit Sbox in Misty1.

8 Conclusions

In some contexts, the notion of parity set provides a powerful tool for representing subsets. We have shown for instance that the division property has a simple formulation in terms of parity sets, which allows to easily deduce some properties of the sets satisfying the division property of a given order. Also, focusing on the parity set, and not only on the minimal weight of its elements, enables the attacker to capture some algebraic properties of the nonlinear functions used in the cipher, besides the algebraic degree. This general view also brings to light the properties of the Sbox which avoid this type of attacks. The counterpart is that computing the whole parity set after many rounds of a cipher is obviously more expensive than considering its minimal weight only, as this is done in the division property. However, a promising technique consists in combining both approaches where the first and last rounds are analysed with the whole parity set, while the propagation through the middle rounds only exploits the degree of the function. Another direction could be to use parity sets for identifying some sets of weak keys. We have focused on distinguishers which hold for all keys. But, the addition of an unknown key increases the size of the parity set, since all words greater than or equal to the words in the input parity set must be considered. A different approach then consists in considering only round keys of a particular form. Then, the resulting parity set after key addition may simplify a lot, and these particular round keys may then be easily detected.

Acknowledgements. We thank the anonymous reviewers for their helpful comments. We also thank Willi Meier and Qingju Wang for insightful discussions.

References

1. Aumasson, J.P., Dinur, I., Meier, W., Shamir, A.: Cube testers and key recovery attacks on reduced-round MD6 and Trivium. In: Dunkelman, O. (ed.) FSE 2009. LNCS, vol. 5665, pp. 1–22. Springer, Berlin, Germany, Leuven, Belgium (Feb 22–25, 2009)
2. Bogdanov, A., Knudsen, L.R., Leander, G., Paar, C., Poschmann, A., Robshaw, M.J.B., Seurin, Y., Vikkelsoe, C.: PRESENT: An ultra-lightweight block cipher. In: Paillier, P., Verbauwhede, I. (eds.) CHES 2007. LNCS, vol. 4727, pp. 450–466. Springer, Berlin, Germany, Vienna, Austria (Sep 10–13, 2007)
3. Borghoff, J., Canteaut, A., Güneysu, T., Kavun, E.B., Knežević, M., Knudsen, L.R., Leander, G., Nikov, V., Paar, C., Rechberger, C., Rombouts, P., Thomsen, S.S., Yalçin, T.: PRINCE - A low-latency block cipher for pervasive computing applications - extended abstract. In: Wang, X., Sako, K. (eds.) ASIACRYPT 2012.

LNCS, vol. 7658, pp. 208–225. Springer, Berlin, Germany, Beijing, China (Dec 2–6, 2012)

4. Borghoff, J., Canteaut, A., Güneysu, T., Kavun, E.B., Knežević, M., Knudsen, L.R., Leander, G., Nikov, V., Paar, C., Rechberger, C., Rombouts, P., Thomsen, S.S., Yalçin, T.: PRINCE - a low-latency block cipher for pervasive computing applications (full version). Cryptology ePrint Archive, Report 2012/529 (2012), `http://eprint.iacr.org/2012/529`
5. Boura, C., Canteaut, A.: On the Influence of the Algebraic Degree of F^{-1} on the Algebraic Degree of $G \circ F$. IEEE Transactions on Information Theory 59(1), 691–702 (Jan 2013), `http://hal.inria.fr/hal-00738398`
6. Boura, C., Canteaut, A., De Cannière, C.: Higher-order differential properties of Keccak and Luffa. In: Joux, A. (ed.) FSE 2011. LNCS, vol. 6733, pp. 252–269. Springer, Berlin, Germany, Lyngby, Denmark (Feb 13–16, 2011)
7. Canteaut, A., Carlet, C., Charpin, P., Fontaine, C.: On cryptographic properties of the cosets of $R(1, m)$. IEEE Transactions on Information Theory 47(4), 1494–1513 (2001)
8. Canteaut, A., Videau, M.: Degree of composition of highly nonlinear functions and applications to higher order differential cryptanalysis. In: Knudsen, L.R. (ed.) EUROCRYPT 2002. LNCS, vol. 2332, pp. 518–533. Springer, Berlin, Germany, Amsterdam, The Netherlands (Apr 28 – May 2, 2002)
9. Daemen, J., Knudsen, L.R., Rijmen, V.: The block cipher Square. In: Biham, E. (ed.) FSE'97. LNCS, vol. 1267, pp. 149–165. Springer, Berlin, Germany, Haifa, Israel (Jan 20–22, 1997)
10. Delsarte, P.: An algebraic approach to the association schemes of coding theory. Ph.D. thesis, Université catholique de Louvain (1973)
11. Dinur, I., Liu, Y., Meier, W., Wang, Q.: Optimized interpolation attacks on LowMC. Cryptology ePrint Archive, Report 2015/418 (2015), `http://eprint.iacr.org/2015/418`
12. Dinur, I., Shamir, A.: Cube attacks on tweakable black box polynomials. In: Joux, A. (ed.) EUROCRYPT 2009. LNCS, vol. 5479, pp. 278–299. Springer, Berlin, Germany, Cologne, Germany (Apr 26–30, 2009)
13. Dinur, I., Shamir, A.: Breaking Grain-128 with dynamic cube attacks. In: Joux, A. (ed.) FSE 2011. LNCS, vol. 6733, pp. 167–187. Springer, Berlin, Germany, Lyngby, Denmark (Feb 13–16, 2011)
14. Khovratovich, D.: Private communication (2016)
15. Knudsen, L.R.: Truncated and higher order differentials. In: Preneel, B. (ed.) FSE'94. LNCS, vol. 1008, pp. 196–211. Springer, Berlin, Germany, Leuven, Belgium (Dec 14–16, 1995)
16. Knudsen, L.R., Wagner, D.: Integral cryptanalysis. In: Daemen, J., Rijmen, V. (eds.) FSE 2002. LNCS, vol. 2365, pp. 112–127. Springer, Berlin, Germany, Leuven, Belgium (Feb 4–6, 2002)
17. Lai, X.: 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. Kluwer Academic Publishers (1994)
18. Leander, G., Poschmann, A.: On the classification of 4 bit s-boxes. In: Carlet, C., Sunar, B. (eds.) Arithmetic of Finite Fields, First International Workshop, WAIFI 2007, Madrid, Spain, June 21-22, 2007, Proceedings. Lecture Notes in Computer Science, vol. 4547, pp. 159–176. Springer (2007), `http://dx.doi.org/10.1007/978-3-540-73074-3_13`
19. MacWilliams, F.J., Sloane, N.J.: The theory of error-correcting codes. North-Holland (1977)

20. Muller, D.E.: Application of Boolean algebra to switching circuit design and to error detection. IEEE Transactions on Computers 3, 6–12 (1954)
21. Reed, I.S.: A class of multiple-error-correcting codes and the decoding scheme. IEEE Transactions on Information Theory 4, 38–49 (1954)
22. Sun, B., Hai, X., Zhang, W., Cheng, L., Yang, Z.: New observation on division property. Cryptology ePrint Archive, Report 2015/459 (2015), `http://eprint.iacr.org/2015/459`
23. Szegö, G.: Orthogonal polynomials. American Mathematical Society Colloquium Publications (1959)
24. Todo, Y.: Integral cryptanalysis on full MISTY1. In: Gennaro, R., Robshaw, M.J.B. (eds.) CRYPTO 2015, Part I. LNCS, vol. 9215, pp. 413–432. Springer, Berlin, Germany, Santa Barbara, CA, USA (Aug 16–20, 2015)
25. Todo, Y.: Structural evaluation by generalized integral property. In: Oswald, E., Fischlin, M. (eds.) EUROCRYPT 2015, Part I. LNCS, vol. 9056, pp. 287–314. Springer, Berlin, Germany, Sofia, Bulgaria (Apr 26–30, 2015)

A Advantages of the distinguishers based on parity sets

In order to estimate the advantages of the distinguishers exhibited in this paper, we need to evaluate the probability that, given an input set X, a randomly chosen permutation π is such that $\pi(X)$ does not satisfy the division property of order 2. For the weaker distinguisher, we similarly need to evaluate the probability that a given u does not belong to $\mathcal{U}(\pi(X))$. Clearly, the probability that $\pi(X)$ satisfies the division property of order 2 (i.e., is balanced) is close to 2^{-n}, while the probability that a given u does not belong $\mathcal{U}(\pi(X))$ is close to 1/2. However, these probabilities may vary with the size of X: for instance, if u is the all-zero word, we have that $u \in \mathcal{U}(\pi(X))$ if and only if $|X|$ is odd. Also, if $|X|$ is odd, $\pi(X)$ cannot satisfy the division property of order 2. A more careful analysis seems therefore needed. The exact values of these probabilities confirming these estimates are then given by the following propositions.

Proposition 11. *Let $u \in \mathbf{F}_2^n$. The probability over all sets $X \subseteq \mathbf{F}_2^n$ of size k that u does not belong to $\mathcal{U}(X)$ is equal to*

$$\frac{1}{2}\left(1 + \frac{P_k(2^{n-wt(u)})}{\binom{2^n}{k}}\right) ,$$

where $P_k(w)$ is the Krawtchouk polynomial

$$P_k(x) = \sum_{i=0}^{k} (-1)^i \binom{x}{i}\binom{2^n - x}{k-i} .$$

In particular, if $wt(u) = 1$, this probability equals

$$\begin{cases} \frac{1}{2} & \text{if } k \text{ is odd} \\ \frac{1}{2}\left(1 + (-1)^{k/2}\frac{\binom{2^{n-1}}{k/2}}{\binom{2^n}{k}}\right) \simeq \frac{1}{2} & \text{if } k \text{ is even.} \end{cases}$$

Proof. From Lemma 2, $u \notin \mathcal{U}(X)$ if and only if the product between the row G_u of index u in Matrix G and the incidence vector of X vanishes. The row G_u is a word of length 2^n and weight $2^{n-wt(u)}$ since $G_{u,v} = 1$ if and only if $u \preceq v$. Then we need to count the number of vectors v_X of weight k such that the scalar product $G_u \cdot v_X = 0$. It is known [10, Theorem 4.1] that, for any vector g of length $N = 2^n$,

$$\sum_{v \in \mathbf{F}_2^N, wt(v)=k} (-1)^{g \cdot v} = P_k(wt(g)) \ .$$

Then,

$$\Pr_{X, |X|=k}[u \notin \mathcal{U}(X)] = \frac{1}{2}\left(1 + \frac{P_k(2^{n-wt(u)})}{\binom{2^n}{k}}\right) \ .$$

In the special case where $wt(u) = 1$, we need to estimate the value of $P_k(2^{n-1})$. The generating function of the Krawtchouk polynomials is [23]

$$(1+z)^{N-i}(1-z)^i = \sum_{\ell=0}^{N} P_\ell(i) z^\ell \ .$$

We deduce that, for $i = N/2$, this generating function is $(1-z^2)^{N/2}$. It contains monomials of even degree only, and then

$$P_{2\ell}(N/2) = (-1)^\ell \binom{N/2}{\ell} \ ,$$

implying that

$$P_k(2^{n-1}) = \begin{cases} 0 & \text{if } k \text{ is odd} \\ (-1)^{k/2}\binom{2^{n-1}}{k/2} & \text{if } k \text{ is even.} \end{cases}$$

The probability that u does not belong to $\mathcal{U}(X)$ is then is very close to $1/2$ in all cases. Indeed, the ratio of the two binomial coefficients satisfies

$$\frac{\binom{2^{n-1}}{k/2}}{\binom{2^n}{k}} = \Theta\left(2^{-2^{n-1} H_2\left(\frac{k}{2^n}\right)}\right)$$

where H_2 is the binary entropy, $H_2(x) = -x \log_2(x) - (1-x)\log_2(1-x)$. □

Similarly, the probability that a set X of given size k satisfies the division property of order d is determined by the number of codewords of weight k in the Reed-Muller code of length 2^n and order $(n-d)$. There is no known formula for this number in general, but it can be computed when $d = 2$, which is the case corresponding to Todo's distinguishers.

Proposition 12. *The probability that a set $X \subseteq \mathbf{F}_2^n$ of size k satisfies the division property of order* 2 *is* 0 *if k is odd and*

$$2^{-n} + (-1)^{k/2}(1-2^{-n})\frac{\binom{2^{n-1}}{k/2}}{\binom{2^n}{k}} \simeq 2^{-n} \ ,$$

if k is even.

Proof. The result comes from the weight distribution of the Reed-Muller code $\mathcal{R}(n-2,n)$. Since this code is the dual of $\mathcal{R}(1,n)$, its weight distribution can be derived from the weight distribution of the dual code by the MacWilliams transform [19, Page 129]: the number of codewords of weight k in $\mathcal{R}(n-2,n)$ is

$$A_k = 2^{-(n+1)} \left(P_k(0) + (2^{n+1}-2)P_k(2^{n-1}) + P_k(2^n) \right) \ ,$$

where the $P_k(i)$ are the previously defined Krawtchouk polynomials. From the generating function, we get that $P_k(0) = \binom{2^n}{k}$ and $P_k(2^n) = (-1)^k \binom{2^n}{k}$, and the value of $P_k(2^{n-1})$ has been computed in the proof of the previous proposition. The result then directly follows. $\square$