Differential Attacks: Using Alternative Operations

Céline Blondeau[1], Roberto Civino[2], and Massimiliano Sala[2]

[1] Aalto University, School of Science, Finland
celine.blondeau@aalto.fi
[2] University of Trento, Italy
roberto.civino@unitn.it, massimiliano.sala@unitn.it

Abstract Is it possible that a block cipher apparently immune to classical differential cryptanalysis can be attacked considering a different operation on the message space? Recently Calderini and Sala showed how to effectively compute alternative operations on a vector space which can serve as message space for a block cipher such that the resulting structure is still a vector space. The latter were used to mount a linearisation attack against a toy cipher. Here we investigate the possibility to design a block cipher which appears to be secure w.r.t. classical differential cryptanalysis, but weaker with respect to our attack which make use of alternative operations. Furthermore we compare the success probabilities of a distinguishing attack.

Keywords: block ciphers, differential cryptanalysis, distinguisher, alternative operations.

1 Introduction

Differential cryptanalysis was introduced in the beginning of the 90's [BS91] as a powerful tool to cryptanalyse some cryptographic primitives, including block ciphers. For these primitives, the difference operation usually taken into consideration by both designers and cryptanalysts is the bitwise addition modulo two, classically called XOR. In this paper we focus on substitution-permutation networks (SPN) similar to the block cipher PRESENT [BKL+07]. These ciphers consist of multiple iterations of confusion layers (non-linear parallel S-boxes), diffusion layers (linear maps) and key additions. To stay in the classical setting we assume that the key is XORed to the state at each round. The aim of this paper is to show that block ciphers may have different levels of resistance against differential attacks, depending on the additive law that is considered on the message space. We propose an example of SPN which is resistant against the classical differential attack, with XOR differences, but it is not resistant against a differential attack which makes use of alternative differences coming from other operations defined on the message space. These operations were described and employed for cryptographic purposes for the first time in [CS17], where they were used to perform a linearisation attack against a toy cipher.

The paper is organised as follows. In Sect. 2 we introduce the notation and describe the general idea of our attack. The way new operations are built is described in Sect. 3. In Sect. 4 we design a 15-bit cipher and perform experiments to study its resistance to differential cryptanalysis. We show that the considered cipher is immune from the classical differential attack and is weak w.r.t. ours. A description of the operation used to mount the attack is also provided. Finally Sect. 5 concludes the paper giving an overview of the obtained results.

2 Notation and preliminaries

Let V be a finite vector space over $\mathbb{F}_2$ which represents the message space and let $\mathcal{K}$ be a key space. Let $\dim(V) = n = mb$ and let us write $V = V_1 \oplus V_2 \oplus \ldots \oplus V_b$ where $\dim(V_j) = m$ for $1 \leq j \leq b$, and $\oplus$ represents the direct sum of subspaces, called bricks. The canonical basis for V is denoted with $\{e_1, e_2, \ldots, e_n\}$. If G is any finite group acting on V, for each $g \in G$ and $v \in V$ we denote the action of g on v as vg, i.e. we use postfix notation for every function evaluation. The identity matrix of size l is denoted by $\mathbb{1}_l$, and the zero matrix of size $l \times h$ is denoted by $\mathbb{0}_{l,h}$. We denote by $\mathrm{Sym}(V)$ the symmetric group acting on V, i.e. the group of all the permutations on the message space.

An r-round *substitution-permutation network*, with round function depicted in Fig. 1, is a family of encryption functions $\{\varphi_k \mid k \in \mathcal{K}\} \subset \mathrm{Sym}(V)$ such that for each $k \in \mathcal{K}$ the map φ_k is the composition of r round functions, i.e. $\varphi_k = \varphi_{1,k}\, \varphi_{2,k} \ldots \varphi_{r,k}$, where $\varphi_{i,k} = \gamma\lambda\sigma_{k_i}$ and

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

$$(x_1, x_2, \ldots, x_n)\gamma = \big((x_1, \ldots, x_m)\gamma', \ldots, (x_{m(b-1)+1}, \ldots, x_n)\gamma'\big) \quad .$$

  The map $\gamma' \in \mathrm{Sym}(V_j)$ is traditionally called an S-box;
- $\lambda \in \mathrm{Sym}(V)$ is a linear map;
- $\sigma_{k_i} : V \to V, x \mapsto x + k_i$ represents the key addition, where $+$ is the usual bitwise XOR on $\mathbb{F}_2$. The round keys $k_i \in V$ are usually derived from the master key k by means of a public algorithm.

For each $1 \leq s \leq r$ we denote by $\varphi_k^{(s)}$ the composition of the first s round functions. In particular $\varphi_k = \varphi_k^{(r)}$.

2.1 Classical differential cryptanalysis

Classical differential cryptanalysis [BS91] and its generalizations [Knu94,BBS99] exploit the fact that some *differences* propagate with unusually high or low probability during the encryption process, leading to a non-uniform distribution of the output differences which can show a non-random behavior of the cipher. Although the term difference may refer to every group operation on the message space, i.e. $\mathrm{diff}(x, y) \stackrel{\text{def}}{=} x * y^{-1}$ where y^{-1} denotes the inverse of y w.r.t. $*$, this

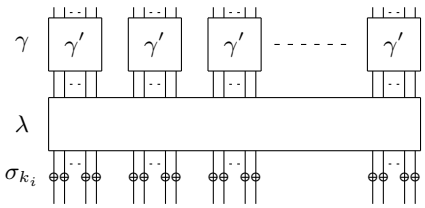

Figure 1. Example of 1-round encryption

operation usually coincides with the inverse of the operation used for the key addition, hence in most of the cases $\mathrm{diff}(x, y) = x + y$.

For a permutation f, a *differential* over f is a pair (δ_I, δ_O), whose corresponding *differential probability* is

$$p_{(\delta_I,\delta_O)} \stackrel{\text{def}}{=} \mathbb{P}_x \left[xf + (x + \delta_I)f = \delta_O\right] \ ,$$

assuming that x is uniformly distributed. It represents the probability that, given two vectors whose difference is δ_I, the difference after applying f is δ_O. Notice that when f is a linear function, for each δ_I only the differential $(\delta_I, \delta_I f)$ is possible, whereas if f is a translation, only the differential (δ_I, δ_I) is possible. When f is an S-box γ', the *difference distribution table* of γ' (DDT) is defined as the integer table where $\mathrm{DDT}[\,\delta_I, \delta_O] \stackrel{\text{def}}{=} p_{(\delta_I,\delta_O)} 2^m$ for each $(\delta_I, \delta_O) \in (V_j)^2$.

Given $1 \leq s \leq r$, we denote by $(\Delta_I, \Delta_O) \in V^2$ an *s-round differential* over $\left\{\varphi_k^{(s)} \mid k \in \mathcal{K}\right\}$, whose corresponding expected probability is

$$p_{(\Delta_I,\Delta_O)} \stackrel{\text{def}}{=} \mathbb{P}_{x,k} \left[x\varphi_k^{(s)} + (x + \Delta_I)\varphi_k^{(s)} = \Delta_O\right] \ ,$$

where x and k are uniformly distributed respectively on V and $\mathcal{K}$. The probability of a given s-round differential (Δ_I, Δ_O) is obtained as the sum of probabilities of its *differential trails*, where a differential trail is an $(s+1)$-tuple $(\beta_0, \beta_1, \ldots, \beta_s)$ of intermediate differences at each round such that $\beta_0 = \Delta_I$ and $\beta_s = \Delta_O$. For each of these s-round trails, only the confusion layer requires a probabilistic analysis, since the diffusion layer is linear and the key-addition layer is a translation. An r-round cipher is considered secure w.r.t. classical differential cryptanalysis if the expected probability of each s-round (s close to r) differential $(\Delta_I, \Delta_O) \in V^2$ is close enough to 2^{-n} that we are not able to distinguish the set of parametrised permutations from a random one.

2.2 Differential cryptanalysis revised

Our goal is to introduce another group operation $\circ$ on the message space and to show that an SPN secure in the classical sense can be distinguished from a random permutation using the new operation considered. Letting $\circ$-*differential,* $\circ$-*differential probability,* $\circ$-*s-round differential*, and $\circ$-*differential trail* be defined similarly to those in Subsect. 2.1 where every occurrence of "+" is replaced

by "$\circ$", we will show that there may exist $\circ$-differentials whose corresponding probabilities are high enough to allow for a distinguishing attack. The remainder of the section contains what is necessary to define some new operations on V and to understand how they have to interact with each component of the cipher in order to make the attack possible.

General idea Let us denote by T_+ the group of translations on V, i.e. $T_+ \stackrel{\text{def}}{=} \{\sigma_a \mid a \in V\}$ and let us stress again that the translation σ_k acts on a vector x in the same way the key addition layer acts on the message x, i.e. $x\sigma_k = x + k$. For this reason SPN's may be also called *translation-based ciphers* [CDVS06]. In order to represent the key addition by means of an action of the translation group on the message space, let us point out that T_+ is 2-elementary abelian regular, and for each $a, b \in V$ it holds $a+b = a\sigma_b$. Our goal is to define alternative operations on the vector space V by means of other 2-elementary abelian regular groups which can play the role of translation groups. Given any 2-elementary abelian regular subgroup $T < \mathrm{Sym}(V)$, we can represent $T = \{\tau_a \mid a \in V\}$, where for a given $a \in V$, τ_a is the unique element in T which maps 0 into a. Then, if we define for each $a, b \in V$ $a \circ b \stackrel{\text{def}}{=} a\tau_b$, it is easy to see that $(V, \circ)$ is an additive group and $\circ$ induces a vector space structure on V, whose corresponding group of translations is $T_\circ = T$.

Design principles Once we have defined new operations $\circ$ on the message space, sufficient conditions for $\circ$-difference propagation during the encryption process have to be investigated. Once again, we stress that we are interested in finding $\circ$-s-round differentials (Δ_I, Δ_O) whose corresponding probabilities

$$p_{(\Delta_I, \Delta_O)} = \mathbb{P}_{x,k}\left[x\varphi_k^{(s)} \circ (x \circ \Delta_I)\varphi_k^{(s)} = \Delta_O\right]$$

are high enough. First of all, although the operation $\circ$ might be *a priori* defined on the whole message space V, studying differential properties of a confusion layer seen as a function with 2^n inputs may be impractical for standard-size ciphers. For this reason, in this paper we choose to focus on operations which are applied in parallel to the different bricks, i.e. $\circ = (\circ^{(1)}, \circ^{(2)}, \ldots, \circ^{(b)})$, where for each $1 \leq j \leq b$, $\circ^{(j)}$ is an operation on V_j. This allow us to independently study each S-box. Furthermore, to limit the impact on the $\circ$-differential probability of a $\circ$-differential trail, we analyse only operations such that the diffusion layer is linear w.r.t. to both $+$ and $\circ$. Indeed, if this is the case, the diffusion layer requires no probabilistic analysis. However, as the chosen operation is different from the XOR, used to add the key at the different rounds of the cipher, differential probabilities have to be introduced when studying the interaction between $\circ$-differences and the key-addition layer. We show how to define a class of operations such that $\circ$-differences resulting from the key-addition layer do not depend on the state considered. In particular we prove that $(x+k) \circ ((x \circ \Delta) + k)$ equals Δ for a subset of keys and does not depend on x for all keys. In Sect. 3 and Sect. 4 we explain how these requirements can be met. For a complete description of the topic the reader is advised to refer also to [CS17,BCS17].

3 New operations and properties

Let $\circ$ be an operation and let us suppose now that $T_\circ < \mathrm{AGL}_+$, where AGL_+ denotes the group of all affine functions on the vector space $(V, +)$. As it has been shown in [CS17], this hypothesis is crucial in the light of being able to give a *practical* way to construct new operations. In this case, for each $a \in V$ there exists $M_a \in \mathrm{GL}_+$ such that $\tau_a = M_a \sigma_a$, where GL_+ is the group of linear functions on $(V, +)$, i.e. for each $x \in V$ it holds $x \circ a = x\tau_a = xM_a + a$. In the following, for any given 2-elementary abelian regular subgroup $T_\circ < \mathrm{AGL}_+$, we denote by $\circ$ the induced operation. Let us define what we call in this context the *weak keys subspace* as

$$W_\circ \stackrel{\text{def}}{=} \{a \mid a \in V, \sigma_a = \tau_a\} = \{k \mid k \in V, \forall x \in V \;\; x \circ k = x + k\} \;\; .$$

It is easy to show that $W_\circ$ is a vector subspace of both $(V, +)$ and $(V, \circ)$. In addition it is worth noting that each $k \in W_\circ$ represents a key that can be added regardless of the operation involved. In this sense, we address those vectors as *weak keys*. The weak key subspace $W_\circ$ is not empty for the following result.

Theorem 1 ([CDVS06,CS17,BCS17]). *Let $T_\circ$ be a 2-elementary abelian regular subgroup of AGL_+. Then $W_\circ$ is a non-trivial subspace of V. Moreover*

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

3.1 On key-addition layer

The standard differential attack exploits the property that each $+$-difference is kept the same after the round key is XORed. It would be desirable to maintain this property for each round key $k \in V$ also in the case of $\circ$-differences. Unfortunately this is never the case. Indeed, for each pair of messages x and $x \circ \Delta$ having $\circ$-difference fixed to Δ, after the addition with the round key k we get

$$(x + k) \circ ((x \circ \Delta) + k) \tag{1}$$

which in general is different from Δ, except for the case when k is a weak key, i.e. $k \in W_\circ$. Indeed, if $k \in W_\circ$ we can replace each "$+k$" with "$\circ k$" and obtain

$$(x + k) \circ ((x \circ \Delta) + k) = (x \circ k) \circ ((x \circ \Delta) \circ k) = \Delta \;\; .$$

This discloses the important role of $W_\circ$: whenever k is a weak key, the key-addition layer σ_k behaves as a translation layer w.r.t. $\circ$-differences. To determine the value of (1) in the general case, a particular subclass of operations is needed. The latter is defined in the light of the following result.

Theorem 2 ([CS17]). *Let $T_\circ$ be a 2-elementary abelian regular subgroup of AGL_+ such that $T_+ < \mathrm{AGL}_\circ$ and $W_\circ = \mathrm{Span}\{e_{n-d+1}, \ldots, e_n\}$, where $\mathrm{AGL}_\circ$*

is the group of affine functions on $(V, \circ)$*. Then for each* $a \in V$ *there exists a matrix* $E_a \in (\mathbb{F}_2)^{(n-d)\times d}$ *such that*

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

Moreover, the canonical vectors $\{e_i\}_{i=1}^n$ *form also a basis for* $(V, \circ)$*.*

Notice that the hypothesis $T_+ < \mathrm{AGL}_\circ$ made the key addition a *key-dependent* affine transformation w.r.t. $\circ$. In the reminder of this paper, *an operation* $\circ$ is always such that $T_+ < \mathrm{AGL}_\circ$, $T_\circ < \mathrm{AGL}_+$, and $W_\circ = \mathrm{Span}\{e_{n-d+1}, \ldots, e_n\}$, where $d \stackrel{\text{def}}{=} \dim(W_\circ)$.

As shown in Theorem 2, fixing such an operation is equivalent to defining the matrices E_{e_i} for each $1 \le i \le n$. Let us denote by $\boldsymbol{b_{i,j}}$ the last d components of the j-th row of M_{e_i} in such a way that we can represent

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

As $W_\circ = \mathrm{Span}\{e_{n-d+1}, \ldots, e_n\}$, we have $E_{e_i} = 0$ for each $n-d+1 \le i \le n$, hence only $n-d$ matrices have to be stored. Moreover, since $T_\circ$ is 2-elementary, for each $1 \le i \le n-d$ it holds $e_i \circ e_i = 0$, which means $\boldsymbol{b_{i,i}} = \mathbf{0}$. In addition since for each $1 \le i, j \le n-d$ we have $e_i \circ e_j = e_j \circ e_i$, then $\boldsymbol{b_{i,j}} = \boldsymbol{b_{j,i}}$. It is also easy to show that for each $a, b, c \in V$ it holds $(a+b) \circ c = a \circ c + b \circ c + c$, hence writing $a = \sum_{i=1}^n \xi_i e_i$ we have

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

where weight(a) denote the Hamming weight of a. From this it follows that $a \circ b$ can be computed in polynomial time.

Example 1. Let $n = 3$. We denote by $\diamond$ the operation on $(\mathbb{F}_2)^3$ defined by the following matrices. This operation is different from the XOR.

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

In Fig. 2, $+$ and the operation $\diamond$ defined above are compared. Each vector is interpreted as a binary number, most significant bit first, and then represented using the hexadecimal notation. As $W_\diamond = \{0, e_3\} = \{\mathtt{0_x}, \mathtt{1_x}\}$, the first two rows and columns in the tables are equal. Different entries are emphasised.

Every operation $\circ$ induces a *product* $\cdot$ defined as follows

$$\forall a, b \in V \quad a \cdot b \stackrel{\text{def}}{=} a + b + a \circ b, \tag{2}$$

Figure 2. Comparison between operation $+$ and $\diamond$

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

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

which helps to simplify computations. The following result has been proven in a more general setting in [CDVS06].

Theorem 3. *Let $\circ$ an operation and let $\cdot$ the product induced. Then $\cdot$ is distributive w.r.t. $+$, i.e. $(V, +, \cdot)$ is an $\mathbb{F}_2$-algebra.*

It is obvious by definition that $x \cdot y$ represents the error committed when confusing $x \circ y$ with $x + y$. Let us define

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

i.e. the set of all the possible products in V, which we will be referring to also as *errors*. Let us fix $x, y \in V$ and see in detail what the algebra product represents. First of all notice that, if $x \in W_\circ$ then $x \cdot y = 0$. In fact in this case $x \cdot y = x + y + x \circ y = x + y + x + y = 0$. In the general case we have

$$\begin{aligned} x \cdot y = x + y + x \circ y = x + y + xE_y + y = x + x \begin{pmatrix} \mathbb{1}_{n-d} & E_y \\ \mathbb{0}_{d,n-d} & \mathbb{1}_d \end{pmatrix} \\ = (\underbrace{0, \ldots, 0}_{n-d}, (x_1, \ldots, x_{n-d})E_y) \in U_\circ \ . \end{aligned}$$

Hence the following is proven:

Theorem 4. *For each $x, y \in V$ there exists $\varepsilon_{x,y} \in U_\circ$ such that*

$$x + y = x \circ y + \varepsilon_{x,y},$$

with $\varepsilon_{x,y} = x \cdot y = (0, \ldots, 0, (x_1, \ldots, x_{n-d})E_y)$. Moreover $U_\circ \subseteq W_\circ$.

Remark 1. From Theorem 4 it follows that $U_\circ$ is composed of all the possible vectors $w \in W_\circ$ whose last d-components are all the possible $\mathbb{F}_2$-linear combinations of the rows of the matrices E_x for each $x \in V$.

Remark 2. From the fact $x \cdot y \in U_\circ \subseteq W_\circ$ always follows $x \cdot y \cdot z = 0$.

We are now ready to determine the value of (1). In particular we can show that the output $\circ$-difference after the key-addition layer does not depend on the state x.

Theorem 5. *Let $\circ$ be an operation. Then for each $x, k, \Delta \in V$*

$$(x + k) \circ ((x \circ \Delta) + k) = \Delta + k \cdot \Delta \in \Delta + U_\circ \ .$$

Proof. Let $x, k, \Delta \in V$. Applying (2) and Theorem 3 we obtain

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

where $x \cdot x = x + x + x \circ x = 0$ and all the triple products vanish because of Remark 2. □

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

Example 2. Let us consider the operation $\diamond$ defined in Example 1 and store all the values $k + k \diamond \Delta$ in a *difference distribution table for the key-addition layer* K, where $K[\Delta_I, \Delta_O] \stackrel{\text{def}}{=} \#\left\{k \mid k \in (\mathbb{F}_2)^3,\ k + k \diamond \Delta_I = \Delta_O\right\}$ (see Fig. 3). For example, considering a $\diamond$-difference $\Delta_I = 2_\text{x}$, the $\diamond$-difference after key-addition layer may be either $\Delta_O = 2_\text{x}$ or $\Delta_O = 3_\text{x}$, each with probability $1/2$. Notice that, from Remark 1, $U_\diamond = W_\diamond = \{0_\text{x}, 1_\text{x}\}$, hence e_3 is in fact the only non-null possible error.

Figure 3. $\diamond$-difference distribution table for the key-addition layer

$K =$

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

3.2 On confusion layer

While in classical differential cryptanalysis differential probabilities are only induced by confusion layers, in the previous section we illustrate that, with new operations, probabilities are also added by the key-addition layer. For the probability of a $\circ$-differential to be larger than the probability of a $+$-differential, we should either have trails with larger probabilities and/or more trails. The first goal can only be achieved if we have larger values in the DDT of the S-box w.r.t. $\circ$. In the following we show through an example that this is possible.

Example 3. Let us consider the following APN [Nyb91] S-box $\gamma' : (\mathbb{F}_2)^3 \to (\mathbb{F}_2)^3$ which is affinely equivalent to the power function $x \mapsto x^3$:

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

.

Figure 4. Difference distribution table of γ' w.r.t. $+$ and $\diamond$

$\mathrm{DDT}_+ =$

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

$\mathrm{DDT}_\diamond =$

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

Tables DDT_+ and $\mathrm{DDT}_\diamond$ in Fig. 4 represent the difference distribution tables of γ' with respect to $+$ and $\diamond$ respectively. Notice that the S-box is no more APN w.r.t. $\diamond$. Although the size of entries of $\mathrm{DDT}_\diamond$ is to be rescaled due to the effect of key-addition layer, the magnitude difference between the entries of DDT_+ and $\mathrm{DDT}_\diamond$ is large enough to allow for a distinguishing attack.

4 Designing a cipher

As an illustration we propose in this section an operation on $V = \oplus_{i=1}^{5}(\mathbb{F}_2)^3$ and a 15-bit SPN. In this section, we consider the operation $\hat{\circ}$ defined by the following matrices

$$M_{e_1} = \left(\begin{array}{c|c} \mathbb{1}_2 & \begin{matrix}0\,0\ldots0\\1\,0\ldots0\end{matrix} \\ \hline \mathbb{0}_{13,2} & \mathbb{1}_{13}\end{array}\right), \; M_{e_2} = \left(\begin{array}{c|c} \mathbb{1}_2 & \begin{matrix}1\,0\ldots0\\0\,0\ldots0\end{matrix} \\ \hline \mathbb{0}_{13,2} & \mathbb{1}_{13}\end{array}\right)$$

and for $3 \leq i \leq 15$, $M_{e_i} = \mathbb{1}_{15}$. The operation $\hat{\circ}$ just now defined is such that $\dim(W_{\hat{\circ}}) = 13$ and it is not hard to notice that such operation acts as the operation $\diamond$ defined in Example 1 on the first brick, and as $+$ on the remaining ones, i.e. $\hat{\circ} = (\diamond, +, +, +, +)$. For example

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

This allows us to attack the first S-box of the cipher, whose differential properties w.r.t. $\hat{\circ}$ are weaker, as already seen in Example 3. Let us notice that from the definition of $\hat{\circ}$ it holds $U_{\hat{\circ}} = \{0, e_3\} \subset W_{\hat{\circ}} = \mathrm{Span}\{e_3, e_4, \ldots, e_{15}\}$.

4.1 On diffusion layer

Let us now explain how to obtain diffusion layers which are linear w.r.t. $+$ and $\circ$. Let us define $H_\circ \stackrel{\text{def}}{=} \mathrm{Hom}(V, +, \cdot)$ the group of homomorphisms of the algebra $(V, +, \cdot)$. It is an easy check that every map $\lambda \in H_\circ$ is linear w.r.t. $+$ and $\circ$, i.e. $H_\circ = \mathrm{GL}_+ \cap \mathrm{GL}_\circ$, where $\mathrm{GL}_\circ$ is the group of linear maps w.r.t. $\circ$. Moreover, the following holds:

Theorem 6 ([CS17])**.** *Let $\circ$ be an operation, $W_\circ$ and $U_\circ$ the weak keys subspace and the set of errors respectively. Then $W_\circ H_\circ = W_\circ$ and $U_\circ H_\circ = U_\circ$.*

Next we present a new result, which provides us with a description of $H_{\hat{\circ}}$. Although this description is valid only for the operation $\hat{\circ}$ defined at the beginning of this section, the result can be generalised and a more general version will be given in the full version of this paper.

Theorem 7. *Let $\lambda \in (\mathbb{F}_2)^{15\times 15}$. The following are equivalent:*

1. *$\lambda \in H_{\hat{\circ}}$;*
2. *there exist $A \in (\mathbb{F}_2)^{2\times 2}$ and $B' \in (\mathbb{F}_2)^{12\times 12}$ invertible matrices, $D \in (\mathbb{F}_2)^{2\times 13}$ and $D' \in (\mathbb{F}_2)^{12\times 1}$ such that*

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

Proof. Firstly let us assume $\lambda \in H_{\hat{\circ}}$ and let us decompose λ in the block form

$$\lambda = \left(\begin{array}{c|c} A & D \\ \hline C & B \end{array}\right),$$

where $A \in (\mathbb{F}_2)^{2\times 2}$, $B \in (\mathbb{F}_2)^{13\times 13}$, $C \in (\mathbb{F}_2)^{13\times 2}$ and $D \in (\mathbb{F}_2)^{2\times 13}$. From Theorem 6, $W_{\hat{\circ}} H_{\hat{\circ}} = W_{\hat{\circ}}$ implies $C = \mathbb{0}_{13,2}$ and consequently A and B are invertible. Moreover, since $U_{\hat{\circ}} = \{0, e_3\}$, from $U_{\hat{\circ}} H_{\hat{\circ}} = U_{\hat{\circ}}$ one has $e_3\lambda = e_3$, which means that

$$B = \left(\begin{array}{cc} 1 & \mathbb{0}_{1,12} \\ D' & B' \end{array}\right),$$

where $B' \in (\mathbb{F}_2)^{12\times 12}$ is invertible and $D' \in (\mathbb{F}_2)^{12\times 1}$. Conversely, let us assume 2 and prove that given $x, y \in V$ it holds $(x \cdot y)\lambda = x\lambda \cdot y\lambda$. If $x \in W_{\hat{\circ}}$, then also $x\lambda \in W_{\hat{\circ}}$, hence there is nothing to prove. For the same reason $(x \cdot y)\lambda = x\lambda \cdot y\lambda$ if and only if $((x_1, x_2, 0, \ldots, 0) \cdot (y_1, y_2, 0, \ldots, 0))\,\lambda = (x_1, x_2, 0, \ldots, 0)\lambda \cdot (x_1, x_2, 0, \ldots, 0)\lambda$, thus it is sufficient to consider the case $x = e_1$ and $y = e_2$. It is easy to check that both the products $e_1 \cdot e_2$ and $e_1\lambda \cdot e_2\lambda$ equal e_3, hence from $e_3\lambda = e_3$ the desired holds. □

Example 4. From Theorem 7, The following 15×15 binary matrix λ is linear w.r.t. $\hat{\circ}$:

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

4.2 Experiments

Let us consider the r-round SPN defined by $\varphi_{i,k} = \gamma\lambda\sigma_{k_i}$, where γ acts on every brick as the S-box γ' defined in Example 3 and λ is the matrix defined in Example 4. For this cipher, we have performed experiments to study its resistance to differential cryptanalysis[3]. While it has not been specified yet, we now assume that for each $k \in \mathcal{K}$, the round keys k_i's are selected uniformly at random in V. In our computation, we used 2^{11} random keys to have a good estimate of the expected differential probability of the best differential on 5 rounds of this cipher. The experimental computations show that the best 5-round differential $(\Delta_{I_+}, \Delta_{O_+}) = (\mathtt{0007}_\mathrm{x}, \mathtt{1301}_\mathrm{x})$ has probability $2^{-14.567}$ where the difference taken into consideration is the classical $+$-difference. Using the $\hat{\circ}$-difference defined in the beginning of Sect. 4, the best 5-round differential is $(\Delta_{I_{\hat{\circ}}}, \Delta_{O_{\hat{\circ}}}) = (\mathtt{3000}_\mathrm{x}, \mathtt{019D}_\mathrm{x})$ with probability $2^{-14.296}$. We can compute the maximal success probability of a distinguishing attack as the probability that at least one pair $(x, x + \Delta_{I_+})$ or $(x, x\hat{\circ}\Delta_{I_{\hat{\circ}}})$ follows the differential, assuming that, when using the full codebook, differentials are binomially distributed over the keys [DR07]. With the probabilities previously given, we find that in more than 50% of the cases the differential is not fulfilled for the $+$-difference and we can conclude that a basic distinguishing attack will not succeed. In the same setting using the $\hat{\circ}$-differences, the differential appears at least once for about 56% of the keys. Consequently this represents an example of small cipher which looks like a classically secure SPN, and for which considering an operation different from the one used for the key-addition produces a successful distinguishing attack.

[3] Note that only the resistance to differential cryptanalysis is considered and we do not claim any resistance criteria for the security of this small cipher.

5 Conclusion

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

References

BBS99. E. Biham, A. Biryukov, and A. Shamir. Cryptanalysis of Skipjack reduced to 31 rounds using impossible differentials. In *International Conference on the Theory and Applications of Cryptographic Techniques*, pages 12–23. Springer, 1999.

BCS17. C. Brunetta, M. Calderini, and M. Sala. Algorithms and bounds for hidden sums in cryptographic trapdoors. *arXiv:1702.08384*, 2017.

BKL+07. A. Bogdanov, L. R. Knudsen, G. Leander, C. Paar, A. Poschmann, M. JB Robshaw, Y. Seurin, and C. Vikkelsoe. PRESENT: An ultra-lightweight block cipher. In *CHES '07*, pages 450–466. Springer, 2007.

BS91. E. Biham and A. Shamir. Differential cryptanalysis of DES-like cryptosystems. *Journal of Cryptology*, 4(1):3–72, 1991.

CDVS06. A. Caranti, F. Dalla Volta, and M. Sala. Abelian regular subgroups of the affine group and radical rings. *Publ. Math. Debrecen*, 69(3):297–308, 2006.

CS17. M. Calderini and M. Sala. Elementary abelian regular subgroups as hidden sums for cryptographic trapdoors. *arXiv:1702.00581*, 2017.

DR07. J. Daemen and V. Rijmen. Probability distributions of correlation and differentials in block ciphers. *Journal of Mathematical Cryptology JMC*, 1(3):221–242, 2007.

Knu94. L. R. Knudsen. Truncated and higher order differentials. In *International Workshop on Fast Software Encryption*, pages 196–211. Springer, 1994.

Nyb91. K. Nyberg. Perfect nonlinear S-boxes. In D. W. Davies, editor, *EUROCRYPT '91*, volume 547 of *LNCS*, pages 378–386. Springer, 1991.