KU LEUVEN

Optimal S-boxes against alternative operations

(with M. Calderini and R. Civino)

Riccardo Invernizzi

WCC 2024 - Perugia

June 17th

Block ciphers

Ingredients

- $n > 0$ such that performing 2^n operations is unfeasible
- $V = \mathbb{F}_2^n$ the message space

Definition

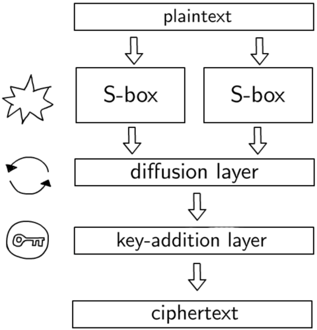
A block cipher is a set of encryption functions indexed by parameters called keys

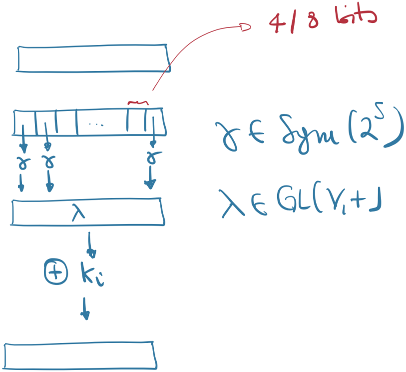
$$\mathcal{C} = \{E_k \mid k \in V\} \subseteq \mathrm{Sym}(V).$$

- $E_k(m)$ is the encryption of a message m with the key k
- there exists an efficient algorithm to compute E_k

Substitution-permutation networks (SPN)

- Structure of AES, PRESENT, ...





KU LEUVEN

Differential Cryptanalysis

- Introduced by Biham and Shamir (1991)
- Analyze how input differences effect output differences:

$$\mathbb{P}[E_k(x) + E_k(x + \Delta_x) = \Delta_y]$$

- in SPN: diffusion and key addition do not alter the difference distribution
  - $\lambda(x) + \lambda(x + \Delta_x) = \lambda(\Delta_x)$, with prob. 1
  - $(x + k) + (x + k + \Delta_x) = \Delta_x$, with prob. 1
- we can reduce the analysis to S-boxes

Differential Cryptanalysis

Definition (Differential uniformity)

The differential uniformity of a function γ is

$$\delta(\gamma) := \max_{a,b\neq 0} |\{x \mid \gamma(x) + \gamma(x+a) = b\}|$$

In order to contrast differential cryptanalysis we need:

- γ with low differential uniformity, in order to reduce the probabilities of certain differences
- λ with "good" diffusion properties, in order to involve as many S-boxes as possible in the analysis

Alternative Operations

We maximize non-linearity w.r.t "classic" + induced by

$$T_+ = \{\sigma_k \mid \sigma_k : x \mapsto x + k\} < \mathrm{Sym}(V)$$

Consider another (elementary abelian regular) group

$$T_\circ = \{\tau_k \mid \tau_k(0) = k\} < \mathrm{Sym}(V)$$

Then

- $a \circ b := \tau_b(a)$
- $(V, \circ) \cong (V, +)$ is a $\mathbb{F}_2$-vector space
- Condition 1: $T_\circ < \mathrm{AGL}(V, +)$ (computational)
- Condition 2: $T_+ < \mathrm{AGL}(V, \circ)$ (cryptanalytic)

Alternative Operations

Important properties:

- Conditions 1 and 2 characterized by [CCS21]
- the weak key space is defined as

$$W_\circ = \{w \in V \mid \sigma_w = \tau_w\}$$

- define $a \cdot b := a + b + a \circ b$; the error space is

$$U_\circ = V \cdot V = \langle a \cdot b \mid a, b \in V \rangle \subset W_\circ$$

- $1 \leq \dim W_\circ \leq n - 2$ ([CDVS06, CCS21])

Alternative cryptanalysis

Question: if $\mathcal{C}$ is a secure block ciphers w.r.t. (classical) differential cryptanalysis, what about $\circ$ operations?

Advantages:

- S-boxes γ are chosen with low (minimal) differential uniformity w.r.t. the classical sum $+$
- higher $\circ$-differential uniformity gives us better trails

Disadvantages:

- mixing layer and key addition may not be affine maps w.r.t $\circ$
- they may impact on the trails

Alternative cryptanalysis - Key addition

- Classically: $(x + k) + (x + k + \Delta) = \Delta$
- in our setting, using condition 2:

$$(x + k) \circ ((x \circ \Delta) + k) = \Delta + \underbrace{\Delta \cdot k}_{\in U_\circ}$$

- if $\dim(W_\circ) = n - 2$, then $\dim(U_\circ) = 1$!
- then

$$(x + k) \circ ((x \circ \Delta) + k) = \begin{cases} \Delta & \text{with pr. } 1/2 \\ \Delta + u & \text{with pr. } 1/2 \end{cases}$$

KU LEUVEN

Alternative cryptanalysis - Mixing layer

- Classically: $x\lambda + (x + \Delta)\lambda = \Delta\lambda$ by linearity
- in our setting:

$$x\lambda \circ (x \circ \Delta)\lambda = \Delta\lambda + (x \cdot \Delta)\lambda + x\lambda \cdot \Delta\lambda + x\lambda \cdot (x \cdot \Delta)\lambda$$

- in general depends on x
- define $H_\circ := \mathrm{GL}(V, +) \cap GL(V, \circ)$
- require $\lambda \in H_\circ$ (compatible maps)

Structure of the mixing layer

- Let $\dim(W_\circ) = n - 2$
- can assume $W_\circ = \langle e_3, ..., e_n \rangle$ and $U_\circ = \{0, (0, 0, \mathbf{b})\}$ with $\mathbf{b} \in \mathbb{F}_2^{n-2} \setminus \{0\}$ ([CCS21])

Theorem (CBS19)

$\lambda \in \mathrm{GL}(V, +) \cap \mathrm{GL}(V, \circ)$ *if and only if*

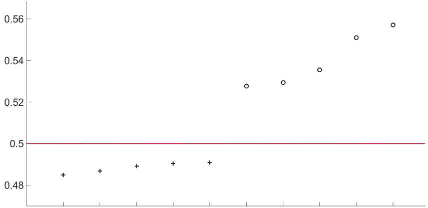
$$\lambda = \begin{pmatrix} A & B \\ 0_{n-2,2} & D \end{pmatrix}$$

for some $A \in GL((\mathbb{F}_2)^2, +)$, $B \in (\mathbb{F}_2)^{2\times n-2}$, *and* $D \in GL((\mathbb{F}_2)^{n-2}, +)$, *with* $\boldsymbol{b}D = \boldsymbol{b}$

A first attack

[CBS19] gave the first example of cipher which is:

- resistant to classical diff. cryptanalysis (APN S-box)
- weak w.r.t. differential attack
- parameters of the cipher: $n = 15$, $s = 3$
- $\circ$ s.t. $\dim(W_\circ) = n - 2$ acts on the first block
- possible to mount a distinguishing attack on 5 rounds



KU LEUVEN

Parallel alternative operation

- Problem: [CBS19] targets only the first S-box
- this requires a "slow" diffusion by λ

Idea: introduce a parallel alternative operation $\circ = (\circ_1, ..., \circ_r)$

- can target each S-box separately
- if $\dim(W_{\circ_i}) = s - 2$, we can assume $\circ_1 = ... = \circ_r$ up to conj. by an element $g \in \mathrm{GL}(V, +)$

First step: determine the structure of $H_\circ$

Structure of $H_\circ$

- Staring point: characterization of [CBS19] for the case $\dim(W_\circ) = n - 2$
- all $\circ_i$ have $\dim(W_{\circ_i}) = n - 2$ and $U_{\circ_i} = \{0, (0, 0, \mathbf{b})\}$
- Consider $\lambda \in GL(V, +)$ and write it as

$$\lambda = \left(\begin{array}{cc|c|cc} A_{11} & B_{11} & \multirow{2}{*}{$\cdots$} & A_{1r} & B_{1r} \\ C_{11} & D_{11} & & C_{1r} & D_{1r} \\ \hline \multicolumn{2}{c|}{\vdots} & \ddots & \multicolumn{2}{c}{\vdots} \\ \hline A_{r1} & B_{r1} & \multirow{2}{*}{$\cdots$} & A_{rr} & B_{rr} \\ C_{r1} & D_{r1} & & C_{rr} & D_{rr} \end{array}\right)$$

Structure of $H_\circ$

Theorem (Calderini, Civino, I.)

$\lambda \in \mathrm{GL}(V,+) \cap \mathrm{GL}(V,\circ)$ *if and only if*

1. $C_{ij} = 0_{(s-2)\times 2}$ *and* $B_{ij} \in (\mathbb{F}_2)^{2\times(s-2)}$;
2. $A_{ij} \in (\mathbb{F}_2)^{2\times 2}$ *such that for each row and each column of blocks there is one and only one non-zero* $A_{ij} \in \mathrm{GL}(\mathbb{F}_2, 2)$;
3. $D_{ij} \in (\mathbb{F}_2)^{(s-2)\times(s-2)}$ *such that if* A_{ij} *is zero* $\boldsymbol{b}D_{ij} = 0$, *and if* A_{ij} *is invertible* $\boldsymbol{b}D_{ij} = \boldsymbol{b}$. *Moreover, the matrix* D *defined by*

$$D = \begin{pmatrix} D_{11} & \cdots & D_{1r} \\ \vdots & \ddots & \vdots \\ D_{r1} & \cdots & D_{rr} \end{pmatrix}$$

must be invertible.

Optimal S-boxes

Second step: study the $\circ$-differential uniformity of optimal functions

- we consider 4-bit S-boxes
- in [LP07] all 4-bit permutations up to affine equivalence (multiplication by maps in $\mathrm{AGL}(V, +)$) are classified
- affine equivalence preserves (among others) differential uniformity
- 302 classes of which 16 are "optimal"
- among the properties of optimal functions we have 4-differential uniformity (best possible for 4-bit permutations)

Optimal S-boxes

	0_x	1_x	2_x	3_x	4_x	5_x	6_x	7_x	8_x	9_x	A_x	B_x	C_x	D_x	E_x	F_x
G_0	0	1	2	D_x	4_x	7_x	F_x	6_x	8_x	B_x	C_x	9_x	3_x	E_x	A_x	5
G_1	0	1	2	D_x	4_x	7_x	F_x	6_x	8_x	B_x	E_x	3_x	5_x	9_x	A_x	12
G_2	0	1	2	D_x	4_x	7_x	F_x	6_x	8_x	B_x	E_x	3_x	A_x	C_x	5_x	9
G_3	0	1	2	D_x	4_x	7_x	F_x	6_x	8_x	C_x	5_x	3_x	A_x	E_x	B_x	9
G_4	0	1	2	D_x	4_x	7_x	F_x	6_x	8_x	C_x	9_x	B_x	A_x	E_x	5_x	3_x
G_5	0	1	2	D_x	4_x	7_x	F_x	6_x	8_x	C_x	B_x	9_x	A_x	E_x	3_x	5
G_6	0	1	2	D_x	4_x	7_x	F_x	6_x	8_x	C_x	B_x	9_x	A_x	E_x	5_x	3_x
G_7	0	1	2	D_x	4_x	7_x	F_x	6_x	8_x	C_x	E_x	B_x	A_x	9_x	3_x	5
G_8	0	1	2	D_x	4_x	7_x	F_x	6_x	8_x	E_x	9_x	5_x	A_x	B_x	3_x	12
G_9	0	1	2	D_x	4_x	7_x	F_x	6_x	8_x	E_x	B_x	3_x	5_x	9_x	A_x	12
G_{10}	0	1	2	D_x	4_x	7_x	F_x	6_x	8_x	E_x	B_x	5_x	A_x	9_x	3_x	12
G_{11}	0	1	2	D_x	4_x	7_x	F_x	6_x	8_x	E_x	B_x	A_x	5_x	9_x	C_x	3_x
G_{12}	0	1	2	D_x	4_x	7_x	F_x	6_x	8_x	E_x	B_x	A_x	9_x	3_x	C_x	5
G_{13}	0	1	2	D_x	4_x	7_x	F_x	6_x	8_x	E_x	C_x	9_x	5_x	B_x	A_x	3_x
G_{14}	0	1	2	D_x	4_x	7_x	F_x	6_x	8_x	E_x	C_x	B_x	3_x	9_x	5_x	10
G_{15}	0	1	2	D_x	4_x	7_x	F_x	6_x	8_x	E_x	C_x	B_x	9_x	3_x	A_x	5_x

$\circ$-differential uniformity of optimal S-boxes

- $\circ$-differential uniformity is not preserved by affine equivalence
- can have different uniformity inside the same class
- $\#$ functions in a single aff. class $\sim 2^{36}$
- $\#$ of alternative sums $\circ = 105$

Proposition

For any $g_1, g_2 \in H_\circ$, $\delta_\circ(f) = \circ(g_1 \cdot f \cdot g_2)$.
For any $\sigma_c \in T_+$, $\delta_\circ(f) = \delta_\circ(\sigma_c \cdot f) = \delta_\circ(f \cdot \sigma_c)$ (under cond. 2).

Consequence: we can restrict to inspect the elements $g_1 G_i g_2$, for $g_1, g_2 \in GL(V,+) \backslash H_\circ$, for each possible sum $\circ$.

Avg. # functions with given ∘-differential uniformity

	2	4	6	8	10	12	14	16
G_0	0	914	7842	3463	420	19	0	14
G_1	0	1019	10352	4226	560	0	0	18
G_2	0	1003	8604	3805	462	21	0	16
G_3	0	1103	7769	1824	177	0	0	0
G_4	0	1101	9295	2715	179	0	0	0
G_5	0	2479	24135	5402	639	0	0	0
G_6	0	1632	10842	3071	218	0	0	0
G_7	0	1257	10679	2994	186	28	0	0
G_8	0	1691	12821	6113	583	93	0	24
G_9	0	1228	7734	2693	154	39	0	0
G_{10}	0	1228	8063	2763	166	41	0	0
G_{11}	0	1637	9940	2941	214	0	0	0
G_{12}	0	2541	16832	5308	352	0	0	0
G_{13}	0	1124	9520	2416	217	15	0	0
G_{14}	0	1207	7641	2584	160	51	0	0
G_{15}	0	1227	7776	2630	163	52	0	0

Experimental results

We tested our attack on some toy ciphers:

- $V = \mathbb{F}_2^{16}$, with 4 S-boxes of 4 bits each
- fix $\circ$ to be the parallel sum defined by $\mathbf{b} = (0, 1)$
- fix the S-box γ to be optimal w.r.t. $+$
- random keys (no key-schedule)

Different choices for the mixing layer:

- first experiment: fixed mixing layer with good diffusion properties
- second experiment: random mixing layers sampled from $H_\circ$

The sum $\circ$

$\circ$	0_x	1_x	2_x	3_x	4_x	5_x	6_x	7_x	8_x	9_x	A_x	B_x	C_x	D_x	E_x	F_x
0_x	0_x	1_x	2_x	3_x	4_x	5_x	6_x	7_x	8_x	9_x	A_x	B_x	C_x	D_x	E_x	F_x
1_x	1_x	0_x	3_x	2_x	5_x	4_x	7_x	6_x	9_x	8_x	B_x	A_x	D_x	C_x	F_x	E_x
2_x	2_x	3_x	0_x	1_x	6_x	7_x	4_x	5_x	A_x	B_x	8_x	9_x	E_x	F_x	C_x	D_x
3_x	3_x	2_x	1_x	0_x	7_x	6_x	5_x	4_x	B_x	A_x	9_x	8_x	F_x	E_x	D_x	C_x
4_x	4_x	5_x	6_x	7_x	0_x	1_x	2_x	3_x	D_x	C_x	F_x	E_x	9_x	8_x	B_x	A_x
5_x	5_x	4_x	7_x	6_x	1_x	0_x	3_x	2_x	C_x	D_x	E_x	F_x	8_x	9_x	A_x	B_x
6_x	6_x	7_x	4_x	5_x	2_x	3_x	0_x	1_x	F_x	E_x	D_x	C_x	B_x	A_x	9_x	8_x
7_x	7_x	6_x	5_x	4_x	3_x	2_x	1_x	0_x	E_x	F_x	C_x	D_x	A_x	B_x	8_x	9_x
8_x	8_x	9_x	A_x	B_x	D_x	C_x	F_x	E_x	0_x	1_x	2_x	3_x	5_x	4_x	7_x	6_x
9_x	9_x	8_x	B_x	A_x	C_x	D_x	E_x	F_x	1_x	0_x	3_x	2_x	4_x	5_x	6_x	7_x
A_x	A_x	B_x	8_x	9_x	F_x	E_x	D_x	C_x	2_x	3_x	0_x	1_x	7_x	6_x	5_x	4_x
B_x	B_x	A_x	9_x	8_x	E_x	F_x	C_x	D_x	3_x	2_x	1_x	0_x	6_x	7_x	4_x	5_x
C_x	C_x	D_x	E_x	F_x	9_x	8_x	B_x	A_x	5_x	4_x	7_x	6_x	0_x	1_x	2_x	3_x
D_x	D_x	C_x	F_x	E_x	8_x	9_x	A_x	B_x	4_x	5_x	6_x	7_x	1_x	0_x	3_x	2_x
E_x	E_x	F_x	C_x	D_x	B_x	A_x	9_x	8_x	7_x	6_x	5_x	4_x	2_x	3_x	0_x	1_x
F_x	F_x	E_x	D_x	C_x	A_x	B_x	8_x	9_x	6_x	7_x	4_x	5_x	3_x	2_x	1_x	0_x

The S-box γ

- γ is an optimal permutation affine equivalent to G_0 (the class of SERPENT's S1)
- $\delta_+(\gamma) = 4$ (optimal), but $\delta_\circ(\gamma) = 16$

x	0_x	1_x	2_x	3_x	4_x	5_x	6_x	7_x	8_x	9_x	A_x	B_x	C_x	D_x	E_x	F_x
$x\gamma$	0_x	E_x	B_x	1_x	7_x	C_x	9_x	6_x	D_x	3_x	4_x	F_x	2_x	8_x	A_x	5_x

The S-box γ

$+$	0_x	1_x	2_x	3_x	4_x	5_x	6_x	7_x	8_x	9_x	A_x	B_x	C_x	D_x	E_x	F_x
0_x	16	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
1_x	·	·	·	·	·	·	·	·	·	·	4	4	·	·	4	4
2_x	·	·	·	·	·	·	·	·	2	2	2	2	2	2	2	2
3_x	·	4	4	·	·	4	·	4	·	·	·	·	·	·	·	·
4_x	·	·	4	·	·	·	·	4	·	·	2	2	·	·	2	2
5_x	·	4	·	·	·	4	·	·	2	2	·	·	2	2	·	·
6_x	·	·	·	·	·	·	4	4	2	2	·	·	2	2	·	·
7_x	·	·	·	·	·	·	4	4	2	2	·	·	2	2	·	·
8_x	·	·	·	4	2	2	·	·	·	·	·	·	·	4	2	2
9_x	·	·	·	4	2	2	·	·	·	·	·	·	4	·	2	2
A_x	·	2	2	·	2	·	2	·	·	2	·	2	·	2	2	·
B_x	·	2	2	·	2	·	2	·	2	·	2	·	2	·	·	2
C_x	·	2	2	·	2	·	2	·	·	2	2	·	·	2	·	2
D_x	·	2	2	·	2	·	2	·	2	·	·	2	2	·	2	·
E_x	·	·	·	4	2	2	·	·	·	4	2	2	·	·	·	·
F_x	·	·	·	4	2	2	·	·	4	·	2	2	·	·	·	·

Figure: DDT of γ w.r.t. $+$

$\circ$	0_x	1_x	2_x	3_x	4_x	5_x	6_x	7_x	8_x	9_x	A_x	B_x	C_x	D_x	E_x	F_x
0_x	16	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
1_x	·	·	·	·	·	·	·	·	·	·	8	·	·	·	8	·
2_x	·	·	·	·	·	·	·	·	4	·	·	4	4	·	·	4
3_x	·	4	4	·	4	·	·	4	·	·	·	·	·	·	·	·
4_x	·	4	4	·	4	·	·	4	·	·	·	·	·	·	·	·
5_x	·	·	·	·	·	·	·	·	4	·	·	4	4	·	·	4
6_x	·	·	·	·	·	·	·	·	·	8	·	·	·	8	·	·
7_x	·	·	·	·	·	·	16	·	·	·	·	·	·	·	·	·
8_x	·	·	·	·	·	·	·	·	·	·	·	·	·	8	8	·
9_x	·	·	·	8	·	8	·	·	·	·	·	·	·	·	·	·
A_x	·	4	4	·	4	·	·	4	·	·	·	·	·	·	·	·
B_x	·	·	·	·	·	·	·	·	4	·	·	4	4	·	·	4
C_x	·	4	4	·	4	·	·	4	·	·	·	·	·	·	·	·
D_x	·	·	·	·	·	·	·	·	4	·	·	4	4	·	·	4
E_x	·	·	·	·	·	·	·	·	·	8	8	·	·	·	·	·
F_x	·	·	·	8	·	8	·	·	·	·	·	·	·	·	·	·

Figure: DDT of γ w.r.t. $\circ$

First experiment

- $\lambda \in H_\circ$ with good diffusion properties
- reminescent of PRESENT's mixing layer

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

First experiment

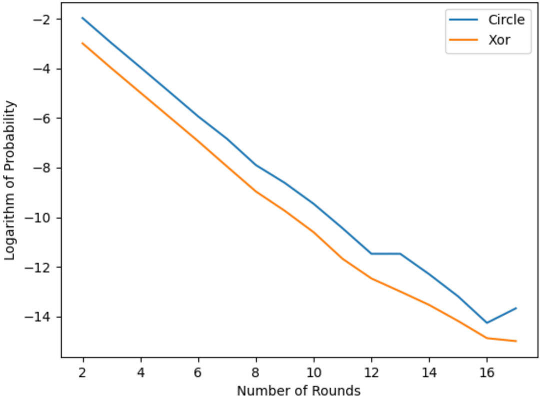


Figure: Best $+$-differential probability vs best $\circ$-differential probability

Second experiment

- Sample random mixing layers in $H_\circ$
- compare trails for different number of rounds

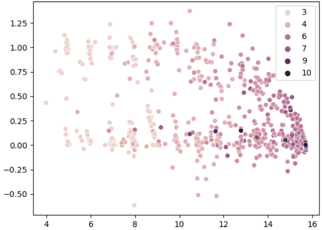


Figure: Best $+$-differential probability vs best $\circ$-differential probability

Concluding remarks

- characterization of parallel $H_\circ$ for $d = n - 2$ (and $n - 3$)
- optimal S-boxes are can have high $\circ$-differentials
- when $\lambda \in H_\circ$ $\circ$-diff. cryptanalysis can give better results
- can purposely create hidden weakness

Some open problems:

- characterization of $H_\circ$ for any d
- cryptanalysis for $d = n - 3$
- can we target key addition and / or key schedule?

Thank you for your attention.

References

- [CCS21] Calderini, Civino, Sala - On properties of translation groups in the affine general linear group with applications to cryptography
- [CDVS06] Caranti, Dalla Volta, Sala - Abelian regular subgroups of the affine group and radical rings
- [CBS19] Civino, Blondeau, Sala - Differential Attacks: Using Alternative Operations
- [LP07] Leander, Poschmann - On the Classification of 4 Bit S-Boxes