Algebraic Insights into the Secret Feistel Network[⋆] (Full Version)[⋆⋆]

Léo Perrin[1], Aleksei Udovenko[2]

[1] leo.perrin@uni.lu, SnT,University of Luxembourg
[2] aleksei.udovenko@uni.lu, SnT,University of Luxembourg

Abstract. We introduce the high-degree indicator matrix (HDIM), an object closely related with both the linear approximation table and the algebraic normal form (ANF) of a permutation. We show that the HDIM of a Feistel Network contains very specific patterns depending on the degree of the Feistel functions, the number of rounds and whether the Feistel functions are 1-to-1 or not. We exploit these patterns to distinguish Feistel Networks, even if the Feistel Network is whitened using unknown affine layers. We also present a new type of structural attack exploiting monomials that cannot be present at round $r-1$ to recover the ANF of the last Feistel function of a r-round Feistel Network. Finally, we discuss the relations between our findings, integral attacks, cube attacks, Todo's division property and the congruence modulo 4 of the Linear Approximation Table.

Keywords: High-Degree Indicator Matrix, Feistel Network, ANF, Linear Approximation Table/Walsh Spectrum, Division Property, Integral Attack

1 Introduction

While the importance of attacks targeting actual primitives is obvious, structural attacks can also lead to interesting development. In fact, the last few years have seen the publications of several such attacks. For example, the attack targeting the SASAS construction has been recently extended to larger constructions [1]. The ASASA structure, which might look weaker at first glance due to its lower number of non-linear layers, has actually proved to be a challenging target; it was even proposed as the basis for public key encryption and white-box scheme [2]. Attacking this generic structure requires sophisticated methods presented in [3] and [4]. Feistel Networks have also been the target of generic attacks in two

[⋆] The work of Léo Perrin is supported by the CORE ACRYPT project (ID C12-15-4009992) funded by the *Fonds National de la Recherche* (Luxembourg). The work of Aleksei Udovenko is supported by the *Fonds National de la Recherche*, Luxembourg (project reference 9037104)

different settings. If the Feistel functions are completely secret, attacks up to 5-rounds are presented in [5]. If the Feistel functions consist in public functions preceded by the addition of a secret key, powerful attacks with very low data complexity are presented in [6].

As illustrated by the usage of the ASASA structure, generic constructions can be applied in white-box cryptography where the aim is to prevent an attacker from having access to some of the inner components of the algorithm to perform some computations. Thus, structural attacks are important in this context. They can also be used to reverse-engineer the secret structure of an S-Box, allowing for example an attacker to enjoy the benefits of a lightweight implementation known a priori only by the designer of the S-Box. The use of small Feistel Networks for lightweight S-Box design is investigated in [7] and, in fact, a secret hardware efficient decomposition[3] was recently discovered for the S-Box of the last Russian standards [8] using such reverse-engineering.

Our Contribution Our results are based on the *high-degree indicator matrix* (HDIM), a new object we introduce. We associate to any n-bit permutation F a $n \times n$ Boolean matrix $\hat{H}(F)$ which can be computed in time $O(n2^{n-1})$ using the full code-book and which is related all at once to the LAT/Walsh spectrum of F, to its algebraic normal form and to the existence of integral distinguishers.

The HDIM provides new attack directions which we illustrate by analysing some generic constructions based on Feistel Networks. In particular, we show the existence of some patterns in the HDIM of $2n$-bit Feistel Networks with r rounds and Feistel functions with degree d depending on $\theta(d, r)$ with

$$\theta(d, r) = d^{\lfloor r/2 \rfloor - 1} + d^{\lceil r/2 \rceil - 1}.$$

These patterns provide efficient distinguishers for such structures. When the round functions are bijective, such patterns always exist in Feistel Networks with up to at least 5 round. We also show that these distinguishers can be interpreted as particular integral distinguishers and describe some relations between our results and Todo's division property [9]. Due to their integral nature, our distinguishers are extremely memory efficient: we only need to store a block containing the sum studied. In contrast, the impossible differential for 5-round Feistel Network [10] and the yoyo-game [5] are the best known distinguishers for 5-rounds FN with bijective Feistel functions and require respectively $O(2^n)$ and $O(2^{2n})$ blocks of memory.

We also present a new type of recovery attack against Feistel Networks with secret round functions which rebuilds the last Feistel function by exploiting the predictable absence of some monomials in the algebraic normal form of the permutation without its last round.

Outline We first describe the definitions and notations that we shall use throughout the paper in Section 2. Then, we investigate in Section 3 the relation between the different rows and columns of a table containing the congruence modulo 4 of

[3] Whether this hidden structure serves another purpose is still an open problem.

R	Type	Power	Restrictions	Time	Data	Ref.
5	Differential	Distinguisher	Non bij. round func.	2^n	2^n	[11]
	Imp. diff.	Distinguisher	Bij. round func.	2^{2n}	2^n	[10]
	SAT-based	Full recovery	$n \leq 7$	Practical	2^{2n}	[12]
	Yoyo	Full recovery	–	2^{2n}	2^{2n}	[5]
	Integral	Full recovery	f_1 or f_3 bij.	$2^{2.81n}$	2^{2n}	[5]
	Guess & Det.	Full recovery	–	$2^{n2^{3n/4}}$	2^{2n}	[5]
	HDIM-based	Distinguisher	Bij. round func.	2^{2n-1}	2^{2n-1}	Sec. 6.1
	Imp. monom.	Full recovery	Bij. round func.	2^{3n}	2^{2n}	Sec. 5.2
r	HDIM-based	Distinguisher	Bij. round func., $\theta(d, r-1) < 2n$	2^{2n-1}	2^{2n-1}	Sec. 6.1
	HDIM-based	Distinguisher	Non bij. round func., $\theta(d, r) < 2n$	2^{2n-1}	2^{2n-1}	Sec. 6.1
	Imp. monom.	Full recovery	$d^{r-3} < n$	2^{3n}	2^{2n}	Sec. 5.3

Table 1: Structural attacks against Feistel Networks. n is the branch size, d is the degree of the Feistel functions.

the biases in the LAT of some n-bit permutation and, in doing so, introduce and study the *high-degree indicator matrix* (HDIM). Section 4 shows that the HDIM of a Feistel Network exhibits very strong patterns depending on the number of rounds, the algebraic degree of the Feistel functions and whether these are bijective or not. We also describe attacks relying on these patterns targeting both Feistel Networks and Feistel Networks whitened using affine layers. In fact, in Section 5, we introduce a new kind of attack rebuilding efficiently the algebraic normal form of secret Feistel functions which exploits the predictable absence of some monomials in the ANF of round-reduced Feistel Networks. Finally, we discuss in Section 6 how our findings can fit in the framework of integral attacks.

Structure	Restrictions	Time	Data	Ref.
$\mathsf{AF}^4\mathsf{A}$	Bij. round func.	2^{6n}	2^{4n}	[8]
$\mathsf{AF}^r\mathsf{A}$	Bij. round func., $\theta(d, r-1) < 2n$	$n2^{2n}$	2^{2n}	Sec. 4.2
	Non bij. round func., $\theta(d, r) < 2n$	$n2^{2n}$	2^{2n}	Sec. 4.2
$\mathsf{AF}^r\mathsf{A}^{-1}$	Bij. round func., $\theta(d, r) < 2n$	$n2^{2n}$	2^{2n}	Sec. 4.2
	Non bij. round func., $\theta(d, r+1) < 2n$	$n2^{2n}$	2^{2n}	Sec. 4.2

Table 2: Structural attacks against Feistel Networks whitened with unknown affine layers. The attacks recover parts of the unknown affine layers. n is the branch size, d is the degree of the Feistel functions.

2 Notations and Boolean Functions Basics

In this section, we introduce the notations and concepts that will be used throughout the paper. A thorough introduction to Boolean functions can be found in [13]. First, let us define some sets and simple operations:

- $\mathbb{F}_2$ denotes the finite field of size 2,
- the exclusive-OR (or XOR) is denoted $\oplus$,
- the logical AND is denoted $\wedge$,
- the hamming weight $\mathrm{hw}(x)$ of a vector x of $\mathbb{F}_2^n$ is the number of ones in x,
- $|S|$ and $\#S$ denote the size of a set S,
- the scalar product of two elements $x = (x_0, ..., x_{n-1})$ and $y = (y_0, ..., y_{n-1})$ of $\mathbb{F}_2^n$ is denoted "$\cdot$" and is equal to $x \cdot y = \bigoplus_{i=0}^{n-1} x_i \wedge y_i$,
- if $x = (x_0, ..., x_{n-1})$ and $u = (u_0, ..., u_{n-1})$ are two elements of $\mathbb{F}_2^n$ then $x^u = \prod_{i=0}^{n-1} x_i^{u_i}$, and
- if $x = (x_0, ..., x_{n-1})$ and $u = (u_0, ..., u_{n-1})$ are two elements of $\mathbb{F}_2^n$ then $x \preccurlyeq u$ is true if and only if $(u_i = 0 \implies x_i = 0)$ is true for all i in $[0, n-1]$. We say that u "covers" x.

We now define some of the key components used in our analysis.

Definition 1 (Boolean Function). *We call Boolean function a function mapping $\mathbb{F}_2^n$ to $\mathbb{F}_2$. A function mapping $\mathbb{F}_2^n$ to $\mathbb{F}_2^m$ is a vectorial Boolean function and its restrictions to each output bit are its coordinates. Finally, for a vectorial Boolean function F, the Boolean functions $x \mapsto c \cdot F(x)$ are its components.*

Note that a coordinate of a Boolean function is one of its components but that the converse is not necessarily true. Let us then introduce the concept of *balancedness*.

Definition 2 (Balanced Boolean Function). *A (vectorial) Boolean function F mapping $\mathbb{F}_2^n$ to $\mathbb{F}_2^m$ is said to be balanced if the size of the preimages of all elements of $\mathbb{F}_2^m$ are equal.*

A Boolean function is balanced if and only if all of its components are balanced.

We also recall the definition of the Algebraic Normal Form of a Boolean function.

Definition 3 (Algebraic Normal Form (ANF)). *Any Boolean function f mapping n bits to 1 can be decomposed into*

$$f(x) = \bigoplus_{u \in \mathbb{F}_2^n} a_u x^u \text{ with } a_u = \bigoplus_{x \preccurlyeq u} f(x),$$

in a unique fashion which is called the Algebraic Normal Form (ANF) of f. The coefficients a_u can be obtained using the so-called Möbius transform. For vectorial Boolean functions, the ANF is the ANF of each of the coordinates.

Definition 4 (Algebraic Degree). *The algebraic degree of a Boolean function is the largest number of variables in a single term of its ANF, i.e. the maximum hamming weight of all u of $\mathbb{F}_2^n$ such that $a_u \neq 0$. The algebraic degree of a vectorial Boolean function is the maximum algebraic degree of its coordinates. The algebraic degree of a (vectorial) Boolean function f is denoted* $\deg(F)$*.*

We observe that the algebraic degree of a permutation of n bits is at most equal to $n-1$.

Our analysis will involve the *LAT* or *Fourier Transform* (related to the *Walsh spectrum* by a constant multiplication) of a Boolean function. These almost identical concepts are introduced below.

Definition 5 (LAT, Fourier Transform, Walsh Spectrum). *The* Linear Approximation Table *of a function f mapping n bits to m is a $2^n \times 2^m$ matrix $\mathcal{L}$ where $\mathcal{L}[a,b] = \#\{x \in \mathbb{F}_2^n, a \cdot x = b \cdot f(x)\} - 2^{n-1}$. We note that the coefficient $\mathcal{L}[a,b]$ can equivalently be expressed as follows:*

$$\mathcal{L}[a,b] \; = \; -\sum_{x \in \mathbb{F}_2^n} \big(b \cdot f(x)\big) \times (-1)^{a \cdot x} \; = \; -\frac{1}{2} \sum_{x \in \mathbb{F}_2^n} (-1)^{a \cdot x \oplus b \cdot f(x)},$$

where the first sum corresponds to the Fourier transform *of $x \mapsto b \cdot f(x)$ and the second to its* Walsh spectrum*. Furthermore, the coefficient $\mathcal{L}[a,b]$ of a LAT $\mathcal{L}$ is called* bias of the approximation $(a \rightsquigarrow b)$*.*

Remark 1. If F is an n-bit permutation then, for all (a,b) in $(\mathbb{F}_2^n)^2$, we have $\mathcal{L}[a,b] \equiv 0 \mod 2$.

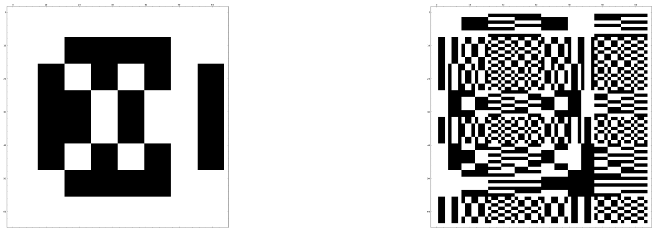
When a Boolean function μ mapping n bits to m is linear, we use μ to represent both the function itself and its matrix representation. The transpose of a matrix μ is denoted μ^t. Finally, we state the following well-known remark regarding the algebraic degree of a (vectorial) Boolean function.

Remark 2. If F is a (vectorial) Boolean function and $\mathcal{V}$ is a vector space of $\mathbb{F}_2^n$ such that $|\mathcal{V}| > 2^{\deg(F)}$, then $\bigoplus_{v \in \mathcal{V}} F(v) = 0$.

3 Patterns in Biases Modulo 4 and HDIM

Our initial goal was to identify new generic attacks against Feistel Networks. As suggested in [12], we looked at a visual representation of the Linear Approximation Table of such permutations. We identified some patterns which turned out to be byproducts of a strong structure in the congruence modulo 4 of the biases. Figures 1a and 1b show the “Pollock representation” of the LAT modulo 4 of a 4- and a 5-round 6-bit Feistel Networks for some bijective Feistel functions picked uniformly at random.

As we can see, the congruence of the biases is constant in each square of dimensions 8×8 for the 4-round Feistel Networks. Furthermore, there seems to be linear patterns for the 5-round structure: if we divide the LAT into 8×8



(a) $r = 4$

(b) $r = 5$

Fig. 1: LAT of r-round Feistel Networks (modulo 4).

squares as before then we find that each square at position (i, j) is the sum of the squares at positions $(i, 0)$ and $(0, j)$ and a square-wise constant.

The reason behind these patterns is two-fold. The first aspect is a generic observation about the linearity (in some sense) of the construction of the LAT modulo 4. Indeed, we show in this section that the function $(a, b) \mapsto (\mathcal{L}[a, b] \mod 4)$ for the LAT $\mathcal{L}$ of a permutation is a bilinear form and that its matrix representation has interesting properties. The second aspect of the justification for the patterns is the probability 1 presence of zeroes in some positions which is discussed later in Section 4.

3.1 The High-Degree Indicator Matrix

We first re-write the congruence modulo 4 of the biases in the LAT of a permutation using Boolean functions.

Lemma 1 (LAT modulo 4). *Let F be a permutation of n bits ($n > 2$) and let $\mathcal{L}$ be its LAT. Then $\mathcal{L}[a, b]$ is such that* $\mathcal{L}[a, b] \equiv 2 \times \left(\bigoplus_{x \in \mathbb{F}_2^n} \big(b \cdot F(x)\big)\big(a \cdot x\big)\right) \mod 4$ *or, equivalently,*

$$\frac{\mathcal{L}[a, b]}{2} \equiv \bigoplus_{x \in \mathbb{F}_2^n} \big(b \cdot F(x)\big)\big(a \cdot x\big) \mod 2.$$

Proof. Since $(-1)^z = 1 - 2z$ (for z in $\{0, 1\}$), the coefficient $\mathcal{L}[a, b]$ is equal to

$$\mathcal{L}[a, b] = -\sum_{x \in \mathbb{F}_2^n} \big(b \cdot F(x)\big) + 2\Big(\sum_{x \in \mathbb{F}_2^n} \big(b \cdot F(x)\big)\big(a \cdot x\big)\Big).$$

The first term in this sum is equal to 2^{n-1} as every component of a permutation is balanced.[4] Thus, if we look at the congruence modulo 4 of $\mathcal{L}[a, b]$, we obtain the following (for any $n > 2$):

$$\mathcal{L}[a, b] \equiv 2\Big(\sum_{x \in \mathbb{F}_2^n} \big(b \cdot F(x)\big)\big(a \cdot x\big)\Big) \mod 4,$$

[4] If F is not a permutation but some function with degree at most $n - 1$, then this term *a priori* does not go away when taking the modulo 4 of the expression.

from which we deduce that

$$\frac{\mathcal{L}[a,b]}{2} \equiv \sum_{x \in \mathbb{F}_2^n} \big(b \cdot F(x)\big)\big(a \cdot x\big) \mod 2$$

As sum and XOR are equivalent modulo 2, this proves the lemma. □

This lemma has several consequences regarding the congruence modulo 4 of the LAT coefficients of F (or, alternatively, the congruence modulo 2 of their half). First, we define $\mathcal{L}_4$ to be a $2^n \times 2^n$ matrix such that $\mathcal{L}_4[a,b] \equiv \mathcal{L}[a,b] \mod 4$ and $\mathcal{L}_4[a,b] \in 0, 2$. Using this, we define $B(\mathcal{L})$ to be a $2^n \times 2^n$ Boolean matrix with $B(\mathcal{L})[a,b] = \mathcal{L}_4[a,b]/2$. This matrix has the following property:

$$B(\mathcal{L})[a \oplus a', b \oplus b'] = B(\mathcal{L})[a,b] \oplus B(\mathcal{L})[a,b'] \oplus B(\mathcal{L})[a',b] \oplus B(\mathcal{L})[a',b'].$$

As consequence, the function $(a,b) \mapsto B(\mathcal{L})[a,b]$ is a bilinear form and can be represented using an $n \times n$ matrix $\hat{H}(F)$.

Definition 6 (High-Degree Indicator Matrix (HDIM)). *Let F be an n-bit permutation and let $B(\mathcal{L})$ be the Boolean matrix representing the congruence modulo 4 of its LAT (as described above). We define the High-Degree Indicator Matrix $\hat{H}(F)$ of F to be the $n \times n$ matrix such that*

$$\hat{H}(F)[i,j] = \bigoplus_{x \in \mathbb{F}_2^n} \big(e_i \cdot F(x)\big)\big(e_j \cdot x\big),$$

where e_k is an all zero n-bit vector with a single 1 at position k. This matrix is such that

$$B(\mathcal{L})[a,b] = b^t \times \hat{H}(F) \times a.$$

Lemma 2. *The coefficients of $\hat{H}(F)$ indicate the presence of the highest degree terms in the coordinates of F. More precisely, $\hat{H}(F)[i,j] = 1$ if and only if the ANF of F_i contains the monomial $\prod_{k \neq j} x_k$ (which has degree $n-1$).*

Proof. Let F be an n-bit permutation. As $\hat{H}(F)[i,j]$ is the sum over of space of size 2^n of the Boolean function $x \mapsto \big(e_i \cdot F(x)\big)\big(e_j \cdot x\big) = F_i(x) \cdot x_j$, it is equal to 0 unless this Boolean function has algebraic degree n. As F has degree $n-1$, this occurs if and only if F_i contains $\prod_{k \neq j} x_k$. Indeed, in this case (and in this case only), the ANF of $x_j \cdot F_i(x)$ contains the only possible degree n term $\prod_{k=0}^{n-1} x_k$. □

This lemma is the reason behind the name "high-degree indicator matrix". Indeed, the HDIM coefficients simply state whether each of the n possible $n-1$ degree terms appear in each coordinate of F or not.

We finally note that the HDIM of a function can be computed much more efficiently than the LAT or the difference distribution table. Indeed, we can compute a column of the HDIM by summing the function over a cube of dimension $n-1$ (see Section 6.1). The complexity for all n columns is therefore $n2^{n-1}$.

3.2 Some Properties of the High-Degree Indicator Matrix

Let us investigate the effect of some simple transformations on the HDIM. First, we point out that due to the fact that the LAT of the inverse of a permutation F is the transpose of the LAT of F, the HDIM of F^{-1} is the transpose of the HDIM of F.

We now show that the HDIM of $\eta \circ f \circ \mu$ can easily be deduced from that of f when η and μ are n-bit linear permutations. The corresponding theorem will be used in Section 4.2 to attack Feistel Networks whitened using affine layers.

Theorem 1. *Let μ, η be linear n-bit mappings, F be an n-bit permutation and let $G = \eta \circ F \circ \mu$. Furthermore, let $\hat{H}(F)$ be the HDIM of f and $\hat{H}(G)$ be that of G. Then it holds that*

$$\hat{H}(G) = \eta \times \hat{H}(F) \times (\mu^t)^{-1}.$$

Proof. We prove this result in two steps. First, the fact that $\hat{H}(F \circ \mu) = \hat{H}(F) \times (\mu^{-1})^t$ can be derived as follows:

$$\begin{aligned} \hat{H}(F \circ \mu)[i,j] &= \bigoplus_{x \in \mathbb{F}_2^n} \big(e_i \cdot F(\mu(x))\big)\big(e_j \cdot x\big) = \bigoplus_{y \in \mathbb{F}_2^n} \big(e_i \cdot F(y)\big)\big(e_j \cdot \mu^{-1}(y)\big) \\ &= \bigoplus_{y \in \mathbb{F}_2^n} \big(e_i \cdot F(y)\big)\big((\mu^t)^{-1}(e_j) \cdot y\big). \end{aligned}$$

We then note that $\hat{H}(\eta \circ F) = \hat{H}(F^{-1} \circ \eta^{-1})^t$ which, using what we just found, is equal to $(\hat{H}(F^{-1}) \times \eta^t)^t = (\hat{H}(F)^t \times \eta^t)^t$, so that $\hat{H}(\eta \circ F) = \eta \times \hat{H}(F)$. This concludes the proof. □

The ANF and the LAT of an n-bit permutation are connected in the sense that it is possible to determine the congruence modulo 4 of the LAT $\mathcal{L}$ of an n-bit permutation F given parts of its ANF. Indeed, as we describe in this section, this congruence only depends on the terms of degree $n-1$ in the ANF of the coordinates of F.

4 The High-Degree Indicator Matrix of Feistel Networks

In what follows, we denote F_d^r an r-round FN with bijective Feistel function of algebraic degree at most d. The structure of a sample is given Figure 2. It is possible to use the HDIM to analyse such generic structures.

4.1 Artifacts in the HDIM of Feistel Networks

The HDIM of a Feistel Network may yield interesting patterns depending on the degree of its Feistel functions, whether they are bijections or not and its number of rounds. These are formalized by Theorem 2 and its corollary (Corollary 1). These results link the maximum degree d of the Feistel functions, the number of

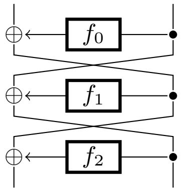


Fig. 2: A sample F_d^3 structure, where $\deg(f_i) \leq d$.

rounds r and the presence or not of some patterns using the function $\theta : \mathbb{Z}^2 \rightarrow \mathbb{Z}$ defined by

$$\theta(d, r) = d^{\lfloor r/2 \rfloor - 1} + d^{\lceil r/2 \rceil - 1},$$

where $\lfloor 2k \rfloor = \lfloor 2k + 1 \rfloor = 2k$ and $\lceil 2k \rceil = \lceil 2k - 1 \rceil = 2k$.

Theorem 2. *Let F be a $2n$-bit F_d^r. Then the HDIM of F is such that $\hat{H}(F)[i, j] = 0$ if $i < n$ **or** $j < n$ under the following conditions:*

- *if the Feistel functions are bijections and $\theta(d, r) < 2n$, or*
- *if the Feistel functions are not bijections and $\theta(d, r + 1) < 2n$.*

The general idea of the proof is to express the sum corresponding to coefficient $\hat{H}(F)[i, j]$ using well-chosen variables (α, β) located in the middle of the encryption. The value of $F(x)$ is then a function of degree $d^{\lceil r/2 \rceil - 1}$ of (α, β) and that of x is a function of degree $d^{\lfloor r/2 \rfloor - 1}$. The coefficients can thus be written as

$$\hat{H}(F)[i, j] = \bigoplus_{(\alpha, \beta) \in (\mathbb{F}_2^n)^2} \big(e_i \cdot F(x(\alpha, \beta))\big)\big(e_j \cdot x(\alpha, \beta)\big)$$

and the result is equal to 0 if $\theta(d, r) = d^{\lfloor r/2 \rfloor - 1} + d^{\lfloor r/2 \rfloor - 1} < 2n$. If the Feistel functions are not bijective then a “trick” used to slightly decrease the degree in (α, β) of the output cannot be used, hence the small discrepancy in this case.

We will now give a complete proof of Theorem 2. This proof requires the following remark which is derived simply by tracking the evolution of the algebraic degree of each branch of the FN.

Remark 3 (FN algebraic degree). Let $F : x \mapsto F_\ell(x) || F_r(x)$ be a $2n$-bit permutation with structure F_d^r and let $G : x \mapsto G_\ell(x) || G_r(x)$ be a $2n$-bit permutation such that $\deg(G_r) = d_G$ and $\deg(G_\ell) \leq d \times d_G$. Then the degree of the left and right words of $F \circ G$ are bounded as follows: $\deg(F_\ell \circ G) \leq d^{r+1} \times d_G$ and $\deg(F_r \circ G) \leq d^r \times d_G$.

Proof (Theorem 2). First of all, note that the inverse of a Feistel Network is also a Feistel Network. Thus, it is sufficient to prove that $\hat{H}(F)[i, j] = 0$ for $i < n$. Indeed, if it is the case for any F_d^r, then it is also the case for F^{-1} so that $\hat{H}(F^{-1})[j, i] = 0$ for $j < n$. As $\hat{H}(F^{-1})[j, i] = \hat{H}(F)[i, j]$, it is also sufficient that $j < n$. Let us thus prove that $\hat{H}(F)[i, j] = 0$ for $i < n$.

Recall that the coefficient $\hat{H}(F)[i,j]$ is equal to $\bigoplus_{x\in\mathbb{F}_2^{2n}} \big(e_i\cdot F(x)\big)\big(e_j\cdot x\big)$. Our proof relies on expressing this sum using another set of variables and showing that the Boolean functions using these variables has an algebraic degree below $2n$, so that it always sums to 0.

Let F be a $2n$-bit F_d^r built using Feistel functions $f_0, ..., f_{r-1}$

We denote γ the input of $f_{\lfloor r/2 \rfloor}$, α the other input of round $\lfloor r/2 \rfloor$ and $\beta = \alpha \oplus f_{\lfloor r/2 \rfloor}(\gamma)$ the output of round $\lfloor r/2 \rfloor$ which is not equal to γ (see Figure 3).

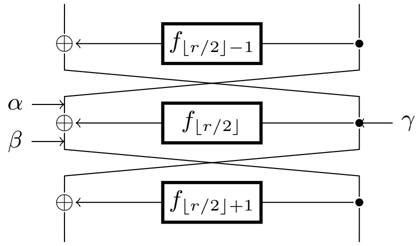

$f_{\lfloor r/2 \rfloor -1}$
α
$f_{\lfloor r/2 \rfloor}$
γ
β
$f_{\lfloor r/2 \rfloor +1}$

Fig. 3: The variables α, β and γ.

Let us denote $\big(x_\ell(\alpha,\beta), x_r(\alpha,\beta)\big)$ the left and right side of the input of F such that the input of round $\lfloor r/2 \rfloor$ is (γ, α) and $\big(y_\ell(\alpha,\beta), y_r(\alpha,\beta)\big)$ the corresponding output. The coefficients of the HDIM of F for $i < n$ can thus be expressed as

$$\bigoplus_{\alpha||\beta\in\mathbb{F}_2^{2n}} \big(e_i \cdot y_r(\alpha,\beta)\big)\big(e_j \cdot x_\ell(\alpha,\beta) \oplus e_j \cdot x_r(\alpha,\beta)\big). \tag{1}$$

It is therefore sufficient to bound the degree in (α, β) of this expression is below n to prove the theorem, which we will achieve by looking separately at the degree of y_r and that of $x_\ell||x_r$.

Our starting point is different depending on whether the Feistel functions are bijective or not. If they are not bijections, then we set $\beta = \alpha \oplus f_1(\gamma)$, sum over $(\alpha||\gamma)$ and look at functions x'_ℓ, x'_r, y'_ℓ and y'_r taking as input α and γ instead. We define $B_b(r) = \deg(y_r) + \deg(x_\ell||x_r)$ (bijective case) and $B_c(r) = \deg(y'_r) + \deg(x'_\ell||x'_r)$ (collisions are allowed).

For $r = 3$, we have:

$$\begin{cases} x_\ell(\alpha,\beta) = f_0(\alpha) \oplus \gamma, x_r(\alpha,\beta) = \alpha \\ y_\ell(\alpha,\beta) = f_1(\beta) \oplus \gamma, y_r(\alpha,\beta) = \beta. \end{cases}$$

If the functions are bijections, we set $\gamma = f_1^{-1}(\alpha \oplus \beta)$ so that the degrees of x_ℓ, x_r, y_ℓ, y_r are upper bounded respectively by $d, 1, d, 1$. If $r = 2k+1$ is odd, we add $k-1$ Feistel rounds before and after $x_\ell||x_r$ and $y_\ell||y_r$. In this case, Remark 3 implies that the degrees become $d^k, d^{k-1}, d^k, d^{k-1}$ so that $B_b(2k{+}1) \leq d^k + d^{k-1} = \theta(d, 2k+1)$. If $r = 2k$, we add $k-2$ rounds at the top and $k-1$ at the bottom which means that the degrees become $d^{k-1}, d^{k-2}, d^k, d^{k-1}$, so that $B_b(2k) \leq d^{k-1} + d^{k-1} = \theta(d, 2k)$.

If they are *not* bijections, the degrees of $x'_\ell, x'_r, y'_\ell, y'_r$ are upper bounded respectively by $d, 1, d^2, d$. The same reasoning as above applies, so that if $r = 2k+1$ then we add $k-1$ rounds above and below and the degrees become $d^k, d^{k-1}, d^{k+1}, d^k$. We deduce that $B_c(2k+1) \leq d^k + d^k = \theta(d, 2k+2)$. Similarly, if $r = 2k$ then we add $k-1$ rounds at the top and $k-2$ at the bottom which implies that the degrees become $d^{k-1}, d^{k-2}, d^{k+1}, d^k$. In this case, we deduce that $B_c(2k) \leq d^{k-1} + d^k = \theta(d, 2k+1)$. □

Corollary 1. *Let F be a $2n$-bit F_d^r. The HDIM of F is such that $\hat{H}(F)[i,j] = 0$ if $i < n$* ***and*** *$j < n$ under the following conditions:*

- *if the Feistel functions are bijections and $\theta(d, r-1) < 2n$, or*
- *if the Feistel functions are not bijections and $\theta(d, r) < 2n$.*

Proof. Let r and d be such that F_d^{r-1} fits the hypothesis of Theorem 2. The right word of the output of a F_d^r structure is the left word output by a F_d^{r-1} structure. As each line of the HDIM corresponds to one output bit, the top n rows of the HDIM of the r-round FN are equal to the bottom n rows of the same permutation reduced to $(r-1)$-round. Because of Theorem 2, this bottom half is such that the first n columns are all 0. Thus, the first n columns of the first n rows of the HDIM of a F_d^r are all equal to 0. □

To illustrate these theorems, we give the HDIM of the 4- and 5-round Feistel with 3-bit bijective Feistel functions picked uniformly at random whose LAT modulo 4 were given in Figures 1a and 1b. The Feistel functions must have an algebraic degree at most equal to 2. Since $\theta(2, 4) = 2^1 + 2^1 = 4 < 6$, these HDIM must exhibit the patterns described in the theorems above. It is the case, as we can see below. The zeroes caused by Theorem 2 and Corrolary 1 are represented in grey:

$$\hat{H}(\mathsf{F}^4) = \begin{bmatrix} 0&0&0&0&0&0 \\ 0&0&0&0&0&0 \\ 0&0&0&0&0&0 \\ 0&0&0&0&1&1 \\ 0&0&0&1&0&1 \\ 0&0&0&1&0&1 \end{bmatrix}, \; \hat{H}(\mathsf{F}^5) = \begin{bmatrix} 0&0&0&1&1&1 \\ 0&0&0&1&1&1 \\ 0&0&0&0&1&1 \\ 0&1&1&0&1&0 \\ 1&0&0&0&0&0 \\ 0&1&1&1&1&1 \end{bmatrix}. \tag{2}$$

Even though a F_d^r structure has an algebraic degree of $2n-1$ in the conditions of Theorem 2, the way in which this high degree is achieved is very structured: only half of the output bits actually have a maximum degree and the monomials of degree $2n-1$ can not contain the product of $n-1$ bits from the right side of the input. Thus, a simple analysis of the algebraic degree can be made more sophisticated by also investigating the possible structure of the monomials of highest degree.

These patterns lead to the existence of distinguishers as long as the conditions necessary for Corollary 1 are satisfied. Table 3 shows the value of the number of rounds for which the conditions of Corollary 1 are satisfied for different values of d, r and n in both the 1-to-1 case and the case where collisions in the Feistel functions are allowed. If real ciphers correspond to these parameters, we specify them. Note that the rotation applied to one of the branches in the round function of LBlock [14] does not change anything. The key-dependent linear FL layers in MISTY1 [15] do not protect from our distinguisher as well and may be included from any side for free.

$(d, 2n)$	Feistel functions	$r_{\max}(d, n)$	Instance
$(2, 32)$	1-to-1	10	—
	collisions	9	SIMON-32 [16]
$(5, 64)$	1-to-1	7	—
	collisions	6	DES [17]
$(31, 64)$	1-to-1	5	MISTY1/KASUMI [15]
	collisions	4	—
$(n-1, 2n)$	1-to-1	5	—
	collisions	4	—

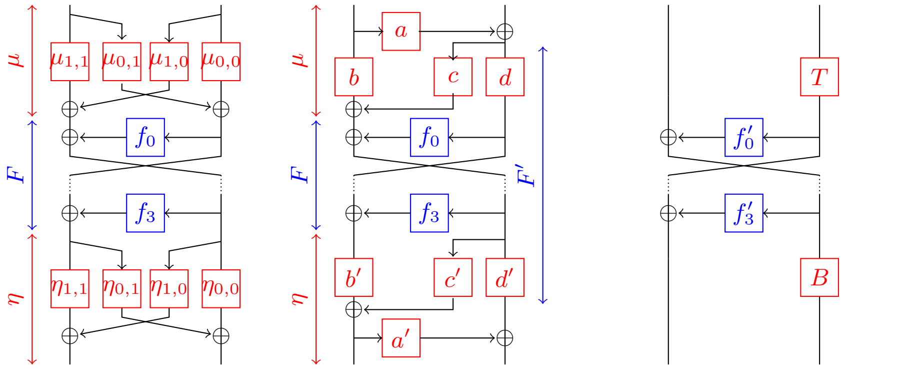
Table 3: If $r = r_{\max}(d, 2n)$ then the $2n$-bit permutation F_d^r exhibits an artifact of size n^2 in its HDIM.

4.2 Bypassing Affine Whitening

In the context of component reverse-engineering/white-box cryptography, it may not be sufficient to be able to attack generic Feistel structure. Indeed, simply whitening a generic structure with secret affine layers can prevent many attacks from succeeding at small cost for the designer. For example, applying affine layers before and after a 5-round Feistel Network would prevent the yoyo-game used in [5] to be exploitable. Similarly, the recent attacks against ASASA [3,4] are much more sophisticated than the attack against SASAS proposed by Biryukov et al. in the first place [18]. We also note that the secret structure of the S-Box of the last Russian standard primitives recently recovered was indeed whitened with seemingly random linear layers [8].

As a consequence, we study the generic construction denoted $\mathsf{AF}_d^r\mathsf{A}$ consisting in a F_d^r construction with secret Feistel functions preceded and followed by the application of independent and secret linear layers[5]. This structure has already been studied in [8] but our attacks are significantly more efficient. Note also that one of the S-Box of ZUC [19] has this structure: it is a 3-round Feistel Network composed with a bit rotation. Let us show how the HDIM and its artifacts we identified in the previous section can be used to attack permutations with $\mathsf{AF}_d^r\mathsf{A}$ structures.

[5] We note that adding constants to make the layers affine is equivalent to replacing the Feistel functions by other ones with identical properties.



(a) $G = \eta \circ F \circ \mu$. (b) G (alt. representation). (c) F' (alt. representation).

Fig. 4: The target of our attack, its result and its alternative representation. In Figure 4c, f'_i is affine equivalent to f_i.

Our attack works for a subset of all possible linear layers. We define $G = \eta \circ F \circ \mu$ where F has a F_d^r structure satifying the conditions of Theorem 2 and μ and η are linear layers. The layer applied first must have a decomposition as follows:

$$\mu = \begin{bmatrix} \mu_{0,0} & \mu_{0,1} \\ \mu_{1,0} & \mu_{1,1} \end{bmatrix} = \begin{bmatrix} d & 0 \\ c & b \end{bmatrix} \times \begin{bmatrix} I & a \\ 0 & I \end{bmatrix} = \begin{bmatrix} d & d \times a \\ c & b + c \times a \end{bmatrix},$$

and the layer applied last must have a similar one:

$$\eta = \begin{bmatrix} \eta_{0,0} & \eta_{0,1} \\ \eta_{1,0} & \eta_{1,1} \end{bmatrix} = \begin{bmatrix} I & a' \\ 0 & I \end{bmatrix} \times \begin{bmatrix} d' & 0 \\ c' & b' \end{bmatrix} = \begin{bmatrix} d' + a' \times c' & a' \times b' \\ c' & b' \end{bmatrix}.$$

It is sufficient for such a decomposition of the first layer to exist that $\mu_{0,0}$ is invertible. Indeed, we can then simply set $d = \mu_{0,0}, c = \mu_{1,0}, a = d^{-1} \times \mu_{0,1}$ and $b = \mu_{1,1} - c \times a$. Note that b has to be invertible since μ is invertible. Similarly, it is sufficient that $\eta_{1,1}$ is invertible to decompose the final layer. We define F' using these decompositions so that G is equal to:

$$G = \begin{bmatrix} I & a' \\ 0 & I \end{bmatrix} \circ \begin{bmatrix} d' & 0 \\ c' & b' \end{bmatrix} \circ F \circ \begin{bmatrix} d & 0 \\ c & b \end{bmatrix} \circ \begin{bmatrix} I & a \\ 0 & I \end{bmatrix} = \begin{bmatrix} I & a' \\ 0 & I \end{bmatrix} \circ F' \circ \begin{bmatrix} I & a \\ 0 & I \end{bmatrix}.$$

A graphical representation of the relation between F, F' and G is provided in Figures 4a and 4b. As F satisfies the condition of Theorem 2, its HDIM is such that $\hat{H}(F)[i,j] = 0$ if $i < n$ or $j < n$. Applying Theorem 1 gives us that the HDIM of F' is equal to

$$\hat{H}(F') = \begin{bmatrix} d' & 0 \\ c' & b' \end{bmatrix} \times \hat{H}(F) \times \begin{bmatrix} d & c \\ 0 & b \end{bmatrix}^{-1} = \begin{bmatrix} 0 & 0 \\ 0 & h' \end{bmatrix} \text{with } h' = b' \times h \times b^{-1},$$

h being the bottom-right part of $\hat{H}(F)$. Like in $\hat{H}(F)$, it holds that $\hat{H}(F')[i,j] = 0$ if $i < n$ or $j < n$. Another way to see why this holds is shown in Figure 4c.

Indeed, F' can be written as a F_d^r structure, like F, where n-bit linear permutations are applied only on two branches and where the Feistel functions f'_i are obtained from compositions of b, b', d, d' and f_i, as well as the addition of c and c' for the first and last rounds. We deduce that if G indeed has a $\mathsf{AF}_d^r\mathsf{A}$ structure satisfying the conditions for Theorem 2, then the following equation with unknowns the $n \times n$ binary matrices a and a' must have at least one solution:

$$\begin{bmatrix} I & a' \\ 0 & I \end{bmatrix} \times \hat{H}(G) \times \begin{bmatrix} I & 0 \\ a & I \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & h_{1,1} \end{bmatrix},$$

where $h_{1,1}$ is the bottom right corner of $\hat{H}(G)$. This system has $2n^2$ unknowns and $3n^2$ equations, meaning that it is unlikely to have solutions if G is a random permutation. However, if it does have a solution then we deduce both that G has an $\mathsf{AF}_d^r\mathsf{A}$ structure and the expression of parts of the linear layers. We summarize these results in the following attack.

Attack 1 (Partial Recovery Against $\mathsf{AF}_d^r\mathsf{A}$) *Let G be a $2n$-bit permutation. It is necessary for G to be in $\mathsf{AF}_d^r\mathsf{A}$ for some (r, d) satisfying Theorem 2 that the equation*

$$\begin{bmatrix} I & a' \\ 0 & I \end{bmatrix} \times \hat{H}(G) \times \begin{bmatrix} I & 0 \\ a & I \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & h_{1,1} \end{bmatrix},$$

where h is an unknown $n \times n$ matrix, has at least one solution. The unknowns are the coefficients of the $n \times n$ matrices a and a', so that $2n^2$ Boolean variables must satisfy $3n^2$ equations corresponding to the zeroes in the right hand side.

This distinguisher requires the full code-book and as much time as is needed to compute the HDIM and solve a system of equations. Since the system is small, the bottle-neck is the computation of the HDIM which can be done in time $\mathrm{O}(n2^{2n})$ where n is the branch size.

We can use the exact same reasoning to attack one more round if the decomposition of η and μ involve the same "linear Feistel function" a. This happens in particular if $\eta = \mu^{-1}$. In this case, we can use the distinguisher obtained from the following attack.

Attack 2 (Partial Recovery Against $\mathsf{A}^{-1}\mathsf{F}_{r+1}^d\mathsf{A}$) *Let G be a $2n$-bit permutation. In order for G to be in $\mathsf{AF}_d^r\mathsf{A}$ for some (r, d) satisfying Corollary 1 in such a way that the linear layers are the inverse of one another, it is necessary that the equation*

$$\begin{bmatrix} I & a \\ 0 & I \end{bmatrix} \times \hat{H}(G) \times \begin{bmatrix} I & 0 \\ a & I \end{bmatrix} = \begin{bmatrix} 0 & h_{0,1} \\ h_{1,0} & h_{1,1} \end{bmatrix},$$

where $h_{0,1}, h_{1,0}$ and $h_{1,1}$ are unknown $n \times n$ matrices, has at least one solution. The unknowns are the coefficients of the $n \times n$ matrices a, so that n^2 Boolean variables must satisfy n^2 equations corresponding to the zero in the right hand side.

Note that if there is a single whitening affine layer applied at some side, we have a similar system with n^2 unknowns. If we consider one more round, we will have n^2 equations as well. Therefore we can attack $\mathsf{F}_d^r\mathsf{A}$, where r is the maximum number of rounds satisfying Corollary 1. Another view on this attack is given in Section 5.3.

5 The Impossible Monomials Attack

In the previous sections we used absent terms of highest degree to recover whitening linear layers from Feistel Networks. In this section we generalize this method to terms of lower degree and, as a result, we present an attack recovering a secret round function from a 5-round Feistel Network with bijections. Furthermore, we generalize this attack to more rounds if the degrees of the round functions are small.

5.1 Impossible monomials in Feistel Networks

Let F be a $2n$-bit F_{n-1}^4 and let F_i be the ith output bit of F (F_0 is the leftmost bit of F). We will denote by $L = \{0, \dots, n-1\}$ and $R = \{n, \dots, 2n-1\}$ the indices from the left and right halves respectively, and F_L and F_R the truncations of the function F to the left and right half respectively. Consider the ANF of F_i:

$$F_i(x_l||x_r) = \bigoplus_{u_l, u_r \in \mathbb{F}_2^n} a_{u_l||u_r}^{F_i} x_l^{u_l} x_r^{u_r}, \tag{3}$$

where x_l and x_r are vectors of input variables from the left and right halves respectively. We will now show that some monomials are impossible, that is, $a_{u_l||u_r}^{F_i} = 0$ for some u_l, u_r independently of the choice of the Feistel functions. To prove it, we will need the following lemmas.

Lemma 3. *Let $a, b \in \mathbb{F}_2^n$ be some vectors of variables and let $f : \mathbb{F}_2^n \to \mathbb{F}_2$ be a Boolean function of degree at most d. Then if some term in the ANF of $f(a \oplus b)$ has degree d_a on variables from a, then it has degree at most $d - d_a$ on variables from b. In particular, there are no terms of degree d on a and non-zero degree on b.*

Proof. Let $s(a, b) = a \oplus b$. Then $\deg s = 1$ and $\deg(f \circ s) \leq d$. Hence a term containing d_a variables from a contains at most $d - d_a$ variables from b.

Lemma 4. *Let $\pi : \mathbb{F}_2^n \to \mathbb{F}_2^n$ be a permutation and let $f : \mathbb{F}_2^n \to \mathbb{F}_2$ be some Boolean function of degree at most $n - 1$. Then* $\deg(f \circ \pi) \leq n - 1$.

Proof. By the Möbius transform, the term of degree n is present in the ANF of $f \circ \pi$ if and only if the sum of $f \circ \pi$ over $\mathbb{F}_2^n$ is equal to 1. Since π is a permutation, we have that $\sum_{x \in \mathbb{F}_2^n} f(\pi(x)) = \sum_{x \in \mathbb{F}_2^n} f(x)$. But this last sum is equal to zero because $\deg f \leq n - 1$. Therefore, there is no term of degree n in the ANF of $f \circ \pi$ and we conclude that $\deg(f \circ \pi) \leq n - 1$.

We now formally describe classes of impossible monomials using the following theorem.

Theorem 3. *Let F and its ANF be as defined before. Then $a^{F_i}_{u_l||u_r} = 0$ if one of the following holds:*

1. *$i \in R$ and $hw(u_l) = n$;*
2. *$i \in R$ and $hw(u_l) = n-1, hw(u_r) = n-1$;*
3. *$i \in R$ and $hw(u_l) = n-1, hw(u_r) = n$;*
4. *$i \in L$ and $hw(u_l) = n, \quad hw(u_r) = n-1$.*

Proof. Points 3-4 are part of Theorem 2 and are presented here for the sake of completeness. It is left to prove points 1 and 2.

1. Consider the 4-round integral characteristic from Figure 5. Let C be any cube which contains the whole left part. From the integral characteristic it follows that the sum of F over the cube C has zero on the right side. Therefore by the Möbius transform the corresponding ANF coefficients are zero.

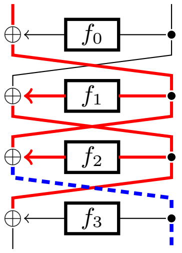


Fig. 5: The 4-round integral characteristic: words taking all values are represented in bold red and balanced words are represented in dashed blue.

2. Let $f_0, f_1, f_2, f_3 : \mathbb{F}_2^n \to \mathbb{F}_2^n$ be the round functions of F. The equation for the right half of the output is then given by:

$$F_R(l||r) = l \oplus f_0(r) \oplus f_2(r \oplus f_1(l \oplus f_0(r))). \quad (4)$$

Clearly, the first two terms do not contain any monomial of degree $n-1$ on l and $n-1$ on r. Consider the expression $f_2(r \oplus f_1(l \oplus f_0(r)))$. Assume that a term with degree $n-1$ on both l and r is present in the ANF of the expression. Then the term is present in the expansion of some product of at most $n-1$ bits, where the bits are output bits of the expression $(r) \oplus f_1(l \oplus f_0(r))$, i.e. in the term each of the $n-1$ factors is either a bit from (r) or from $f_1(l \oplus f_0(r))$. Note that the term may not be generated by choosing bits *only* from (r), because in that case there will be no variables from l in it. Therefore there are at most $n-2$ bits taken from the outer (r); $n-1$ variable from l and at least one variable r_i are taken from $f_1(l \oplus f_0(r))$. It means that there exists a monomial function π such that $\pi \circ f_1(l \oplus f_0(r))$ contains term of degree $n-1$ on l and degree at least 1 on r. By Lemma 4, $\pi \circ f_1$ has degree at most $n-1$ and by Lemma 3 there can not be such term in $\pi \circ f_1(l \oplus f_0(r))$.

5.2 An Attack on 5-round Feistel Network

In this section we use the impossible monomials to attack 5-round Feistel Network built from permutations. The key idea is to observe the presence of some 4-round impossible monomials in the 5-round network and extract some information about the last round function. Consider some monomial x^u which is impossible at the right side of a 4-round Feistel Network. We now add the 5th round. If we observe x^u on the left side, then we know that this monomial has come from the last round function. Otherwise, we know that it has *not* come from the last round function and it gives us some information as well. Using these observations we build a system of linear equations where the unknowns are the ANF coefficients of the coordinates of the last round function. By solving the system we recover the ANF coefficients and hence the function itself. Note that in order to compute the ANF, we have to obtain the full codebook.

Let F^5 be a $2n$-bit F_d^5, F^4 be its first 4 rounds and f be the last round function. Let a_u^g be the coefficient of term x^u in the ANF of the Boolean function g. Consider the equation of the ith bit of F^5 for $i \in L$:

$$F_i^5(x) = F_{i+n}^4(x) \oplus f_i(F_R^5(x)) = \bigoplus_{u \in \mathbb{F}_2^{2n}} a_u^{F_{i+n}^4} x^u \oplus \bigoplus_{u \in \mathbb{F}_2^{2n}} a_u^{f_i} (F_R^5(x))^u.$$

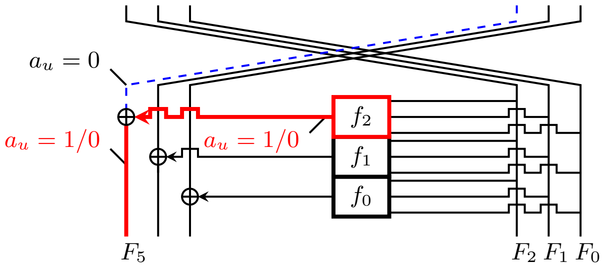


Fig. 6: Impossible monomials in the last round of a 5-round FN with 3-bit branches. The wire with 4-round impossible monomials is in dashed blue, the path of the observed monomials is highlighted with bold red. a_u is the ANF coefficient of some 4-round impossible monomial.

The ANF of F_i^5 with $i \in L$ contains some monomial from the first or the second group from Theorem 3 if and only if the ANF of $f_i \circ F_R^5$ does. Since we can compute the ANF of F_R^5, we can check which possible terms from the ANF of f_i generate the impossible monomial. Then from the presence of the impossible monomial in the ANF of F_{i+n}^5 we deduce if the number of such terms in the ANF of f_i is odd or even. This gives us a linear equation over $\mathbb{F}_2$ where the unknowns are the ANF coefficients of f_i. For an illustration see Figure 6.

Note that the 4-round impossible monomials which are still impossible in a 5-round Feistel Network do not leak any information about f. For example, since Feistel Network is a bijection, the monomial of degree $2n$ is impossible for any number of rounds but it can not be used in the attack. However it is the only such monomial. Therefore we can use $2^n - 1$ impossible monomials from the first group of Theorem 3 and n^2 ones from the second group. Each such monomial yields an equation per each bit of f. There are 2^n unknown coefficients in the ANF of f_i so the number of equations will be enough to recover f_i for all i and hence f with high probability. Note that we can recover f only up to xor with a constant because the constant may propagate through the Feistel Network and merge with other round functions (see the introduction of [5] for a more detailed explanation of this phenomenon).

The complexity of the attack is $O(2^{3n})$ and is dominated by generating the equation matrix, which is the same for all output bits (the only difference is the target vector). For each of the 2^n possible terms in the ANF of f_i we compute the ANF of the term applied after F in time $O(2^{2n})$ and then we check if this term generates the impossible monomials. The next step is to solve the systems. Since the equation matrix is the same for all output bits, we can do some precomputation (for example the LU-decomposition) once and solve all n systems of equations very fast. Computing the target vectors is dominated by computing the ANF of F_i^5 for $i \in L$ which takes total time of $O(n2^{2n})$.

As a consequence of the algebraic nature of the attack, if the round function has lower degree, then the complexity decreases. Indeed, there are less unknowns and therefore both steps of generating the equation matrix and solving the systems take less time. As an edge case, consider the F^5A structure where the affine layer can be seen as the 6th round with a function of degree 1. The complexity of recovering the affine round is $O(n2^{2n})$, as was shown in Section 4.2.

Note that the attack can be run in the reverse direction as well, so that we recover the first round function instead of the last one.

We have implemented the attack in Sage [20]. We successfully attacked a 5-round Feistel Network with bijections and branch size of up to 9 bits and recovered the outer secret round functions in a few minutes on a modern laptop.

5.3 A Generalization of the Attack on Feistel Networks with Low Degree Round Functions

When the round functions in a Feistel Network have low degree, the degree deficiency is decreasing slowly and as a result impossible monomials may exist for more than 5 rounds. Moreover, since there are less unknowns to recover, we need less impossible monomials to mount the attack.

In the following theorem we give a lower bound on the maximum number of Feistel rounds for which the large class of monomials is impossible. Namely, this class is point 1 from Theorem 3. The size of the class is 2^n, which is enough to recover a round function of full degree. Therefore this is the lower bound on maximum number of rounds that can be attacked using the ANF recovery technique from Section 5.2.

Theorem 4. *Let F be a $2n$-bit F_d^r with arbitrary functions and let its ANF be as in the Equation 3. Then $a^{F_i}_{u_l||u_r} = 0$ if $d^{r-2} < n, i \in R$ and $hw(u_l) = n$.*

Proof. Let $l||r$ be the input to F. Consider the degrees on the variables from l at the intermediate states. Initially, the degrees are 1 on the left and 0 on the right. After the first round the degrees are the same, because input to the round function has no variables from l. Now if we have the respective degrees d_1, d_2 at some point and we add a swap and xor with the round function, the degrees become $max(d_2, d \cdot d_1), d_1$. Then for 2 rounds the degrees are $d, 1$, for 3 rounds - d^2, d, and, in general, for r rounds the degrees are d^{r-1}, d^{r-2}. Therefore, when $d^{r-2} < n$, the r-round Feistel Network has no monomials with degree n on l in the right branch of the output.

As a corollary of the theorem, we can attack a $2n$-bit F_d^r if $d^{r-3} < n$. Note that for the 5-round Feistel with bijections which we attacked in the previous section this bound is not satisfied (for $n \geq 3$): $d^{5-3} = (n-1)^2 > n$, i.e. we attacked more rounds than we could attack by Theorem 4. Though we expect that the bound is tight for the specified class of monomials in FN with non-bijective round functions, there are another classes of impossible monomials for Feistel Networks with more rounds. Moreover, if the degree is low, there are less ANF coefficients to recover and, therefore, smaller classes of impossible monomials may be enough for attack. As an edge case, consider an additional round function of degree 1 (a linear function). The impossible monomials of degree $2n-1$ from Corollary 1 can be used to recover such round function, as was shown in attacks from Section 4.2. The maximal number of rounds (without the last linear one) for this attack is given by the condition $\theta(d, r) = d^{\lfloor r/2 \rfloor - 1} + d^{\lceil r/2 \rceil - 1} < 2n$ (or 1 more round if the Feistel functions are bijections). In general case, if the Feistel functions are bijections, we can attack 5 normal rounds plus 1 linear round.

6 Relationship Between Our Results and Other Attacks

6.1 Integral Attacks

The HDIM has a simple integral interpretation. Indeed, its coefficients correspond to the presence or not of some monomials in the ANF of its coordinates. They thus correspond to coefficients in said ANF which can be computed using the Möbius transform:

$$\hat{H}(F)[i, j] = \bigoplus_{x \preccurlyeq \overline{e_j}} F_i(x)$$

where $\overline{e_j}$ is the vector where all elements are equal to 1 except in position j. This has two consequences.

1. we can compute the HDIM of an n-bit permutation in time $\mathrm{O}(n2^{n-1})$, and
2. zeroes in column j imply the existence of an integral distinguisher.

In light of this, we state the following corollary of Corollary 1.

Corollary 2 (Integral Distinguisher for F_d^r). *Let F be a $2n$-bit F_d^r and suppose that one of the following conditions holds:*

- *the Feistel functions are bijections and $\theta(d, r-1) < 2n$, or*
- *the Feistel functions are not bijections and $\theta(d, r) < 2n$.*

Then there exists an integral distinguisher with data and time complexity 2^{2n-1} for this structure, namely

$$\bigoplus_{x \preccurlyeq \overline{e_j}} \big(e_i \cdot F(x)\big) = 0$$

for all $i < n$ and $j < n$. In other words, the sum of the right words of $F(x)$ is equal to 0 over a cube where one bit of the input right word is fixed to 0.

We notice a relation between our attacks and the so-called *division property*. This tool for finding integral attacks was introduced by Todo in [9] and later used by the same author to attack the full MISTY1 [21]. In his seminal paper, Todo gives some integral distinguishers against Feistel Network for various block sizes, number of rounds, degree of the Feistel functions for both bijective and non-bijective Feistel functions. Interestingly, his results are extremely similar to ours! Indeed, while there is no generic formula in Todo's paper, the application of his algorithm shows the existence of cubes of size $2n-1$ whose sum is equal to 0 for a number of rounds identical to the ones we predicted. In fact, results about the division property of the output of a Feistel Network can be extracted from its HDIM. To explain this, we first recall the definition of the division property.

Definition 7 (Division Property). *Let $\mathbb{X}$ be a multiset of $\mathbb{F}_2^n$ and k be an integer of $[0, n]$. We say that $\mathbb{X}$ has the division property $\mathcal{D}_k^n$ if, for all u in $\mathbb{F}_2^n$ such that $hw(u) \leq k$, $\bigoplus_{x \in \mathbb{X}} x^u = 0$.*

This property is further generalized into the *vectorial division property* which we define in the particular case of a Feistel Network.

Definition 8 (Vectorial Division Property (for Feistel Networks)). *Let $\mathbb{X}$ be a multiset of $(\mathbb{F}_2^n)^2$ and k^L, k^R be integers of $[0, n]$. We say that $\mathbb{X}$ has the collective division property $\mathcal{D}_{(k^L,k^R)}^n$ if, for all u, v in $\mathbb{F}_2^n$ such that $hw(u) \leq k^L$ and $hw(v) \leq k^R$, $\bigoplus_{(x,y) \in \mathbb{X}} x^u y^v = 0$.*

In particular, Todo applied his technique to $2n$-bit F_d^r. The integral distinguisher against the highest number of rounds correspond to integrals over cubes of size $2n-1$ were the constant bit has to be on the left side.[6] As we have seen, summing over such a cube is equivalent to computing half of the lines of the HDIM of the function.

Let F be a $2n$-bit F_d^r, x denote the left input bits, y denote the right ones and F_L and F_R denote its left and right output halves so that $F(x||y) = F_L(x||y)||F_R(x||y)$. Suppose that the top left corner of the HDIM of F is all zero. We deduce that the following holds for any cube $\mathcal{C}_k$ of dimension $2n-1$ where the bit at index $k \leq n$ is fixed and for any $i \leq n$: $\bigoplus_{x \in \mathcal{C}_k} F(x) \cdot e_i(x) = 0$. This can also be written as

$$\bigoplus_{x \in \mathcal{C}_k} \left(F_L(x)\right)^{u_i} \left(F_R(x)\right)^0 = \bigoplus_{x \in \mathcal{C}_k} \left(F_L(x)\right)^{u_i} = 0,$$

where u_i is the element of $\mathbb{F}_2^n$ equal to 0 except at position i where it is equal to 1. In other words, for all u in $\mathbb{F}_2^n$, $\mathrm{hw}(u) \leq 1$ implies that $\bigoplus_{x \in \mathcal{C}_k} \left(F_L(x)\right)^u = 0$, which means that the image of $\mathcal{C}_k$ has vectorial division property $\mathcal{D}_{1,0}^n$. The HDIM of Feistel Networks can thus be interpreted as describing the vectorial division property of each output half!

The relation between the ANF and integral attacks is further stressed by the attack we described in Section 5. Indeed, the complexity of this attack is very similar to that of the integral attack against 5-round FN with bijective Feistel functions described in [5].

7 Conclusion

Investigating surprising visual patterns in the LAT of Feistel Network lead us to interesting results. To explain them, we introduced the high-degree indicator matrix (HDIM). It causes a form of linearity of the LAT modulo 4 and is related to the presence (or lack thereof) of some monomials in the ANF of the permutation. We identified patterns in the distribution of these monomials for Feistel Networks and provided theorems allowing us to predict the existence of these patterns (Theorem 2 and Corollary 1). More generally, we showed how the predictable absence of some monomials can be leveraged to attack a Feistel Network in an impossible monomial attack. We also drew some connections between our results and integral distinguisher.

8 Acknowledgment

We thank Alex Biryukov and Dmitry Khovratovich for helpful discussions. We also thank the anonymous reviewers for their helpful comments. The work of Léo Perrin is supported by the CORE ACRYPT project (ID C12-15-4009992) funded by the *Fonds National de la Recherche* (Luxembourg). The work of Aleksei Udovenko is supported by the *Fonds National de la Recherche*, Luxembourg (project reference 9037104).

[6] It is actually on the right side in Todo's paper. Unlike in our paper, the Feistel function is XORed in the right branch in his case.

References

1. Biryukov, A., Khovratovich, D.: Decomposition attack on SASASASAS. IACR Cryptology ePrint Archive **2015** (2015) 646
2. Biryukov, A., Bouillaguet, C., Khovratovich, D.: Cryptographic schemes based on the ASASA structure: Black-box, white-box, and public-key (extended abstract). In: Advances in Cryptology - ASIACRYPT 2014 - Proceedings, Part I. (2014) 63–84
3. Dinur, I., Dunkelman, O., Kranz, T., Leander, G.: Decomposing the ASASA Block Cipher Construction. Cryptology ePrint Archive, Report 2015/507 (2015) http://eprint.iacr.org/.
4. Minaud, B., Derbez, P., Fouque, P.A., Karpman, P.: Key-Recovery Attacks on ASASA. In Iwata, T., Cheon, J.H., eds.: Advances in Cryptology - ASIACRYPT 2015. Volume 8270 of Lecture Notes in Computer Science. Springer Berlin Heidelberg (2015) To appear
5. Biryukov, A., Leurent, G., Perrin, L.: Cryptanalysis of Feistel Networks with Secret Round Functions. In Dunkelman, O., Keliher, L., eds.: Selected Areas in Cryptography – SAC 2015. Lecture Notes in Computer Science. Springer International Publishing (2015) To appear
6. Dinur, I., Dunkelman, O., Keller, N., Shamir, A.: New Attacks on Feistel Structures with Improved Memory Complexities. In: Advances in Cryptology – CRYPTO 2015. Lecture Notes in Computer Science. Springer Berlin Heidelberg (2015) (to appear)
7. Canteaut, A., Duval, S., Leurent, G.: Construction of Lightweight S-Boxes using Feistel and MISTY structures (Full Version). Cryptology ePrint Archive, Report 2015/711 (2015) http://eprint.iacr.org/.
8. Biryukov, A., Perrin, L., Udovenko, A.: Reverse-Engineering the S-Box of Streebog, Kuznyechik and STRIBOBr1. In: Advances in Cryptology-Eurocrypt 2016, Springer (2016) To appear
9. Todo, Y.: Structural evaluation by generalized integral property. In: Advances in Cryptology–EUROCRYPT 2015. Springer (2015) 287–314
10. Knudsen, L.R.: Deal-a 128-bit block cipher, aes submission (1998)
11. Patarin, J.: Generic attacks on feistel schemes. Cryptology ePrint Archive, Report 2008/036 (2008) http://eprint.iacr.org/.
12. Biryukov, A., Perrin, L.: On Reverse-Engineering S-Boxes with Hidden Design Criteria or Structure. In Gennaro, R., Robshaw, M., eds.: Advances in Cryptology – CRYPTO 2015. Volume 9215 of Lecture Notes in Computer Science. Springer Berlin Heidelberg (2015) 116–140
13. Carlet, C.: Boolean functions for cryptography and error correcting codes. Boolean Models and Methods in Mathematics, Computer Science, and Engineering **2** (2010) 257–397
14. Wu, W., Zhang, L.: LBlock: a lightweight block cipher. In: Applied Cryptography and Network Security, Springer (2011) 327–344
15. Matsui, M.: New block encryption algorithm MISTY. In: Fast Software Encryption, Springer (1997) 54–68
16. Beaulieu, R., Shors, D., Smith, J., Treatman-Clark, S., Weeks, B., Wingers, L.: The SIMON and SPECK Families of Lightweight Block Ciphers. IACR Cryptology ePrint Archive **2013** (2013) 404
17. U.S. DEPARTMENT OF COMMERCE/National Institute of Standards and Technology: Data encryption standard. Federal Information Processing Standards Publication (1999)

18. Biryukov, A., Shamir, A.: Structural cryptanalysis of SASAS. In Pfitzmann, B., ed.: Advances in Cryptology – EUROCRYPT 2001. Volume 2045 of Lecture Notes in Computer Science. Springer Berlin Heidelberg (2001) 395–405
19. ETSI/Sage: Specification of the 3GPP Confidentiality and Integrity Algorithms 128-EEA3 & 128-EIA3. Document 4 : Design and Evaluation Report (http://www.gsma.com/aboutus/wp-content/uploads/2014/12/EEA3_EIA3_Design_Evaluation_v2_0.pdf). Technical report, ETSI/Sage (September 2011)
20. The Sage Developers: Sage Mathematics Software (Version 6.8). (2015) http://www.sagemath.org.
21. Todo, Y.: Integral Cryptanalysis on Full MISTY1. In: Advances in Cryptology–CRYPTO 2015. Springer (2015) 413–432