IACR Transactions on Symmetric Cryptology
ISSN 2519-173X, Vol. 2023, No. 4, pp. 1–27. DOI:10.46586/tosc.v2023.i4.1-27

Multidimensional Linear Cryptanalysis of Feistel Ciphers

Betül Aşkın Özdemir[1], Tim Beyne[1] and Vincent Rijmen[1,2]

[1] COSIC, KU Leuven, Leuven, Belgium
name.lastname@esat.kuleuven.be
[2] University of Bergen, Bergen, Norway

Abstract. This paper presents new generic attacks on Feistel ciphers that incorporate the key addition at the input of the non-invertible round function only. This feature leads to a specific vulnerability that can be exploited using multidimensional linear cryptanalysis. More specifically, our approach involves using key-independent linear trails so that the distribution of a combination of the plaintext and ciphertext can be computed. This makes it possible to use the likelihood-ratio test as opposed to the χ^2 test. We provide theoretical estimates of the cost of our generic attacks and verify these experimentally by applying the attacks to CAST-128 and LOKI91. The theoretical and experimental findings demonstrate that the proposed attacks lead to significant reductions in data-complexity in several interesting cases.

Keywords: Multidimensional linear cryptanalysis · Likelihood-ratio test · Generic attack · Feistel ciphers · CAST-128 · LOKI91

1 Introduction

Linear cryptanalysis, proposed by Matsui [Mat94], is one of the most general methods in cryptanalysis. It relies on probabilistic linear approximations, which are linear relations between a cipher's input and output bits. A traditional linear key recovery attack recovers information about a cipher's key using only one linear approximation. Matsui proposed two methods: Algorithm 1 and Algorithm 2. Algorithm 1 extracts one bit of information about the secret key, whereas Algorithm 2 can recover multiple key bits.

Matsui's attack was extended significantly by later work. One improvement is to use many linear approximations simultaneously, known as multiple linear cryptanalysis. Kaliski and Robshaw [KR94] used several linear approximations to reduce the data-complexity of Matsui's algorithms. Biryukov, De Cannière, and Quisquater [BDCQ04] later analyzed the complexity of extracting several key bits with generalized versions of Algorithm 1 and Algorithm 2. Hermelin et al. [HCN08, HCN09, HCN19] introduced multidimensional linear cryptanalysis as a special case of multiple linear cryptanalysis where the masks form a vector space. The latter property leads to an equivalent description of the distinguisher in terms of the non-uniformity of the probability distribution of a combination of the plaintext and ciphertext.

One difficulty in multidimensional linear cryptanalysis is the key-dependence of the signs of trail correlations, leading to incomplete knowledge about the aforementioned probability distribution. A potential solution is to guess the correlation signs, but this often requires guessing too many key bits. Alternatively, Vaudenay proposed the Chi-squared (χ^2) test [Vau96] to test for non-uniformity instead of a specific distribution. In the case of known signs and therefore known probability distributions, the likelihood-ratio test [BJV04] was proposed as an optimal distinguisher. Indeed, compared to the case with

Received: 2023-09-01 Accepted: 2023-11-01 Published: 2023-12-08

unknown signs, the data-complexity in the case with known signs is lower by a factor equal to the square root of the number of approximations.

The Feistel structure is a well-known and widely analyzed design pattern for block ciphers. Hence, there is considerable interest in determining the minimal number of rounds to ensure the security of a Feistel cipher. Patarin proposed generic attacks on Feistel ciphers in a series of papers [Pat92, Pat01, Pat04] and these results were summarized in 2017 by Nachef et al. in [NPV17]. Several papers also focus on special cases of the generic Feistel construction. In this paper, we focus on Feistel ciphers that incorporate key addition at the input of the round function, as shown in Figure 1. Isobe and Shibutani [IS13a, IS13b] propose generic key-recovery attacks using an enhanced meet-in-the-middle approach on such Feistel ciphers.

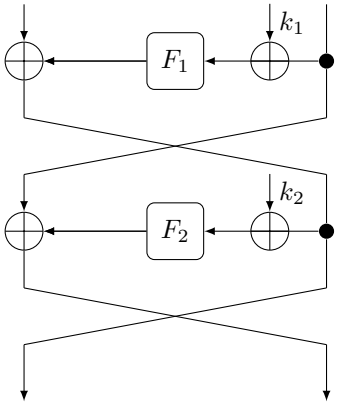

Figure 1: Two-round Feistel cipher with key additions at the input of the round function.

The block ciphers CAST-128 and LOKI91 are examples of Feistel ciphers that follow the structure shown in Figure 1. They have been extensively analyzed, as seen in Table 2 and Table 3. Isobe and Shibutani [IS13b, IS13a] applied their generic attacks to 7 and 8 rounds of CAST128. The first linear attack on reduced-round CAST-128 was proposed by Nakahara et al. [JR07]. With 2^{37} known plaintexts, their attack recovers 37 subkey bits for 4-round CAST-128. Wang et al. [WWH09] then gave a known-plaintext attack on 6-round CAST-128 using $2^{53.96}$ plaintexts as well as a ciphertext-only attack on 4-round CAST-128 using $2^{33.38}$ ciphertexts. The first linear cryptanalysis on LOKI91 was carried out by Tokita et al. [TSM95]. The authors evaluated reduced-round versions of LOKI91 and retrieved 13 key bits for 4-round LOKI91 using 2^{23} known plaintexts. Knudsen et al. improved these results by using multiple non-linear approximations [KR96]. Their analysis showed that seven more bits can be retrieved using less than a quarter of the plaintexts required in [TSM95]. Lastly, Sakurai et al. [SF97] introduced a technique to enhance the linear cryptanalysis method and applied this technique to improve existing attacks on LOKI91. According to [SF97], breaking 4-round LOKI91 with $2^{18.51}$ known plaintexts is possible. In previous works, other reduced round versions of LOKI91 were also attacked, as shown in Table 3.

Contribution In this work, we use key-independent multidimensional linear approximations to reduce the data-complexity of linear attacks on Feistel ciphers that add round keys before the round function as in Figure 1. Our approach leads to new generic attacks, as well as new concrete attacks on the ciphers CAST-128 and LOKI91.

The multidimensional linear property is turned into a distinguisher using the likelihood-ratio test. We also investigate the χ^2 test to highlight the difference between both methods. For $r \geq 2$ rounds, the data-complexities of the generic χ^2 and likelihood-ratio

Table 1: Comparison of generic attacks on r-round Feistel ciphers of the type shown in Figure 1 with block size $2n$ and key length k. All work efforts and data-complexities are given up to constant factors. All distinguishers are in the known-plaintext model.

Types	k	r	Data	Ref.	Work effort
Distinguisher†		2	$2^{n/2}$ 1	[NPV17] Section 3	$2^{n/2}$ 1
		3	$2^{n/2}$ 1	[NPV17] Section 3	$2^{n/2}$ 1
		4	2^{n} 2^{n}	[NPV17] Section 3	2^{n} 2^{n}
		5	$2^{3n/2}$ 2^{n}	[NPV17] Section 3	$2^{3n/2}$ 2^{n}
		6	2^{2n} 2^{2n}	[NPV17] Section 3	2^{2n} 2^{2n}
		7	2^{3n} 2^{2n}	[NPV17] Section 3	2^{3n} 2^{2n}
Key recovery	$2n$	4	1 1	[IS13b] Section 5.1	2^{n} 2^{n}
		5	$2^{n/2}$	[IS13b]	2^{n}
		6	2^{n} 2^{n}	[IS13a] Section 5.1	$2^{3n/2}$ $n2^{n}$
	$4n$	4	1 1	[IS13b] Section 5.1	2^{n} 2^{n}
		5	$2^{n/2}$ 1	[IS13b] Section 5.1	2^{n} 2^{2n}
		6	1 2^{n}	Section 5.1 Section 5.1	2^{3n} $n2^{n}$
		7	2^{n} 2^{n}	[IS13a] Section 5.1	$2^{3n/2}$ $n2^{2n}$
		8	1 2^{n}	[IS13b] Section 5.1	2^{3n} $n2^{3n}$
		9	$2^{5n/6}$	[IS13b]	2^{3n}

† Only distinguishers in the known-plaintext model are listed.

Table 2: Key recovery results of r-round CAST-128.

r	P_S	Data	Ref.	Work effort
3	$1^{\dagger}$	2^{37}	[JR07]	2^{91}
4	$1^{\dagger}$	$2^{33.38}$	[WWH09]	2^{91}
	1	$2^{5.80}$	Section 5.2	2^{37}
5	1	$2^{6.80}$	Section 5.2	2^{74}
6	$1^{\dagger}$	$2^{53.96}$	[WWH09]	2^{91}
	1	$2^{7.39}$	Section 5.2	$2^{118.39}$
	1	2^{35}	Section 5.2	2^{40}
7	1	6	[IS13a]	2^{114}
	1	2^{35}	Section 5.2	2^{77}
8	1	8	[IS13b]	2^{118}
	1	2^{35}	Section 5.2	2^{114}

† Derived from Matsui's estimate [Mat94].

Table 3: Key recovery results of r-round LOKI91.

r	P_S	Data	Ref.	Work effort
4	0.90	$2^{21.77}$	[KR96]	2^{49}
	0.86	$2^{20.86}$	[KR96]	2^{44}
	$\ddagger$	$2^{18.51}$	[SF97]	2^{40}
	$1^{\dagger}$	$2^{19.60}$	[TSM95]	2^{51}
	1	$2^{5.55}$	Section 5.2	2^{37}
5	$1^{\dagger}$	$2^{23.20}$	[TSM95]	2^{51}
6	$1^{\dagger}$	$2^{31.70}$	[TSM95]	2^{51}
	$\ddagger$	$2^{27.58}$	[SF97]	2^{40}
	0.90	2^{23}	Section 5.2	2^{37}
7	$1^{\dagger}$	$2^{40.10}$	[TSM95]	2^{51}
	0.90^{*}	$2^{39.21}$	[KR96]	2^{51}
	0.86^{*}	$2^{37.88}$	[KR96]	2^{44}
	$\ddagger$	$2^{36.67}$	[SF97]	2^{40}
9	$1^{\dagger}$	$2^{49.80}$	[TSM95]	2^{51}
	$\ddagger$	$2^{45.74}$	[SF97]	$2^{45.74}$
10	$1^{\dagger}$	$2^{58.30}$	[TSM95]	2^{51}
	0.90^{*}	$2^{57.21}$	[KR96]	2^{51}
	0.86^{*}	$2^{55.88}$	[KR96]	2^{44}
	$\ddagger$	$2^{54.83}$	[SF97]	$2^{54.83}$

* Extrapolated from experimental results.
† Derived from Matsui's estimate [Mat94].
‡ Not explicitly provided.

distinguishers are $\mathcal{O}(2^{n\lfloor r/2\rfloor - n/2})$ and $\mathcal{O}(2^{n\lfloor r/2\rfloor - n})$ respectively, with $2n$ the block size. The data-complexities obtained using the likelihood-ratio test are comparable to the best-known attacks on generic Feistel ciphers, which are also the best-known results on Feistel ciphers of the type shown in Figure 1 without guessing round keys[1]. However, our approach yields known-plaintext distinguishers, whereas previous work relies on chosen plaintexts or adaptively chosen ciphertexts. This is particularly useful when extending these distinguishers to key-recovery attacks. As shown in Table 1, our distinguishers improve over previous work in the known-plaintext model.

Moreover, we propose generic key-recovery attacks that combine the multidimensional linear distinguisher with two key-recovery methods: the naive approach and a variant of the Fast-Fourier transform method [CSQ07]. As can be seen in Table 1, our key-recovery attacks improve over previous work in some – but not all – cases. In particular, when the key is more than twice as long as the block length, the meet-in-the-middle attacks of [IS13a, IS13b] usually perform better. However, our approach is beneficial for shorter key lengths and leads to the best-known attack on six rounds.

Finally, we apply our attacks to CAST-128 and LOKI91 and support our analysis with experimental results. Key-recovery results are presented in Table 2 and Table 3. Our findings show that we can distinguish 3- and 5-round CAST-128 with improved data and time complexities compared to prior dedicated attacks. Indeed, the likelihood-ratio distinguisher for 5-round CAST-128 requires fewer plaintexts than the dedicated 3-round distinguisher presented in [JR07]. Moreover, we enhanced the time and data-complexities of existing key recovery attacks on 4-, 5- and 6-round CAST-128. For 7- and 8-round CAST-128, we also have an improvement in time complexities. Note that this is despite the fact that the meet-in-the-middle attack from [IS13a, IS13b] is generically better for these parameters. The reason is that each round of CAST-128 involves five additional key bits that control a secret rotation. Previous studies on LOKI91 have primarily focused on key recovery. Our key recovery results demonstrate a notable improvement over the previous findings on 4- and 6-round LOKI91 regarding data and time complexities. However, as the number of rounds increases, better linear attacks on LOKI91 are available. This is because our generic linear trails are suboptimal for many rounds of LOKI91.

2 Preliminaries

Standard linear cryptanalysis relies on probabilistic linear approximations, which are linear relations between a cipher's input and output bits. In linear cryptanalysis, one attempts to find appropriate linear approximations with high absolute correlations. A vectorial Boolean function $F : \mathbb{F}_2^m \to \mathbb{F}_2^n$ maps a binary vector of length m to a vector of length n. A function $F : \mathbb{F}_2^n \to \mathbb{F}_2^n$ is called iterative if there are r vectorial Boolean functions such that $F = F_r \circ F_{r-1} \circ \cdots \circ F_1$. The correlation matrix of the function F_i is denoted by C^{F_i} [DGV95], which is a $2^n \times 2^n$ matrix with coordinates

$$C^{F_i}_{v,u} = \mathsf{C}(v^T F_i(x) + u^T x) = \frac{1}{2^n} \sum_{x \in \mathbb{F}_2^n} (-1)^{v^T F_i(x) + u^T x} \tag{1}$$

for each input mask $u \in \mathbb{F}_2^n$ and output mask $v \in \mathbb{F}_2^n$.

To analyze the linear approximations of iterative functions, we need to specify the input and output mask for each round function: a tuple of $r+1$ masks α_i such that $(\alpha_1, \alpha_2, \cdots, \alpha_{r+1})$, is called a linear trail. Hence, the i^{th} round function has the input mask α_i and output mask α_{i+1}. Then, the correlation of the trail can be obtained by

[1] This can be formalized by allowing the key to be randomized between queries.

multiplying correlations of the individual functions F_i:

$$C^F = \prod_{i=1}^{r} C^{F_i}. \tag{2}$$

The correlation of a linear approximation can be obtained by summing the correlations of all linear trails that share the same input and output masks as the approximation

$$C^F_{\alpha_{r+1},\alpha_1} = \sum_{\alpha_2,\cdots,\alpha_r} \prod_{i=1}^{r} C^{F_i}_{\alpha_{i+1},\alpha_i}. \tag{3}$$

Thus, the piling-up lemma [Mat94] is valid when a single trail is dominant, which means the correlation is approximately equal to the correlation of the dominant trail.

In the case of the key-alternating cipher, the round functions are parameterized by round keys such that $F_i(x) = G(x) \oplus k_i$. Then, we have

$$C^F_{\alpha_{r+1},\alpha_1} = \sum_{\alpha_2,\ldots,\alpha_r} \prod_{i=1}^{r} (-1)^{\alpha_{i+1}^T k_i} C^G_{\alpha_{i+1},\alpha_i}. \tag{4}$$

In other words, the key only affects the signs of the trail correlations. Multidimensional linear cryptanalysis involves multiple linear approximations whose respective masks collectively form a vector space. Multidimensional linear approximations hold particular significance as they are associated with distinguishers that rely on the probability distribution of linear projections of plaintext and ciphertext. The key theoretical consequence of multidimensional linear cryptanalysis is Theorem 1, known as the Poisson summation formula in Fourier analysis. In [HCN08], authors discuss the idea of Theorem 1, but it is presented in a different manner that requires a change-of-basis.

Theorem 1 (Theorem 4.1. [Bey21])**.** *Let* $\mathbf{z}$ *be a random variable on* $\mathbb{F}_2^d$*,* $V \subseteq \mathbb{F}_2^d$ *a vector space and* $V^\perp$ *an orthogonal complement of the subspace* V*. For any* $\eta \in \mathbb{F}_2^d$ *it holds that*

$$p(\eta) = \Pr(\mathbf{z} = \eta \bmod V^\perp) = \frac{1}{|V|} \sum_{v \in V} (-1)^{v^T \eta} \mathsf{C}(v^T \mathbf{z})$$

where $\mathsf{C}(v^T\mathbf{z}) = 2\Pr[v^T\mathbf{z} = 0] - 1$ is the correlation of $v^T\mathbf{z}$.

2.1 Statistical hypothesis testing

A statistical hypothesis test [NP92] aims to determine the extent to which the observed data support a given hypothesis. This involves formulating two opposing hypotheses, namely the null hypothesis and the alternative hypothesis:

$$\begin{aligned} \text{null hypothesis } H_0 &: \text{ data comes from the real primitive} \\ \text{alternative hypothesis } H_1 &: \text{ data comes from an ideal primitive} \end{aligned} \tag{5}$$

The acceptance region $\mathcal{A}$ is a range of test statistic values that support the null hypothesis. If the test statistic t_{real} (t when interacting with the real cipher) falls within this range, then the null hypothesis is accepted; otherwise, the null hypothesis is rejected in favor of the alternative hypothesis. The success probability P_{S} represents the proportion of correctly identified positive instances in a hypothesis test, while the false positive rate P_{F} is the fraction of negative cases inaccurately identified as positive. Adversaries aim to maximize Adv, which is defined as $\text{Adv} = |P_{\text{S}} - P_{\text{F}}|$.

In statistics, several hypothesis testing methods are available, each presenting unique strengths and weaknesses. The data-complexity for given P_{S} and P_{F} of a distinguishing attack depends on the statistical test employed, such as the χ^2 and the likelihood-ratio tests.

2.2 Multidimensional linear distinguisher

The objective of a cryptanalyst is to present a distinguisher and/or to retrieve the cryptographic key. The distinguisher operates by estimating the correlation of the approximation using q pairs of known plaintext and ciphertext. Given a function $F : \mathbb{F}_2^m \to \mathbb{F}_2^n$, Theorem 1 can be applied to the case where $\mathbf{z} = (\mathbf{x}, F(\mathbf{x}))$ is a random variable, with $\mathbf{x}$ uniformly distributed on $\mathbb{F}_2^m$ and Λ a vector space of mask pairs. In other words, if we know the correlations $C_{v,u}^F$ for approximations $(v, u) \in \Lambda$ including their signs, we can fully determine the probability distribution of linear projections of the plaintext and ciphertext. Consequently, we have

$$\Pr\left((\mathbf{x}, F(\mathbf{x})) = (t_1, t_2) \bmod \Lambda^\perp\right) = \frac{1}{|\Lambda|} \sum_{(u,v)\in\Lambda} (-1)^{u^T t_1 + v^T t_2} C_{v,u}^F. \tag{6}$$

The required amount of data q for the distinguisher can be estimated, by choosing q large enough to ensure that the distributions of t_{real} and t_{ideal} exhibit a significant difference in means. For a successful attack, the data-complexity should be large enough to ensure the inequality $\mu_{\text{real}} - \mu_{\text{ideal}} \gg \sigma_{\text{ideal}}$. However, the values of μ and σ are subject to change based on the statistical test, leading to different data complexities. Evaluating the inequality for both known and unknown sign cases reveals that the data complexities of both cases are inversely proportional to the capacity of the multidimensional linear approximation Λ over the function $F : \mathbb{F}_2^m \to \mathbb{F}_2^n$, which is defined as $\mathsf{Cap}(\Lambda) = \sum_{(u,v)\in\Lambda} (C_{v,u}^F)^2$. In the case of unknown correlation signs, precise probability distribution estimation requires guessing. To mitigate this challenge, Vaudenay [Vau96] proposed using the χ^2 test to detect non-uniformity rather than a specific distribution. In this case, the rough data-complexity of the multidimensional distinguisher is $q \approx \sqrt{|\Lambda|}/\mathsf{Cap}(\Lambda)$. On the other hand, the exact probability distribution can be obtained if correlations are known. For this case, Vaudenay et al. [BJV04] have proposed the likelihood-ratio test as an optimal distinguisher, effectively employed when two distributions are known. The rough data-complexity of the known sign case is $q \approx 1/\mathsf{Cap}(\Lambda)$. Thus, the data-complexity in the case with known signs is lower by a factor equal to the square root of the number of approximations of the data-complexity of the case with unknown signs.

In the standard linear cryptanalysis, the data-complexity is inversely proportional to the squared correlation of a single approximation. Therefore, the known sign case of multidimensional linear cryptanalysis implies a significant reduction in data-complexity compared to the unknown sign case and standard linear cryptanalysis.

3 Generic multidimensional linear distinguishers

This section presents the multidimensional linear distinguisher that the key-recovery attacks in Section 5 rely on. As discussed in the introduction, our attacks are applicable to Feistel ciphers with round functions of the form $x \mapsto F_i(x + k_i)$. The analysis in this section is generic, *i.e.* we assume that the public functions $F_1, F_2, \ldots, F_r$ have been chosen independently and uniformly at random. Particular aspects for the concrete ciphers CAST-128 and LOKI91 will be investigated in Section 4.

Equation (4) shows that the signs of the correlations of linear trails in a key-alternating cipher with independent round keys all depend on the key. However, as illustrated by the linear trail in Figure 2, trails with a key-independent correlation can exist in Feistel ciphers even if round keys are independent. Although the existence of such trails is of little consequence in ordinary linear cryptanalysis, it can be exploited to significantly reduce the data-complexity of multidimensional linear attacks.

To determine the correlation of the iterative trail in Figure 2, we use the following result due to Daemen and Rijmen [DR07]. The formulation in Theorem 2 is due to [Bey21].

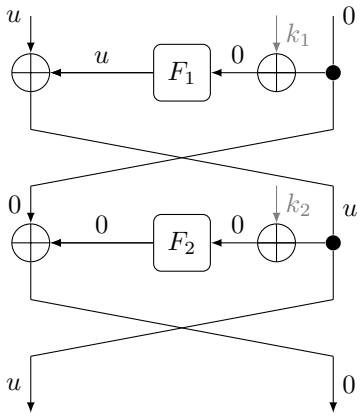

Figure 2: Two-round linear trail for a Feistel cipher with key-independent correlation.

Theorem 2 (Theorem 3.1. [Bey21])**.** *Let $\boldsymbol{c}$ be the correlation of a nontrivial linear approximation for a uniform random function $F : \mathbb{F}_2^n \to \mathbb{F}_2^n$. The random variable $2^{n-1}(\boldsymbol{c}+1)$ follows a binomial distribution with mean 2^{n-1} and variance 2^{n-2}. In particular, as n approaches infinity, the distribution of $2^{n/2}\boldsymbol{c}$ converges to the standard normal distribution $\mathcal{N}(0, 1)$.*

Let $\mathbf{c}_i$ be the correlation of the linear approximation with input mask zero and output mask u over the (uniformly random) function F_{2i-1}. By Theorem 2, this is a random variable with an asymptotically normal distribution centered around zero and with standard deviation $2^{-n/2}$. Since the trail in Figure 2 only activates the functions in odd-numbered rounds, its correlation for $r \geq 2$ rounds is equal to

$$\mathbf{c}_u = \prod_{i=1}^{\lfloor r/2 \rfloor} \mathbf{c}_i \,.$$

Note that, in the expression above, we round down $r/2$ since if r is odd, one can choose the input mask equal to $(0, u)$ to skip the first round. This is also a good approximation for correlation of the corresponding r-round linear approximation, since the trail in Figure 2 is dominant. Since the random variables are independent by the strong assumption that the random functions $F_1, \ldots, F_r$ are independent, the variance of $\mathbf{c}_u$ is equal to $\mathbb{E}\big(\mathbf{c}_u^2\big) = 2^{-\lfloor r/2 \rfloor n}$.

A multidimensional linear distinguisher can be set up by simultaneously working with all possible masks u. Specifically, the vector space of input-output mask pairs is equal to

$$\Lambda = \Big\{ \big((u, 0), (u, 0)\big) \mid u \in \mathbb{F}_2^n \Big\} \,.$$

As explained in Section 2.2, the correlations of the linear approximations in Λ completely determine the distribution of $(\mathbf{x}, F(\mathbf{x})) \mod \Lambda^\perp$ with uniform random $\mathbf{x}$ and F the r-round Feistel cipher. Note that for odd r, the value $(\mathbf{x}, F(\mathbf{x})) \mod \Lambda^\perp$ is equal to the sum of the right half of the plaintext $\mathbf{x}$ and the left half of the ciphertext $F(\mathbf{x})$. To implement this multidimensional distinguisher, one can estimate the empirical distribution of $(\mathbf{x}, F(\mathbf{x}))$ mod $\Lambda^\perp$ by sampling plaintext-ciphertext pairs and observing the number of occurrences for each value of the sum of the left halves (assuming r is even) of the plaintext and ciphertext.

As discussed in Section 2.2, the data-complexity of multidimensional linear distinguishers depends on the sum of the squared correlations of the approximations in Λ. The same is true in particular for the distinguishers we present in Sections 3.1 and 3.2. The sum

of the squared correlations is well approximated by the sum of the squares of the trail correlations, which can be calculated in time $\mathcal{O}(n2^n)$ given the round functions $F_1, \ldots, F_r$. Additional details will be given in Section 4. In the generic case, the sum of squared correlation is a random variable. However, since its distribution is strongly concentrated around the mean, its average is a good estimate:

$$\mathbb{E}\left(\sum_{u\in\mathbb{F}_2^n\setminus\{0\}} \mathbf{c}_u^2\right) = 2^n\mathbb{E}\left(\mathbf{c}_{u^\star}^2\right) = 2^{n-\lfloor\frac{r}{2}\rfloor n}\,,$$

where $u^\star$ is an arbitrary nonzero mask.

The following two sections show how the χ^2- and likelihood-ratio tests can be used to detect the multidimensional linear approximation with mask space Λ. The χ^2 approach is only discussed for background and comparison because the choice of linear approximations implies that we know the signs of all correlations. This makes it possible to use the more efficient likelihood-ratio test.

3.1 Distinguisher using the χ^2 test

The χ^2 test is a frequently employed statistical hypothesis test that compares an empirical distribution with an ideal distribution using the sum of the squared differences between the observed and expected frequencies divided by the expected frequencies. In our case, the ideal distribution is the uniform distribution, and the empirical distribution is that of the sum of the left halves of the plaintext and ciphertext for uniform random plaintext-ciphertext pairs.

Let (x_1^i, x_2^i) and (y_1^i, y_2^i) be q plaintext and corresponding ciphertext samples respectively, for $i \in \{1, 2, \ldots, q\}$. Store the number of occurrences of each of the 2^n possible values of $x_2^i + y_2^i$ (assuming r is odd) in a table T. The χ^2 statistic is then calculated by

$$\chi^2 = \sum_{i=1}^{2^n} \frac{\left(T[x_2^i + y_2^i] - q/2^n\right)^2}{q/2^n}\,.$$

Pseudo-code to calculate the χ^2 statistic can be found in Appendix C. One can show that (see Appendix B) the χ^2 statistic is closely related to the number of collisions N_{coll} between the observed values $x_2^i + y_2^i$. Hence, counting these collisions provides an alternative approach to implementing the χ^2 test. Furthermore, up to a constant factor, the average number of collisions equals the sum of the squared correlations of the linear approximations in Λ. This leads to Proposition 1.

Proposition 1. *Given q samples $z_1, \ldots, z_q$ in $\mathbb{F}_2^n$, let N_{coll} be the number of unordered collisions between them. The χ^2 statistic (for uniform expected frequencies) computed using the values $z_1, \ldots, z_q$ is related to N_{coll} by*

$$\chi^2 = \frac{2^{n+1}}{q} N_{\text{coll}} + 2^n - q.$$

Furthermore, if the samples are independent and identically distributed, then

$$\mathbb{E}(\mathbf{N}_{\text{coll}}) = \frac{1}{2^n}\binom{q}{2} \sum_{u\in\mathbb{F}_2^n} \mathsf{C}(u^T\mathbf{z})^2\,,$$

where $\mathbf{z}$ is a random variable with distribution equal to the distribution of the sample.

Proof. See Appendix B.

In Section 3, it was shown that the sum of the squared correlations is equal to $1 + 2^{n-n\lfloor r/2 \rfloor}$. Hence, by Proposition 1, the multidimensional linear approximation leads to approximately $2^{-n\lfloor r/2 \rfloor} q^2/2$ additional collisions on average. The standard deviation of the number of collisions is approximately $q/\sqrt{2^n}$, so that the data-complexity of the multidimensional χ^2 distinguisher is roughly

$$q = \mathcal{O}\left(2^{n\lfloor r/2 \rfloor - n/2}\right).$$

Hence, the data-complexity is $2^{n/2}$ and $2^{3n/2}$ for 2 & 3- and 4 & 5- round distinguishers, respectively.

In attacks on concrete primitives, such as those in Section 4, finer estimates are desired. In Section 4, an experimental approach is used to estimate the success probability in terms of q. However, this still requires choosing the decision threshold t of the test given the desired false-positive rate. Let $\boldsymbol{\chi}^2_{\text{ideal}}$ be the random variable equal to the χ^2 statistic when the data is uniformly random. In an asymptotic sense, when 2^n and q both approach infinity, the distribution of $\boldsymbol{\chi}^2_{\text{ideal}}$ converges towards the normal distribution $\mathcal{N}(\nu, 2\nu)$ where $\nu = 2^n - 1$ is the number of degrees of freedom. Hence, the false-positive rate satisfies

$$P_{\text{F}} = \Pr\left[\boldsymbol{\chi}^2_{\text{ideal}} \geq t\right] = 1 - \Phi\left(\frac{t-\nu}{\sqrt{2\nu}}\right).$$

where $\Phi(\text{x})$ is the cumulative normal distribution function. Hence, the threshold equals

$$t = (2^n - 1) + \sqrt{2(2^n - 1)}\,\Phi^{-1}(1 - P_{\text{F}}). \tag{7}$$

Once the threshold value t is determined for fixed P_{F}, the success probability can be computed experimentally for a given data-complexity by counting the number of times the χ^2_{real} statistic (χ^2 statistic when interacting with the real cipher) exceeds t.

3.2 Distinguisher using the likelihood-ratio test

The Neyman-Pearson lemma [NP92] states that the uniformly most powerful statistical method for distinguishing between two hypotheses is the likelihood-ratio test. Since the correlations of the trails in Figure 2 are key-independent, the probability distribution of the sum of the left half of the plaintext and ciphertext can be fully determined using Theorem 1. This makes it possible to use the likelihood-ratio test.

Let p_0 be the distribution of the sum of the right half of the plaintext and ciphertext for the real cipher with uniform random input, and let p_1 be the uniform distribution on $\mathbb{F}_2^n$. Given samples $z_1, \ldots, z_q$ defined by $z_i = x_2^i + y_2^i$, the logarithmic likelihood-ratio (LLR) statistic is equal to

$$\lambda = \log \prod_{i=1}^{q} \frac{p_{\text{real}}(z_i)}{p_{\text{ideal}}(z_i)} = \sum_{i=1}^{q} \log \frac{p_{\text{real}}(z_i)}{p_{\text{ideal}}(z_i)},$$

when testing to reject the null hypothesis (interaction with the real cipher). Swapping p_{real} and p_{ideal} in the formula above yields the statistic for testing to reject the alternative hypothesis (equal to $-\lambda$). Note that this form of the test assumes that samples are independent. Pseudo-code to calculate the LLR test statistic can be found in Appendix C.

As mentioned in Section 2.2, the asymptotic data-complexity of the LLR test is inversely proportional to the sum of squared correlations. That is,

$$q = \mathcal{O}\left(2^{\lfloor r/2 \rfloor n - n}\right)$$

Thus, the data-complexities are constant, 2^n and 2^{2n} for 2 & 3, 4 & 5 and 6 & 7 rounds. Note that this is a factor $2^{n/2}$ lower than for the χ^2 test.

For the 3-round distinguisher, it is likely that when interacting with a random permutation, values of z_i such that $p_{\text{real}}(z_i) = 0$ will be observed. In this case, the test statistic is undefined — although formally, it could be taken to equal $-\infty$ and the test would give the correct result. In fact, this case leads to an interesting simplification of the distinguisher. Since zero likelihoods signify that one is not interacting with the real cipher, a decision can be made immediately. More specifically, we can define two cases:

1. If $p_{\text{real}}(z_i) = 0$ for some i, we immediately decide not to interact with the real cipher (reject the null hypothesis).

2. If $p_{\text{real}}(z_i) \neq 0$ for all i, then compare λ to a threshold value t to reach a decision. If it does not exceed the threshold, reject the null hypothesis.

For concrete attacks, finer estimates of the data-complexity are necessary. Hence, below we provide such estimates for the above procedure. In Section 4, experiments will be done to estimate the true success probability for a given false-positive rate and data-complexity. For the standard LLR test, estimates are given by Baignères, Junod, and Vaudenay [BJV04]. For the case-by-case approach, the false-positive rate equals

$$P_{\text{F}} = \Pr[p_{\text{real}}(\mathbf{z}_i) = 0 \text{ for some } i] P_{\text{F}_1} + \big(1 - \Pr[p_{\text{real}}(\mathbf{z}_i) = 0 \text{ for some } i]\big) P_{\text{F}_2} \tag{8}$$

where P_{F_1} and P_{F_2} are the conditional false-positive rates for the cases above. In particular, $P_{\text{F}_1} = 0$. If N_{zero} is the number values of z such that $p_{\text{real}}(z) = 0$, then Equation (8) implies

$$P_{\text{F}} \leq 1 - \Pr[p_{\text{real}}(\mathbf{z}_i) = 0 \text{ for some } i] = \left(1 - \frac{N_{\text{zero}}}{2^n}\right)^q .$$

For the conditional false-positive rate P_{F_2} of the second case, the results of [BJV04] can be applied. In particular, asymptotically the distribution of $\boldsymbol{\lambda}_{\text{real}}$ (the test-statistic based on random samples from the cipher) converges to $\mathcal{N}(\mu_{\text{real}}, \sigma^2)$ and the distribution of $\boldsymbol{\lambda}_{\text{ideal}}$ (the test-statistic based on uniform random samples) converges to $\mathcal{N}(\mu_{\text{ideal}}, \sigma^2)$, where

$$\mu_{\text{real}} = D_{\text{KL}}(p_{\text{real}} \parallel p_{\text{ideal}}) := \sum_{z \in \mathbb{F}_2^n} p_{\text{real}}(z) \log \frac{p_{\text{real}}(z)}{p_{\text{ideal}}(z)}$$

$$\begin{aligned} \mu_{\text{ideal}} &\approx -\mu_{\text{real}} \\ \sigma^2 &\approx 2\mu_{\text{real}} , \end{aligned}$$

where D_{KL} is the relative entropy or Kullback-Leibler divergence between p_{real} and p_{ideal}. Hence, the false-positive rate P_{F_2} satisfies

$$P_{\text{F}_2} = \Pr[\boldsymbol{\lambda}_{\text{ideal}} \geq t] \approx 1 - \Phi\left(\frac{t + \mu_{\text{real}}}{\sqrt{2\,\mu_{\text{real}}}}\right) .$$

Hence, the required threshold to achieve false-positive rate P_{F_2} equals

$$t = -\mu_{\text{real}} + \sqrt{2\mu_{\text{real}}}\,\Phi^{-1}(1 - P_{\text{F}}) . \tag{9}$$

Once the threshold value is calculated for fixed P_{F}, the success probability can be empirically estimated for a given data-complexity. Finally, it is worth noting that the success probability can be estimated as

$$P_{\text{S}} = \Pr[\boldsymbol{\lambda}_{\text{real}} \geq t] \approx 1 - \Phi\left(\frac{t - \mu_{\text{real}}}{\sqrt{2\,\mu_{\text{real}}}}\right) .$$

4 Application to CAST-128 and LOKI91

The Feistel ciphers CAST-128 and LOKI91 are designed to include a key addition at the round function's input, allowing for the application of our proposed method. In this section, we briefly introduce the target ciphers and present the experimental verification of our approach in detail.

4.1 CAST-128

CAST is a family of block ciphers. CAST-128 [Ada97] is a Feistel network with 64-bit blocks, and the key sizes range from 40 to 128 in 8-bit increments. The cipher employs 12 rounds for key sizes up to and including 80 bits, whereas for key sizes that exceed 80 bits, the cipher utilizes the complete 16 rounds. The algorithm employs three types of round functions F_1, F_2, and F_3 identical in structure but differ in the operations used: addition, subtraction, and XOR. For F_1, $\circledast$, $\odot$, and $\circledcirc$ in Figure 3 correspond $\oplus$, $\boxplus$, and $\boxminus$ respectively. For F_2 and F_3, the same operations are used in changing positions. The designers of the algorithm aimed to achieve efficiency in both software and hardware implementations, and it was later standardized by ISO/IEC [ISO].

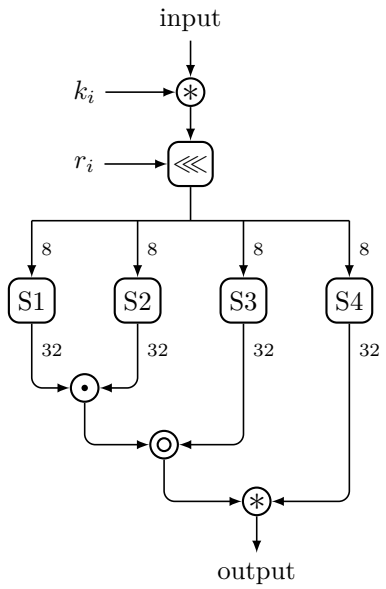

Figure 3: The round function of CAST-128 algorithm.

4.2 LOKI91

LOKI is a family of block ciphers that uses the Feistel network structure. LOKI91, a variant of the LOKI family, is designed with a block size and key size of 64 bits. The LOKI91 block cipher consists of 16 rounds that employ a substitution-permutation network for its round function, similar to the Data Encryption Standard (DES). The round function in LOKI91 receives a 32-bit branch and a 32-bit round key as input and performs a bitwise XOR operation between them. The resulting 32-bit value is expanded to 48 bits and divided into 12-bit blocks. These blocks are then processed by four S-boxes, with each S-box generating an 8-bit output. The round function output is obtained by permuting the 32-bit value.

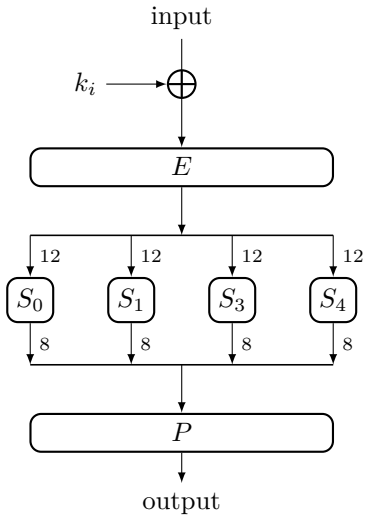

Figure 4: The round function of LOKI91.

4.3 Experimental verification and results

This section includes information regarding the χ^2 and the LLR distinguisher for reduced-round CAST-128 and LOKI91. Recall that the data-complexity of the r-round multidimensional distinguisher is proportional to $2^{n\lfloor r/2 \rfloor - n/2}$ using the χ^2 test and $2^{n\lfloor r/2 \rfloor - n}$ using the LLR test, for $r \geq 2$.

Table 4: Capacity and approximate data-complexity q for the r-round LLR distinguishers of CAST-128 and LOKI91.

Cipher	r	Round functions	Capacity	q
CAST-128	3	F_1	1	1
		F_2	1	1
		F_3	1	1
	5	F_1, F_2	2.3200×10^{-10}	$2^{31.998}$
		F_1, F_3	2.3290×10^{-10}	$2^{31.999}$
		F_2, F_3	2.3297×10^{-10}	$2^{31.999}$
	7	F_1, F_2, F_3	5.4308×10^{-20}	$2^{63.997}$
LOKI91	3	F	1	1
	5	F, F	2.8332×10^{-6}	$2^{21.434}$
	7	F, F, F	6.3678×10^{-11}	$2^{39.212}$
	9	F, F, F, F	9.1587×10^{-16}	$2^{56.346}$

For CAST-128 and LOKI91, the block size is 64-bit, so the expected data-complexities are a constant close to one, 2^{32} and 2^{64} for 3-, 5- and 7-round respectively. However, these estimates are based on the assumption that the round functions are sampled independently and uniformly at random. Hence, it is worthwhile to experimentally verify these estimates by calculating the capacity of the distinguisher.

To compute the sum of the squared correlations efficiently, we use the Fast-Fourier Transform. Firstly, all possible 2^n values are encrypted by the active round functions, and histograms of the output values of each function are stored in separate vectors. Then,

the Fast-Fourier Transform is applied to each vector. Finally, the sum of the squared correlations is calculated by taking the sum of the squares of the pointwise product of the obtained vectors. Note that the 5-round CAST-128 distinguisher includes two active functions: F_1 and F_2. Similarly, in the 7-round distinguisher, F_1, F_2, and F_3 are active functions. Since LOKI91 employs the same round function in each round, only one function has to be analyzed. Detailed calculations are given in Appendix A.

As shown in Table 4, experimental results show that capacities for the 3-, 5-, and 7-round CAST-128 distinguishers are roughly as expected: constant, $2^{-31.99}$, and $2^{-63.99}$, respectively. Hence, in this respect, CAST-128 round functions behave like random functions. For LOKI91, capacities for the 3-, 5-, 7-, and 9-round distinguishers are constant, $2^{-21.43}$, $2^{-39.21}$, and $2^{-56.34}$, respectively. Note that these values are significantly higher than for the generic case. The related pseudo-code is given in Appendix C.

More precise estimates of the data-complexity can be obtained experimentally in terms of the false-positive rate P_{F} and success probability P_{S}. To demonstrate the effectiveness of the likelihood ratio method, we also test the χ^2 distinguisher for three rounds. A summary of experimental data-complexities of the reduced-round multidimensional distinguishers for CAST-128 and LOKI91 is presented in Table 5.

Table 5: Data-complexities of multidimensional distinguishers of r-round CAST-128 and LOKI91 for $P_{\mathrm{S}} \approx 1$ and $P_{\mathrm{F}} \approx 0$.

Cipher	r	Type	Data
CAST-128	3	χ^2	2^{18}
		LLR	$2^{3.32}$
	5	LLR	2^{35}
LOKI91	3	χ^2	2^{18}
		LLR	$2^{3.80}$
	5	LLR	$2^{24.51}$

More detailed results are presented below.

3-round χ^2 distinguisher Based on Table 4 and Section 3.1, the data-complexity of the 3-round χ^2 distinguishers on CAST-128 and LOKI91 is close to 2^{16}. To experimentally determine the success probability for a given false-positive rate, we determine the decision threshold based on Equation (7):

$$t = (2^{32} - 1) + \sqrt{2(2^{32} - 1)}\Phi^{-1}(1 - P_{\mathrm{F}}).$$

After obtaining the threshold value t, we calculate the data-complexity for a given success probability experimentally by comparing 2^{10} χ^2 test statistics to t. Figure 6a and Figure 6b show the results for three rounds of CAST-128 and LOKI91, respectively. Further details can be found in Table 8 and Table 9 in Appendix B.

3-round LLR distinguisher As the probability mass function p_{real} contains a rather large number of zeros for 3-rounds of CAST-128 and LOKI91, considering only the first case of the analysis in Section 3.2 already yields a good distinguisher with $P_{\mathrm{S}} = 1$. The number of zeros is equal to 1580108708 and 1638446404 for CAST-128 and LOKI91, respectively. The data-complexity can be calculated for fixed false-positive rate P_{F} using Equation (8). Hence, the advantage of the distinguisher can be estimated as

$$1 - \left(1 - \frac{N_{\mathrm{zero}}}{2^{32}}\right)^q.$$

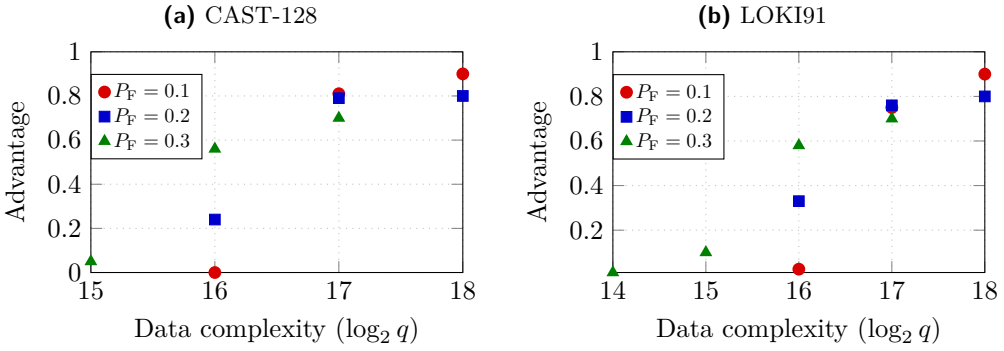

Figure 5: Advantage as a function of data-complexity for 3-round χ^2 distinguishers.

Figure 7 presents the experimental advantage for both target algorithms in terms of the data-complexity with 2^{10} attempts.

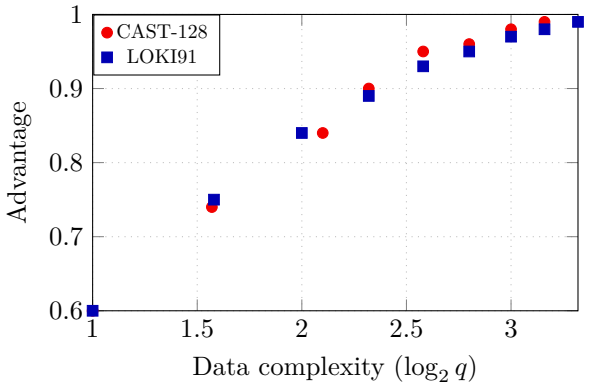

Figure 7: Advantage as a function of data-complexity for 3-round LLR distinguishers.

5-round LLR distinguisher Examining the probability distribution p_{real} for 5-round CAST-128 and LOKI91 shows that no zeros are present. As a result, only the second case in the analysis from Section 3.2 needs to be considered. Consequently, the threshold value t can be obtained using Equation (9), after calculating the value of μ_{real}. The success probability is estimated by comparing the computed LLR test statistics with t, similar to the experiments for the χ^2 test. The experimental results about the data-complexity for target ciphers are provided in Figure 8.

5 Key-recovery attacks

The distinguisher can be extendable to a key recovery attack by appending one or more rounds to the distinguisher. Let r' be the number of rounds added to the r-round distinguisher for $r + r'$ round key recovery. Then, one can estimate the correlation for each key guess by means of partial decryption of the last r' rounds. A test statistic is derived for each guess, and when an incorrect key is utilized, it is assumed that the test statistics behave similarly to those for a random permutation.

Maintaining a low false-positive rate is essential to increase the number of guessed key bits ℓ. To reduce the candidates to one or a few keys, one should have $P_F \times 2^\ell \approx 1$.

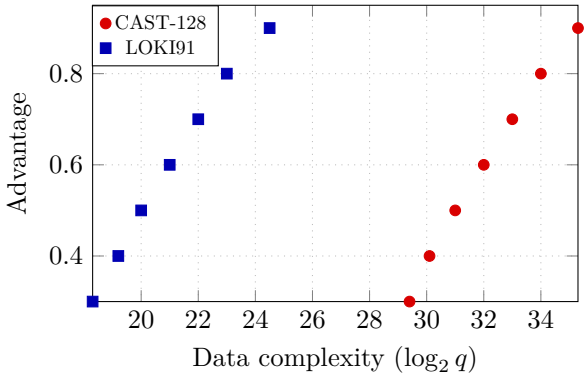

Figure 8: Advantage as a function of data-complexity for 5-round LLR distinguishers.

Once the false-positive rate is determined, a suitable decision threshold can be calculated. The data-complexity to achieve a given P_S is then computed. Note that the total time complexity (T) is equal to the sum of the time complexity required to recover part of the subkey (T_{check}) and to guess the remaining bits (T_{guess}). If only a few candidates remain for the last round key, the remaining key bits can be determined by an $r-1$ round attack so that $T_{\text{guess}} \leq T_{\text{check}}$. Hence, the total time complexity of the naive key-recovery method is $\mathcal{O}(q2^{\ell})$, where q is the data-complexity to recover ℓ bits of the key with the fixed P_F.

5.1 Generic key-recovery attacks

The generic multidimensional distinguishers from Section 3 can be used to obtain a generic multidimensional key recovery attack. Hence, we can use the data-complexity results for the r-round distinguishers that we calculated in Section 3 to determine the data-complexity of an $r+r'$ rounds key recovery attack. Recall that the data-complexity is $2^{n\lfloor r/2 \rfloor - n}$ for an $r \geq 2$ round multidimensional distinguisher using the likelihood-ratio test.

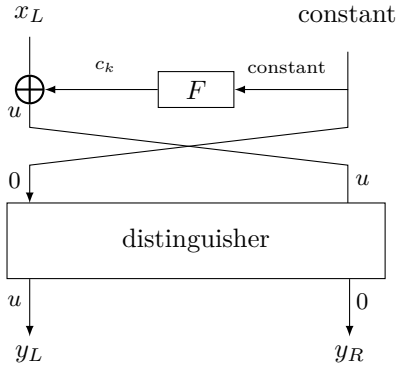

Figure 9: Key recovery with FFT method.

To improve the work effort of the naive key-recovery method, we now describe a variant of the Fast Fourier Transform method [CSQ07] for key-recovery. Suppose we prepend one round to the distinguisher as in Figure 9. Fix the right half of the input to a constant.

Table 6: Generic key recovery attack on an r-round Feistel cipher with block size $2n$ and k-bit master key.

k	$r+r'$	r'	Data	Method	Work effort
2n	4	1	1	naive	2^n
		1	1	FFT	$n2^n$
	6	1	2^n	naive	2^{2n}
		1	2^n	FFT	$n2^n$
4n	4	1	1	naive	2^n
		1	1	FFT	$n2^n$
	5	2	1	naive	2^{2n}
		2	1	FFT	$n2^{2n}$
	6	3	1	naive	2^{3n}
		3	1	FFT	$n2^{3n}$
		1	2^n	naive	2^{2n}
		1	2^n	FFT	$n2^n$
	7	2	2^n	naive	2^{3n}
		2	2^n	FFT	$n2^{2n}$
	8	3	2^n	FFT	$n2^{3n}$

The LLR statistic for the key-dependent constant c_k is equal to

$$\begin{aligned}\lambda(x_L + y_L + c_k) &= \sum_{i=1}^{q} \lambda(x_{iL} + y_{iL} + c_k) \\ &= \sum_{i=1}^{q} \sum_{z \in \mathbb{F}_2^n} \lambda(x_{iL} + y_{iL} + c_k)\delta_{z, x_{iL} + y_{iL}} \\ &= \sum_{z \in \mathbb{F}_2^n} \lambda(z + c_k) \sum_{i=1}^{q} \delta_{z, x_{iL} + y_{iL}} \\ &= \sum_{z \in \mathbb{F}_2^n} M_{c_k, z}\, w_z\,,\end{aligned} \tag{10}$$

where w is a vector of occurrences of $z = x_{iL} + y_{iL}$ values and M is a matrix with entries $M_{c_k,z} = \lambda(z + c_k)$. The time complexity of computing the matrix M is $\mathcal{O}(2^{2n})$ without a trick. We exploit the circulant structure of M using the FFT method to reduce the time complexity. Since M is a circulant matrix, it is enough to compute $M_{0,z}$. Indeed, we compute the matrix-vector product by taking the pointwise product of the FFT of the first row and the FFT of w, followed by taking the inverse Fourier transform. Then the complexity of calculating matrix M becomes $\mathcal{O}(n2^n)$. The complexity of computing the vector w is $\mathcal{O}(2^n + q)$. Thus the total complexity becomes $\mathcal{O}(n2^n + q)$, corresponding to $\mathcal{O}(n2^n)$ when $q \leq n2^n$. If we incorporate additional rounds after the distinguisher for the key recovery, the same operations must be performed for each key guess. The data-complexity must remain below the input size 2^n for a successful attack since the right half of the input is constant. Thus, our proposed $r + r'$-round generic key recovery attack is viable with an $r \leq 5$ round distinguisher.

Table 7: Key recovery results of r-round CAST-128 and LOKI91.

Cipher	$r + r'$	r'	q	T
CAST-128	4	1	$2^{5.80}$	2^{37}
	5	2	$2^{6.80}$	2^{74}
	6	3	$2^{7.39}$	2^{111}
		1	2^{35}	2^{40}
	7	2	2^{35}	2^{77}
	8	3	2^{35}	2^{114}
LOKI91	4	1	$2^{5.55}$	2^{37}
	6	1	2^{24}	2^{37}

5.2 Application to CAST-128 and LOKI91

In the case of CAST-128 and LOKI91, the block size is 64 bits, and the subkey size is 32 bits, meaning that $m = \ell = n = 32$. To achieve 4-round and 6-round key recovery, we can utilize a 3- and 5-round distinguisher, respectively. We use the FFT method since the data-complexity q is always greater than 32 in practice. As both cases satisfy $q < n2^n$, the time complexity is determined as 2^{37} for the 4-round key-recovery attack. The same operations must be repeated for each key guess when any additional rounds are appended, resulting in a time-complexity of $2^{37+32(r'-1)}$ for the r-round key recovery attack, where $r = 3 + r'$ or $r = 5 + r'$ using a 3-round or 5-round distinguisher, respectively. In the case of CAST-128, due to the subkey rotation before the round function, one needs to guess an additional 5 bits for every round and the time complexity becomes $\mathcal{O}(2^{37(r')})$. However, guessing 5 more key bits in the first round is not required. Moreover, unlike the MITM attack, which requires key guessing in every round, our approach only involves key guessing in a limited number of rounds for CAST-128, usually one or two. Hence, our work efforts are better than existing key recovery results on 7- and 8-round CAST-128. On the other hand, the data-complexity of the MITM attacks from [IS13b, IS13a] is better than our findings for 7- and 8-round CAST-128. On the bad side, more than 8-round key recovery attacks on CAST-128 are not feasible due to the key-dependent rotation. Another particular feature of CAST-128 is its utilization of three distinct round functions, F_1, F_2, and F_3. When we add one round before the distinguisher, F_3 is active instead of F_2. However, as shown in Section 4, the effect on the data-complexity is negligible. Key recovery results using the FFT method are given in Table 7.

Which values of r' are admissible depends on the key size. For instance, in the case of CAST-128, which utilizes a 128-bit key size, $37 \times r'$ bits of a subkey from $3 + r'$ rounds can be successfully retrieved by exploiting the 3-round distinguisher, where $r' = 1, 2, 3$. However, in the case of LOKI91, which employs a 64-bit key, the only viable option for r' is 1. Consequently, we can recover a 32-bit subkey of the 4-round and 6-round LOKI91 by utilizing the 3-round and 5-round distinguisher, respectively.

For the attacks on 6, 7 and 8 round CAST-128, the amount of data exceeds 2^{32} despite the fact that we need to fix the right half of the plaintext when using the FFT method for key-recovery (so only 2^{32} plaintexts are available). Since the false-positive rate will not be sufficiently low to filter the keys, it is necessary to repeat the attack several times (we estimate 8 times should be enough) and to combine the test statistics. To do this, it is necessary to translate the list of candidate constants to a longer list of candidates for the key of the first round. This can be done in time 2^{37}. Overall, the work effort increases by a factor approximately equal to the number of repetitions of the attack.

6 Conclusion

This work presents a new approach to reduce the data complexity of linear attacks on Feistel ciphers that incorporate the key addition to the input of the round function only. We introduced efficient generic multidimensional linear attacks on Feistel ciphers by exploiting key-independent trails. Our work also provides a detailed comparison between the χ^2 and likelihood-ratio tests in linear cryptanalysis. In particular, our theoretical and experimental findings support that the likelihood-ratio method is more powerful. Based on our distinguishers, efficient generic key recovery attacks were obtained using a variant of the FFT method.

Theoretical estimates of the costs of our generic attacks were provided and subsequently validated through experiments on the concrete Feistel ciphers CAST-128 and LOKI91. In several cases, our approach improves over previous dedicated attacks.

Overall, our results emphasize the effectiveness of key-independent multidimensional approximations and the likelihood-ratio test in the context of multidimensional linear attacks.

Acknowledgments

This work was supported by CyberSecurity Research Flanders with reference number VR20192203. In addition, this work was partially supported by the Research Council KU Leuven, C16/18/004 through the IF/C1 on New Block Cipher Structures and by the Flemish Government through FWO Project Locklock G0D3819N. Tim Beyne is funded by an FWO fellowship.

References

[Ada97] Carlisle M. Adams. The CAST-128 Encryption Algorithm. *RFC*, 2144:1–15, 1997. https://api.semanticscholar.org/CorpusID:46673380.

[BDCQ04] Alex Biryukov, Christophe De Cannière, and Michaël Quisquater. On Multiple Linear Approximations. In Matt Franklin, editor, *Advances in Cryptology – CRYPTO 2004*, pages 1–22. Springer Berlin Heidelberg, 2004.

[Bey21] Tim Beyne. Linear Cryptanalysis of FF3-1 and FEA. In Tal Malkin and Chris Peikert, editors, *Advances in Cryptology – CRYPTO 2021*, pages 41–69, Cham, 2021. Springer International Publishing. https://doi.org/10.1007/978-3-030-84242-0_3.

[BG09] Céline Blondeau and Benoît Gérard. On the data complexity of statistical attacks against block ciphers (full version). Cryptology ePrint Archive, Paper 2009/064, 2009. https://eprint.iacr.org/2009/064.

[BJV04] Thomas Baignères, Pascal Junod, and Serge Vaudenay. How Far Can We Go Beyond Linear Cryptanalysis? In *Advances in Cryptology - ASIACRYPT 2004*, volume 3329 of *Lecture Notes in Computer Science*, pages 432–450. Springer, 2004. https://doi.org/10.1007/978-3-540-30539-2_31.

[CSQ07] Baudoin Collard, F. X. Standaert, and Jean-Jacques Quisquater. Improving the Time Complexity of Matsui's Linear Cryptanalysis. In Kil-Hyun Nam and Gwangsoo Rhee, editors, *Information Security and Cryptology - ICISC 2007*, pages 77–88. Springer Berlin Heidelberg, 2007. https://doi.org/10.1007/978-3-540-76788-6_7.

[DGV95] Joan Daemen, René Govaerts, and Joos Vandewalle. Correlation Matrices. In Bart Preneel, editor, *Fast Software Encryption*, pages 275–285. Springer Berlin Heidelberg, 1995. https://doi.org/10.1007/3-540-60590-8_21.

[DR07] Joan Daemen and Vincent Rijmen. Probability distributions of correlation and differentials in block ciphers. *J. Math. Cryptol.*, 1(3):221–242, 2007.

[HCN08] Miia Hermelin, Joo Yeon Cho, and Kaisa Nyberg. Multidimensional Linear Cryptanalysis of Reduced Round Serpent. In Yi Mu, Willy Susilo, and Jennifer Seberry, editors, *Information Security and Privacy*, pages 203–215. Springer Berlin Heidelberg, 2008. https://doi.org/10.1007/978-3-540-70500-0_15.

[HCN09] Miia Hermelin, Joo Yeon Cho, and Kaisa Nyberg. Multidimensional Extension of Matsui's Algorithm 2. In Orr Dunkelman, editor, *Fast Software Encryption*, pages 209–227. Springer Berlin Heidelberg, 2009. https://doi.org/10.1007/978-3-642-03317-9_13.

[HCN19] Miia Hermelin, Joo Yeon Cho, and Kaisa Nyberg. Multidimensional Linear Cryptanalysis. *Journal of Cryptology*, 32(1):1–34, 2019. https://doi.org/10.1007/s00145-018-9308-x.

[IS13a] Takanori Isobe and Kyoji Shibutani. All subkeys recovery attack on block ciphers: Extending meet-in-the-middle approach. In Lars R. Knudsen and Huapeng Wu, editors, *Selected Areas in Cryptography*, pages 202–221. Springer Berlin Heidelberg, 2013. https://doi.org/10.1007/978-3-642-35999-6_14.

[IS13b] Takanori Isobe and Kyoji Shibutani. Generic key recovery attack on feistel scheme. In Kazue Sako and Palash Sarkar, editors, *Advances in Cryptology - ASIACRYPT 2013*, pages 464–485. Springer Berlin Heidelberg, 2013. https://doi.org/10.1007/978-3-642-42033-7_24.

[ISO] ISO/IEC. The International Organization for Standardization and The International Electrotechnical Commission. https://www.iso.org/obp/ui/#iso:std:iso-iec:18033:-3:ed-2:v1:en:sec:4.4.

[JR07] Jorge Nakahara Jr and Mads Rasmussen. Linear Analysis of Reduced-Round CAST-128 and CAST-256. In *Anais do VII Simpósio Brasileiro em Segurança da Informação e de Sistemas Computacionais*, pages 15–25, Porto Alegre, RS, Brasil, 2007. SBC.

[KR94] Burton S. Kaliski and M. J. B. Robshaw. Linear Cryptanalysis Using Multiple Approximations. In Yvo G. Desmedt, editor, *Advances in Cryptology — CRYPTO '94*, pages 26–39. Springer Berlin Heidelberg, 1994. https://doi.org/10.1007/3-540-48658-5_4.

[KR96] Lars R. Knudsen and M. J. B. Robshaw. Non-Linear Approximations in Linear Cryptanalysis. In Ueli Maurer, editor, *Advances in Cryptology — EUROCRYPT '96*, pages 224–236. Springer Berlin Heidelberg, 1996. https://doi.org/10.1007/3-540-68339-9_20.

[Mat94] Mitsuru Matsui. Linear Cryptanalysis Method for DES Cipher. In Tor Helleseth, editor, *Advances in Cryptology — EUROCRYPT '93*, pages 386–397. Springer Berlin Heidelberg, 1994. https://doi.org/10.1007/3-540-48285-7_33.

[NP92] J. Neyman and E. S. Pearson. On the Problem of the Most Efficient Tests of Statistical Hypotheses. In Samuel Kotz and Norman L. Johnson, editors, *Breakthroughs in Statistics: Foundations and Basic Theory*. Springer New York, 1992. https://doi.org/10.1007/978-1-4612-0919-5_6.

[NPV17] Valerie Nachef, Jacques Patarin, and Emmanuel Volte. *Generic Attacks on Classical Feistel Ciphers*, pages 65–73. Springer International Publishing, Cham, 2017. https://doi.org/10.1007/978-3-319-49530-9_6.

[Pat92] Jacques Patarin. New Results on Pseudorandom Permutation Generators Based on the DES Scheme. In Joan Feigenbaum, editor, *Advances in Cryptology — CRYPTO '91*, pages 301–312. Springer Berlin Heidelberg, 1992. https://doi.org/10.1007/3-540-46766-1_25.

[Pat01] Jacques Patarin. Generic Attacks on Feistel Schemes. In Colin Boyd, editor, *Advances in Cryptology — ASIACRYPT 2001*, pages 222–238. Springer Berlin Heidelberg, 2001. https:/[h]/doi.org/10.1007/3-540-45682-1_14.

[Pat04] Jacques Patarin. Security of Random Feistel Schemes with 5 or More Rounds. In Matt Franklin, editor, *Advances in Cryptology – CRYPTO 2004*, pages 106–122. Springer Berlin Heidelberg, 2004. https://doi.org/10.1007/978-3-540-28628-8_7.

[SF97] Kouichi Sakurai and Souichi Furuya. Improving Linear Cryptanalysis of LOKI91 by Probabilistic Counting Method. In Eli Biham, editor, *Fast Software Encryption*, pages 114–133. Springer Berlin Heidelberg, 1997. https://doi.org/10.1007/BFb0052340.

[TSM95] Toshio Tokita, Tohru Sorimachi, and Mitsuru Matsui. Linear Cryptanalysis of LOKI and s2DES. In Josef Pieprzyk and Reihanah Safavi-Naini, editors, *Advances in Cryptology — ASIACRYPT'94*, pages 293–303. Springer Berlin Heidelberg, 1995. https://doi.org/10.1007/BFb0000442.

[Vau96] Serge Vaudenay. An Experiment on DES Statistical Cryptanalysis. In *Conference on Computer and Communications Security*, 1996. https://api.semanticscholar.org/CorpusID:15041839.

[WWH09] Meiqin Wang, Xiaoyun Wang, and Changhui Hu. New Linear Cryptanalytic Results of Reduced-Round of CAST-128 and CAST-256. In Roberto Maria Avanzi, Liam Keliher, and Francesco Sica, editors, *Selected Areas in Cryptography*, pages 429–441. Springer Berlin Heidelberg, 2009. https://doi.org/10.1007/978-3-642-04159-4_28.

A Data complexity for CAST-128

In principle, the data complexity for CAST-128 depends on which round functions are active. The data complexities for multidimensional distinguishers with (F_2, F_1, F_2) and (F_1, F_2, F_3) correspond to 3, 5, and 7 rounds distinguishers for CAST-128, respectively.

One of the reasons for calculating the data complexities of other cases is the possibility of obtaining better data complexity for different combinations of round functions. These combinations should be carefully considered when a distinguisher is expanded from the beginning for key recovery. Hence, it is worthwhile evaluating every possible combination. However, our calculations show that the data complexities are close for all cases. In all

cases, the data complexities for the 3-, 5-, and 7-round CAST-128 distinguishers are:

$$q_{3r} \approx 1 \Big/ \sum_u \left(C_{u,0}^{F_2}\right)^2 = 1$$
$$q_{5r} \approx 1 \Big/ \sum_u \left(C_{u,0}^{F_2} \times C_{u,0}^{F_1}\right)^2 = 2^{31.99}$$
$$q_{7r} \approx 1 \Big/ \sum_u \left(C_{u,0}^{F_3} \times C_{u,0}^{F_2} \times C_{u,0}^{F_1}\right)^2 = 2^{63.99}$$

In LOKI91, all round functions are the same, so data complexities of 3-, 5-, 7- and 9-round distinguishes are

$$q_{3r} \approx 1 \Big/ \sum_u \left(C_{u,0}^{F}\right)^2 = 1$$
$$q_{5r} \approx 1 \Big/ \sum_u \left(C_{u,0}^{F}\right)^4 = 2^{21.434}$$
$$q_{7r} \approx 1 \Big/ \sum_u \left(C_{u,0}^{F}\right)^6 = 2^{39.212}$$
$$q_{9r} \approx 1 \Big/ \sum_u \left(C_{u,0}^{F}\right)^8 = 2^{56.346}.$$

B Proof of Proposition 1

Proof. Let O_z denote the number of observations of value z. The number of collisions N_{coll} is related to these values by

$$\begin{aligned} N_{\mathrm{coll}} &= \sum_{z \in \mathbb{F}_2^n} \frac{\mathrm{O}_z(\mathrm{O}_z - 1)}{2} \\ &= \frac{1}{2} \sum_{z \in \mathbb{F}_2^n} \mathrm{O}_z^2 - \frac{1}{2} \sum_{z \in \mathbb{F}_2^n} \mathrm{O}_z \\ &= \frac{1}{2} \sum_{z \in \mathbb{F}_2^n} \mathrm{O}_z^2 - \frac{1}{2} q \,. \end{aligned} \tag{11}$$

Hence, we obtain

$$\sum_{z \in \mathbb{F}_2^n} \mathrm{O}_z^2 = 2N_{\mathrm{coll}} + q. \tag{12}$$

Similarly, the χ^2 test statistic can be expressed using O_z and the expected number of observations of value z, i.e. E_z:

$$\begin{aligned} \chi^2 &= \sum_{z \in \mathbb{F}_2^n} \frac{(\mathrm{O}_z - \mathrm{E}_z)^2}{\mathrm{E}_z} \\ &= \underbrace{\sum_{z \in \mathbb{F}_2^n} \frac{\mathrm{O}_z^2}{\mathrm{E}_z}}_{\mathrm{E}_z = \frac{q}{2^n}} - 2 \underbrace{\sum_{z \in \mathbb{F}_2^n} \mathrm{O}_z}_{= q} + \underbrace{\sum_{z \in \mathbb{F}_2^n} \mathrm{E}_z}_{= q} \\ &= \frac{2^n}{q} \sum_{z \in \mathbb{F}_2^n} \mathrm{O}_z^2 - q \,, \end{aligned} \tag{13}$$

where the expected values E_z are calculated for a uniform distribution. Eventually, the link is demonstrated replacing $\sum_{z\in\mathbb{F}_2^n} \mathrm{O}_z^2$ with $2N_{\text{coll}} + q$:

$$\chi^2 = \frac{2^{n+1}}{q} N_{\text{coll}} + 2^n - q.$$

□

In addition to its relation with the χ^2 statistic, the number of collisions can also be employed as an independent method for distinguishing a cipher. The probability of obtaining a collision can be computed using the following equation:

$$\Pr[x_i = x_j] = \sum_{x\in\mathbb{F}_2^n} p(x)^2 = 2^{-n} \sum_{(u,v)\in\Lambda} (C_{v,u}^F)^2.$$

where p is the probability mass function and assuming $|\Lambda| = 2^n$. Hence, the average number of collisions in a sample of size q equals

$$N_{\text{coll}} = \binom{q}{2} \Pr[x_i = x_j] = 2^{-n} \binom{q}{2} \sum_{(u,v)\in\Lambda} (C_{v,u}^F)^2. \tag{14}$$

As $C_{0,0}^F = 1$, Equation (14) becomes

$$N_{\text{coll}} = 2^{-n}\binom{q}{2}\left(1 + \sum_{(u,v)\neq(0,0)} (C_{v,u}^F)^2\right) = \underbrace{2^{-n}\binom{q}{2}}_{\text{same as random}} \underbrace{+\,2^{-n}\binom{q}{2} \sum_{(u,v)\neq(0,0)} (C_{v,u}^F)^2}_{\text{extra collisions}}. \tag{15}$$

For a successful attack, it is necessary that the difference between the means of the real and ideal distributions of valid pairs surpass the standard deviation of the ideal distribution [BG09]: $\mu_{\text{real}} - \mu_{\text{ideal}} = \mu_{\text{real}} \gg \sigma_{\text{ideal}}$. Then we have $D(\mathrm{p}_{\text{real}} - \mathrm{p}_{\text{ideal}}) \gg \sqrt{D\mathrm{p}_{\text{ideal}}}$ where $D = \binom{q}{2}$. When we rewrite the equation, we get

$$\binom{q}{2} \geq \frac{p_{\text{ideal}}}{(p_{\text{real}} - p_{\text{ideal}})^2}, \tag{16}$$

where p_{real} and p_{ideal} are the probability of obtaining collisions for the real primitive and a random permutation, respectively. For 3-round CAST-128 and LOKI91, $p_{\text{ideal}} = 2^{-32}$ and $p_{\text{real}} = 2^{-32} \sum_{u\in\mathbb{F}_2^{32}} (C_{u,0}^F)^2 \approx 2^{-31}$. So, the amount of data is close to 2^{16} theoretically. Table 8 and Table 9 contain the average number of collisions, estimated using 2^{10} different keys.

Table 8: The average number of collisions for 3-round CAST-128.

	Total		Random	
Dataset	Theoretical	Experimental	Theoretical	Experimental
2^{14}	0.062	0.056	0.032	0.027
2^{15}	0.249	0.223	0.125	0.121
2^{16}	0.999	1.011	0.5	0.496
2^{17}	3.999	3.914	2	1.943
2^{18}	15.999	15.922	8	7.932
2^{19}	63.999	63.94	32	31.92
2^{20}	256	256.014	128	127.451
2^{21}	1024	1024.561	512	511.387

Table 9: The average number of collisions for 3-round LOKI91.

	Total		Random	
Dataset	Theoretical	Experimental	Theoretical	Experimental
2^{14}	0.065	0.061	0.032	0.032
2^{15}	0.259	0.304	0.125	0.111
2^{16}	1.038	1.051	0.5	0.503
2^{17}	4.155	4.034	2	1.982
2^{18}	16.422	16.124	8	8.048
2^{19}	65.486	62.94	32	32.232
2^{20}	265.945	262.014	128	127.451
2^{21}	1024	1030.456	512	511.651

C Pseudo-code

Algorithm 1: Computation the theoretical data complexity of multidimensional linear distinguisher for an r-round Feistel cipher.

Input : input mask: $(0 \parallel u)$
output mask: $(0 \parallel u)$

Set $n \leftarrow 2^{32}$;
for $i = 0$ *to* 2^n **do**
  $++\, p_0[f_2(i)]$;
  $++\, p_1[f_4(i)]$;
  $\vdots$
  $++\, p_k[f_{2k}(i)]$; $//k = \lfloor \frac{r}{2} \rfloor$
end
for $i = 0$ *to* k **do**
  $p_i \leftarrow FFT(p_i)$;
end
for $i = 1$ *to* n **do**
  $p_0 \leftarrow p_0[i]p_1[i] \dots p_k[i]$;
  $c^2 + \leftarrow p_0^2$;
end
$q \leftarrow \frac{1}{c^2}$;

Algorithm 2: Computation the χ^2 test statistic for an r-round Feistel cipher with a multidimensional approximations.

Input : input mask: $(0 \parallel u)$
output mask: $(0 \parallel u)$

Set $e \leftarrow q2^{-32}$;
for $i = 0$ *to* q **do**
  x^i: $x_1^i \parallel x_2^i$;
  $F_r(x) \leftarrow y^i$: $y_1^i \parallel y_2^i$;
  $z_j \leftarrow x_2^i + y_2^i$;
  $++\, o[z_j]$;
end
for *each* z_j **do**
  $\chi^2 + \leftarrow (o[z_j] - e)^2/e$;
end

Algorithm 3: Computation the LLR test statistic for an r-round multidimensional distinguisher.

Input : input mask: $(0 \parallel u)$
output mask: $(0 \parallel u)$

Set $n \leftarrow 2^{32}$;
Set $p \leftarrow 2^{-32}$;
for $i = 0$ *to* q **do**
  $x^i : x_1^i \parallel x_2^i$;
  $F_r(x) \leftarrow y^i : y_1^i \parallel y_2^i$; $//F_r = f_1 \circ \cdots \circ f_r$
  $++\,\eta[x_2^i \oplus y_2^i]$;
end
for $i = 0$ *to* n **do**
  $++\,p_0[f_2(i)]$;
  $++\,p_1[f_4(i)]$;
  $\vdots$
  $++\,p_k[f_{2k}(i)]$; $//k = \lfloor \frac{r}{2} \rfloor$
end
for $i = 0$ *to* k **do**
  $p_i \leftarrow FFT(p_i)$;
end
for $i = 0$ *to* n **do**
  $p_0[i] \leftarrow p_0[i]p_1[i] \ldots p_k[i]$;
end
if $r \geq 4$ **then**
  $p_0 \leftarrow FFT(p_0)$;
end
for $i = 0$ *to* n **do**
  $\lambda_0 + \leftarrow \log \frac{p_0[\eta[i]]}{p}$;
end

Algorithm 4: FFT-based key recovery attack using r-round multidimensional distinguisher.

Input : input mask: $(0 \parallel u)$
output mask: $(0 \parallel u)$

Set $n \leftarrow 2^{32}$;
Set $p \leftarrow 2^{-32}$;
Set master key $\leftarrow k$;
Set number of runs $\leftarrow t$; *//Run the code t times*
for $i = 0$ *to* n **do**
  $x^i : i \parallel c$;
  $F_r(x) = y^i : y_1^i \parallel y_2^i$; *//* $F_{r+1} = f_1 \circ \cdots \circ f_r$
  $z_i = i \oplus y_2^i$;
end
for $i = 0$ *to* n **do**
  $++ p_0[f_3(i)]$;
  $++ p_1[f_5(i)]$;
  $\vdots$
  $++ p_\ell[f_{2\ell+1}(i)]$; *//* $\ell = \lfloor \frac{r}{2} \rfloor$
end
for $i = 0$ *to* ℓ **do**
  $p_i \leftarrow FFT(p_i)$;
end
for $i = 0$ *to* n **do**
  $p_0[i] \leftarrow p_0[i]p_1[i] \ldots p_k[i]$;
end
if $r \geq 4$ **then**
  $p_0 \leftarrow FFT(p_0)$;
end
for $i = 0$ *to* n **do**
  $\lambda_0[i] \leftarrow \log \frac{p_0[i]}{p}$;
end
$\lambda_0 \leftarrow FFT(\lambda_0)$;
for $i = 0$ *to* n **do**
  $++ w[z_i]$;
end
for $i = 0$ *to* n **do**
  $w[i] \leftarrow w[i]/n$;
end
$w \leftarrow FFT(w)$;
for $i = 0$ *to* n **do**
  $w[i] \leftarrow w[i] * \lambda_0[i]$;
end
$w \leftarrow FFT(w)$;
Sort w in descending order with indeces s[i];
$c_{k_1} \leftarrow f_1(k_1, c)$; *//*$k_i :$ *subkey of* $i^{th} -$ *round function*
$n_a \leftarrow \#$ *of attempts*;
for $i = 0$ *to* n **do**
  if $c_{k_1} = s[i]$ & $\ell = \{\#i \mid s[i] \geq P_{\mathrm{F}} * 2^n\}$ **then**
    $P_{\mathrm{S}} \leftarrow \frac{\ell}{n_a}$;
    The data complexity $\leftarrow n_a * 2^n$;
  end
end