Generic Attacks on Lai-Massey Structures

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 work introduces new generic attacks on FOX-style Lai-Massey schemes that use multidimensional linear approximations. A multidimensional linear property is transformed into a distinguisher. Our research primarily focuses on extending this distinguisher to message- and key-recovery attacks on Lai-Massey schemes. Indeed, our results demonstrate that a 3-round distinguisher can be extended to a 4-round message-recovery attack (and, depending on key lengths, a key-recovery attack), using roughly the same amount of data as a 3-round distinguisher. We propose a generic key-recovery attack that combines the multidimensional distinguisher with the Fast Fourier transform method. Notably, our 5-round distinguisher can be extended to an 8-round key recovery attack, the best generic attack on FOX-style Lai-Massey schemes to the best of our knowledge. Finally, we apply our generic attack to the FOX64 and SPC block ciphers.

Keywords: Multidimensional linear cryptanalysis · χ^2 test · Generic attack · Lai-Massey structure · FOX · SPC.

1 Introduction

Generic attacks on symmetric cryptographic primitives do not rely on specific design details of a cipher. They typically reveal inherent security limitations based on factors like key size, number of rounds, and overall architecture. Generic attacks can be devised using linear cryptanalysis, which was introduced by Matsui [24] in the early 1990s. A version of this technique, known as multiple linear cryptanalysis, uses several linear approximations to improve distinguishing advantage. Kaliski and Robshaw [17] showed that employing multiple approximations can lower the data complexity of an attack. Building on this idea, Hermelin et al. [11,12,13] introduced multidimensional linear cryptanalysis, a refinement where the set of masks is a vector space.

In 1991, the IDEA (International Data Encryption Algorithm) [19] was designed by Lai and Massey as a modified version of the PES (Proposed Encryption Standard) block cipher [20]. At Asiacrypt'99 [32], Vaudenay introduced a construction based on the IDEA block cipher, known as the Lai–Massey scheme. Numerous ciphers, including the WIDEA [15], FOX [16], MESH [25], and FLY [18], have been inspired by the Lai-Massey structure. FOX is a family of block ciphers

based on the Lai–Massey scheme, supporting block sizes of 64 bits and 128 bits. Its round function is a substitution-permutation network. More recently, the SPC cipher [1] was introduced, based on the Lai–Massey structure like FOX, but with a round function adapted from SipHash [2] that operates on 64-bit instead of 32-bit words. This paper focuses on FOX-like Lai-Massey schemes, as shown in Figure 1, which are based on XOR operations rather than modular addition and subtraction as in IDEA.

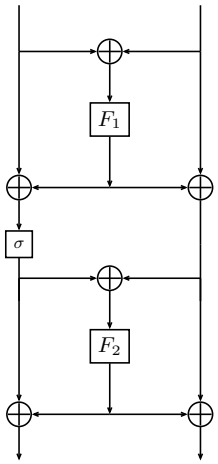


Fig. 1: 2-round FOX-like Lai-Massey scheme.

Although some research has addressed the (strong) pseudorandomness of the Lai-Massey scheme, several questions remain regarding its resilience to generic attacks. In [32], the pseudorandomness and strong pseudorandomness of 3- and 4-round Lai-Massey schemes were evaluated. Later, it was proven in [22] that 4 rounds are not only sufficient but also necessary for a Lai-Massey scheme to achieve strong pseudorandomness for $\mathcal{O}(2^{n/2})$ queries, assuming the round function $F_i : \mathbb{F}_2^n \to \mathbb{F}_2^n$ is ideal. In later research [23], Patarin's works [27,28,29] on Feistel networks are adopted by developing reduced-round attacks on Lai-Massey schemes with bijective and non-bijective round functions. They concluded that attacking 4- and 5-round Lai–Massey ciphers requires greater computational effort than attacking 4- and 5-round Feistel schemes. In the literature, there are also works on multidimensional linear attacks against Feistel ciphers, such as Beyne [6] and by Özdemir et al. [26]. While Feistel and Lai–Massey structures share certain similarities, their differences are significant from a cryptanalytic perspective. In this paper, we analyze the security of Lai-Massey constructions against multidimensional linear attacks. A key distinction from similar work by [6] and [26] is that trails were constructed for Feistel ciphers with non-invertible round functions, whereas in our setting, the round functions are permutations.

A further difference is that the analyses in [26] relied on key-independent trails, which only exist for Feistel ciphers with key additions at the input of the round functions. In our case, the round function is key-dependent. These structural differences lead to fundamentally different multidimensional linear approximations.

Our Contribution: This work uses multidimensional linear approximations to obtain generic attacks on FOX-style Lai-Massey schemes, where σ is a linear operation given by $\sigma(x_1, x_2) = (x_2, x_1 \oplus x_2)$. The multidimensional linear property is transformed into a distinguisher using the χ^2 test. For $r \geq 2$ rounds, the data complexity of the generic multidimensional distinguisher (in the KPA model) is $\mathcal{O}(2^{n\lfloor r/2 \rfloor - n/2})$, where n is the input size of the round function. Prior studies have analyzed the number of rounds required to achieve pseudorandomness in Lai–Massey constructions. Unlike Luo et al. [22], we demonstrate that the 3-round multidimensional distinguisher, with data complexity $2^{n/2}$, can be extended to a 4-round message recovery attack with the same data complexity. Using the same idea employed for message recovery, we obtain a key recovery attack by fixing the input of the first-round function to a constant.

In Table 1, we summarize the cost of generic attacks and compare them with existing distinguisher results. As seen in Table 1, we reduce the data complexity of 3- and 5-round distinguishers compared to prior generic attacks. Furthermore, we present a 6-round generic distinguishing attack in the CPA model, which, to the best of our knowledge, is the best distinguishing attack. Additionally, our method enables 6-round and 8-round key-recovery attacks for master-key lengths $2n$ and $4n$, respectively, given a round-key size of n. To validate our generic attacks, we apply the proposed method to FOX64 and SPC. The results for the data complexity of the χ^2 multidimensional distinguisher are given in Figure 8. They align with our theoretical predictions. The data complexities of the FOX cipher are slightly better than the generic estimates. Additionally, the experimental findings for the message recovery attack on SPC, presented in Figure 10, are compatible with theoretical findings. Finally, a comparison of our key recovery attack complexities of FOX64 with those derived from integral attacks, summarized in Table 5, indicates improved time complexities in many cases.

2 Preliminaries

This section presents the concepts required for understanding our generic attacks on FOX-style Lai–Massey schemes using multidimensional linear cryptanalysis.

2.1 Security model

A statistical hypothesis test aims to evaluate how observed data support a specific hypothesis. This involves proposing two hypotheses: the null hypothesis, which claims that the data originate from the real primitive, and the alternative hypothesis, which asserts that the data come from an ideal primitive. The acceptance region $\mathcal{A}$ is a range of test statistic values that support the null hypothesis.

Table 1: Summary of generic attacks on r-round FOX-like Lai-Massey cipher with $2n$-bit block, n-bit round key and k-bit master key sizes. All work efforts and data complexities are given up to constant factors. Key recovery attacks leverage a variant of the Fast Fourier Transform method.

Types	k	r	Data	Ref.	Work effort
KP based distinguisher		2	$2^{n/2}$	§ 3	$2^{n/2}$
			$2^{n/2}$	[23]	$2^{n/2}$
		3	$2^{n/2}$	§ 3	$2^{n/2}$
			$2^{n\dagger}$	[23]	$2^{5n/4}$
		4	$2^{3n/2}$	§ 3	$2^{3n/2}$
			$2^{3n/2}$	[23]	$2^{3n/2}$
		5	$2^{3n/2}$	§ 3	$2^{3n/2}$
			$2^{7n/4}$	[23]	$2^{7n/4}$
CP based distinguisher		2	1	§ 3	1
			1	[22]	1
		3	$2^{n/2}$	§ 3	$2^{n/2}$
			$2^{n/2\dagger}$	[23]	$2^{n/2}$
		4	$2^{n/2}$	§ 3	$2^{n/2}$
			$2^{n\dagger}$	[23]	2^{2n}
		5	$2^{3n/2}$	§ 3	$2^{3n/2}$
			$2^{3n/2}$	[23]	$2^{3n/2}$
		6	2^{2n}	§ 3	2^{2n}
Key recovery	$2n$	4	$2^{n/2}$	§ 5.1	$n2^n$
		6	$2^{3n/2}$	§ 5.1	$2^{3n/2}$
	$4n$	4	$2^{n/2}$	§ 5.1	$n2^n$
		5	$2^{n/2}$	§ 5.1	$n2^{2n}$
		6	$2^{n/2}$	§ 5.1	$n2^{3n}$
			$2^{3n/2}$	§ 5.1	$2^{3n/2}$
		7	$2^{3n/2}$	§ 5.1	$n2^{2n}$
		8	$2^{3n/2}$	§ 5.1	$n2^{3n}$

† The results are given for bijective round functions.

If the test statistic t_{real} (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} is the rate of correctly identified positive instances, while the false positive rate P_{F} is

the rate of negative cases incorrectly identified as positive. The data complexity required for a distinguishing attack to achieve specific values of P_S and P_F depends on the statistical test used. Adversaries aim to maximize the advantage of the attack, which is defined as $\text{Adv} = P_S - P_F$ where $P_S > P_F$.

A distinguisher can be extended into a message-recovery attack by following the formalization of Bellare et al. [5]. In this model, an adversary queries the encryption of multiple distinct message pairs associated with a secret message and uses the resulting ciphertexts to infer information about that secret message.

2.2 Linear approximations

In linear cryptanalysis [24], one attempts to find 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. The correlation matrix of the function F_i is denoted by C^{F_i} [9], which is a $2^n \times 2^m$ matrix with coordinates

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

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

A function $F : \mathbb{F}_2^n \to \mathbb{F}_2^n$ is called iterative if there are r vectorial Boolean functions such that $F = F_r \circ F_{r-1} \circ \cdots \circ F_1$. To analyze the linear approximations of iterative functions, we can specify the input and output mask for each function F_i: 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 function F_i has input mask α_i and output mask α_{i+1}. Then, the correlation of the trail is the product of the correlations of the individual functions F_i. 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, as follows:

$$C^{\mathsf{F}}_{\alpha_{r+1},\alpha_1} = \sum_{\alpha_2,\ldots,\alpha_r} \prod_{i=1}^{r} C^{F_i}_{\alpha_{i+1},\alpha_i} \,. \tag{1}$$

A linear distinguisher operates by estimating the correlation of the approximation using q pairs of known plaintexts and ciphertexts. The data complexity of the distinguisher must be sufficiently large to ensure that the distributions of t_{real} and t_{ideal} exhibit a significant difference in their means. For a successful attack, the data complexity must be sufficiently large to satisfy the inequality $\mu_{\text{real}} - \mu_{\text{ideal}} \gg \sigma_{\text{ideal}}$. However, the mean μ and standard deviation σ of the corresponding distribution vary depending on the statistical test, leading to different data complexities.

2.3 Multidimensional linear cryptanalysis

One extension of linear cryptanalysis is multidimensional linear cryptanalysis. It is based on multiple linear approximations whose masks collectively form a

vector space Λ. The data capacity of the multidimensional linear approximation Λ over the function $F : \mathbb{F}_2^m \to \mathbb{F}_2^n$, defined as

$$\mathsf{Cap}(\Lambda) = \sum_{(u,v)\in\Lambda\setminus\{(0,0)\}} (C_{v,u}^F)^2.$$

where $C_{v,u}^F$ denotes the correlation of the linear approximation (u, v).

The main theoretical result of multidimensional linear cryptanalysis is Theorem 1, also known as the Poisson summation formula in Fourier analysis. In the context of cryptanalysis, the underlying idea is originally from [3] and is further discussed in [11]. However, the formulation given in Theorem 1 differs in that it applies to an arbitrary vector space of masks. Consequently, using this version requires a change of basis to express the problem in the general form needed for multidimensional attacks.

Theorem 1 (Theorem 4.1. [6]). *Let* $\mathbf{z}$ *be a random variable on* $\mathbb{F}_2^d$*,* $V \subseteq \mathbb{F}_2^d$ *a vector space, and* $V^\perp$ *the orthogonal complement of the subspace* V*. For all* $\eta \in \mathbb{F}_2^d$ *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* $v^T\mathbf{z}$*.*

For multidimensional linear cryptanalysis, Theorem 1 can be applied to the vector space $V = \Lambda$ and the random variable $\mathbf{z} = (\mathbf{x}, \mathrm{F}(\mathbf{x}))$, where $\mathbf{x}$ is uniform random on $\mathbb{F}_2^n$. This gives the following result [7]:

$$\Pr\left((\mathbf{x}, F(\mathbf{x})) = (t_1, t_2) \bmod \Lambda^\perp\right) = \frac{1}{|\Lambda|} \sum_{(u,v)\in\Lambda} (-1)^{u^Tt_1+v^Tt_2} C_{v,u}^F\,. \tag{2}$$

where $\Lambda^\perp = \left\{(\mathbf{x}, F(\mathbf{x})) \in \mathbb{F}_2^n \times \mathbb{F}_2^n \mid u^T\mathbf{x} + v^TF(\mathbf{x}) = 0 \text{ for all } (u, v) \in \Lambda\right\}$. In other words, the correlations $C_{v,u}^{\mathrm{F}}$ with $(u, v) \in \Lambda$ completely determine the probability distribution of the linear projections of plaintext and ciphertext, which is a linear map from $\mathbb{F}_2^n \times \mathbb{F}_2^n$ to $\mathbb{F}_2^n \times \mathbb{F}_2^n/\Lambda^\perp$. The quotient space $(\mathbb{F}_2^n \times \mathbb{F}_2^n)/\Lambda^\perp$ is a vector space of the same dimension as Λ. In light of Equation (2), instead of estimating the correlations of linear approximations in Λ, one can estimate the probability distribution of $\pi_{\Lambda^\perp}(\mathbf{x}, F(\mathbf{x}))$ for uniform random $\mathbf{x}$.

A multidimensional linear approximation Λ can be used to construct a distinguisher based on Pearson's χ^2 test. The use of this test in cryptanalysis goes back to Vaudenay [31]. The corresponding data complexity is [13,7]

$$q \approx \frac{\sqrt{|\Lambda|}}{\mathsf{Cap}(\Lambda)}\,. \tag{3}$$

In this setting, the relevant capacity captured by the difference between the real and ideal capacities $\mathsf{Cap}_{\mathrm{real}}(\Lambda) - \mathsf{Cap}_{\mathrm{ideal}}(\Lambda)$. In general, the ideal capacity is negligible, so the overall capacity is reduced to the real capacity.

Further details on the asymptotic data complexities of the distinguishers are provided in Section A. Equation (3) represents the expected data complexity, but a more precise estimation can be obtained for a fixed false positive probability P_{F} and a given success probability P_{S}. The χ^2 test statistic is calculated as the sum of the squared differences between the observed frequencies $T[z]$ of values z in $\mathbb{F}_2^n$ and their expected frequencies (uniform distribution), divided by the expected frequencies:

$$\chi^2 = \sum_{z \in \mathbb{F}_2^n} \frac{\left(T[z] - q/2^n\right)^2}{q/2^n} \,. \tag{4}$$

By the central limit theorem, the $\boldsymbol{\chi}^2_{\mathrm{ideal}}$ asymptotically approaches the normal distribution $\mathcal{N}(\nu, 2\nu)$ as 2^n and q approach infinity. The false-positive rate is calculated by

$$P_F = \Pr\left[\chi^2_{\mathrm{ideal}} \geq t\right] = \Pr\left[\underbrace{\frac{\chi^2_{\mathrm{ideal}} - \nu}{2\nu}}_{\mathbf{Z} \sim \mathcal{N}(0,1)} \geq \frac{t-\nu}{2\nu}\right] = 1 - \Phi\left[\frac{t-\nu}{2\nu}\right] ,$$

where $\Phi(x)$ is the cumulative distribution function. Hence, the threshold value t is $t = \nu + \sqrt{2\nu}\,\Phi^{-1}(1 - P_{\mathrm{F}})$. Once t is set for a fixed P_{F}, the success probability can be calculated by how often the χ^2_{real} statistic (the χ^2 statistic based on samples from the real cipher) exceeds t for a given data complexity.

2.4 Lai-Massey Structure

The Lai-Massey framework employs addition and subtraction operations in a finite Abelian group G and incorporates an orthomorphism permutation denoted by $\sigma : G \to G$. That is $x \to \sigma(x)$ and $x \to \sigma(x) - x$ are both permutations. The FOX [16] algorithm uses addition and subtraction operations in the additive group $G = \mathbb{F}_2^n$, and we have

$$\sigma(x_1, x_2) = (x_2 \parallel x_1 + x_2) \,.$$

This study focuses on attacking the FOX-like Lai-Massey structure, as shown in Figure 1. For clarity, we provide a summary of the targeted Lai–Massey ciphers. However, the details are not essential for this work.

FOX family of block cipher FOX is available in two versions, with a variable number of rounds depending on the key size. The first version, denoted as FOX64/k/r, operates on 64-bit blocks and supports key lengths that are multiples of 8 bits, with a maximum length of 256 bits. The second version, FOX128/k/r, uses a 128-bit block size and allows the same key lengths. The designers recommend using 16 rounds for FOX64 with a 128-bit key and FOX128 with a 256-bit key.

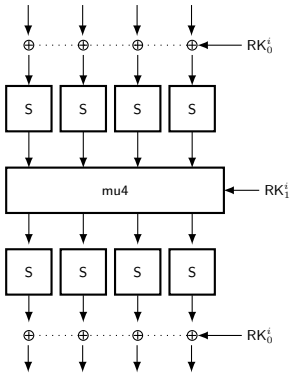


Fig. 2: The round function of FOX64.

The round function of FOX64/k/r consists of three main components: a substitution layer S, a diffusion layer $\mathsf{mu4}$ based on matrix multiplication, and a round key addition (see Figure 2). Formally, the function takes a 32-bit input and a 64-bit round key $\mathsf{RK}_{(64)} = \mathsf{RK}_{0(32)} \| \mathsf{RK}_{1(32)}$, and produces a 32-bit output. The round function of FOX128/k/r is structurally analogous. It takes a 64-bit input and a 128-bit round key $\mathsf{RK}_{(128)} = \mathsf{RK}_{0(64)} \| \mathsf{RK}_{1(64)}$, and returns a 64-bit output.

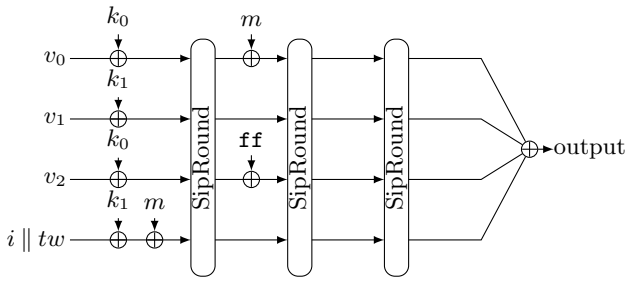


Fig. 3: The round function of SPC.

SPC block cipher SPC [1] is a tweakable block cipher that employs the FOX-style Lai-Massey architecture with a 128-bit key and block size. The round function is a variant of the SipHash-1-2 function [2]. It processes a 128-bit key $k = k_0 \| k_1$, an 8-bit round counter i, a 56-bit tweak tw, 64-bit constant values v_i and a 64-bit input $m$tance as seen in Figure 3. SPC is a 4-round Lai-Massey construction.

3 Generic multidimensional linear distinguisher

The analysis in this section is generic, *i.e.* we assume that the round functions $F_1, F_2, \ldots, F_r$ have been chosen independently and uniformly at random. Figure 4 represents a 2-round iterative trail, which extends to 3-round and 5-round trails as shown in Figure 5a and Figure 5b, respectively. In FOX-style Lai-Massey schemes, the map σ is a linear operation. For a state $x = (x_1, x_2)$, the σ is defined as $\sigma(x_1, x_2) = (x_2, x_1 \oplus x_2)$. Equivalently, σ can be expressed as the matrix–vector product Ax, where

$$A = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} \quad \text{and} \quad x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \implies Ax = \begin{bmatrix} x_1' \\ x_2' \end{bmatrix}.$$

Note that $A^3 = \mathbb{I}$. Therefore, applying σ three times returns to the original value, *i.e.* $\sigma(x) = y$, $\sigma(y) = z$, and $\sigma(z) = x$. Hence, the permutation σ has order 3.

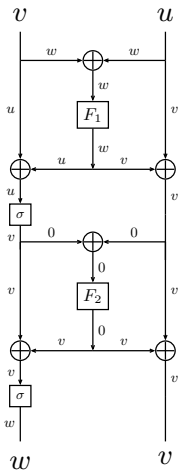


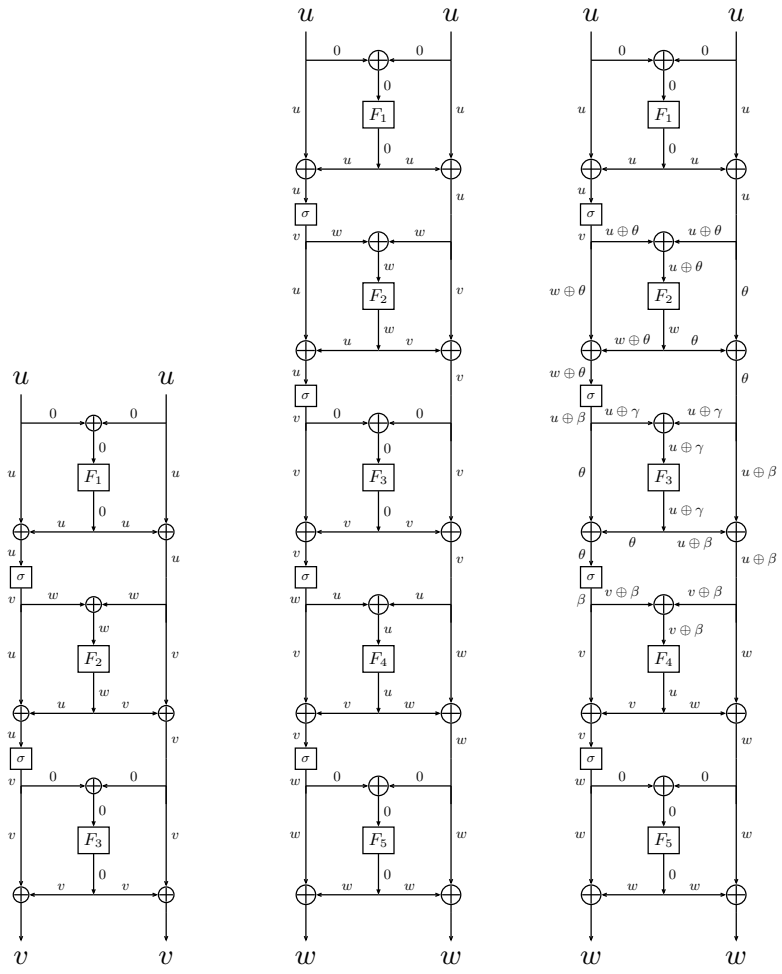
Fig. 4: Two-round iterative linear trail for a Lai-Massey scheme.

3.1 Linear trails

The relationship between masks u, v and w in Figure 4 can be determined as follows:

$$u^T x \oplus v^T \sigma(x) = (u \oplus \mathrm{A}^{\mathrm{T}} v)^T x = 0 \implies u \oplus \mathrm{A}^{\mathrm{T}} v = 0,$$

for all $x \in \mathbb{F}_2^n$. Since $\mathrm{A}^{\mathrm{T}} = \mathrm{A}$, we obtain $v = \mathrm{A}^{-1} u$, which implies $v = \sigma^{-1}(u)$. Similarly, we have $w = \sigma^{-1}(v)$ and $u = \sigma^{-1}(w)$ where $w = u \oplus v$. To compute the correlation of the multidimensional linear trails, we use the following result due to Daemen and Rijmen [10].



(a) 3-round linear trail (b) 5-round linear trail (c) 5-round linear trail with 3 active round functions where $\theta \oplus \beta = \gamma$.

Fig. 5: Multidimensional linear trails for a Lai-Massey scheme in the KPA model.

Theorem 2. *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 an arbitrary input and output mask over the uniformly random function F_{2i}. By Theorem 2, $\mathbf{c}_i$ is a random variable with an asymptotically normal distribution centered around

zero and with standard deviation $2^{-n/2}$. Then, the correlation of the linear trail, based on an iterative trail of $r \geq 2$ rounds in Figure 5, is equal to

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

Here w is the input and output mask of the first active round function, and $u = \sigma^{-1}(w)$ and $v = u \oplus w$. Hence, w determines the trail and the masks u and v uniquely. We therefore refer to $\mathbf{c}(w)$ as the correlation of the trail with $\lfloor r/2 \rfloor$ active round functions. Note that, in the expression above, we round down $r/2$ since there is one active round function in every two rounds. This provides a good approximation for the correlation of the corresponding r-round linear approximation. Since the random variables $c_1, c_2, \cdots, c_{\lfloor r/2 \rfloor}$ are independent by the strong assumption that the random functions $F_1, \ldots, F_r$ are independent, the variance of $\mathbf{c}(w)$ is equal to $\mathbb{E}\big(\mathbf{c}^2(w)\big) = 2^{-\lfloor r/2 \rfloor n}$.

5-round trails. The correlation of the corresponding linear approximation can be written as a sum over all linear trails, which have the same input and output masks

$$\mathbf{c}(w) = \sum_{\tau} X_\tau \,, \tag{5}$$

where each trail contribution X_τ is the product of correlations of the active round functions. For $r = 5$, $\mathbb{E}\big(\mathbf{c}^2(w)\big) = 2^{-2n}$, which is too low to obtain a valid attack. However, there are extra trails with the same input and output masks for 5-round Lai-Massey ciphers. The linear trails are obtained by the choice of the input mask for the fork operation; hence, different masks can be chosen as input masks for middle-round functions, as shown in Figure 5c. Hence, 5-round Lai–Massey construction with three active round functions, each trail contributes $X_\tau = \mathbf{c}_2 \cdot \mathbf{c}_3 \cdot \mathbf{c}_4$, where $\mathbf{c}_i$ denotes the correlation of the linear approximation through round function F_i. The squared correlation of the approximation satisfies

$$\mathbb{E}\left(\mathbf{c}(w)^2\right) = \mathbb{E}\left(\left(\sum_{\tau} X_\tau\right)^2\right) = \sum_{\tau} \mathbb{E}\left(X_\tau^2\right) + \sum_{\tau \neq \tau'} \mathbb{E}\left(X_\tau X_{\tau'}\right). \tag{6}$$

The approximation for the 5-round Lai–Massey scheme with the input mask (u, u) and the output mask (w, w) receives contributions from one principal trail shown in Figure 5b, and approximately 2^n additional trails, because of 2^n possible values for θ, described in Figure 5c. Since the trails in Figure 5b and Figure 5c are dominant, the sum of the correlations of these trails is a good estimate for the correlation of the corresponding approximation. While other linear trails exist, their individual correlations are sufficiently small that their aggregate contribution remains negligible relative to the dominant terms [6]. Therefore, the total contribution equals

$$\sum_{\tau} \mathbb{E}\left(X_\tau^2\right) \approx 2^{-2n} + 2^n \cdot 2^{-3n} = 2 \cdot 2^{-2n}. \tag{7}$$

For two distinct parameter choices, the corresponding trails differ in at least one active-round mask pair (see Figure 5c). The covariance between the correlations of distinct trails is zero for independent uniform random functions [7], so that the term $\mathbb{E}\left(X_\tau X_{\tau'}\right)$ becomes 0. Hence, we have $\mathbb{E}\left(\mathbf{c}^2(w)\right) \approx 2 \cdot 2^{-2n}$. For $r \leq 4$, contributions of these additional trails are negligible.

3.2 Distinguisher in the KPA model

The multidimensional linear distinguisher presented in this section is derived under the known-plaintext attack model, *i.e.*, attackers have access to q plaintexts and their corresponding ciphertexts. A multidimensional linear distinguisher can be constructed by simultaneously estimating the correlations for all choices of mask w. Note that there is a unique u and a unique v for every $w \in \mathbb{F}_2^n$. Specifically, the vector space of input-output mask pairs for three rounds is given by

$$\Lambda = \left\{ \big((u,u),(v,v)\big) \mid w \in \mathbb{F}_2^n \right\}.$$

As discussed in Section 2.3, the data-complexity of multidimensional linear distinguishers depends on the sum of the squared correlations of the approximations in Λ. The sum of the squared correlations is well approximated by the sum of the squares of the trail correlations. Since the distribution of the sum of squared correlations is strongly concentrated around its mean, the average is a reasonable estimate

$$\mathbb{E}\left(\sum_{w \in \mathbb{F}_2^n \setminus \{0\}} \mathbf{c}^2(w)\right) = (2^n - 1)\mathbb{E}\Big(\mathbf{c}^2(w)\Big) \approx 2^{n - n\lfloor r/2 \rfloor}.$$

As demonstrated by the linear trail in Figure 4, the input mask of even-indexed round functions is non-zero, leading to an unknown probability distribution. In such a scenario, as discussed in Section 2.3, the χ^2 test provides the appropriate method for constructing a multidimensional linear distinguisher. Hence, by Equation (3), the data complexity of the multidimensional χ^2 distinguisher is roughly

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

for $r \leq 4$.

5-round distinguisher. For the 5-round scenario, it is already explained in Section 3.1 that $\mathbb{E}\big(\mathbf{c}^2(w)\big) \approx 2(2^{-2n})$, which implies

$$\mathsf{Cap}_{\text{real}}(\Lambda) = \sum_{w \in \mathbb{F}_2^n \setminus \{0\}} \mathbf{c}(w)^2 = 2 \cdot 2^{-2n} |\Lambda|. \tag{8}$$

It has already been mentioned in Section 2.3, the data complexity is inversely proportional to $\mathsf{Cap}_{\text{real}}(\Lambda) - \mathsf{Cap}_{\text{ideal}}(\Lambda)$. For $r \leq 4$, the ideal capacity is negligible, i.e., $\mathsf{Cap}(\Lambda) \approx \mathsf{Cap}_{\text{real}}(\Lambda)$. On the other hand, the additional trails are

essential for the correctness of the 5-round attack, so the data complexity becomes

$$q \approx \frac{\sqrt{|\Lambda|}}{\mathsf{Cap}_{\text{real}}(\Lambda) - \mathsf{Cap}_{\text{ideal}}(\Lambda)} ,$$

where $\mathsf{Cap}_{\text{real}}(\Lambda) - \mathsf{Cap}_{\text{ideal}}(\Lambda) \approx 2^{-2n}|\Lambda|$ since $\mathsf{Cap}_{\text{ideal}}(\Lambda) = 2^{-2n}|\Lambda|$. The resulting data-complexity becomes $q = \mathcal{O}(2^{3n/2})$ since $|\Lambda| = 2^n$.

3.3 Distinguisher in the CPA model

The distinguishing strategy of Section 3.2 extends naturally to the chosen–plaintext (CPA) setting. By selecting plaintexts such that two Lai–Massey branches satisfy $x_R = x_L + c$ for some constant c, the attacker can control the XOR for an input of the first round and thereby skip the initial encryption layer for the distinguisher, as illustrated in Figure 6. Hence, in total, the attacker can skip 2 rounds before the first active round function. Under this condition, the expected capacity of the multidimensional approximation becomes

$$\mathbb{E}\left(\sum_{w \in \mathbb{F}_2^n \setminus \{0\}} \mathbf{c}^2(w)\right) \approx 2^{-n\lfloor (r-1)/2 \rfloor}|\Lambda| \approx 2^{n-n\lfloor (r-1)/2 \rfloor} . \tag{9}$$

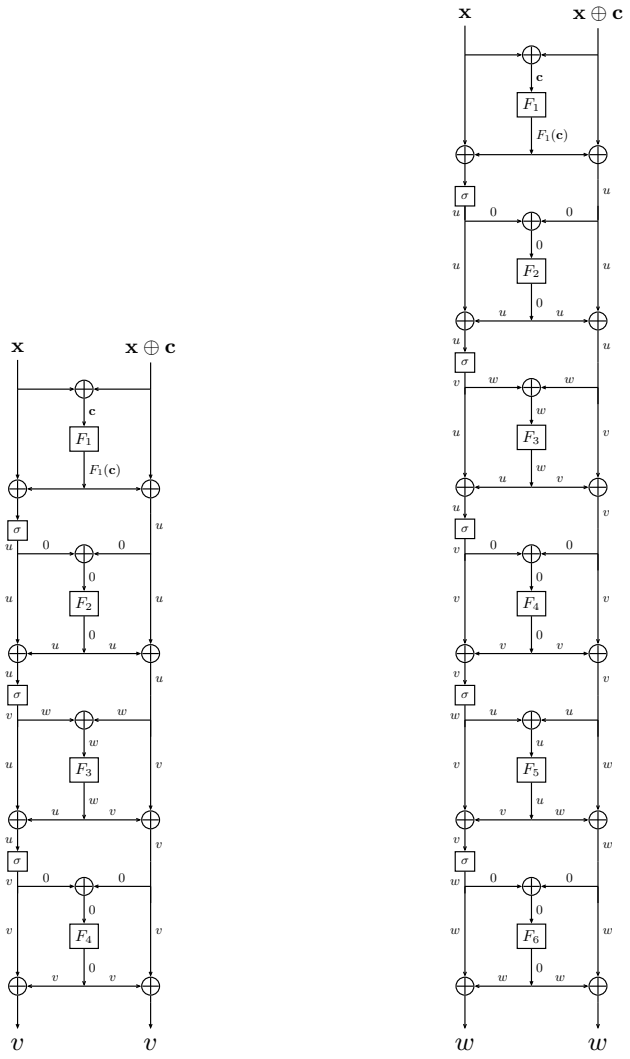
The data complexity of the distinguisher is

$$q = \mathcal{O}\left(2^{n\lfloor (r-1)/2 \rfloor - n/2}\right) .$$

This yields, in particular, a 4-round CPA distinguisher with the same data complexity as the 3-round KPA distinguisher. However, the CPA setting restricts the number of inputs to at most 2^n, which requires additional considerations for 6-round attacks. Hence, we must repeat the attack for several values of c.

Although different choices of c modify the value of $F_1(c)$ entering the second round, this modification is deterministic and affects both branches symmetrically. For any fixed c, the input to the 5-round distinguisher is therefore just shifted by a constant. Changing c simply produces different constant shifts of the same input. From the perspective of the multidimensional distinguisher, different choices of the constant correspond to using a larger set of approximations. The corresponding capacities therefore add, so that both the real and ideal capacities scale linearly with the number of distinct constants c. This follows from the same covariance argument used above: correlations associated with distinct trail decompositions are modeled as uncorrelated for independent uniform random functions [7], and hence the covariance between the correlations of distinct trails is zero.

6-round distinguisher: From the perspective of the multidimensional χ^2 distinguisher, using l distinct constants can be viewed as enlarging the set of observations from $|\Lambda|$ to $l|\Lambda|$.



(a) 4-round linear trail

(b) 6-round linear trail

Fig. 6: Multidimensional linear trails for the Lai–Massey scheme in the CPA setting with one prepended encryption round; x and c denote the plaintext and the constant.

However, this interpretation implicitly assumes that the corresponding statistics are independent. Hence, the data complexity is approximately

$$q \approx \frac{\sqrt{l|\Lambda|}}{\mathsf{Cap}(\Lambda)}\,, \tag{10}$$

where $\mathsf{Cap}(\Lambda) = \mathsf{Cap}_{\text{real}}(\Lambda) - \mathsf{Cap}_{\text{ideal}}(\Lambda)$. The key point is that, once the difference between two branches is fixed for the 6-round distinguisher, the squared correlation of linear approximations in Λ is the same as the case of the 5-round distinguisher. The first round can then be skipped without contributing to the overall correlation, and there are additional linear trails for the remaining 5 rounds. In particular, due to the mapping $x \mapsto (x, x + c)$ at the beginning of the construction, any two masks on the left and right branch whose sum equals the input mask on x yield a valid trail. Under this assumption, both the real and ideal capacities scale by a factor l. As a result, the real and ideal capacities satisfy

$$\mathsf{Cap}_{\text{real}}(\Lambda) \approx (2 \cdot 2^{-2n}) l|\Lambda|, \qquad \mathsf{Cap}_{\text{ideal}}(\Lambda) = (2^{-2n}) l|\Lambda|$$

based on Equation (8) and Equation (9), respectively. Hence, $\mathsf{Cap}(\Lambda) \approx 2^{-2n} l|\Lambda|$. Since at most 2^n plaintexts are available per constant c and the size of the vector space is 2^n, Equation (10) becomes

$$2^n \approx \frac{\sqrt{l|\Lambda|}}{2^{-2n} l|\Lambda|} \implies \sqrt{l} \approx \frac{1}{2^{-n/2}}\,.$$

Thus, using $l \approx 2^n$ different choices of c results in a data complexity of $lq \approx 2^{2n}$. This estimate assumes that the additional trails contribute approximately 2^{-n} to the capacity of each approximation.

Table 2: Theoretical data complexities of r-round distinguisher on Lai-Massey ciphers. The results in this work are based on the χ^2 test. All complexities are given up to constant factors.

r	KPA	Ref.	CPA	Ref.
2	$2^{n/2}$	§ 3	1	§ 3
	$2^{n/2}$	[23]	1	[22]
3	$2^{n/2}$	§ 3	$2^{n/2}$	§ 3
	2^{n}	[23]	$2^{n/2}$	[23]
4	$2^{3n/2}$	§ 3	$2^{n/2}$	§ 3
	$2^{3n/2}$	[23]	2^{n}	[23]
5	$2^{3n/2}$	§ 3	$2^{3n/2}$	§ 3
	$2^{7n/4}$	[23]	$2^{3n/2}$	[23]
6	—		2^{2n}	§ 3

3.4 Comparison with existing results

Despite some structural similarities, Lai–Massey and Feistel ciphers differ in that many classical attacks yield different results [23] or cannot be applied directly to the Lai–Massey structure. Patarin's attacks [27,28,29] on Feistel ciphers rely

on collision-based techniques, which are conceptually similar to the approach adopted in this work. Luo et al. [23] adopted Patarin's works to Lai-Massey ciphers, thereby making their work comparable to ours. In Table 2, we compare our results with those of [23] for the case of the bijective round function. Our contributions reduce the data complexity of the generic attacks. Specifically, we enhance the data complexity of 3- and 5-round generic attacks in the KPA model and the 4-round generic attack in the CPA model, as shown in Table 2. Furthermore, we present a 6-round generic distinguishing attack in the CPA model, which, to the best of our knowledge, covers most rounds.

3.5 Applications to SPC and FOX

We experimentally validate the asymptotic capacities using randomly generated round functions over $\mathbb{F}_2^{16}$ to reflect the generic setting. Experimental results show that capacities are approximately 1, 2^{16}, and 2^{32} for 3-, 5-, and 7-round distinguishers, respectively. The results are consistent with the theoretical predictions. This agreement supports the accuracy of our capacity estimates and the underlying assumptions used in the analysis.

We also apply χ^2 distinguishing attacks to the SPC and FOX64 ciphers. Recall that the data complexity of an r-round multidimensional distinguisher, with random functions over $\mathbb{F}_2^n$, based on the χ^2 test for Lai-Massey ciphers is approximately $2^{n\lfloor r/2 \rfloor - n/2}$ as stated in Section 3.2.

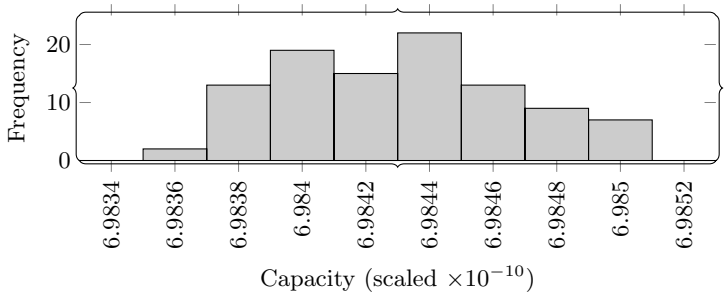
Algorithm 1: Computing the capacity of a multidimensional linear distinguisher for an r-round Lai-Massey cipher.

Input: input mask: $(u \parallel u)$
output mask: $(v \parallel v)$
Set $n \leftarrow 2^{32}$;
for $i = 0$ *to* n **do**
  $p_0[F_2(i)] \leftarrow p_0[F_2(i)] + 1$;
  ⋮
  $p_j[F_{2j}(i)] \leftarrow p_j[F_{2j}(i)] + 1$; //$j = r/2$
end
for $i = 0$ *to* j **do**
  $p_j \leftarrow FFT(p_j)$;
end
for $i = 1$ *to* n **do**
  $p_0 \leftarrow p_0[i]p_1[i] \dots p_j[i]$;
  $\text{Cap} \leftarrow \text{Cap} + p_0^2$;
end

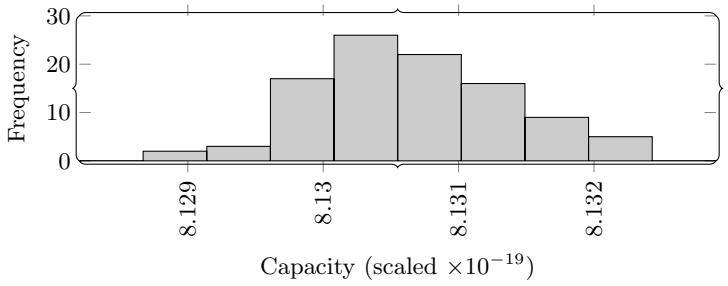
For the targeted ciphers, the number of approximations is $|\Lambda| = 2^{32}$. Assuming ideal round functions, the expected data complexities are 2^{16} for 2 and 3 rounds, and 2^{48} for 4 and 5 rounds by Section 3.2. We follow the capacity computation algorithm in Algorithm 1 to validate these values. Although Algorithm 1 supports varying round functions per round, FOX uses the same round function with different round keys, so we compute the distribution only once.

Specifically, all 2^{32} input values are encrypted, the output frequencies are stored in vectors, and the FFT is applied to vectors. Then, the sum of squared correlations is obtained by taking the pointwise product of the vectors. For a single active round function, the distribution can be computed without FFT; however, computing the pointwise product of Fourier transformations (equivalently, convolution of probability mass functions) becomes computationally more efficient when multiple active round functions exist.

Additionally, we evaluate the impact of varying key values on the data complexity by calculating the corresponding capacities for the reduced-round FOX cipher. For the 3-round cipher, it must be 1. For the 5-round cipher, capacities range from 6.98353×10^{-10} to 6.98506×10^{-10} as seen in Figure 7a, which are approximately $2^{30.41}$, corresponding to a data complexity of approximately $2^{46.41}$. For the 7-round case, capacities range from 8.1289×10^{-19} to 8.13235×10^{-19} as seen in Figure 7b, which are approximately $2^{60.09}$, indicating a data complexity near $2^{76.09}$.



(a) 5-round FOX cipher.



(b) 7-round FOX cipher.

Fig. 7: Frequency distribution of capacities for reduced-round FOX cipher evaluated across approximately 100 distinct keys.

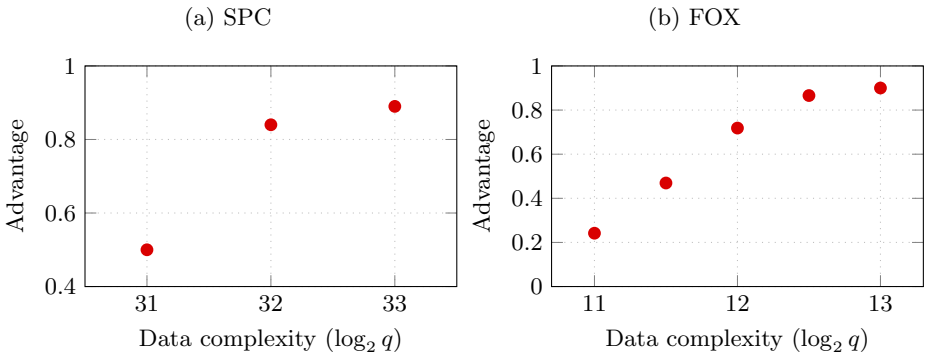
A more precise estimate of the data complexity can be obtained using the χ^2 test when the data complexity is not too high. Let (x_L^i, x_R^i) and (y_L^i, y_R^i) be

q plaintext and corresponding ciphertext samples, respectively, with the fixed tweak. To compute χ^2 test statistic of the distinguisher, first, we store the frequencies of $z_i = x_L^i \oplus x_R^i \oplus \sigma^{-1}(y_L^i \oplus y_R^i)$ in a vector T for all 2^{32} possible values of $x_L^i \oplus x_R^i$. Then, the χ^2 test statistic can be computed using Equation (4). As detailed in Section 2.3, to experimentally determine the success probability for a given false-positive rate, we define the decision threshold as

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

After obtaining the threshold value t using, we experimentally calculate the data-complexity for a given success probability by comparing 2^5 χ^2 test statistics with the corresponding threshold t. Figure 8 provides a summary of the experimental data complexities of 3-round multidimensional χ^2 distinguishers applied to SPC and FOX ciphers.

Fig. 8: Advantage as a function of data-complexity for 3-round χ^2 distinguishers where $P_{\mathrm{F}} = 0.1$.



4 Message recovery

The message recovery attack in this section is in the message-recovery model proposed by Bellare et al. [5]. Briefly, it is assumed that the adversary can non-adaptively query the encryption of distinct message pairs associated with a secret message.

We extend a 3-round distinguisher into a 4-round message recovery attack by prepending an additional round. The message recovery attack is initiated by prepending a single round of constant-value encryption before the distinguisher, ensuring that the 4-round attack maintains the same data complexity as the original 3-round distinguisher. The underlying principle of the attack is that changes in the plaintext affect the distribution of the ciphertext because it is not

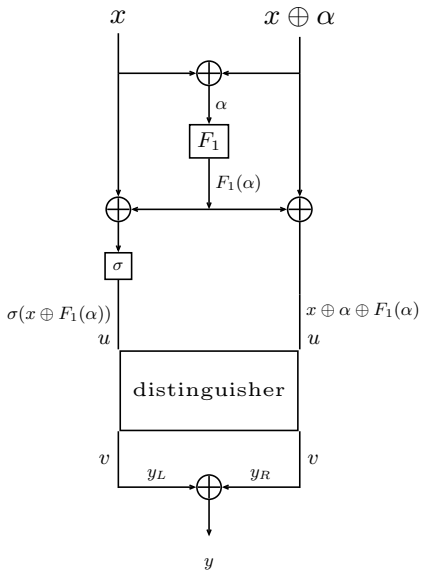
uniform. Thus, the objective is not full plaintext recovery, but distinguishing a difference between plaintext halves.

4.1 Distinguishing constant values

The distinguisher is based on comparing the ciphertext distributions resulting from two specific plaintexts, and the goal is to determine whether $\alpha_1 = \alpha_2$ by exploiting statistical differences in the ciphertext distributions where α_1 and α_2 are the differences between the two branches.

To extend the 3-round distinguisher to a 4-round message recovery attack using the same amount of data, we prepend one round to the distinguisher, fixing the input to the first round to a constant value, as illustrated in Figure 9.

Fig. 9: 4-round message recovery attack on FOX-style Lai-Massey schemes.



Let (x_L, x_R) and (y_L, y_R) be plaintext and corresponding ciphertext samples. Hence, $\alpha = x_L \oplus x_R$ and $y = y_L \oplus y_R$, *i.e.*, the XOR of two branches of the plaintext and ciphertext, respectively. Following the approach in [6], we have $c_2(u) \approx (-1)^{u^T \alpha} c_1(u)$ where $c_1(u)$ and $c_2(u)$ denote the correlation of the linear approximation. Then Theorem 1 implies that $p_2(y \oplus \alpha) \approx p_1(y)$. The attack proceeds by estimating the probability distribution of the sum of the left and right halves of the ciphertext in two scenarios when encrypting the modified plaintext with the constant values α_1 and α_2: $x^i \parallel (x^i \oplus \alpha_1)$, and $x^i \parallel (x^i \oplus \alpha_2)$ where $i \in 1, 2, \ldots q/2$. The goal is to distinguish the constants α_j. Let p_1 and p_2 denote the distributions of the sum of the left and right halves of the ciphertext

for the first and the second cases, respectively. Similarly, $c_1(u)$ and $c_2(u)$ denote the correlation of the linear approximation corresponding to the mask u when the plaintext is fixed to $x^i \,\|\, x^i \oplus \alpha_j$. The two considered functions are the same up to the XOR of two halves.

If the distributions are sufficiently close, then it is likely that $\alpha_1 = \alpha_2$; conversely, if the distributions differ, then $\alpha_1 \neq \alpha_2$. The data complexity of the 4-round message recovery attack on the Lai-Massey scheme is based on the data complexity of the 3-round distinguisher.

Algorithm 2: Computing the test statistic for distinguishing constant values in the 4-round algorithm.

Input: input : sample x^i uniformly at random
for $i = 0, j = 0$ *to* $q/2$ **do**
  $f_1[E_k(x^i, x^i \oplus \alpha_1)] \leftarrow f_1[E_k(x^i, x^i \oplus \alpha_1)] + 1$; //$y_L \oplus y_R \leftarrow E_k(x_L, x_R)$
  $f_2[E_k(x^j, x^j \oplus \alpha_2)] \leftarrow f_2[E_k(x^j, x^j \oplus \alpha_2)] + 1$;
end
for $i = 0$ *to* $q/2$ **do**
  $s \leftarrow \sum_i \frac{(f_1(z_i) - f_2(z_i))^2}{f_1(z_i) + f_2(z_i)}$;
end

While classical cryptanalysis often uses the one-sample χ^2 test to detect biases relative to uniformity, the two-sample test provides an extension to compare two empirical distributions directly. Hence, we consider the two-sample χ^2 test [30] in our statistical evaluation. The χ^2 divergence between the two distributions is computed using element-wise differences. So, the test statistic simplifies to

$$\boldsymbol{s}_{\alpha'} = \sum_i \frac{(f_1(z_i) - f_2(z_i))^2}{f_1(z_i) + f_2(z_i)}\,, \tag{11}$$

where $\alpha_1 = \alpha_2$. Once the threshold t is set for a given false positive probability P_{F} as $t = (2^n - 1) + \sqrt{2(2^n - 1)}\,\Phi^{-1}(1 - P_{\mathrm{F}})$, the success probability can be empirically evaluated for a given data complexity by comparing the test statistic $\boldsymbol{s}_{\alpha'}$ with this threshold.

In the case where $\alpha' \neq \alpha$, one can follow the same strategy given in [6], where it is shown that

$$\mathbb{E}\,\boldsymbol{s}_{\alpha'} - \mathbb{E}\,\boldsymbol{s}_{\alpha} \approx q \sum_u \mathbb{E}\,|c_1(u)|^2\,.$$

If the data complexity satisfies $q \gg \sqrt{2^n} / \sum_u \mathbb{E}\,|\boldsymbol{c}_1(u)|^2$ then $\mathbb{E}\,\boldsymbol{s}_{\alpha'} - \mathbb{E}\,\boldsymbol{s}_{\alpha} \gg \sqrt{2^n}$. In this case, the distribution of $\mathbb{E}\,\boldsymbol{s}_{\alpha'}$ converges asymptotically to the χ^2 distribution with $2^n - 1$ degrees of freedom. To achieve a low false-positive probability, the decision threshold t must be chosen larger than the standard deviation of the ideal distribution, i.e., $t \gg 2^{n/2}$. This shows that $\mathcal{O}\big(2^{n\lfloor r/2 \rfloor - n/2}\big)$ data is sufficient.

4.2 Application to SPC

Note that the expected data complexity of the 3-round multidimensional distinguisher for the Lai-Massey scheme is $2^{n/2}$ over round functions $F : \mathbb{F}_2^n \to \mathbb{F}_2^n$. Hence, the expected data complexity of the message recovery attack on the 4-round Lai-Massey cipher, outlined in Algorithm 2, is $2^{n/2}$. The algorithm initializes two vectors f_1 and f_2, then encrypts $q/2$ plaintext samples of each case using the 4-round Lai-Massey cipher, where $\alpha_j = x_L + x_R$, and stores the frequencies of the sum of two halves of ciphertexts in f_1 and f_2, respectively.

To check whether $\alpha_1 = \alpha_2$, the test statistic $\boldsymbol{s}_{\alpha'}$ is computed using Equation (11). For a chosen false positive probability, a threshold t is determined accordingly as discussed in Section 4.1. The success probability is then evaluated empirically by comparing the computed test statistic $\boldsymbol{s}_{\alpha'}$ to this threshold t for a given data complexity. The experimental results on the data complexity of SPC are shown in Figure 10.

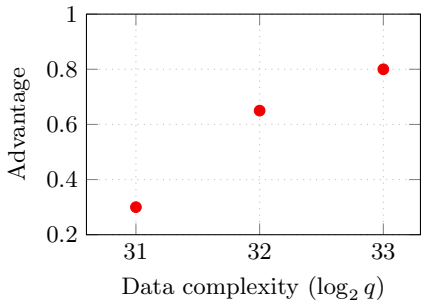


Fig. 10: Advantage as a function of data-complexity of the message recovery attack on SPC.

5 Key recovery

The distinguisher can be extended to a key recovery attack by adding one or more rounds. Let r' represent the number of additional rounds added to the r-round distinguisher, resulting in a total of $r + r'$ rounds in the key recovery process. The correlation for each key guess can be estimated by decrypting r' rounds. A test statistic is then calculated for each guess, and it is assumed that when a correct key is used, the test statistic can be distinguished. To increase the number of guessed key bits (ℓ), it is essential to maintain a low false-positive rate. The candidates can be reduced to one or a few keys by ensuring $P_{\mathrm{F}} \times 2^{\ell} \approx 1$. The total time complexity of the naive key-recovery method is $\mathcal{O}(q2^{\ell})$, where q is the data complexity required to compute the statistic. Matsui's technique reduces the time complexity to $\mathcal{O}(q + 2^{2\ell})$ by separating the computation into

two parts—one dependent on the data and one on the round key. Indeed, the estimated correlation for any key guess k can be expressed as

$$\widehat{c}_k = \frac{1}{q}\sum_{i=1}^{q}(-1)^{u^\top x_i + v^\top \mathrm{G}_{r'}^{-1}(k,y_i)} = \frac{1}{q}\sum_{z\in\mathbb{F}_2^\ell}(-1)^{v^\top \mathrm{G}_{r'}^{-1}(k,z)}\sum_{i=1}^{q}(-1)^{u^\top x_i}\delta_{z,y_i}\,,$$

where $\mathrm{G}_{r'}$ corresponds to the appended or prepended rounds. In multidimensional linear cryptanalysis, the same computation must be performed for every approximation. The notion of separability is therefore crucial, as without it the entire expression must be recomputed for every pair of masks.

5.1 Generic key recovery attack

In the generic setting, round functions are modeled as independent uniformly random permutations whose inputs depend on secret keys. We assume that each round function is parameterized as $F_i(x) = F(x + k_i)$, where k_i denotes the round subkey and F is a uniformly random permutation. In particular, we focus on recovering the value $F_1(c)$ first and then the corresponding round key.

The generic multidimensional distinguisher in the KPA model, which is described in Section 3, can be extended to obtain a generic multidimensional key recovery attack. Although slightly higher data complexity may be expected in practice, the results derived for the r-round distinguishers in Section 3 can be used to estimate the data complexity of an $r + r'$ rounds key recovery attack similar to the strategy in [26] with improved time complexities compared to the naive and Matsui's approaches. Generic attacks are often analyzed in terms of both data and time complexity. For example, Patarin's generic attacks on Feistel schemes [28], and similar considerations appear in works of Leurent et al. [21,4]

Recovering $F_1(c)$***:*** In this section, the aim is to recover $F_1(c)$ by constructing inputs where the right branch is set as the XOR of the left branch with the constant c. Rewriting the expression into a matrix–vector product precisely enables an efficient extension, which motivates defining a matrix M with coordinates indexed by $(F_1(c), k)$ where $F_1(c) \in \mathbb{F}_2^n$ and a vector w indexed by tested values z so that the vector of estimated correlations is calculated as matrix-vector product. In this formulation, the entries $M_{F_1(c),z}$ capture the key-dependent values, while the vector w depends only on the observed data. A variant of the FFT method [8] can be used to improve the time complexity. By adding an extra round before the distinguisher as shown in Figure 9 and encrypting the values $x^i \,\|\, (x^i + c)$, the χ^2 statistic for the key-dependent constant $F_1(c)$ is equal to

$$
\begin{aligned}
\chi^2_{F_1(c)} &= \sum_{i=1}^{q} \chi^2\big(\sigma(x^i \oplus F_1(c)) \oplus x^i \oplus c \oplus F_1(c) \oplus \sigma^{-1}(y_L^i) \oplus \sigma^{-1}(y_R^i)\big) \\
&= \sum_{i=1}^{q} \sum_{z \in \mathbb{F}_2^n} \chi^2\big(\sigma(x^i) \oplus \sigma^{-1}(y_L^i) \oplus x^i \oplus \sigma^{-1}(y_R^i) \oplus \sigma(F_1(c)) \oplus c \oplus F_1(c)\big) \delta_{z,(\sigma(x^i) \oplus \sigma^{-1}(y_L^i) \oplus x^i \oplus \sigma^{-1}(y_R^i))} \\
&= \sum_{z \in \mathbb{F}_2^n} \chi^2\big(z \oplus \sigma(F_1(c)) \oplus c \oplus F_1(c)\big) \sum_{i=1}^{q} \delta_{z,(\sigma(x^i) \oplus \sigma^{-1}(y_L^i) \oplus x^i \oplus \sigma^{-1}(y_R^i))} \\
&= \sum_{z \in \mathbb{F}_2^n} M_{F_1(c),z}\, w_z \,,
\end{aligned}
\tag{12}
$$

where w is a vector containing the occurrences of observed values

$$z = \sigma(x^i) \oplus \sigma^{-1}(y_L^i) \oplus x^i \oplus \sigma^{-1}(y_R^i)\,,$$

and M is a matrix with entries

$$M_{F_1(c),z} = \chi^2(z \oplus \sigma(F_1(c)) \oplus c \oplus F_1(c))\,.$$

We have $(u^T\sigma(x_i \oplus F_1(c))) \oplus (u^T(x_i \oplus c \oplus F_1(c)))$. Since $v = \sigma^{-1}(u)$, we have

$$v^T y_L^i \oplus v^T y_R^i = (u^T\sigma^{-1}(y_L^i) \oplus u^T\sigma^{-1}(y_{i_R}))\,.$$

Table 3: Generic key recovery attack on $(r + r')$-round Lai-Massey schemes with $2n$-bit blocks, n-bit keys and a k-bit master key.

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

Without any optimizations, the time complexity of computing the matrix M is $\mathcal{O}(2^{2n})$. However, we can leverage the circulant structure of M using the

FFT method to reduce the complexity of the matrix-vector product. M is a circulant matrix since the key k is XORed directly to the plaintext, so it suffices to compute $M_{0,z}$. We compute the test statistic by performing the matrix-vector product through the pointwise multiplication of the FFT of the first row of M and the FFT of w, followed by an inverse Fourier transform. This reduces the complexity of calculating matrix M to $\mathcal{O}(n2^n)$. Therefore, the total complexity is $\mathcal{O}(n2^n + q)$, which simplifies to $\mathcal{O}(n2^n)$ if $q \leq n2^n$. Consequently, the overall time complexity is strictly improved over both the naive method $\mathcal{O}(q2^n)$ and Matsui's approach $\mathcal{O}(q + 2^{2n})$ as seen in Table 3.

Recovering the round key of F_1***:*** The attack can continue with recovering the round key of F_1. It is essential to maintain a low false-positive rate to reduce the number of key candidates based on $P_{\mathrm{F}}2^\ell$ where ℓ is the number of key bits in the round function. Then, the time complexity becomes $\mathcal{O}(n2^n + q + 2^\ell)$ where $P_{\mathrm{F}} \approx 1/2^\ell$.

Key recovery attack with additional rounds: For a key-recovery attack that extends the distinguisher with r' additional rounds, these operations must be performed for each key guess. However, instead of guessing the key for each candidate, a lookup table can be precomputed to map the round-function outputs to their corresponding subkey candidates. Then each candidate is resolved by a single table lookup with 2^ℓ entries, each of size n bits, and time complexity $\mathcal{O}(2^\ell)$. The overall complexity is then dominated by three components: the recovery of key-dependent value $F_1(c)$, the recovery of the round key of F_1, and the recovery of the additional rounds' keys. This yields $\mathcal{O}(n2^{n(r'-1)\ell} + q + 2^\ell)$. In Section 3.5, r can be either 3 or 5, corresponding to data complexities of $\mathcal{O}(2^{n/2})$ and $\mathcal{O}(2^{3n/2})$, respectively. Substituting these values, the time complexity simplifies to

$$T = \begin{cases} \mathcal{O}\left(n2^{\,n+(r'-1)\ell} + 2^\ell\right), & r = 3, \\ \mathcal{O}\left(n2^{\,n+(r'-1)\ell} + 2^{3n/2} + 2^\ell\right), & r = 5, \end{cases}$$

where $P_{\mathrm{F}} \approx 1/2^{r'\ell}$. The term 2^ℓ accounts for the precomputation of the lookup table. When fixing the guessed key size ℓ, these expressions can be further simplified. For the $(3+r')$-round key-recovery attack, the time complexity is $\mathcal{O}(n2^{nr'})$ for $\ell = n$, and $\mathcal{O}(n2^{2nr'-n} + 2^{2n})$ for $\ell = 2n$. Similarly, for the $(5 + r')$-round key-recovery attack, the time complexity is $\mathcal{O}(n2^{nr'} + 2^{3n/2})$ for $\ell = n$, and $\mathcal{O}(n2^{2nr'-n} + 2^{2n})$ for $\ell = 2n$.

The maximum feasible number of additional rounds, $r'_{\max}$, depends on the master key size, k, and the number of guessed key bits, ℓ. Since the time complexity is upper-bounded by exhaustive key search, i.e., $T \leq 2^k$, the value of r' can be computed accordingly. The results for $r'_{\max}$ are provided in Table 4.

Table 4: Maximum feasible $r'_{\max}$ and time complexities for key recovery attacks.

r	ℓ	k	$r'_{\max}$	T
3 or 5	n	$2n$	1	$n2^{n}$
		$4n$	3	$n2^{3n}$
	$2n$	$4n$	2	$n2^{3n}$

5.2 Application to FOX64

We can apply our key recovery technique described in Section 5.1 to the FOX64 cipher. FOX64 operates on a 64-bit block and uses a round key structured as $\mathsf{RK}_{(64)} = \mathsf{RK}_{0(32)} \parallel \mathsf{RK}_{1(32)}$, as illustrated in Figure 2. Hence, $\ell = 2n$ where $n = 32$, resulting in time complexity of $2^{5+32r'} + 2^{64}$ for $3 + r'$ and $5 + r'$ distinguishers. For a block cipher with k-bit key size, a practical key recovery attack is expected to have a time complexity below 2^k, corresponding to 2^{64} and 2^{128} for FOX64/64/16 and FOX64/128/16, respectively. Hence, valid values for r are 3 and 5, with $r' = 1$ for FOX64/64/16 and $r' \in \{1, 2\}$ for FOX64/128/16. This yields key recovery attacks on 4- and 6-round FOX64/64/16 with expected time complexity 2^{64}. Similarly, a 7-round key recovery attack is possible for FOX64/128/16, with a time complexity of 2^{101}, where $r = 5$. The data and time complexities of these key recovery attacks on FOX64 are summarized in Table 5. More details about the implementation of the FFT-based key recovery attack are provided in Section B.

Table 5: Key recovery results of $r + r'$-round FOX64.

$r + r'$	r'	Data	Ref.	Work effort
4	1	$2^{15.8}$	§ 5.1	2^{64}
		2^{9}	[33]	$2^{45.4}$
5	2	$2^{16\dagger}$	§ 5.1	2^{101}
		2^{9}	[33]	$2^{109.4}$
6	1	$2^{32\dagger}$	§ 5.1	2^{64}
		15	[14]	2^{124}
7	2	$2^{32\dagger}$	§ 5.1	2^{101}
		$2^{30.9}$	[14]	2^{124}

† Asymptotic results.

Acknowledgements 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 theFlemish Government through FWO Project Locklock G0D3819N. Tim Beyne is funded byan FWO fellowship.

A Asymptotic data complexity

When the signs of correlations are unknown, precise probability distribution estimation requires guessing, and Vaudenay [31] proposed using the χ^2 test. For a successful attack, the data complexity must be sufficiently large to guarantee that $\mu_{\text{real}} - \mu_{\text{ideal}} \gg \sigma_{\text{ideal}}$. The bound in Equation (3) follows from analyzing the condition $\mu_{\text{real}} - \mu_{\text{ideal}} \gg \sigma_{\text{ideal}}$. Using the multidimensional χ^2 test, the expected value of the test statistic under the real cipher is

$$\mu_{\text{real}} = \mathbb{E}\left[\chi^2_{\text{real}}\right] \approx |\Lambda| - 1 + q \sum_{(u,v)\in\Lambda\setminus(0,0)} \mathbb{E}\left[(C^F_{v,u})^2\right],$$

The central limit theorem states that the sum of many independent and identically distributed random variables follows a normal distribution in the limit. In particular, as the sample size grows, the distribution of the sample mean approaches a normal distribution. For uniformly random data, the random variable $\boldsymbol{\chi}^2_{\text{ideal}}$ (the test statistic under uniform random samples) asymptotically approaches the normal distribution $\mathcal{N}(\mu_{\text{real}}, 2\mu_{\text{real}})$. The χ^2 distribution with μ_{real} degrees of freedom represents the distribution of the sum of squares of μ_{real} independent standard normal random variables, so $\mu_{\text{real}} = \mathbb{E}[\chi^2_{\text{ideal}}] = |\Lambda| - 1$ and $\sigma^2_{\text{ideal}} = 2\mu_{\text{real}} = 2(|\Lambda| - 1)$, which implies $\mathbb{E}\left[\chi^2_{\text{real}}\right] - \mathbb{E}\left[\chi^2_{\text{ideal}}\right] \gg \sqrt{2(|\Lambda| - 1)}$. Consequently, a successful χ^2 distinguisher requires $q\mathsf{Cap}(\Lambda) \gg \sqrt{|\Lambda|}$, which yields the data-complexity bound $q \approx \frac{\sqrt{|\Lambda|}}{\mathsf{Cap}(\Lambda)}$.

B Pseudo-code

Algorithm 3: FFT-based key recovery attack using r-round multidimensional distinguisher.

input mask: $(u \parallel u)$
output mask: $(v \parallel v)$
Set $N \leftarrow 2^n$ where n represents the input size of the round function;
Set master key $\leftarrow k$;
for $i = 0$ *to* N **do**
  $input : x \parallel x + c$;
  $output : y_L^i \parallel y_R^i$;
  $z_i = \sigma(x^i) + \sigma^{-1}(y_L^i) + x^i + \sigma^{-1}(y_R^i)$;
end
for $i = 0$ *to* N **do**
  $++ p_0[f_3(i)]$; *//f_j are active round functions whr $j \in \{0, \ldots, \ell\}$*
  $\vdots$
  $++ p_\ell[f_{2\ell+1}(i)]$; *//$\ell = \lfloor \frac{r}{2} \rfloor$*
end
for $i = 0$ *to* N **do**
  $p_0[i] \leftarrow p_0[i]p_1[i] \ldots p_\ell[i]$;
end
for $i = 0$ *to* N **do**
  $M[i] \leftarrow \frac{(p_0[i] - q/2^n)^2}{q/2^n}$; *//$M[i]$ is the entry of $M_{F_1(c),z}$*
end
$M \leftarrow FFT(M)$;
for $i = 0$ *to* q **do**
  $++ w[z_i]$; *//w is a vector of the occurrences of 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**
  $w[i] \leftarrow w[i] * M[i]$;
end
$\chi^2_{F_1(c)} \leftarrow FFT(w)$; *//$\chi^2$ statistic for constant $F_1(c)$*
Computing success probability of the attack for a given false-positive rate:
Sort $\chi^2_{F_1(c)}$ in descending order with indices $s[i]$;
$c_{k_1} \leftarrow f_1(k_1, c)$; *//$k_i$ subkey of i^{th}-round function*
$n_a \leftarrow \#$ *of attempts*;
for $i = 0$ *to* N **do**
  if $c_{k_1} = s[i]$ & $\ell = \{\#i \mid s[i] \geq P_F * N\}$ **then**
    $P_S \leftarrow \frac{\ell}{n_a}$;
    The data complexity $\leftarrow n_a * N$;
  end
end

References

1. Aumasson, J.P.: SPC Tweakable Block Cipher (2021), https://github.com/veorq/spc, https://github.com/veorq/spc

2. Aumasson, J.P., Bernstein, D.J.: SipHash: A Fast Short-Input PRF. In: Galbraith, S., Nandi, M. (eds.) Progress in Cryptology - INDOCRYPT 2012. pp. 489–508. Springer Berlin Heidelberg, Berlin, Heidelberg (2012)
3. Baignères, T., Junod, P., Vaudenay, S.: How Far Can We Go Beyond Linear Cryptanalysis? In: Advances in Cryptology - ASIACRYPT 2004. Lecture Notes in Computer Science, vol. 3329, pp. 432–450. Springer (2004), https://doi.org/10.1007/978-3-540-30539-2_31
4. Bao, Z., Dinur, I., Guo, J., Leurent, G., Wang, L.: Generic attacks on hash combiners. J. Cryptol. **33**(3), 742–823 (Jul 2020). https://doi.org/10.1007/s00145-019-09328-w, https://doi.org/10.1007/s00145-019-09328-w
5. Bellare, M., Hoang, V.T., Tessaro, S.: Message-recovery attacks on feistel-based format preserving encryption. In: Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security. p. 444–455. CCS '16, Association for Computing Machinery, New York, NY, USA (2016). https://doi.org/10.1145/2976749.2978390, https://doi.org/10.1145/2976749.2978390
6. Beyne, T.: Linear Cryptanalysis of FF3-1 and FEA. In: Malkin, T., Peikert, C. (eds.) Advances in Cryptology – CRYPTO 2021. pp. 41–69. Springer International Publishing, Cham (2021), https://doi.org/10.1007/978-3-030-84242-0_3
7. Beyne, T., Rijmen, V.: Linear Cryptanalysis. Cambridge University Press (2025)
8. Collard, B., Standaert, F.X., Quisquater, J.J.: Improving the Time Complexity of Matsui's Linear Cryptanalysis. In: Nam, K.H., Rhee, G. (eds.) Information Security and Cryptology - ICISC 2007. pp. 77–88. Springer Berlin Heidelberg (2007), https://doi.org/10.1007/978-3-540-76788-6_7
9. Daemen, J., Govaerts, R., Vandewalle, J.: Correlation Matrices. In: Preneel, B. (ed.) Fast Software Encryption. pp. 275–285. Springer Berlin Heidelberg (1995), https://doi.org/10.1007/3-540-60590-8_21
10. Daemen, J., Rijmen, V.: Probability distributions of correlation and differentials in block ciphers. J. Math. Cryptol. **1**(3), 221–242 (2007). https://doi.org/10.1515/JMC.2007.011, https://doi.org/10.1515/JMC.2007.011
11. Hermelin, M., Cho, J.Y., Nyberg, K.: Multidimensional Linear Cryptanalysis of Reduced Round Serpent. In: Mu, Y., Susilo, W., Seberry, J. (eds.) Information Security and Privacy. pp. 203–215. Springer Berlin Heidelberg (2008), https://doi.org/10.1007/978-3-540-70500-0_15
12. Hermelin, M., Cho, J.Y., Nyberg, K.: Multidimensional Extension of Matsui's Algorithm 2. In: Dunkelman, O. (ed.) Fast Software Encryption. pp. 209–227. Springer Berlin Heidelberg (2009), https://doi.org/10.1007/978-3-642-03317-9_13
13. Hermelin, M., Cho, J.Y., Nyberg, K.: Multidimensional Linear Cryptanalysis. Journal of Cryptology **32**(1), 1–34 (2019), https://doi.org/10.1007/s00145-018-9308-x
14. Isobe, T., Shibutani, K.: Improved all-subkeys recovery attacks on fox, katan and shacal-2 block ciphers. In: Cid, C., Rechberger, C. (eds.) Fast Software Encryption. pp. 104–126. Springer Berlin Heidelberg, Berlin, Heidelberg (2015)
15. Junod, P., Macchetti, M.: Revisiting the IDEA Philosophy. In: Dunkelman, O. (ed.) Fast Software Encryption. pp. 277–295. Springer Berlin Heidelberg, Berlin, Heidelberg (2009)
16. Junod, P., Vaudenay, S.: FOX : A New Family of Block Ciphers. In: Handschuh, H., Hasan, M.A. (eds.) Selected Areas in Cryptography. pp. 114–129. Springer Berlin Heidelberg, Berlin, Heidelberg (2005), https://link.springer.com/chapter/10.1007/978-3-540-30564-4_8

17. Kaliski, B.S., Robshaw, M.J.B.: Linear Cryptanalysis Using Multiple Approximations. In: Desmedt, Y.G. (ed.) Advances in Cryptology — CRYPTO '94. pp. 26–39. Springer Berlin Heidelberg (1994), https://doi.org/10.1007/3-540-48658-5_4
18. Karpman, P., Grégoire, B.: The Littlun S-box and the Fly Block Cipher. Gaithersburg, United States (Oct 2016), https://hal.science/hal-03543497
19. Lai, X., Massey, J.L.: A Proposal for a New Block Encryption Standard. In: Advances in Cryptology - EUROCRYPT '90, Workshop on the Theory and Application of of Cryptographic Techniques, Aarhus, Denmark, May 21-24, 1990, Proceedings. Lecture Notes in Computer Science, vol. 473, pp. 389–404. Springer (1990). https://doi.org/10.1007/3-540-46877-3_35
20. Lai, X., Massey, J.L.: A proposal for a new block encryption standard. In: Damgård, I.B. (ed.) Advances in Cryptology — EUROCRYPT '90. pp. 389–404. Springer Berlin Heidelberg, Berlin, Heidelberg (1991)
21. Leurent, G., Sibleyras, F.: Low-memory attacks against two-round even-mansour using the 3-xor problem. In: Boldyreva, A., Micciancio, D. (eds.) Advances in Cryptology – CRYPTO 2019. pp. 210–235. Springer International Publishing, Cham (2019)
22. Luo, Y., Lai, X., Gong, Z.: Pseudorandomness Analysis of the (extended) Lai-Massey Scheme. Inf. Process. Lett. **111**, 90–96 (2010), https://api.semanticscholar.org/CorpusID:205482820
23. Luo, Y., Lai, X., Zhou, Y.: Generic Attacks on the Lai–Massey Scheme. Designs, Codes and Cryptography **83**, 407 – 423 (2016), https://api.semanticscholar.org/CorpusID:35851771
24. Matsui, M.: The First Experimental Cryptanalysis of the Data Encryption Standard. In: Desmedt, Y.G. (ed.) Advances in Cryptology — CRYPTO '94. pp. 1–11. Springer Berlin Heidelberg (1994), https://doi.org/10.1007/3-540-48658-5_1
25. Nakahara, J., Rijmen, V., Preneel, B., Vandewalle, J.: The MESH Block Ciphers. In: Chae, K.J., Yung, M. (eds.) Information Security Applications. pp. 458–473. Springer Berlin Heidelberg, Berlin, Heidelberg (2004), https://link.springer.com/chapter/10.1007/978-3-540-24591-9_34
26. n Özdemir, B.A., Beyne, T., Rijmen, V.: Multidimensional Linear Cryptanalysis of Feistel Ciphers. IACR Transactions on Symmetric Cryptology **023 No. 4**, 1–27 (2023). https://doi.org/10.46586/tosc.v2023.i4.1-27, https://tosc.iacr.org/index.php/ToSC/article/view/11277
27. Patarin, J.: New Results on Pseudorandom Permutation Generators Based on the DES Scheme. In: Feigenbaum, J. (ed.) Advances in Cryptology — CRYPTO '91. pp. 301–312. Springer Berlin Heidelberg (1992), https://doi.org/10.1007/3-540-46766-1_25
28. Patarin, J.: Generic Attacks on Feistel Schemes. In: Boyd, C. (ed.) Advances in Cryptology — ASIACRYPT 2001. pp. 222–238. Springer Berlin Heidelberg (2001), https:/[h]/doi.org/10.1007/3-540-45682-1_14
29. Patarin, J.: Security of Random Feistel Schemes with 5 or More Rounds. In: Franklin, M. (ed.) Advances in Cryptology – CRYPTO 2004. pp. 106–122. Springer Berlin Heidelberg (2004), https://doi.org/10.1007/978-3-540-28628-8_7
30. Press, W.H., Teukolsky, S.A., Flannery, B.P., Vetterling, W.T.: Numerical Recipes in FORTRAN: The Art of Scientific Computing. Cambridge University Press, USA, 2nd edn. (1992)
31. Vaudenay, S.: An Experiment on DES Statistical Cryptanalysis. In: Conference on Computer and Communications Security (1996), https://api.semanticscholar.org/CorpusID:15041839

32. Vaudenay, S.: On the Lai-Massey Scheme. In: Lam, K.Y., Okamoto, E., Xing, C. (eds.) Advances in Cryptology - ASIACRYPT'99. pp. 8–19. Springer Berlin Heidelberg, Berlin, Heidelberg (1999)
33. Wu, W., Zhang, W., Feng, D.: Integral cryptanalysis of reduced fox block cipher. In: Won, D.H., Kim, S. (eds.) Information Security and Cryptology - ICISC 2005. pp. 229–241. Springer Berlin Heidelberg, Berlin, Heidelberg (2006)