Linear Cryptanalysis of FF3-1 and FEA

Tim Beyne

imec-COSIC, KU Leuven, Belgium
name.lastname@esat.kuleuven.be

Abstract. Improved attacks on generic small-domain Feistel ciphers with alternating round tweaks are obtained using linear cryptanalysis. This results in practical distinguishing and message-recovery attacks on the United States format-preserving encryption standard FF3-1 and the South-Korean standards FEA-1 and FEA-2. The data complexity of the proposed attacks on FF3-1 and FEA-1 is $\widetilde{\mathcal{O}}(N^{r/2-1.5})$, where N^2 is the domain size and r is the number of rounds. For example, FF3-1 with $N = 10^3$ can be distinguished from an ideal tweakable block cipher with advantage $\geq 1/10$ using 2^{23} encryption queries. Recovering the left half of a message with similar advantage requires 2^{24} data. The analysis of FF3-1 serves as an interesting real-world application of (generalized) linear cryptanalysis over the group $\mathbb{Z}/N\mathbb{Z}$.

Keywords: Linear cryptanalysis · FF3-1 · FEA-1 · FEA-2 · Format-preserving encryption

1 Introduction

Format-preserving encryption enables the encryption of plaintext with a specific format, while ensuring that the ciphertext has the same format. For example, in some applications it is convenient to be able to encrypt nine-digit integers (such as social security numbers) to nine-digit integers.

Several generic techniques such as cycle walking [5, 7] can be used to transform (tweakable) block ciphers into format-preserving ciphers. However, these techniques are inefficient when there is a significant size difference between the domain of the underlying block cipher and the target domain. Consequently, a number of dedicated constructions based on small-domain tweakable Feistel ciphers were introduced. The best known examples are the United States standards FF1 and FF3-1 [12] (NIST SP800-38G rev. 1). The South-Korean standards FEA-1 and FEA-2 [16] (TTAK.KO-12.0275) follow a similar design but with lighter round functions.

Small-domain Feistel ciphers are known to be vulnerable to a number of generic attacks. In a series of papers, Patarin [18–20] analyzed the security of r-round Feistel ciphers with uniform random round functions. In particular, Patarin [20, §8] describes a distinguisher with data and time complexity $\widetilde{\mathcal{O}}(N^{r-4})$ for Feistel ciphers with domain size N^2. At CCS 2016, Bellare, Hoang and Tessaro [4] presented a message-recovery attack with a data complexity of

$\widetilde{\mathcal{O}}(N^{r-2})$ or $\widetilde{\mathcal{O}}(N^{r-3})$ (to recover the left half of the message) queries. Subsequent improvements were obtained by Hoang, Tessaro and Trieu [15].

The applicability of these attacks to FF3 in part motivated the US National Institute of Standards and Technology (NIST) to revise the FF3 standard [12]. In particular, the revised standard FF3-1 includes the requirement that the domain size must be at least one million, *i.e.* $N \geq 10^3$. Furthermore, the revision decreased the size of the tweak from 64 to 56 bits. This change was introduced to prevent a powerful slide-type attack presented by Durak and Vaudenay [11] at CRYPTO 2017 that was subsequently improved by Hoang *et al.* [14] and Amon *et al.* [1]. These attacks were the consequence of a weakness in the tweak-schedule of FF3 that is resolved by the changes in FF3-1.

Recently, Dunkelman *et al.* [10] have proposed new distinguishers for FEA, FF1 and FF3-1. The data complexity of these attacks is $\widetilde{\mathcal{O}}(N^{r-4})$, which is comparable to the attack of Patarin [20]. The time complexity is $\widetilde{\mathcal{O}}(N^{r-3})$.

Contribution. This paper develops new distinguishing and message-recovery attacks on small-domain Feistel ciphers with alternating round tweaks. The attacks are based on linear cryptanalysis, but go beyond standard methods in several ways. In particular, the role of the tweak input is analyzed, properties of small uniform random functions are exploited, and for FF3-1 a generalization of linear cryptanalysis to the group $\mathbb{Z}/N\mathbb{Z}$ is used. Furthermore, the principle behind the message-recovery attacks is novel.

If the round tweaks alternate between two values, as in FEA-1 and FF3-1, the data and time complexity of these attacks is $\widetilde{\mathcal{O}}(N^{r/2-1.5})$. For FEA-2, which has a different tweak schedule, distinguishing and message-recovery respectively require $\widetilde{\mathcal{O}}(N^{r/3-1.5})$ and $\widetilde{\mathcal{O}}(N^{r/3-0.5})$ data and time. The new attacks are not applicable to FF1. For many instances of FF3-1, FEA-1 and FEA-2, the data and time complexity are well within the reach of real-world adversaries.

The proposed distinguishers only need weak access to the block cipher: it is sufficient to have ciphertext-only access to encryptions of an arbitrary constant message under many half-constant tweaks. In fact, access to the complete ciphertext is not necessary. The message-recovery attacks follow the security model introduced by Bellare *et al.* [4]. Specifically, given the encryption (with FF3-1 or FEA-1) of a secret message and a known message with the same right-hand side under $\widetilde{\mathcal{O}}(N^{r/2-1.5})$ tweaks, the attack recovers the left half of the secret message. With $\widetilde{\mathcal{O}}(N^{r/2-0.5})$ queries, full messages can also be recovered. For FEA-1, the message-recovery attack can be used to set up a key-recovery attack. If q is the concrete data cost of the left-half message-recovery attack, then the key-recovery attack requires less than $16\lceil 8/\log_2 N\rceil\, q + 8q$ data and time equivalent to at most $2^{69}/N + 16\lceil 8/\log_2 N\rceil\, q + 8q$ evaluations of FEA-1.

Table 1 summarizes the cost of the main attacks from the literature and some of the new attacks proposed in this paper. In addition, the bottom part of the table reports concrete costs for the smallest instances of FEA-1, FEA-2 ($N = 16$) and FF3-1 ($N = 10^3$). Detailed cost-estimates for previous attacks on the same instances are not always available, but the improvement is substantial. For

example, the attacks on FF3-1 with $N = 10^3$ require data and time comparable to previous attacks for $N = 2^5$ [4, 15] that led to the requirement $N \geq 10^3$. The numbers in Table 1 have been experimentally verified by performing each attack many times. Source code to reproduce this is provided as supplementary material[1]. Further experiments and cost calculations are given in the indicated sections.

As with previous attacks on tweakable small-domain Feistel ciphers, the maximum value of N for which the attacks are applicable is typically determined by the tweak length rather than by the length of the key. For FEA-1 and FEA-2 the main interest of these attacks is for small N, so the tweak is long enough for most practical purposes. For FF3-1, the upper bounds are similar to those for previous attacks: naive estimates are $N < 2^{19}$ for distinguishing and right-half message-recovery and $N < 2^{12}$ for left-half recovery. The latter bound is quite close to the required $N \geq 10^3$ for FF3-1, However, as discussed in Section 5, it is not a hard limit.

Early notification. Prior to the submission of this paper, both NIST (for FF3-1) and ETRI (for FEA-1 and FEA-2) were notified about these results. Both parties have acknowledged the attacks and have indicated their intention to revise their standards. Modifying the tweak schedule seems to be the most promising approach to thwart the attacks.

Organization. After revisiting the overall structure of FEA-1, FEA-2 and FF3-1 in Section 2, the basic idea behind the attacks is introduced in Section 3. It is shown that there exists a linear trail through FEA-1 (and similarly for FEA-2) with high correlation. The novelty of this trail is the fact that it requires considering the tweak as a proper part of the input of the cipher, and its reliance on the properties of small random functions. An analogous $\mathbb{Z}/N\mathbb{Z}$-linear trail is then obtained for FF3-1. This result is an application of a generalization of linear cryptanalysis to other finite Abelian groups [3, 6].

Section 4 combines the linear approximations identified in Section 3 to obtain multidimensional linear approximations. These approximations are subsequently used to construct a χ^2-distinguisher. The formalism of (generalized) multidimensional linear cryptanalysis is applied to justify the attack and to obtain initial estimates of the data complexity. Finally, Section 5 shows how the χ^2-distinguisher can be turned into a message-recovery attack. Each attack comes with a detailed analysis of the advantage and data complexity, and an experimental verification of the theoretical analysis.

2 Preliminaries

The attacks in this paper are applicable to tweakable small-domain Feistel ciphers with alternating round tweaks. The South-Korean format-preserving encryption standards FEA-1 and FEA-2 [16] and the NIST standard FF3-1 [12] all follow such a design.

[1] https://homes.esat.kuleuven.be/~tbeyne/fpe

Table 1: Summary of attacks on FEA-1, FEA-2 and FF3-1. The costs in the top half of the table are up to polylogarithmic factors in N (all of which are small in practice). Time is expressed in encryption operations. Memory requirements are small for all attacks. All of the message-recovery attacks listed in this table recover the left half of a message.

		Data	Time	Advantage	Reference
Generic	Distinguisher	N^{r-4}	N^{r-3}	Constant	[10]
		N^{r-4}	N^{r-4}	Constant	[20]
		$N^{r/2-1}$	$N^{r/2-1}$	Constant	Section 3†
		$N^{r/2-1.5}$	$N^{r/2-1.5}$	Constant	Section 4†
		$N^{r/3-1}$	$N^{r/3-1}$	Constant	Section 3‡
		$N^{r/3-1.5}$	$N^{r/3-1.5}$	Constant	Section 4‡
	Message recovery	N^{r-3}	N^{r-3}	Constant	[4, 15]
		$N^{r/2-1.5}$	$N^{r/2-1.5}$	Constant	Section 5†
		$N^{r/3-0.5}$	$N^{r/3-0.5}$	Constant	Section 5‡
FEA-1 $N = 16, r = 12$	Distinguisher	2^{22}	2^{22}	0.1	Section 3
		2^{17}	2^{17}	0.1	Section 4
		2^{22}	2^{22}	0.6	Section 4
	Message recovery	2^{17}	2^{17}	0.1	Section 5
		2^{24}	2^{24}	0.6	Section 5
FEA-2 $N = 16, r = 18$	Distinguisher	2^{20}	2^{20}	0.1	Section 3
		2^{17}	2^{17}	0.1	Section 4
		2^{21}	2^{21}	0.6	Section 4
FF3-1 $N = 10^3, r = 8$	Distinguisher	2^{29}	2^{29}	0.1	Section 3
		2^{23}	2^{23}	0.1	Section 4
		2^{26}	2^{26}	0.6	Section 4
	Message recovery	2^{24}	2^{24}	0.1	Section 5
		2^{27}	2^{27}	0.6	Section 5

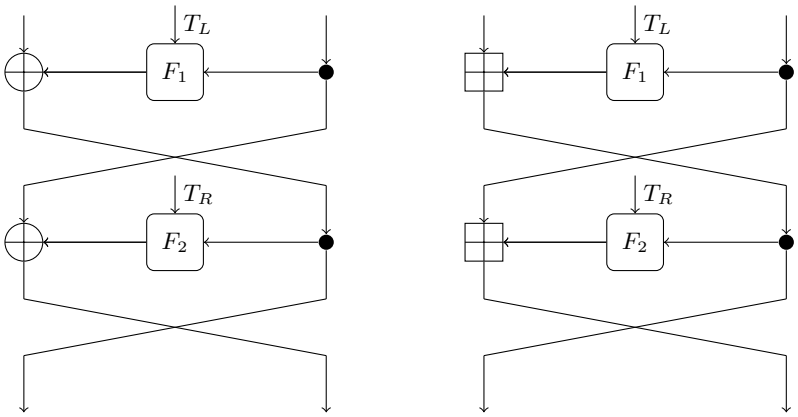
† Assuming the round tweaks alternate between two values, as in FEA-1 and FF3-1.
‡ Assuming the round tweaks alternate between three values, as in FEA-2.

Figure 1 depicts two rounds of the overall structure of FEA-1 and FF3-1. For simplicity, it will be assumed that both branches have the same size. In both designs, the tweak is divided into two equal halves, which will be denoted by T_L and T_R for convenience. A crucial property that will be exploited by the new attacks is that the round tweak alternates between T_L and T_R. The round functions $F_1, F_2, \ldots$ can nevertheless be arbitrary.

As shown in Figure 1a, FEA-1 is a regular Feistel cipher over $\mathbb{F}_2^m \oplus \mathbb{F}_2^m$ with $m = \log_2 N$, where $\oplus$ denotes the direct sum. For 128 bit keys, it has a total of 12 rounds. The tweaks T_L and T_R consist of $64-m$ bits. The round functions F_i are truncations of a two-round SHARK-like construction [21], but can be considered to be uniform random functions for all attacks discussed in this paper except for the key-recovery attack in Section 6. The necessary details of the round function will be reproduced in Section 6.

The design of FEA-2 is very similar to that of FEA-1. The main difference is that it uses three distinct round tweaks (repeating with period three), one of which is constant. In addition, for FEA-2, both tweaks have a length of 64 bits and the number of rounds is 18 for 128 bit keys.

FF3-1 is an eight-round Feistel cipher over $\mathbb{Z}/N\mathbb{Z} \oplus \mathbb{Z}/N\mathbb{Z}$. The round functions $F_1, F_2, \ldots$ are defined as truncations of the AES with the round tweak and a unique round counter as the input; the details are not important for this work as these functions will be modelled as uniform random. The tweaks T_L and T_R are bitstrings of length 28.



(a) FEA-1. (b) FF3-1.

Fig. 1: Two rounds of a tweakable Feistel cipher with alternating round tweaks.

Sections 3 and 4 introduce distinguishers for full-round FEA-1, FEA-2 and FF3-1. The *advantage* of a distinguisher is equal to the difference between its success-probability P_{S} and false-positive rate P_{F} and provides a convenient measure for its statistical quality. The distinguishers discussed in Sections 3 and 4

allow for a trade-off between success-probability and false-positive rate. Since they are ultimately simple hypothesis tests, the trade-off is determined by the choice of some threshold parameter t. The advantage that will be considered in this paper is thus the maximum achievable advantage for some value of t:

$$\mathsf{Adv} = \max_t \ |P_{\mathsf{S}}(t) - P_{\mathsf{F}}(t)|.$$

For message-recovery attacks, it is also meaningful to define a similar measure of quality. Given a list of possible messages, one is interested in narrowing it down to some fraction P_{F} with a given probability P_{S}. Clearly, P_{F} and P_{S} are dependent quantities. The advantage of a message-recovery attack will be defined as the maximum achievable value of $|P_{\mathsf{S}} - P_{\mathsf{F}}|$ for a given amount of data. This corresponds to the notion of key-recovery advantage that is often used in linear and differential cryptanalysis [22]. For the attacks in this paper, it also coincides with the message-recovery advantage defined by Bellare *et al.* [4]

Concepts related to linear and multidimensional linear cryptanalysis will be introduced in Sections 3 and 4 respectively.

3 Linear Distinguishers

In this section, linear distinguishers for FEA-1, FEA-2 and FF3-1 are introduced. Section 3.1 summarizes the main concepts from linear cryptanalysis, but some familiarity with these ideas is necessarily assumed.

Since the attacks on FEA-1 and FEA-2 are based on ordinary $\mathbb{F}_2$-linear cryptanalysis, these are described first in Section 3.2. Section 3.3 then transfers these results to Feistel ciphers defined over $\mathbb{Z}/N\mathbb{Z}$. Finally, the data complexity of the attacks is analyzed in detail and verified experimentally in Section 3.4.

3.1 Linear Approximations

Linear cryptanalysis was introduced by Matsui [17] and is based on probabilistic linear relations or *linear approximations*, a concept introduced by Tardy-Corfdir and Gilbert [23]. Let $F : \mathbb{F}_2^n \to \mathbb{F}_2^m$ be a function, possibly depending on a key. Linear distinguishers are based on linear approximations with large absolute correlation. A linear approximation for F is defined by a pair of masks $(u_1, u_2) \in \mathbb{F}_2^m \oplus \mathbb{F}_2^n$ and its *correlation* is equal to

$$C^F_{u_1,u_2} = 2\Pr\left[u_1^\top F(\boldsymbol{x}) = u_2^\top \boldsymbol{x}\right] - 1 = \frac{1}{2^n} \sum_{x \in \mathbb{F}_2^n} (-1)^{u_1^\top F(x) + u_2^\top x},$$

where the probability is over a uniform random $\boldsymbol{x}$ on $\mathbb{F}_2^n$. If $u_1 \neq 0$, then the correlation for a uniform random function is concentrated around zero with a standard deviation of $2^{-n/2}$. A more detailed result is given in Theorem 3.1 in Section 3.2 below. Hence, if the correlation c is significantly larger than $2^{-n/2}$, a distinguisher is obtained by estimating the correlation using $q = \Theta(1/c^2)$

queries and comparing the result to some threshold t. As discussed in Section 3.2 below, this description is somewhat simplified since the correlation is usually key-dependent.

For FF3-1, $\mathbb{F}_2$-linear approximations are inconvenient because the FF3-1 Feistel structure operates on the ring $\mathbb{Z}/N\mathbb{Z} \oplus \mathbb{Z}/N\mathbb{Z}$. Instead, Section 3.3 will rely on a generalization of linear cryptanalysis to arbitrary finite Abelian groups. Such a generalization was first proposed by Baignères, Stern and Vaudenay [3]. A more general perspective that includes the multidimensional case (which will be used in Section 4) was introduced in [6].

Let $F : G \to H$ be a function between finite Abelian groups G and H. A linear approximation corresponds to a pair of *group characters* (ψ_1, ψ_2) of H and G respectively. A group character ψ_1 is a group homomorphism $\psi_1 : H \to \mathbb{C}^\times$. The characters of H themselves also form a group of order $|H|$ under pointwise multiplication. The correlation of the linear approximation (ψ_1, ψ_2) is equal to

$$C^F_{\psi_1,\psi_2} = \frac{1}{|G|} \sum_{x \in G} \overline{\psi_1(F(x))} \psi_2(x) \,.$$

In the above, $\overline{\psi_1}$ denotes the complex-conjugate of ψ_1. For $H = \mathbb{F}_2^m$, every character ψ_1 corresponds to a vector $u \in \mathbb{F}_2^m$ such that $\psi_1(x) = (-1)^{u^\top x}$. If $H = \mathbb{Z}/N\mathbb{Z}$, then for each character ψ_1, there exists a non-negative integer $k < N$ such that $\psi_1(x) = \exp(2\pi\sqrt{-1}\, kx/N)$. This essentially covers all cases, since any finite Abelian group is a direct sum of cyclic groups. An important property of group characters is that they are orthogonal functions. That is,

$$\sum_{x \in G} \overline{\chi(x)} \psi(x) = \begin{cases} |G| & \text{if } \chi = \psi, \\ 0 & \text{otherwise.} \end{cases}$$

for any two characters χ and ψ of G. Additional background on group characters and Fourier analysis may be found in [24].

For a sequence of functions $F_1, \ldots, F_l$, the *piling-up principle* can be used to estimate the correlation of linear approximations over the composition $F = F_l \circ \cdots \circ F_1$. The idea is that, for an approximation with characters (ψ_1, ψ_{l+1}), there may exist a dominant sequence of approximations $(\psi_1, \psi_2), \ldots, (\psi_l, \psi_{l+1})$ such that

$$C^F_{\psi_1,\psi_{l+1}} \approx \prod_{i=1}^{l} C^{F_i}_{\psi_i,\psi_{i+1}} \,.$$

The sequence of approximations $(\psi_1, \psi_2), \ldots, (\psi_l, \psi_{l+1})$ is called a *trail* and the right-hand side of the above equation is called the correlation of the trail. The sum of the correlations of all trails over $F_l \circ \cdots \circ F_1$ equals the correlation of the approximation (ψ_1, ψ_{l+1}) of F [8].

3.2 FEA-1 and FEA-2

At first sight, both FEA-1 and FEA-2 seem to be robust against linear cryptanalysis, especially when their round functions $F_1, F_2, \ldots$ are replaced by uniform

random functions. The key observation behind the attacks in this paper is that this is not the case when (part of) the tweak is considered as a proper part of the input.

Figure 2 shows linear trails over two rounds of FEA-1 and three rounds of FEA-2[2]. In these trails, the tweak T_L is an arbitrary constant and T_R is considered to be a variable part of the input. Note that the tweak T_R is not active, so it need not be known to perform the attack. The idea behind these trails is that the absolute correlation of a linear approximation over round function F_i (chosen uniformly at random) exceeds $1/\sqrt{N} = 2^{-m/2}$ with fairly large probability. This becomes meaningful when the tweak is included in the input, because the domain of the function which maps the tweak and the plaintext to the ciphertext is large. Indeed, the correlation of linear approximations over a random function with the same input size (including T_R of length $64-m$) as FEA-1 is centered around zero with a standard deviation of $2^{-32-m/2}$. More specifically, we have the following result.

Theorem 3.1 (Daemen and Rijmen [9]). *Let $\boldsymbol{c}$ denote the correlation of a nontrivial linear approximation for a uniform random function $\mathbb{F}_2^n \to \mathbb{F}_2^m$. The random variable $2^{n-1}(\boldsymbol{c}+1)$ is binomially distributed with mean 2^{n-1} and variance 2^{n-2}. In particular[3], as $n \to \infty$, the distribution of $2^{n/2}\boldsymbol{c}$ converges to the standard normal distribution $\mathcal{N}(0,1)$.*

Let $r \geq 2$ be an even integer. By the piling-up principle, the correlation of the r-round trail from Figure 2a is equal to $\boldsymbol{c} = \prod_{i=1}^{r/2} \boldsymbol{c}_i$, where $\boldsymbol{c}_i \sim \mathcal{N}(0, 1/N)$ holds asymptotically due to Theorem 3.1. The random variables $\boldsymbol{c}_i$ will be assumed to be independent, which follows for instance from the strong assumption that the round functions $F_1, F_3 \ldots F_{r-1}$ are independent. One can verify that the other trails through FEA-1 and FF3-1 have negligible correlation.

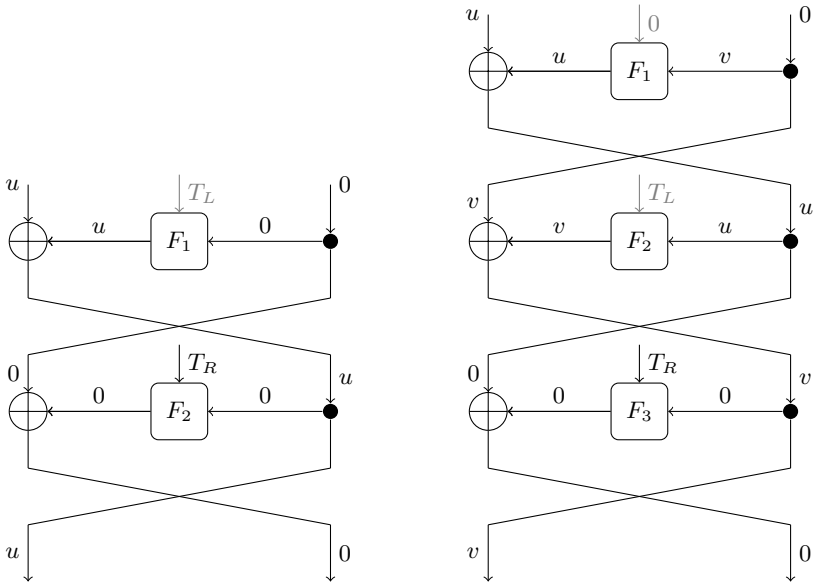
As mentioned above, the data complexity of a constant-advantage linear distinguisher based on an approximation with correlation c is $\Theta(1/c^2)$. In this case, the correlation varies strongly with the key so this result can not be applied directly to estimate the data complexity. A commonly used heuristic estimate is given by $1/\mathbb{E}\boldsymbol{c}^2$, where $\mathbb{E}\boldsymbol{c}^2$ is the average squared trail correlation for a uniform random key. For FEA-1, this yields $1/\mathbb{E}\boldsymbol{c}^2 = N^{r/2}$. The data complexity is analyzed in considerably more detail in Section 3.4.

For FEA-2 with r divisible by three, the expected squared correlation of each trail is equal to $N^{-2r/3}$. However, the number of trails for a given choice of input and output masks is $(N-1)^{r/3-1}$. Recall that the correlation of a linear approximation is equal to the sum of the correlations over all possible trails. Hence, since the trails in Figure 2b are indeed dominant, the sum $\boldsymbol{c}$ of the correlations of these trails is a good estimate for the correlation of the corresponding approximation. Since the covariance between the correlations of distinct trails is zero for independent uniform random round functions, it follows that

$$1/\mathbb{E}\boldsymbol{c}^2 = N^{2r/3}/(N-1)^{r/3-1} \sim N^{r/3+1}.$$

[2] I thank Dongyoung Roh for bringing the trails with $u \neq v$ to my attention.

[3] This result is a useful approximation even when n is small (for example, when $n \geq 8$).



(a) Two-round trail for FEA-1.

(b) Three-round trail for FEA-2.

Fig. 2: Linear trails for FEA-1 and FEA-2. The tweak T_R is considered part of the input and the value of T_L should be fixed.

The fact that the covariance terms are zero is somewhat nontrivial, but it can be easily deduced from the definition of correlation for a uniform random function. Neglecting the covariance between the correlation of different trails is, in general, inaccurate. Finally, note that any other trail through FEA-2 necessarily has a much smaller average squared correlation.

Before continuing with the analysis of FF3-1, a simple but significant improvement to the correlation of the aforementioned linear approximation should be pointed out. If the right part of the plaintext is fixed to an arbitrary constant, then after two rounds the left branch of the state is equal to the left part of the plaintext up to addition by some constant. Consequently, the first two rounds can be effectively skipped. This decreases the data complexity by a factor N to $N^{r/2-1}$ for FEA-1. By fixing both halves of the plaintext, the first three rounds of FEA-2 can similarly be avoided. In addition, since the input mask is then no longer fixed, the number of trails within one approximation increases to $(N-1)^{r/3}$. Hence, the resulting data complexity estimate becomes $N^{r/3-1}$. A more detailed estimate of the data complexity will be given in Section 3.4.

3.3 FF3-1

The analysis of FF3-1 proceeds analogously to that of FEA-1, but with linear cryptanalysis over the additive group $\mathbb{Z}/N\mathbb{Z}$ rather than $\mathbb{F}_2^m$. An iterative two-

round trail is shown in Figure 3. In the figure, ψ denotes an arbitrary nontrivial character of $\mathbb{Z}/N\mathbb{Z}$ and $\mathbb{1}$ is the trivial character, *i.e.* $\mathbb{1}(x) = 1$ for all $x \in \mathbb{Z}/N\mathbb{Z}$.

In order to characterize the correlation of this trail, an analog of Theorem 3.1 is required. This is provided by Theorem 3.2 below. Recall that a complex-valued random variable $\boldsymbol{z}$ has a standard complex normal distribution $\mathcal{CN}(0,1)$ if its real part $\Re\{\boldsymbol{z}\} \sim \mathcal{N}(0,1/2)$ and its imaginary part $\Im\{\boldsymbol{z}\} \sim \mathcal{N}(0,1/2)$ are independent random variables.

Theorem 3.2. *Let G and H be finite Abelian groups and let $\boldsymbol{c}$ denote the correlation of a nontrivial linear approximation for a uniform random function $G \to H$ corresponding to non-real characters. The correlation $\boldsymbol{c}$ has mean zero and variance $1/|G|$. Furthermore, as $|G| \to \infty$, the distribution of $\sqrt{|G|}\,\boldsymbol{c}$ converges to the standard complex normal distribution $\mathcal{CN}(0,1)$.*

Proof. Recall that a linear approximation corresponds to a pair of group characters $(\overline{\psi_1}, \psi_2)$. The random variable $\boldsymbol{c}$ can then be written as

$$\boldsymbol{c} = \frac{1}{|G|} \sum_{i=1}^{|G|} \psi_1(\boldsymbol{y}_i)\psi_2(x_i),$$

where $x_1, \ldots, x_{|G|}$ are the elements of G and $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_{|G|}$ are independent uniform random variables on H. The mean of $\boldsymbol{c}$ is zero, since $\mathbb{E}\psi_1(\boldsymbol{y}_i) = 0$ by the orthogonality relations for group characters. In addition, it follows from $\mathbb{E}|\psi_1(\boldsymbol{y}_i)|^2 = 1$ that $\mathbb{E}|\boldsymbol{c}|^2 = 1/|G|$. Finally, the convergence to a normal distribution follows from the central limit theorem for the sum of independent identically distributed random variables. □

By Theorem 3.2, the average squared correlation of the r-round trail from Figure 3 is equal to $N^{-r/2}$. As in the case of FEA-1, the right part of the plaintext can be fixed in order to obtain a trail with average squared correlation $N^{1-r/2}$. This gives a corresponding data complexity estimate of $N^{r/2-1}$.

3.4 Cost Analysis and Experimental Verification

As mentioned in Sections 3.2 and 3.3 above, the data complexity of a distinguisher based on a linear approximation with correlation c is roughly $1/|c|^2$. By *heuristically* plugging in the average squared trail correlation, the approximation $1/\mathbb{E}|\boldsymbol{c}|^2$ was obtained. This resulted in an estimated data complexity of $N^{r/2-1}$ for FEA-1 and FF3-1 and $N^{r/3-1}$ for FEA-2. This section analyzes the data complexity in more detail, along with the advantage achieved by the distinguisher. Broadly speaking, the detailed analysis confirms the heuristic estimates from Sections 3.2 and 3.3.

The distinguisher performs a hypothesis test, with null-hypothesis that the data comes from an ideal tweakable block cipher and alternative hypothesis that the data comes from the real cipher. If the absolute value of the estimated correlation exceeds a predetermined threshold, then the null-hypothesis is rejected.

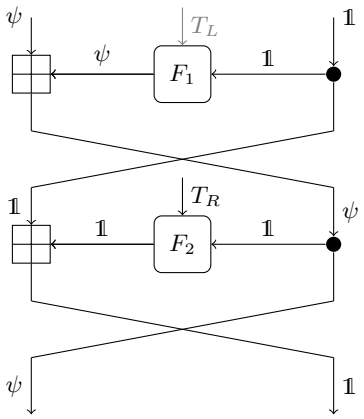


Fig. 3: Iterative two-round trail for FF3-1. The tweak T_L is fixed.

Like any hypothesis test, linear distinguishers allow for a trade-off between success probability P_{S} and false-positive rate P_{F}. Both probabilities are determined by the threshold parameter t. The distinguisher is successful if the estimated correlation exceeds $t\sqrt{q}$ when interacting with the true block cipher after q queries. If the estimated correlation exceeds this threshold for an ideal tweakable block cipher, then a false-positive occurs. Note that $P_{\mathsf{S}}(t)$ and $P_{\mathsf{F}}(t)$ are key-averaged quantities.

Figure 4 depicts the estimates of the maximum advantage $\max_t |P_{\mathsf{S}}(t) - P_{\mathsf{F}}(t)|$ which are derived below. Importantly, for large N, the curve is essentially independent of N. This will be shown below. The red dots correspond to experimental verifications of the estimates for full-round instances of FEA-1, FEA-2 and FF3-1. Each point corresponds to 1024 (FEA-1 and FF3-1) or 512 (FEA-2) evaluations of the distinguisher. For FF3-1, the experiments were performed for $N = 100 < 1000$ to limit the computational cost. The verification of the more efficient χ^2-distinguishers in Section 4 will be performed for $N = 1000$.

The false-positive rate is easily computed. Assume the correlation is estimated using q independent queries. If the input space is sufficiently large[4], then by Theorems 3.1 and 3.2 the variance of the ideal correlation is negligible. Hence, if the number of queries q is moderately large, then the estimated correlation $\widehat{\boldsymbol{c}}_{\mathsf{ideal}}$ is approximately distributed as $\mathcal{N}(0, 1/q)$ for FEA-1 and FEA-2 or $\mathcal{CN}(0, 1/q)$ for FF3-1. The false-positive rate is then

$$P_{\mathsf{F}}(t) = \Pr\left[|\widehat{\boldsymbol{c}}_{\mathsf{ideal}}| \geq t/\sqrt{q}\right] \approx 1 - \chi_\nu(\sqrt{\nu}t),$$

where χ_ν is the cumulative distribution function of the χ-distribution with ν degrees of freedom. For FEA-1 and FEA-2, $\nu = 1$ since $\boldsymbol{c}$ is real. For FF3-1, $\nu = 2$.

[4] Relative compared to the required number of queries q.

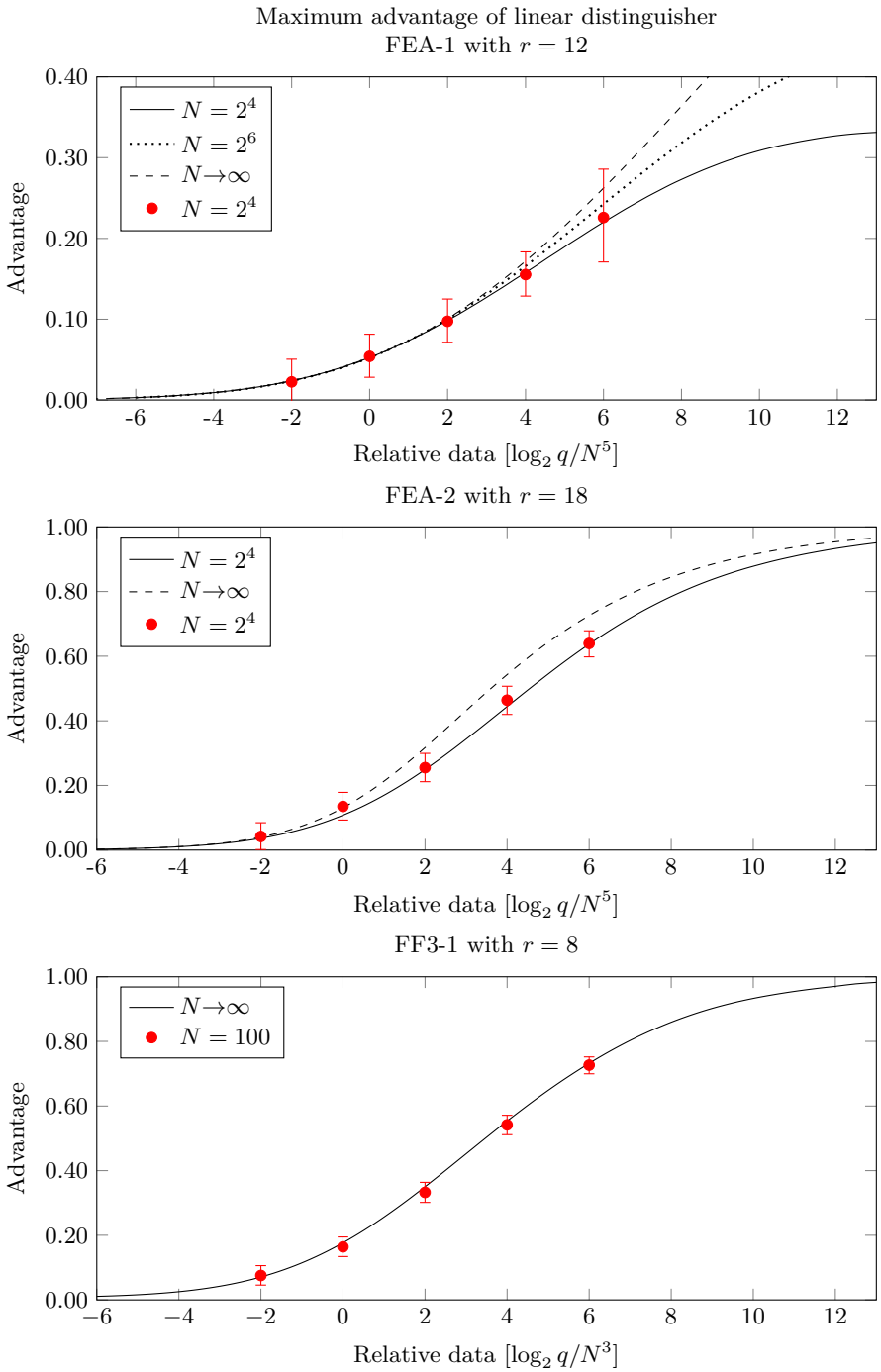


Fig. 4: Theoretical and experimentally observed maximum advantage of the linear distinguishers for full-round FEA-1, FEA-2 and FF3-1. The error bars correspond to 95% Clopper-Pearson confidence intervals.

The calculation of the success rate P_{S} is more complicated, because the absolute correlation $|\boldsymbol{c}_{\mathsf{real}}|$ is not as strongly concentrated around its mean. Let $\widehat{\boldsymbol{c}}_{\mathsf{real}}$ denote the estimated correlation for a particular choice of the key. If the underlying correlation for this key is equal to c_{real}, then $\widehat{\boldsymbol{c}}_{\mathsf{real}}$ is approximately distributed as $\mathcal{N}(c_{\mathsf{real}}, 1/q)$ for FEA-1 and FEA-2 or $\mathcal{CN}(c_{\mathsf{real}}, 1/q)$ for FF3-1 if q is large enough and $c_{\mathsf{real}}^2 \ll 1$. The average success probability can be approximated as

$$P_{\mathsf{S}}(t) \approx \mathop{\mathbb{E}}_{\boldsymbol{c}_{\mathsf{real}}} \Pr\left[|\boldsymbol{z}_\nu - \boldsymbol{c}_{\mathsf{real}}\sqrt{q}| \geq t\right]$$

where $\boldsymbol{c}_{\mathsf{real}}$ is the trail correlation assuming uniform random round functions and $\boldsymbol{z}_\nu$ a standard (complex if $\nu = 2$) normal random variable. To compute the average with respect to $\boldsymbol{c}_{\mathsf{real}}$, a Monte-Carlo approach was used. The implementation is provided as supplementary material. Importantly, the success probability curve (and consequently the maximum advantage) has essentially the same shape for all sufficiently large values of N. Indeed, by Theorems 3.1 and 3.2, the distribution of the round correlations converges to a (complex) normal distribution for large N. Hence, for $q_0 = 1/\mathbb{E}|\boldsymbol{c}_{\mathsf{real}}|^2$, the distribution of $\boldsymbol{c}_{\mathsf{real}}\sqrt{q_0}$ will be approximately the same for all large values of N. Consequently, the success probability curves tend to a constant function of q/q_0.

4 χ^2 Distinguishers

This section introduces additional distinguishers on FEA-1, FEA-2 and FF3-1, based on Pearson's χ^2-test for goodness-of-fit between distributions. Vaudenay [25] proposed χ^2-distinguishers as a method for distinguishing non-uniform distributions in cryptanalysis when precise knowledge about these distributions is lacking.

The distinguishers in Section 3 are based on individual linear approximations. A natural improvement to these attacks is to exploit all approximations simultaneously. Multidimensional linear cryptanalysis provides a convenient framework to describe such attacks.

As shown in Section 4.2 below, the existence of a multidimensional linear approximation implies that a particular probability distribution related to the ciphertext is highly non-uniform. Pearson's χ^2-test can then be used to verify this property, resulting in a distinguisher.

Sections 4.1 and 4.2 explain the distinguisher in detail. The data complexity is estimated and experimentally verified in Section 4.3.

4.1 Multidimensional Linear Approximations

A multidimensional $\mathbb{F}_2$-linear approximation can be defined as a collection of linear approximations such that the set of pairs of input and output masks is a vector space [13]. This generalizes to arbitrary groups, by requiring that the set of pairs of input and output characters is a group under pointwise multiplication. A general description of this approach can be found in [6].

To obtain a uniform description of the attacks on FEA-1, FEA-2 and FF3-1, denote the half-domain by $\mathcal{D}$ and the space of tweaks T_R by $\mathcal{T}$. The ciphertext space is then $H = \mathcal{D} \oplus \mathcal{D}$. The input space G is either $\mathcal{D} \oplus \mathcal{T}$ or $\mathcal{T}$, depending on whether or not the left half of the plaintext is kept fixed (the right half always is).

Any character ψ of $H \oplus G$ uniquely determines a linear approximation of the cipher. Specifically, the restriction of ψ to H corresponds to the output character of the approximation, and the restriction to G corresponds to the complex conjugate of the input character. The need for complex conjugation is due to technical reasons. Let Z^0 be the set of all such characters ψ corresponding to the linear approximations that were investigated in Section 3. The choice of notation for Z^0 will be motivated in Section 4.2. Concretely, with $\widehat{\mathcal{D}}$ the group of characters of the domain, let

$$Z^0 = \begin{cases} \{\psi : (y_L, y_R, x_L, T_R) \mapsto \overline{\chi(x_L)}\chi(y_L) \mid \chi \in \widehat{\mathcal{D}}\} & \text{for FEA-1 and FF3-1,} \\ \{\psi : (y_L, y_R, T_R) \mapsto \chi(y_L) \mid \chi \in \widehat{\mathcal{D}}\} & \text{for FEA-2.} \end{cases}$$

Note that for all three ciphers, Z^0 is a group under pointwise multiplication of functions. Hence, the collection of these approximations is a multidimensional linear approximation. Finally, let $c : Z^0 \to \mathbb{C}$ be a function that assigns to a group character $\psi \in Z^0$ the correlation $c(\psi)$ of the corresponding linear approximation.

The data complexity of an optimal distinguisher based on a multidimensional linear approximation is inversely proportional to the capacity of the approximation [2], which is defined as the quantity

$$\sum_{\psi \neq \mathbb{1}} |c(\psi)|^2,$$

where the sum is over all nontrivial characters in Z^0. However, as pointed out in Section 3, the correlations $c(\psi)$ are heavily key-dependent and this will affect the optimal data complexity. Nevertheless, by linearity of expectation, it is easy to compute the key-averaged capacity:

$$\mathbb{E} \sum_{\psi \neq \mathbb{1}} |\boldsymbol{c}(\psi)|^2 \approx \begin{cases} N^{2-r/2} & \text{for FEA-1 and FF3-1,} \\ N^{2-r/3} & \text{for FEA-2.} \end{cases}$$

The above calculation suggests a data complexity of $N^{r/2-2}$ for FEA-1 and FF3-1 and $N^{r/3-2}$ for FEA-2. However, as will be shown below, this is somewhat optimistic because the result that relates the capacity to the data complexity of an optimal distinguisher assumes that the correlations $c(\psi)$ are known exactly.

The multidimensional linear approximation can be turned into a distinguisher by directly estimating the capacity. It will be shown in Section 4.3 that the data complexity of this approach can be heuristically estimated as $\sqrt{N}/\sum_{\psi \neq \mathbb{1}} \mathbb{E}|\boldsymbol{c}(\psi)|^2$. However, there exists an equivalent but more direct distinguisher in terms of Pearson's χ^2-statistic.

4.2 Distinguisher Based on Pearson's χ^2 Statistic

The relation between χ^2-distinguishers and multidimensional linear approximations is due to the link between correlations and the Fourier transformation of the probability distribution of the active parts of the input and output state. In particular, the existence of a strong multidimensional approximation can be used to show that a distribution related to the approximations is highly non-uniform.

Pearson's χ^2-statistic will be used as a measure of goodness-of-fit between an estimated (empirical) probability distribution $\widehat{p} : S \to [0, 1]$ and the uniform distribution on S. For this particular case, the χ^2-statistic with q samples satisfies

$$\chi^2/q = M \left\| \widehat{p} - \mathbb{1}/M \right\|_2^2,$$

where $\| \cdot \|_2$ is the Euclidean norm, $M = |S|$ and $\mathbb{1}(x) = 1$ for all $x \in S$. The χ^2-distinguisher succeeds in identifying the real cipher when the χ^2-statistic exceeds some threshold. Indeed, as $q \to \infty$, the estimated distribution $\widehat{p}$ tends to the true distribution p and χ^2/q tends to $M \left\| p - \mathbb{1}/M \right\|_2^2$. In particular, if the tested distribution is uniform, then χ^2/q tends to zero as $q \to \infty$. Statistical aspects will be discussed in Section 4.3.

The link between multidimensional linear approximations and probability distributions is provided by the following result, which generalizes a classical result for G and H vector spaces over $\mathbb{F}_2$ [3, 13]. Below, this result will be used to show that the existence of large correlations leads to highly non-uniform distributions.

Theorem 4.1. *Let $F : G \to H$ be a function between finite Abelian groups G and H. Let Z be a subgroup of the group $H \oplus G$ and let Z^0 be the group of characters of $H \oplus G$ with kernel Z. If $\boldsymbol{x}$ is a uniform random variable on G, then*

$$\Pr\left[(F(\boldsymbol{x}), \boldsymbol{x}) \equiv z \bmod Z\right] = \frac{1}{|Z^0|} \sum_{\psi \in Z^0} C^F_{\psi_H, \overline{\psi}_G} \, \psi(z) \,,$$

where ψ_H is the restriction of ψ to H and ψ_G similarly for G.

Proof. The result is a straightforward consequence of the coordinate-free characterization of multidimensional linear approximations given in [6]. For the sake of completeness, a self-contained proof is given here. Let $S = \{(F(x), x) \mid x \in G\}$ be the graph of F. By the definition of correlation given in Section 3.1,

$$C^F_{\psi_H, \overline{\psi}_G} = \frac{1}{|G|} \sum_{z' \in S} \overline{\psi(z')} \,.$$

It follows that for any $z \in H \oplus G$,

$$\sum_{\psi \in Z^0} C^F_{\psi_H, \overline{\psi}_G} \, \psi(z) = \frac{1}{|G|} \sum_{z' \in S} \sum_{\psi \in Z^0} \overline{\psi(z')} \psi(z) = \frac{1}{|G|} \sum_{z' \in S} \sum_{\psi \in Z^0} \psi(z - z').$$

If $z - z' \in Z$, then $\psi(z - z') = 1$ by the definition of Z^0. If $z - z' \notin Z$, then there exists some character $\chi \in Z^0$ such that $\chi(z - z') \neq 1$. However, since Z^0 is a group under pointwise multiplication, we have

$$\sum_{\psi \in Z^0} \psi(z - z') = \chi(z - z') \sum_{\psi \in Z^0} \psi(z - z') \,.$$

It follows that

$$\sum_{\psi \in Z^0} \psi(z - z') = \begin{cases} |Z^0| & \text{if } z - z' \in Z \\ 0 & \text{otherwise} \,. \end{cases}$$

Since $z - z' \in Z$ is equivalent to $z \equiv z' \bmod Z$, the above implies that

$$\sum_{\psi \in Z^0} C^F_{\psi_H, \overline{\psi}_G} \, \psi(z) \; = |Z^0| \, \Pr\left[(F(\boldsymbol{x}), \boldsymbol{x}) \equiv z \bmod Z\right] .$$

Dividing both sides by $|Z^0|$ gives the result. □

Theorem 4.1 can be applied to the multidimensional linear approximations that were discussed in Section 4.1. For FEA-1 and FEA-2, Z can be taken as the orthogonal complement of the $\mathbb{F}_2$-vector space consisting of the masks in the multidimensional linear approximation. For both FEA-1 and FF3-1, the right half of the plaintext is fixed and reduction modulo Z corresponds to taking the difference of the left half of the ciphertext and the plaintext. More explicitly, if $\mathcal{D}$ is the half-domain of the cipher and $\mathcal{T}$ the space of half-tweaks T_R, then $H = \mathcal{D} \oplus \mathcal{D}$, $G = \mathcal{D} \oplus \mathcal{T}$ and

$$Z = \{(y_L, y_R, x_L, T_R) \in \mathcal{D} \oplus \mathcal{D} \oplus \mathcal{D} \oplus \mathcal{T} \mid y_L - x_L = 0\}.$$

For FEA-2, the full plaintext will be fixed, so $G = \mathcal{T}$. Consequently, reduction modulo Z will amount to truncating the ciphertext to its left half.

As in Section 4.1, let $c(\psi)$ denote the correlation of the approximation corresponding to $\psi \in Z^0$. For all three ciphers, Theorem 4.1 then shows that

$$\Pr\left[(F(\boldsymbol{x}), \boldsymbol{x}) \equiv z \bmod Z\right] = \frac{1}{|Z^0|} \sum_{\psi \in Z^0} c(\psi)\psi(z) \,,$$

where $\boldsymbol{x}$ is uniform random on the input domain (which includes half of the tweak) and F is the mapping to the ciphertext. In fact, the right hand side above is the inverse Fourier transformation of the function $\psi \mapsto c(\psi)$ [24].

A χ^2-distinguisher can now be set up based on the non-uniformity of $(F(\boldsymbol{x}), \boldsymbol{x})$ modulo Z. Denote the probability mass functions of this random variable by $p(z)$ and denote the size of its domain by $M = |G|/|Z| = |Z^0|$. As the number of queries q increases, the empirical distribution approaches p and the χ^2/q statistic approaches the value

$$M \, \|p - \mathbb{1}/M\|_2^2 = \|c - \delta_{\mathbb{1}}\|_2^2 = \sum_{\psi \neq \mathbb{1}} |c(\psi)|^2 \,. \qquad (\dagger)$$

The first equality above follows from the fact that characters are orthogonal functions (as noted in Section 3.1) and is also known as Parseval's theorem [24]. This shows that the χ^2-statistic can be interpreted as an alternative method to estimate the sum of the squared correlations $|c(\psi)|^2$ for $\psi \in Z^0$ with $\psi \neq \mathbb{1}$. As discussed in the next section, this result suggests that the data complexity of the χ^2-distinguisher can be heuristically estimated as $\sqrt{M}/\sum_{\psi\neq\mathbb{1}} \mathbb{E}|\boldsymbol{c}(\psi)|^2$ with $\boldsymbol{c}(\psi)$ the correlation $c(\psi)$ for a uniform random key and $M = N$ for the choices of Z discussed above.

Using the estimates of $\sum_{\psi\neq\mathbb{1}} \mathbb{E}|\boldsymbol{c}(\psi)|^2$ from Section 4.1, the data complexity of the χ^2-distinguishers for r-round FEA-1 and FF3-1 can be estimated as $N^{r/2-1.5}$. For FEA-2, the data complexity estimate becomes $N^{r/3-1.5}$. This is a significant improvement over the linear attacks from Section 3. Furthermore, by considering smaller choices of the group Z, it is still possible to set up χ^2-distinguishers even if only part of the ciphertext is available.

4.3 Cost Analysis and Experimental Verification

As in Section 4.2, consider the χ^2-statistic for the empirical probability distribution of $(F(\boldsymbol{x}), \boldsymbol{x})$ modulo Z, where $\boldsymbol{x}$ is a uniform random input (consisting of the tweak T_R and possibly the right half of the plaintext). Before going into detailed calculations of the advantage of the distinguisher, the heuristic estimate that was used in the previous section will be derived.

Let $\boldsymbol{\chi}^2_{\mathsf{ideal}}$ be the χ^2-statistic when the true distribution is uniform random. This is a good model for the distribution that would be observed for an ideal tweakable block cipher. Likewise, denote the χ^2-statistic for the real cipher by $\boldsymbol{\chi}^2_{\mathsf{real}}$. It is well known that $\boldsymbol{\chi}^2_{\mathsf{ideal}}$ follows a χ^2 distribution with $N-1$ degrees of freedom when the number of queries q is sufficiently large. Hence, $\mathbb{E}\boldsymbol{\chi}^2_{\mathsf{ideal}} = N-1$. For $\boldsymbol{\chi}^2_{\mathsf{real}}$, taking the Fourier transformation (as in (†)) yields

$$\mathbb{E}\boldsymbol{\chi}^2_{\mathsf{real}} = q \sum_{\psi\neq\mathbb{1}} \mathbb{E}\,|\widehat{\boldsymbol{c}}(\psi)|^2$$

where the average is taken with respect to a uniform random key and the random empirical correlations $\widehat{\boldsymbol{c}}(\psi)$ based on q samples. The expected value of $|\widehat{\boldsymbol{c}}(\psi)|^2$ for a fixed key is approximately equal to $|c(\psi)|^2 + 1/q$ when $|c(\psi)|^2$ is negligible compared to one. For a uniform random key, the true correlation $\boldsymbol{c}(\psi)$ is itself a random variable and hence

$$\mathbb{E}\boldsymbol{\chi}^2_{\mathsf{real}} \approx N - 1 + q \sum_{\psi\neq\mathbb{1}} \mathbb{E}\,|\boldsymbol{c}(\psi)|^2 \approx \mathbb{E}\boldsymbol{\chi}^2_{\mathsf{ideal}} + q \sum_{\psi\neq\mathbb{1}} \mathbb{E}\,|\boldsymbol{c}(\psi)|^2 \,.$$

To obtain a low false-positive rate, the decision threshold t should be larger than the standard deviation of $\boldsymbol{\chi}^2_{\mathsf{ideal}}$. That is, $t \geq \sqrt{2(N-1)}$. Hence, a constant advantage can be expected when $\mathbb{E}\boldsymbol{\chi}^2_{\mathsf{real}} - \mathbb{E}\boldsymbol{\chi}^2_{\mathsf{ideal}} \gg \sqrt{N}$. That is,

$$q \gg \sqrt{N}/\textstyle\sum_{\psi\neq\mathbb{1}} \mathbb{E}|\boldsymbol{c}(\psi)|^2 \,.$$

Since the main purpose of this section is to obtain accurate estimates of the advantage for concrete values of N, the above heuristic reasoning will not be formalized here.

It is relatively easy to estimate the average false-positive rate $P_{\mathsf{F}}(t)$ of the χ^2-distinguisher. Indeed, as mentioned above, the statistic $\boldsymbol{\chi}^2_{\mathsf{ideal}}$ follows a χ^2 distribution with $N-1$ degrees of freedom when the number of queries q is sufficiently large. Consequently,

$$P_{\mathsf{F}}(t) = \Pr\left[\boldsymbol{\chi}^2_{\mathsf{ideal}} \geq t\right] \approx 1 - \chi^2_{N-1}(t).$$

The average success-probability $P_{\mathsf{S}}(t)$ is significantly harder to compute. If $\boldsymbol{\chi}^2_{\mathsf{real}}$ denotes the χ^2-statistic for a random sample and a random key, then

$$P_{\mathsf{S}}(t) = \Pr\left[\boldsymbol{\chi}^2_{\mathsf{real}} \geq t\right].$$

To accurately estimate this probability, a Monte-Carlo approach was used to sample from $\boldsymbol{\chi}^2_{\mathsf{real}}$. Sampling from the correlation distribution can be done efficiently, provided that the piling-up approximation is used. A detailed exposition of the sampling strategy is beyond the goals of this paper, but an implementation is provided as supplementary material.

Figure 5 shows the estimated maximum achievable advantage for the χ^2-distinguishers for full-round FEA-1 and FEA-2 with $N = 16$ and FF3-1 with $N = 1000$. The red dots correspond to experimental verifications of the advantage by performing each attack 512 times. These figures confirm the rough data complexity estimate of $N^{r/2-1.5}$.

5 Message Recovery Attacks

In this section, it is shown how the χ^2-distinguishers from Section 4 can be turned into message-recovery attacks. These attacks should be situated in the message-recovery security model of Bellare *et al.* [4]. Informally, this model assumes that the adversary is allowed to (non-adaptively) query the encryption of many *distinct* tweak-message pairs related to a secret message. The distinctness requirement is sufficient to ensure that a trivial guessing attack cannot achieve a nontrivial advantage.

Section 5.1 shows how the left-half of a message encrypted using FEA-1 or FF3-1 can be recovered. The assumptions of the attack are very similar to previous work: the attacker is given the encryption of a target message and a second message with the same right half under many tweaks. Contrary to previous work [4,15], it is not necessary that both messages are encrypted under exactly the same set of tweaks. Instead, part of each tweak (T_L) must be constant. The data complexity of the attack is computed and experimentally verified in Section 5.2.

With more data, it is also possible to recover the right half of messages. This is discussed in Section 5.3. When combined with the left-half recovery attack, this results in recovery of entire messages. The same idea is used to extend the attacks to FEA-2.

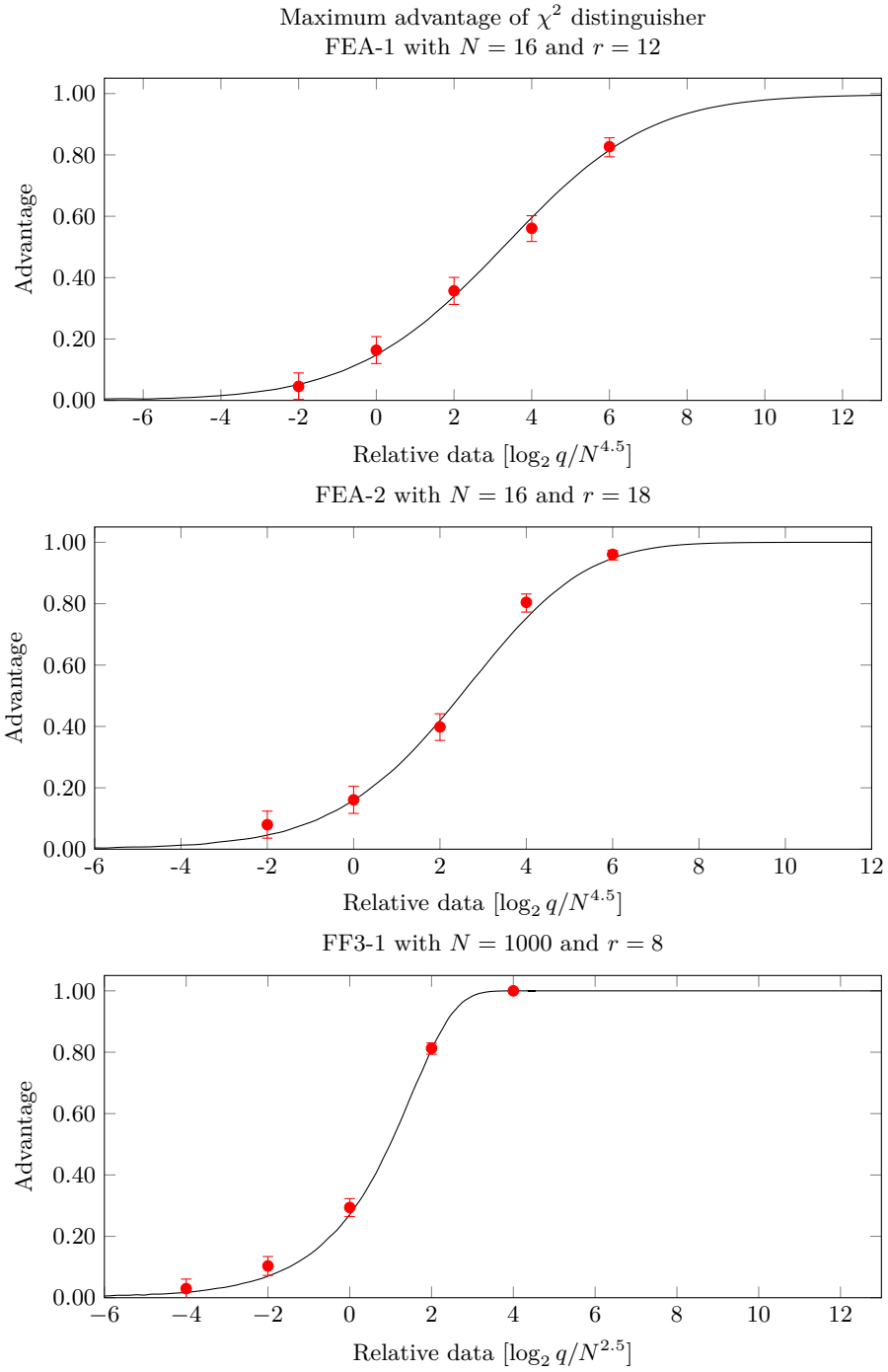


Fig. 5: Theoretical and experimental maximum advantage of the χ^2-distinguishers for full-round FEA-1, FEA-2 and FF3-1. The error bars correspond to 95% Clopper-Pearson confidence intervals.

5.1 Left-Half Recovery for FEA-1 and FF3-1

Consider FEA-1 or FF3-1 with a fixed plaintext input. In this scenario, the χ^2-distinguisher from Section 4.2 is still applicable by using only the left part of the output. That is, $Z = \{(y_L, y_R, T_R) \in \mathcal{D} \oplus \mathcal{D} \oplus \mathcal{T} \mid y_L = 0\}$. The capacity of this multidimensional approximation is the same as before.

The idea behind the message-recovery attack is that a change in the plaintext affects the distribution of the left half of the ciphertext (for uniform random tweaks T_R) in a predictable way. Let $c_1(\psi)$ denote the correlation of the linear approximation corresponding to the character ψ when the plaintext is fixed to (x_L, x_R). Similarly, denote the correlation for a second plaintext (x'_L, x_R) by $c_2(\psi)$. Following the piling-up principle, $c_1(\psi)$ and $c_2(\psi)$ are well-approximated by the correlations of the trails given in Section 3. The two considered functions are the same up to the subtraction of a constant $\Delta = x_L - x'_L$ in the first round of the trail (the third round of the cipher). Hence,

$$c_2(\psi) \approx \psi_{\mathcal{D}}(\Delta) c_1(\psi)$$

with $\psi_{\mathcal{D}}$ the restriction of ψ to the half-domain $\mathcal{D}$. This approximation is highly accurate in practice, since the trails in Figures 2a and 3 are strongly dominant. Denote the probability distribution of the left half of the ciphertext in the first and second case by p_1 and p_2 respectively. Theorem 4.1 implies that

$$p_2(y_L) = \frac{1}{N} \sum_{\psi \in Z^0} c_2(\psi)\psi(y_L) \approx \frac{\psi_{\mathcal{D}}(\Delta)}{N} \sum_{\psi \in Z^0} c_1(\psi)\psi(y_L) = p_1(y_L + \Delta)\,.$$

In other words, the distributions p_1 and p_2 are (nearly) shifted over a distance Δ. It should be emphasized that this is a property of the ciphertext distributions and *not* of individual ciphertexts. As shown in Section 4.2, the distributions p_1 and p_2 are highly non-uniform. This is what makes it possible to recover Δ.

The message-recovery attack begins by estimating the probability distribution (for uniform random tweaks T_R) of the left half of the ciphertext twice: once for the secret plaintext (x_L, x_R) with fixed tweak T_L, and once for an arbitrary message (x'_L, x_R) with the same right half and for the same fixed tweak T_L. Next, for each candidate value Δ_{g} for Δ, compute the statistic

$$r(\Delta_{\mathsf{g}}) = qN/4 \, \|\widehat{p}_1 - \widehat{p}_{\mathsf{g}}\|_2^2,$$

where $\widehat{p}_{\mathsf{g}}(y_L) = \widehat{p}_2(y_L - \Delta_{\mathsf{g}})$ with $\widehat{p}_1$ and $\widehat{p}_2$ the empirical estimates of p_1 and p_2 based on $q/2$ samples each. The statistics $r(\Delta_{\mathsf{g}})$ with $\Delta_{\mathsf{g}} \in \mathcal{D}$ can then be ranked in ascending order. If the number of samples used to obtain the empirical distributions is large enough, the values of Δ_{g} corresponding to the top of the list are likely to be good candidates for Δ.

5.2 Cost Analysis and Experimental Verification

The data complexity of the message-recovery attack can be estimated using a heuristic argument similar to the one that was used for the χ^2-distinguisher in

Section 4.2. For a random sample, the statistic $\boldsymbol{r}(\Delta_{\mathrm{g}})$ satisfies

$$\boldsymbol{r}(\Delta_{\mathrm{g}}) = \frac{q}{4} \sum_{\psi \neq \mathbb{1}} |\widehat{\boldsymbol{c}}_1(\psi) - \overline{\psi_{\mathcal{D}}(\Delta_{\mathrm{g}})}\widehat{\boldsymbol{c}}_2(\psi)|^2,$$

where $\widehat{\boldsymbol{c}}_1(\psi)$ and $\widehat{\boldsymbol{c}}_2(\psi)$ are the empirical correlations and the sum is over all nontrivial $\psi \in Z^0$. When the fixed-key correlation $|c_i(\psi)|^2$ is small, averaging over the sample gives $\mathbb{E}|\widehat{\boldsymbol{c}}_i(\psi)|^2 \approx |c_i(\psi)|^2 + 2/q$. Hence, the average of $\boldsymbol{r}(\Delta_{\mathrm{g}})$ over the sample and over a uniform random key is equal to

$$\begin{aligned}
\mathbb{E}\,\boldsymbol{r}(\Delta_{\mathrm{g}}) &= \frac{q}{4} \sum_{\psi \neq \mathbb{1}} \mathbb{E}\left(|\widehat{\boldsymbol{c}}_1(\psi)|^2 + |\widehat{\boldsymbol{c}}_2(\psi)|^2 - 2\Re\Big\{ \overline{\psi_{\mathcal{D}}(\Delta_{\mathrm{g}})}\, \overline{\widehat{\boldsymbol{c}}_1(\psi)}\widehat{\boldsymbol{c}}_2(\psi) \Big\} \right) \\
&\approx \frac{q}{4} \sum_{\psi \neq \mathbb{1}} \left(\frac{4}{q} + \mathbb{E}|\boldsymbol{c}_1(\psi)|^2 + \mathbb{E}|\boldsymbol{c}_2(\psi)|^2 \right) - \frac{q}{2}\,\Re\Big\{ \sum_{\psi \neq \mathbb{1}} \overline{\psi_{\mathcal{D}}(\Delta_{\mathrm{g}})}\, \mathbb{E}\overline{\boldsymbol{c}_1(\psi)}\boldsymbol{c}_2(\psi) \Big\} \\
&\approx N - 1 + \frac{q}{2} \sum_{\psi \neq \mathbb{1}} \mathbb{E}|\boldsymbol{c}_1(\psi)|^2 - \frac{q}{2} \sum_{\psi \neq \mathbb{1}} \Re\big\{ \psi_{\mathcal{D}}(\Delta - \Delta_{\mathrm{g}}) \big\}\, \mathbb{E}|\boldsymbol{c}_1(\psi)|^2 \,.
\end{aligned}$$

where the third step follows from $c_2(\psi) \approx \psi_{\mathcal{D}}(\Delta) c_1(\psi)$. In fact, $\mathbb{E}|\boldsymbol{c}_1(\psi)|^2$ is nearly constant in ψ. If $\Delta_{\mathrm{g}} \neq \Delta$, then $\sum_{\psi \neq \mathbb{1}} \psi_{\mathcal{D}}(\Delta - \Delta_{\mathrm{g}}) = -1$ and it follows that

$$\mathbb{E}\,\boldsymbol{r}(\Delta_{\mathrm{g}}) - \mathbb{E}\,\boldsymbol{r}(\Delta) \approx q \sum_{\psi \neq \mathbb{1}} \mathbb{E}|\boldsymbol{c}_1(\psi)|^2 \,.$$

In particular, if $q \gg \sqrt{N} / \sum_{\psi \neq \mathbb{1}} \mathbb{E}|\boldsymbol{c}_1(\psi)|^2$, then $\mathbb{E}\,\boldsymbol{r}(\Delta_{\mathrm{g}}) - \mathbb{E}\,\boldsymbol{r}(\Delta) \gg \sqrt{N}$. This is sufficient to obtain a constant advantage since the standard deviation of $\boldsymbol{r}(\Delta_{\mathrm{g}})$ is of the order $\sqrt{N}$. This can be motivated by noting that, for a uniform output distribution, the distribution of $\boldsymbol{r}(\Delta_{\mathrm{g}})$ would be asymptotically χ^2 with $N - 1$ degrees of freedom. Hence, $\widetilde{\mathcal{O}}(N^{r/2-1.5})$ data should suffice to obtain a constant message-recovery advantage.

No attempt will be made here to make the above argument rigorous. Instead, accurate estimates of the message-recovery advantage for specific values of N can be computed using a Monte-Carlo approach. The main ingredient is a method to sample from the correlation distributions, which is identical to the one used for the calculations in Section 4.3. Results for full-round FEA-1 with $N = 16$ and FF3-1 with $N = 1000$ are shown in Figure 6, along with experimental estimates of the advantage.

Observe that for FF3-1 with $q = 4 \times \lfloor 2N^{2.5} \rceil \approx 2^{28}$, the theoretical advantage is an overestimate. This is due to the fact that only 2^{28} data is available for a fixed choice of the plaintext and tweak T_L. Once the variations in the ideal distribution (which was assumed to be uniform in the analysis) are of the same order as the sampling error, the advantage begins to flatten off. However, this does not imply that the advantage of the FF3-1 message-recovery attack cannot be made close to one. Indeed, one can simply perform the attack for a different choice of T_L. Of course, for even larger N, the maximum advantage that can be achieved using one choice of T_L decreases and the attack eventually becomes infeasible. Based

on the estimated data complexity of the attack and Figure 6, this is expected to occur for $N > 2^{12}$. The right-half recovery attack from Section 5.3 avoids this problem and can be used for all $N < 2^{19}$, but it has a higher overall data complexity.

5.3 Right-Half Recovery and Application to FEA-2

The left-half recovery attack on FEA-1 and FF3-1 could also be applied for two messages (x_L, x_R) and (x'_L, x'_R) with $x_R \neq x'_R$. However, the recovered difference would then be $\Delta = x_L - x'_L + F_1(x_R) - F_1(x'_R)$. If $x_L - x'_L$ is known, then the adversary can recover Δ to obtain the difference $F_1(x_R) - F_1(x'_R)$. This is useful because it leads to a right-half recovery attack. In addition, the output differences will be directly used in the key-recovery attack on FEA-1 that is described in Section 6. It is also possible to apply the same attack with a different choice of Z that includes the left half of the plaintext. In this case, the recovered difference would simply be $F_1(x_R) - F_1(x'_R)$ due to reduction modulo Z. The main advantage of this approach is that it increases the amount of available data per choice of the right half by a factor of N. This extends the reach of the attack to $N < 2^{19}$, compared to $N < 2^{12}$ for left-half recovery.

The right-half can be recovered by guessing x'_R until the recovered difference is zero. This does not violate the distinctness requirement of the message-recovery framework, since the tweaks T_R and the left halves of the guessed messages can be different from those of the secret message. The attack proceeds by computing the statistics $r(0)$ from Section 5.1 with $\widehat{p}_1$ the empirical distribution for the secret message and $\widehat{p}_2$ the empirical distribution with right-half guess x'_R. If these statistics are ranked in ascending order, the values of x'_R corresponding to the top of the list are the most promising candidates for x_R. By the analysis in Section 5.2, this attack requires $\widetilde{\mathcal{O}}(N^{r/2-0.5})$ data. A simulation of the maximum advantage is shown in the bottom of Figure 6, along with experimental results. Note that the error bars are wider than for the left-half recovery experiments because each data point was estimated using only 40 runs of the attack (to limit the time complexity of the experiment).

The same idea as above can be used to extend the message-recovery attack to FEA-2. For example, consider left-half recovery. In this case, the adversary queries the encryption of the secret message (x_L, x_R) under many tweaks with constant T_L. In addition, for each guess of x'_L, similar queries are made for (x'_L, x_R). The same process as above can be used to identify the values of x_L for which

$$F_2(x_L + F_1(x_R)) + x_R = F_2(x'_L + F_1(x_R)) + x_R\,.$$

However, there is an additional issue that must be addressed: since the approximation shown in Figure 2b does not have equal input and output masks, the effect of changing the plaintext input on the correlations is more complicated. Nevertheless, one can still use the same approach (with roughly the same data complexity) to check for equality between the two output distributions.

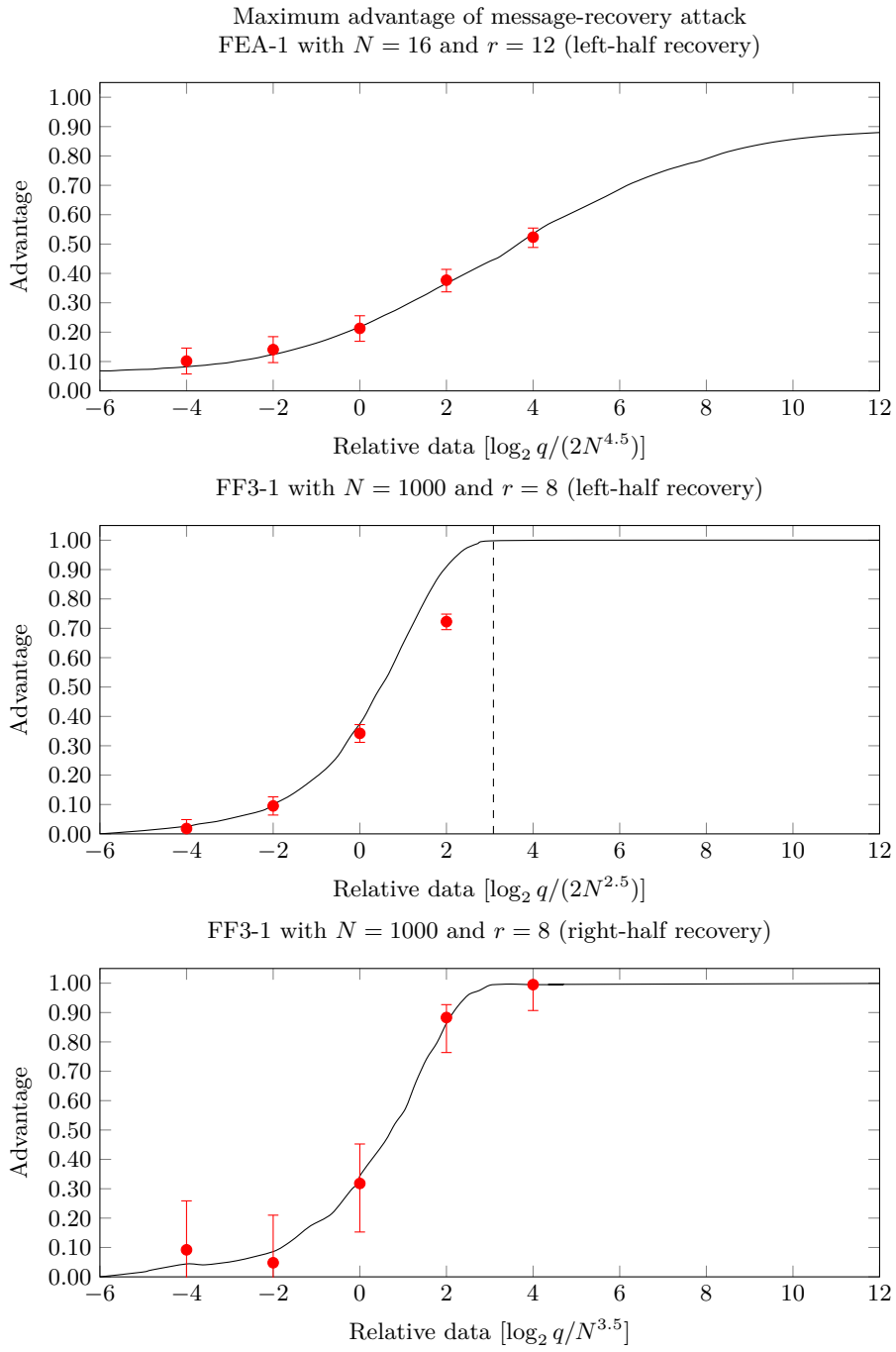


Fig. 6: Theoretical and experimental maximum advantage of the message-recovery attacks for full-round FEA-1 and FF3-1. The error bars correspond to 95% Clopper-Pearson confidence intervals. The dashed vertical line corresponds to a data complexity of 2×2^{28}.

6 Key-Recovery Attack on FEA-1

This section shows how the left-half message-recovery attack on FEA-1 from Section 5.1 can be used for key-recovery. Naturally, the attack heavily depends on the internal details of the round function F_1. For FF3-1, key-recovery is not feasible since the round functions are truncations of the AES.

The FEA-1 round function is illustrated in Figure 7. It consists of two iterations of a key-addition layer, an S-box layer and a linear layer with branch number nine. Each of these layers acts on a state in a vector space $\mathbb{F}_{2^8}^8$. The round keys will be denoted by K_a and K_b. The round function F_1 is defined as the truncation of this structure to m bits.

The exact choice of the matrix representation M of the linear layer is not important. The only property of M that will be used is the fact that it has branch number nine (equivalently, is MDS). The S-box is based on inversion in $\mathbb{F}_{2^8}$, but the details are not important. However, it is important that for all nonzero Δ_1 and Δ_2, the equation $S(x + \Delta_1) = S(x) + \Delta_2$ has either no, two or four solutions in x. For each $\Delta_1 \neq 0$, the case with four solutions occurs for exactly one choice of Δ_2.

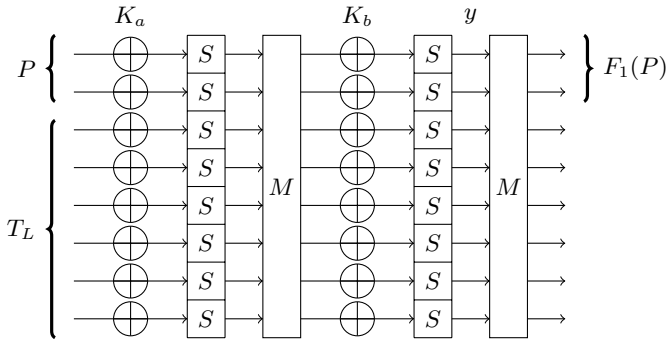


Fig. 7: Round function of FEA-1 with round keys K_a and K_b.

Recall from Section 5.3 that it is possible to recover output differences $F_1(P) + F_1(P')$ for an arbitrary choice of P and P'. The idea behind the key-recovery attack is to guess parts of the internal state of the round function and to check the validity of these guesses using such output differences. After recovering the relevant parts of the internal state, the round keys can be recovered.

Let x denote the first byte of the round function input $P \| T_L$. Observe that byte i of the internal state y (indicated in Figure 7) can be written as

$$y_i = S(\gamma_i + M_{i,1}\, S(K_{a,1} + x))\,,$$

where $\gamma_1, \ldots, \gamma_8 \in \mathbb{F}_{2^8}$ are constants depending on the round keys K_a and K_b (but not on the first byte $K_{a,1}$) and on the tweak T_L. Importantly, $\gamma_1, \ldots, \gamma_8$ do

not depend on x. Specifically,

$$\gamma_i = K_{b,i} + \sum_{j=2}^{8} M_{i,j} S([P \| T_L]_j + K_{a,j}) \,.$$

In Section 6.1, it will be shown how $K_{a,1}$ and γ_i can be recovered using a limited number of output differences. Section 6.2 then shows how the entire round keys K_a and K_b can be extracted from these constants and a few additional output differences.

6.1 Recovering $K_{a,1}$ and the Internal Constants γ_i

It is clear from Figure 7 that the output difference is a linear function of the difference between the internal states y and y' (corresponding to two inputs x and x'). Furthermore, since M is an invertible matrix, this function is of rank m. Hence, $y + y'$ can take $2^{64-m} = 2^{64}/N$ possible values. By computing an echelon form for the linear function that maps $y + y'$ to the output difference, these candidate solutions can easily be enumerated. For each guess of $y + y'$, one obtains the values

$$y_i + y'_i = S(\gamma_i + M_{i,1} S(K_{a,1} + x)) + S(\gamma_i + M_{i,1} S(K_{a,1} + x')) \,.$$

For each $i = 1, \ldots, 8$, one can determine the set of possible input differences $S(K_{a,1} + x) + S(K_{a,1} + x')$ that can lead to the known difference $y_i + y'_i \neq 0$. Due the properties of S, there are 127 possible input differences. Hence, each i reduces the number of candidate differences by a factor $127/255 < 1/2$. It follows that the difference $S(K_{a,1} + x) + S(K_{a,1} + x')$ can be uniquely determined. However, since the difference $x + x'$ is known, two candidates for $K_{a,1}$ can be computed from the difference equation. The case with four solutions is unlikely to occur and does not significantly affect the overall time and data complexity of the attack.

Once $K_{a,1}$ has been determined (as one of two possible values), the constants γ_i can also be obtained by solving a difference equation. In particular, since the case with four solutions is rare, one usually ends up with two candidates for each γ_i. To check the validity of these candidates, additional output differences will be used. To save data, one of x or x' can be reused. For each of the 2^9 candidate values, the expected output difference should then be computed and compared to the observed difference. This requires roughly 2^{12} S-box evaluations. If the candidate values are wrong, the output difference will match in roughly $1/N$ of the cases. Hence, the computational cost is dominated by the calculation of the expected output difference for the first pair.

The total number of candidates for the difference $y + y'$, the internal constants and the first byte of K_a is $2^{64+9}/N = 2^{73}/N$. Hence, $\lceil 73/m - 1 \rceil$ pairs are sufficient to obtain a unique solution. For $m = 4$, the number of available input differences is too small to obtain a unique candidate. However, this is not a major

issue since the time complexity of the round key recovery procedure described in Section 6.2 is small enough that it can be repeated several times.

The data complexity of the above process is $(\lceil 73/m - 1 \rceil + 1)q/2$ queries, where q is the data complexity of the left-half message-recovery attack. This comes with an equal computational cost, measured in FEA-1 evaluations. The remaining computational cost is dominated by $2^{64+12}/N$ S-box evaluations. Since the cipher contains 12×16 S-boxes, one can conservatively estimate that this takes less time than $2^{68}/N$ evaluations of full-round FEA-1.

6.2 Recovering the Round Keys

Once the constants $\gamma_1, \ldots, \gamma_8$ have been recovered, obtaining the round keys K_a and K_b is relatively easy. In particular, recall that

$$\gamma_i = K_{b,i} + \sum_{j=2}^{8} M_{i,j} S([P \| T_L]_j + K_{a,j}) \,.$$

Suppose $P \| T_L$ and $P' \| T'_L$ differ only in byte $j \in \{2, \ldots, 8\}$ and let γ'_i be the new value of γ_i for input $P' \| T'_L$. It is easy to see that

$$\gamma_i + \gamma'_i = M_{i,j} S([P \| T_L]_j + K_{a,j}) + M_{i,j} S([P' \| T'_L]_j + K_{a,j}) \,.$$

Hence, after guessing $K_{a,j}$, one can compute the new constants γ'_i and the expected output differences for pairs with tweak T'_L. To obtain a unique (up to a constant) candidate for $K_{a,j}$, a total of $\lceil 8/m \rceil$ differences are sufficient. Recovering all of the bytes of K_a thus requires $7 \times \lceil 8/m \rceil$ differences. Once K_a is recovered, K_b can be computed directly.

To conclude, the data complexity of this step is $7q/2 \times (\lceil 8/m \rceil + 1)$ with q the data complexity of the left-half message-recovery attack. A few additional pairs will be required to filter spurious candidates for $K_{a,j}$, or if no unique solution for the constants $\gamma_1, \ldots, \gamma_8$ was obtained in the first step of the attack ($m = 4$). The time complexity, excluding the time required for message-recovery, is negligible compared to that of the first step.

6.3 Recovering All Round Keys

By the results in Sections 6.1 and 6.2, the round keys K_a and K_b of the first round function can be recovered using at most $\lceil 73/m - 1 \rceil + 7\lceil 8/m \rceil \leq 16\lceil 8/m \rceil$ evaluations of the left-half message-recovery attack and additional time equivalent to at most $2^{68}/N$ FEA-1 evaluations. If q is the amount of data required for the left-half recovery attack, this amounts to a total of less than $8\lceil 8/m \rceil q + 4q$ queries. However, the FEA-1 key-schedule is a Lai-Massey structure that generates two round keys per iteration. Hence, the remaining round keys can not be obtained by iterating the key-schedule without knowing the round keys for the second round. To obtain these keys, it suffices to perform the same key-recovery attack on F_2. Hence, the total cost is less than $16\lceil 8/\log_2 N \rceil q + 8q$ data for left-half recoveries and additional time equivalent to less than $2^{69}/N$ evaluations of FEA-1.

7 Conclusion

It was shown that the format-preserving encryption standards FF3-1, FEA-1 and FEA-2 are all vulnerable to linear cryptanalysis. More generally, the analysis in this paper is applicable to any small-domain Feistel cipher with alternating round tweaks.

The attacks rely on the ability to vary the tweaks in even-numbered rounds (FF3-1 and FEA-1) or rounds numbered by a multiple of three (FEA-2), while keeping the tweaks in the other rounds fixed. Combined with the observation that the variance of the correlation of a nontrivial linear approximation over a small random function is quite large, this results in strong linear trails through the cipher. The analysis of FF3-1 is also of theoretical interest as an application of the theory of linear cryptanalysis over the group $\mathbb{Z}/N\mathbb{Z}$.

The data requirements of the basic linear distinguishers were reduced using multidimensional linear cryptanalysis. Based on the same principle, efficient message-recovery attacks were obtained. For FEA-1, the message-recovery attack was in turn extended to a key-recovery attack.

For many instances of FF3-1, FEA-1 and FEA-2, the data requirements of the new attacks are small enough to be a practical concern for users of these standards.

Acknowledgments. I thank Dongyoung Roh (ETRI) and Morris Dworkin (NIST) for useful comments on an early draft of this work, and Vincent Rijmen for proofreading the paper. The author is supported by a PhD Fellowship from the Research Foundation – Flanders (FWO).

References

1. Amon, O., Dunkelman, O., Keller, N., Ronen, E., Shamir, A.: Three third generation attacks on the format preserving encryption scheme ff3. Cryptology ePrint Archive, Report 2021/335 (2021), https://eprint.iacr.org/2021/335
2. Baignères, T., Junod, P., Vaudenay, S.: How far can we go beyond linear cryptanalysis? In: Lee, P.J. (ed.) ASIACRYPT 2004. LNCS, vol. 3329, pp. 432–450. Springer, Heidelberg (Dec 2004). https://doi.org/10.1007/978-3-540-30539-2_31
3. Baignères, T., Stern, J., Vaudenay, S.: Linear cryptanalysis of non binary ciphers. In: Adams, C.M., Miri, A., Wiener, M.J. (eds.) SAC 2007. LNCS, vol. 4876, pp. 184–211. Springer, Heidelberg (Aug 2007). https://doi.org/10.1007/978-3-540-77360-3_13
4. Bellare, M., Hoang, V.T., Tessaro, S.: Message-recovery attacks on feistel-based format preserving encryption. In: Weippl, E.R., Katzenbeisser, S., Kruegel, C., Myers, A.C., Halevi, S. (eds.) Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, Vienna, Austria, October 24-28, 2016. pp. 444–455. ACM (2016). https://doi.org/10.1145/2976749.2978390, https://doi.org/10.1145/2976749.2978390
5. Bellare, M., Ristenpart, T., Rogaway, P., Stegers, T.: Format-preserving encryption. In: Jr., M.J.J., Rijmen, V., Safavi-Naini, R. (eds.) Selected Areas in Cryptography, 16th Annual International Workshop, SAC 2009, Calgary, Alberta, Canada,

August 13-14, 2009, Revised Selected Papers. Lecture Notes in Computer Science, vol. 5867, pp. 295–312. Springer (2009). https://doi.org/10.1007/978-3-642-05445-7_19, https://doi.org/10.1007/978-3-642-05445-7_19
6. Beyne, T.: Linear Cryptanalysis in the Weak Key Model. Master's thesis, KU Leuven (2019), https://homes.esat.kuleuven.be/~tbeyne/masterthesis/thesis.pdf
7. Black, J., Rogaway, P.: Ciphers with arbitrary finite domains. In: Preneel, B. (ed.) CT-RSA 2002. LNCS, vol. 2271, pp. 114–130. Springer, Heidelberg (Feb 2002). https://doi.org/10.1007/3-540-45760-7_9
8. Daemen, J., Govaerts, R., Vandewalle, J.: Correlation matrices. In: Preneel, B. (ed.) FSE'94. LNCS, vol. 1008, pp. 275–285. Springer, Heidelberg (Dec 1995). https://doi.org/10.1007/3-540-60590-8_21
9. 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
10. Dunkelman, O., Kumar, A., Lambooij, E., Sanadhya, S.K.: Cryptanalysis of feistel-based format-preserving encryption. Cryptology ePrint Archive, Report 2020/1311 (2020), https://eprint.iacr.org/2020/1311
11. Durak, F.B., Vaudenay, S.: Breaking the FF3 format-preserving encryption standard over small domains. In: Katz, J., Shacham, H. (eds.) CRYPTO 2017, Part II. LNCS, vol. 10402, pp. 679–707. Springer, Heidelberg (Aug 2017). https://doi.org/10.1007/978-3-319-63715-0_23
12. Dworkin, M.: Recommendation for block cipher modes of operation: methods for format-preserving encryption. NIST Special Publication **800 38Gr1** (February 2019). https://doi.org/10.6028/NIST.SP.800-38Gr1-draft
13. Hermelin, M., Cho, J.Y., Nyberg, K.: Multidimensional linear cryptanalysis of reduced round Serpent. In: Mu, Y., Susilo, W., Seberry, J. (eds.) ACISP 08. LNCS, vol. 5107, pp. 203–215. Springer, Heidelberg (Jul 2008)
14. Hoang, V.T., Miller, D., Trieu, N.: Attacks only get better: How to break FF3 on large domains. In: Ishai, Y., Rijmen, V. (eds.) EUROCRYPT 2019, Part II. LNCS, vol. 11477, pp. 85–116. Springer, Heidelberg (May 2019). https://doi.org/10.1007/978-3-030-17656-3_4
15. Hoang, V.T., Tessaro, S., Trieu, N.: The curse of small domains: New attacks on format-preserving encryption. In: Shacham, H., Boldyreva, A. (eds.) CRYPTO 2018, Part I. LNCS, vol. 10991, pp. 221–251. Springer, Heidelberg (Aug 2018). https://doi.org/10.1007/978-3-319-96884-1_8
16. Lee, J.K., Koo, B., Roh, D., Kim, W.H., Kwon, D.: Format-preserving encryption algorithms using families of tweakable blockciphers. In: Lee, J., Kim, J. (eds.) Information Security and Cryptology - ICISC 2014. pp. 132–159. Springer International Publishing, Cham (2015)
17. Matsui, M.: Linear cryptanalysis method for DES cipher. In: Helleseth, T. (ed.) EUROCRYPT'93. LNCS, vol. 765, pp. 386–397. Springer, Heidelberg (May 1994). https://doi.org/10.1007/3-540-48285-7_33
18. Patarin, J.: New results on pseudorandom permutation generators based on the DES scheme. In: Feigenbaum, J. (ed.) CRYPTO'91. LNCS, vol. 576, pp. 301–312. Springer, Heidelberg (Aug 1992). https://doi.org/10.1007/3-540-46766-1_25
19. Patarin, J.: Generic attacks on Feistel schemes. In: Boyd, C. (ed.) ASIACRYPT 2001. LNCS, vol. 2248, pp. 222–238. Springer, Heidelberg (Dec 2001). https://doi.org/10.1007/3-540-45682-1_14

20. Patarin, J.: Security of random Feistel schemes with 5 or more rounds. In: Franklin, M. (ed.) CRYPTO 2004. LNCS, vol. 3152, pp. 106–122. Springer, Heidelberg (Aug 2004). https://doi.org/10.1007/978-3-540-28628-8_7
21. Rijmen, V., Daemen, J., Preneel, B., Bosselaers, A., De Win, E.: The cipher SHARK. In: Gollmann, D. (ed.) FSE'96. LNCS, vol. 1039, pp. 99–111. Springer, Heidelberg (Feb 1996). https://doi.org/10.1007/3-540-60865-6_47
22. Selçuk, A.A.: On probability of success in linear and differential cryptanalysis. Journal of Cryptology **21**(1), 131–147 (Jan 2008). https://doi.org/10.1007/s00145-007-9013-7
23. Tardy-Corfdir, A., Gilbert, H.: A known plaintext attack of FEAL-4 and FEAL-6. In: Feigenbaum, J. (ed.) CRYPTO'91. LNCS, vol. 576, pp. 172–181. Springer, Heidelberg (Aug 1992). https://doi.org/10.1007/3-540-46766-1_12
24. Terras, A.: Fourier analysis on finite groups and applications. Cambridge University Press (1999)
25. Vaudenay, S.: An experiment on DES statistical cryptanalysis. In: Gong, L., Stern, J. (eds.) ACM CCS 96. pp. 139–147. ACM Press (Mar 1996). https://-doi.org/10.1145/238168.238206