KFC - The Krazy Feistel Cipher

Thomas Baignères* and Matthieu Finiasz

EPFL
CH-1015 Lausanne – Switzerland
http://lasecwww.epfl.ch

Abstract. We introduce KFC, a block cipher based on a three round Feistel scheme. Each of the three round functions has an SPN-like structure for which we can either compute or bound the advantage of the best d-limited adaptive distinguisher, for any value of d. Using results from the decorrelation theory, we extend these results to the whole KFC construction. To the best of our knowledge, KFC is the first practical (in the sense that it can be implemented) block cipher to propose tight security proofs of resistance against large classes of attacks, including most classical cryptanalysis (such as linear and differential cryptanalysis, taking hull effect in consideration in both cases, higher order differential cryptanalysis, the boomerang attack, differential-linear cryptanalysis, and others).

1 Introduction

Most modern block ciphers are designed to resist a wide range of cryptanalytic techniques. Among them, one may cite linear cryptanalysis [19,20,23], differential cryptanalysis [7,8], as well as several variants such as impossible differentials [5], the boomerang attack [27] or the rectangle attack [6]. Proving resistance against all these attacks is often tedious and does not give any guarantee that a subtle new variant would not break the construction. Rather than considering all known attacks individually, it would obviously be preferable to give a *unique* proof, valid for a family of attacks.

In [26], Vaudenay shows that the decorrelation theory provides tools to prove security results in the Luby-Rackoff model [18], i.e., against adversaries only limited by the number of plaintext/ciphertext pairs they can access. Denoting d this number of pairs, the adversaries are referred to as *d-limited distinguishers*. Unfortunately, this class of adversaries does not capture the most widely studied statistical attacks such as linear and differential cryptanalysis. Instead, these attacks are formalized by so-called *iterated attacks of order d* [25]. This class of attacks was initially inspired by linear and differential cryptanalysis and actually formalizes most of the possible statistical attacks against block ciphers. For example, linear cryptanalysis is an iterated attack of order 1, differential cryptanalysis is of order 2, and higher order differential cryptanalysis [15, 16] of order i is an iterated attack of order $d = 2^i$.

It is proven that resistance against all $2d$-limited distinguishers is sufficient to resist iterated attacks of order d [26]. Consequently, designing a block cipher

* Supported by the Swiss National Science Foundation, 200021-107982/1.

resistant to d-limited distinguishers for a large d is enough to resist most standard attacks against block ciphers. Obviously, this is not a trivial task as, to the best of our knowledge, no *efficient* block cipher was ever designed to resist d-limited distinguishers for $d > 2$ [14, 26].

In a previous article entitled "Dial C for Cipher" [1], we presented a block cipher construction provably resistant against (among others) linear and differential cryptanalysis (where the linear hull [21] and differentials [17] effects are taken into account, which is unfortunately not usual in typical proofs of security of block ciphers), several of their variants, 2-limited distinguishers and thus, all iterated attacks of order 1. Our aim in this article, is to design a block cipher based on the same principles as C but provably secure against d-limited distinguishers for large values of d. We call this construction KFC as it is based on a Feistel scheme. KFC is practical in the sense that it can be implemented and reach a throughput of a few Mbits/s. Just as the typical security proofs of block ciphers do not compare to ours, the encryption speed reached by KFC does not compare to those of nowadays block ciphers.

Constructions based on the decorrelation theory have already been proposed. COCONUT98 [24] was one of the first efficient block cipher based on decorrelation concepts. It resists 2-limited distinguishers but can be attacked by David Wagner's boomerang attack [27], which is an iterated attack of order 4. Of course this does not prove that the decorrelation theory is useless, but only that decorrelation results do not prove more than what they claim. KFC is designed to resist d-limited distinguishers (and consequently, iterated attacks up to a given order), nothing more.

High Overview and Outline of the Paper. Before building a provably secure block cipher, we need to define precisely against which class of attacks it should be resistant. The adversary model and some reminders about the decorrelation theory are given in Section 2. Then, in Section 3, we give some hints about why we chose to use a Feistel scheme [13] for KFC. A description of the structure of the random functions we use in the Feistel scheme is then given in Section 4. The exact advantage of the best 2-limited distinguisher is computed in Section 5, and in Section 6, we bound the advantage of higher order adversaries.

2 Security Model

In this paper, a *perfectly random function* (resp. *permutation*) denotes a random function (resp. permutation) uniformly distributed among all possible functions (resp. permutations). Consequently, when referring to a *random function* or a *random permutation*, nothing is assumed about its distribution.

The Luby-Rackoff Model [18]. We consider an adversary $\mathcal{A}$ with unbounded computational power, only limited by its number of queries d to an oracle $\mathcal{O}$ implementing a random permutation. The goal of $\mathcal{A}$ is to guess whether $\mathcal{O}$ is implementing an instance drawn uniformly among the permutations defined by a block cipher C or among all possible permutations, knowing that these two events have probability $\frac{1}{2}$ and that one of them is eventually true. Such an adversary is referred to as a *d-limited adaptive distinguisher* when he adaptively chooses his queries depending on previous answers from the Oracle or as a *d-limited non-adaptive distinguisher* when all the queries are made at once. In both cases, the ability of $\mathcal{A}$ to succeed is measured by mean of its *advantage*.

Definition 1. *The advantage of $\mathcal{A}$ of distinguishing two random functions F_0 and F_1 is defined by* $\mathrm{Adv}_{\mathcal{A}}(F_0, F_1) = \big| \Pr[\mathcal{A}(F_0) = 0] - \Pr[\mathcal{A}(F_1) = 0] \big|$.

Informally, a secure block cipher C (i.e., a random permutation) should be indistinguishable from a perfectly random permutation C^*, i.e., the advantage $\mathrm{Adv}_{\mathcal{A}}(C, C^*)$ of any adversary $\mathcal{A}$ should be negligible. A secure random function F should be indistinguishable from a perfectly random function F^*, i.e., the advantage $\mathrm{Adv}_{\mathcal{A}}(F, F^*)$ of any adversary $\mathcal{A}$ should be negligible. Apart from very specific (and usually non-practical) constructions, computing the exact advantage of the best d-limited distinguisher is not straightforward. The decorrelation theory [26] gives some tools that will allow us to compute (or at least bound) this advantage for KFC.

Reminders on the Decorrelation Theory. Let $F : \{0,1\}^n \to \{0,1\}^n$ be a random function. The *distribution matrix* $[F]^d$ of F at order d is a $2^{nd} \times 2^{nd}$ matrix defined by $[F]^d_{(x_1,\ldots,x_d),(y_1,\ldots,y_d)} = \Pr_F[F(x_1) = y_1, \ldots, F(x_d) = y_d]$. If F_1 and F_2 are two independent random functions, we have $[F_2 \circ F_1]^d = [F_1]^d \times [F_2]^d$. The advantage of the best distinguisher between F and F^* only depends on the *distance* between $[F]^d$ and $[F^*]^d$, whose exact definition will depend on whether the considered distinguisher is adaptive or not.

Definition 2. *Let $A \in \{0,1\}^{nd} \times \{0,1\}^{nd}$ be a matrix indexed by d-tuples of elements in $\{0,1\}^n$. We let:*

$$|||A|||_\infty = \max_{x_1,\ldots,x_d} \sum_{y_1,\ldots,y_d} \left| A_{(x_1,\ldots,x_d),(y_1,\ldots,y_d)} \right| \quad \textit{and}$$
$$\|A\|_a = \max_{x_1} \sum_{y_1} \cdots \max_{x_d} \sum_{y_d} \left| A_{(x_1,\ldots,x_d),(y_1,\ldots,y_d)} \right| .$$

Property 3 (Theorems 10 and 11 in [26]). *Let F be a random function and F^* be a perfectly random function. The advantage of the best d-limited* non-adaptive *distinguisher $\mathcal{A}$ is such that* $\mathrm{Adv}_{\mathcal{A}}(F, F^*) = \frac{1}{2}|||[F]^d - [F^*]^d|||_\infty$ *whereas the advantage of the best d-limited* adaptive *distinguisher $\mathcal{A}_a$ is such that* $\mathrm{Adv}_{\mathcal{A}_a} = \frac{1}{2}\|[F]^d - [F^*]^d\|_a$.

An iterated attack of order d consists in iterating independent non-adaptive d-limited attacks with random inputs. The algorithm of Fig. 1 gives a more formal definition of this concept. For example, linear cryptanalysis is an iterated attack of order 1 where $\mathcal{T}(X, Y) = a \bullet X \oplus b \bullet Y$ (where a and b respectively denote the input and output masks) and where X is an uniformly distributed random variable on text space. Similarly, differential cryptanalysis is an iterated attack of order 2 where $\mathcal{T}((X_1, X_2), (Y_1, Y_2))$ is 1 when $Y_1 \oplus Y_2 = b$ and 0 otherwise and where X_1 is a uniformly distributed random variable and $X_2 = X_1 \oplus a$.

Parameters: a complexity n, a distribution on X, a test function $\mathcal{T}$ outputting one bit, a set $\mathcal{S}$
Oracle: a permutation C

```
for i = 1, ..., n do
    pick X = (X_1, ..., X_d) at random
    get Y = (C(X_1), ..., C(X_d))
    set T_i = T(X, Y)
end for
if (T_1, ..., T_n) ∈ S then output 1 else output 0 end if
```

Fig. 1. Iterated attack of order d.

As proved in Theorem 18 in [26] bounding the advantage of the best $2d$-limited non-adaptive adversary is sufficient to bound the advantage of any adversary performing an iterated attack of order d. Roughly speaking, a block cipher C with a negligible order $2d$ decorrelation $|||[C]^{2d}-[C^*]^{2d}|||_\infty$ is resistant to iterated attacks of order d.

3 From the SPN of C to the Feistel Scheme of KFC

The block cipher C (introduced in [1, 2]) is based on the same substitution-permutation network (SPN) as the AES [11], except that the fixed substitution boxes are replaced by mutually independent and perfectly random permutations. It achieves goals similar to those we want to achieve with KFC: being resistant against 2-limited adversaries, it is secure against all iterated attacks of order 1. These results were obtained by exploiting strong symmetries (induced by intrinsic symmetries of the confusion and diffusion layers) in the order 2 distribution matrix of C. Unfortunately, we were not able to exhibit similar symmetries for higher orders. It appears that layers of perfectly random permutations are suitable for proving security results at order 2, not above.

Instead of explicitly computing the advantage of a d-limited distinguisher we will try to bound it by a function of the advantage of the best $(d-1)$-limited distinguisher, and apply this bound recursively down to order 2 (which we know how to compute). This seems clearly impossible with layers of random permutations as two distinct inputs will always lead to two correlated outputs. However, this is not the case anymore when considering a layer of mutually independent and perfectly random *functions*. For instance, two distinct inputs of a perfectly random function yield two independent outputs. Similarly, if the two inputs of a layer of functions are distinct on each function input, the outputs are independent. This extends well to a set of d texts: if one text is different from *all* the others on *all* function inputs, the corresponding output is independent from all other outputs. A formal treatment of this idea is given in Section 4.

However, layers of random functions cannot always be inverted and thus do not fit in a classical SPN structure. The straightforward solution is to use a

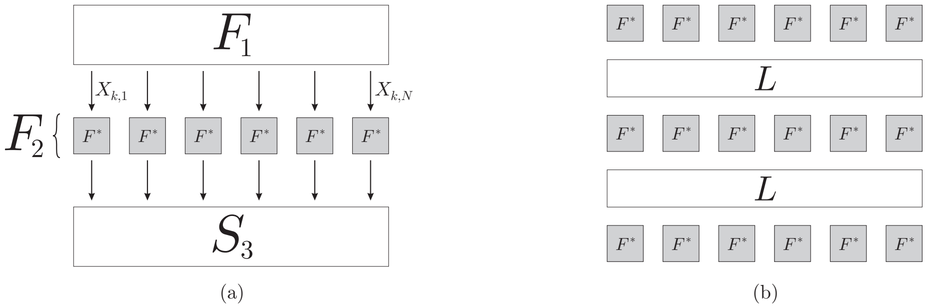

Fig. 2. Increasing the decorrelation order using a layer made of small *independent* and *perfectly random* functions.

Feistel scheme [13]. Moreover, decorrelation results on the round functions of a Feistel scheme extend well to the whole construction.

Theorem 4 (Theorem 21 in [26]). *Let F^* be a uniformly distributed random function on $\{0,1\}^n$. Let $F_1, \ldots, F_r$ be r independent random functions on $\{0,1\}^n$ such that $\mathrm{Adv}_{\mathcal{A}}(F_i, F^*) \leq \epsilon$ $(i = 1, \ldots, r)$ for any adversary $\mathcal{A}$. Let $C = \Psi(F_1, \ldots, F_r)$ be an r round Feistel cipher on $\{0,1\}^{2n}$. For any adversary $\mathcal{A}$ limited to d queries and for any integer $k \geq 3$, we have:*

$$\mathrm{Adv}_{\mathcal{A}}(C, C^*) \leq \frac{1}{2}\left(2k\epsilon + \frac{2d^2}{2^{n/2}}\right)^{\lfloor r/k \rfloor}.$$

This theorem shows that if we can instantiate *independent* random functions secure against all d-limited distinguishers, we can obtain a block cipher provably secure against any d-limited distinguisher. In the following sections, we focus on building a round function F_{KFC} following the ideas we have introduced here.

4 A Good Round Function F_{KFC} for the Feistel Scheme

To analyze the behavior of a layer of random functions, we analyze the construction $F = S_3 \circ F_2 \circ F_1$ where $F_1 : \{0,1\}^n \to \{0,1\}^n$ is a random function, S_3 is a random permutation, and F_2 is a layer made of small independent and perfectly random functions (see Fig. 2(a)). We assume that F_1, F_2, and S_3 are mutually independent. We obtain an interesting property, making it possible to relate the order d decorrelation of F to its order $d-1$ decorrelation. We consider a set of d inputs of the function F and denote the corresponding random outputs of F_1 by $X_1, \ldots, X_d$, where $X_k = (X_{k,1}, \ldots, X_{k,N})$ for $k = 1, \ldots, d$. Let α be the event $\{\exists k \text{ s.t. } \forall j\ X_{k,j} \notin \{X_{1,j}, \ldots, X_{k-1,j}, X_{k+1,j}, \ldots, X_{d,j}\}\}$, that is, α is the event that one of the inputs is different from all the others on the N blocks. If α occurs, at least one of the outputs of the functions layer is a uniformly distributed random variable independent from the others. More formally, if we denote $\mathcal{A}_d$

the best d-limited adversary trying to distinguish F from F^*, we have:

$$\begin{aligned}
\mathrm{Adv}_{\mathcal{A}_d}(F, F^*) &= \left|1 - 2 \cdot \Pr[\mathcal{A}_d(F) = 1]\ \right| \\
&= |1 - 2 \cdot (\Pr[\mathcal{A}_d(F) = 1|\alpha] \Pr[\alpha] + \Pr[\mathcal{A}_d(F) = 1|\overline{\alpha}] \Pr[\overline{\alpha}])\ | \\
&\leq \mathrm{Adv}_{\mathcal{A}_{d-1}}(F, F^*) \Pr[\alpha] + |1 - 2 \cdot \Pr[\mathcal{A}_d(F) = 1|\overline{\alpha}]\ | \Pr[\overline{\alpha}] \\
&\leq \mathrm{Adv}_{\mathcal{A}_{d-1}}(F, F^*) + \Pr[\overline{\alpha}], \qquad (1)
\end{aligned}$$

where the first inequality comes from the fact that if α occurs, at least one output of F is completely independent from all the others. As S_3 is a permutation, it preserves this independence. Therefore, when α occurs, a d-limited distinguisher cannot be more efficient than the best $(d-1)$-limited distinguisher (this is formally proven in Appendix A by looking at the definition of the decorrelation norms).

Why this is not Enough. From the previous inequality, it seems natural to consider a substitution-permutation-like construction made of an alternance of layers of independent and perfectly random functions and layers of linear diffusion (as shown on Fig. 2(b)). Intuitively, one could think that (as it is the case when iterating random permutations) iterating random functions is sufficient to decrease the advantage of a distinguisher. However, this is definitely *not* the case. Indeed, consider a 2-limited attack where the two plaintexts are equal on $N-1$ blocks and different on the last block. There is a non-negligible probability $2^{-\ell}$ that, after the first layer of functions, both outputs are completely equal, thus leading to a distinguisher with advantage $2^{-\ell}$. For practical values of ℓ (e.g., $\ell = 8$), this is not acceptable. This means that we need a good resistance against 2-limited adversaries to initialize the recurrence relation of equation (1).

The Sandwich Technique. As proven in [1], an SPN using layers of mutually independent and perfectly random permutations is efficient against 2-limited distinguishers. Intuitively, this means that any set of d inputs will lead to a set of d *pairwise independent* outputs. As we will see in Section 6, pairwise independence is exactly what we need to apply the recursive relation (1).

For these reasons the construction we chose for F_{KFC} consists in sandwiching the construction sketched on Figure 2(b) between two SPN using layers of mutually independent and perfectly random permutations.

Description of F_{KFC}. The round function F_{KFC} used in the Feistel scheme defining KFC is based on three different layers:

- a substitution layer S made of N mutually independent and perfectly random ℓ bit permutations,
- a function layer F made of N mutually independent and perfectly random ℓ bit functions,
- a linear layer L which is a $N \times N$ matrix of elements in $\mathrm{GF}(2^\ell)$ defining an MDS code (for optimal diffusion), which requires $N \leq 2^{\ell-1}$.

Let r_1 and r_2 be two integers. The round function F_{KFC} of KFC is defined as:

$$F_{\mathsf{KFC}} = F_{\mathsf{KFC}[r_1,r_2]} = \mathsf{S} \circ (\mathsf{L} \circ \mathsf{F})^{r_2} \circ (\mathsf{L} \circ \mathsf{S})^{r_1}.$$

5 Computing the Advantage of the Best 2-limited Distinguisher against F_{KFC}

As all layers of F_{KFC} are mutually independent, the order 2 distribution matrix $[F_{\mathsf{KFC}}]^2$ can be expressed as

$$[F_{\mathsf{KFC}}]^2 = [\mathsf{S} \circ (\mathsf{L} \circ \mathsf{F})^{r_2} \circ (\mathsf{L} \circ \mathsf{S})^{r_1}]^2 = ([\mathsf{S}]^2 \times [\mathsf{L}]^2)^{r_1} \times ([\mathsf{F}]^2 \times [\mathsf{L}]^2)^{r_2} \times [\mathsf{S}]^2.$$

Each of these matrices is a $2^{2n} \times 2^{2n}$ square matrix, which makes direct computations impossible for practical parameters. In the rest of this Section we will exploit symmetries in order to reduce the computation to a product of $(N+1) \times (N+1)$ square matrices. For simplicity, we respectively denote by S, F, and L the distribution matrices $[\mathsf{S}]^2$, $[\mathsf{F}]^2$, and $[\mathsf{L}]^2$ and let $q = 2^\ell$.

5.1 Conversion Matrices

Definition 5. *Considering $a \in \{0,1\}^n$ as a N-tuple of elements in $\{0,1\}^\ell$, the* support *of a is the binary N-tuple with 1's at the non-zero positions of a and 0 elsewhere. It is denoted $\textsc{supp}(a)$. The* weight *of the support, denoted $w(\textsc{supp}(a))$ or $w(a)$, is the Hamming weight of the support. When considering a pair $x, x' \in \{0,1\}^n$, the support of the pair is $\textsc{supp}(x \oplus x')$.*

Distribution matrices at order 2 are indexed by pairs of texts. Using symmetries at two levels, we will first shrink them to $2^N \times 2^N$ matrices indexed by supports of pairs and then to $(N+1) \times (N+1)$ matrices indexed by weights. To do so, we define the following conversion matrices.

Pair of texts ↔ Support of pair. We let PS (resp. SP) denote the matrix that converts a pair of texts into a support (resp. a support into a pair of texts) in a uniform way. That is:

$$PS_{(x,x'),\gamma} = \mathbf{1}_{\gamma=\textsc{supp}(x\oplus x')} \quad \text{and} \quad SP_{\gamma',(y,y')} = \mathbf{1}_{\gamma'=\textsc{supp}(y\oplus y')} q^{-N}(q-1)^{-w(\gamma')},$$

where $x, x', y, y' \in \{0,1\}^n$ and $\gamma, \gamma' \in \{0,1\}^N$. One can note that $SP \times PS = Id$.

Support of pair ↔ Weight. Similarly, we let WS (resp. SW) denote the matrix that converts a support into a weight (resp. a weight into a support) in a uniform way. That is:

$$SW_{\gamma,w} = \mathbf{1}_{w(\gamma)=w} \quad \text{and} \quad WS_{w',\gamma'} = \mathbf{1}_{w(\gamma')=w'} \binom{N}{w'}^{-1},$$

where $\gamma, \gamma' \in \{0,1\}^N$ and $w, w' \in \{0, \ldots, N\}$. We have $WS \times SW = Id$.

Pair of texts ↔ Weight. Finally we let $PW = PS \times SW$ and $WP = WS \times SP$ so that we obtain:

$$PW_{(x,x'),w} = \mathbf{1}_{w(x\oplus x')=w} \quad \text{and} \quad WP_{w',(y,y')} = \mathbf{1}_{w(y\oplus y')=w'} \binom{N}{w'}^{-1} q^{-N}(q-1)^{-w'}.$$

Again, we have $WP \times PW = Id$.

5.2 Shrinking F and S, the First Step

Let $x, x', y, y' \in \mathrm{GF}(q)^N$. As the N random functions of the F layer are mutually independent, we can express the coefficients of the distribution matrix F as

$$\mathsf{F}_{(x,x'),(y,y')} = q^{-q \cdot N} \prod_{i=1}^{N} \#\{f_i : \mathrm{GF}(q) \to \mathrm{GF}(q) \ : \ f_i(x_i) = y_i \, , \ f_i(x'_i) = y'_i\}.$$

In the case where $\textsc{supp}(y \oplus y') \not\subseteq \textsc{supp}(x \oplus x')$, we have $\mathsf{F}_{(x,x'),(y,y')} = 0$. When $\textsc{supp}(y \oplus y') \subseteq \textsc{supp}(x \oplus x')$, the uniform distribution of the f_i's leads to:

$$\mathsf{F}_{(x,x'),(y,y')} = q^{-q \cdot N} q^{-w(x \oplus x') + q \cdot N - N} = q^{-w(x \oplus x') - N},$$

and we see that F only depends on support of pairs. Consequently,

$$\begin{aligned}
\mathsf{F}_{(x,x'),(y,y')} &= \mathbf{1}_{\textsc{supp}(y \oplus y') \subseteq \textsc{supp}(x \oplus x')} q^{-w(x \oplus x') - N} \\
&= \sum_{\gamma, \gamma'} \mathbf{1}_{\gamma = \textsc{supp}(x \oplus x')} \mathbf{1}_{\gamma' = \textsc{supp}(y \oplus y')} \mathbf{1}_{\gamma' \subseteq \gamma} q^{-w(\gamma) - N} \\
&= \sum_{\gamma, \gamma'} PS_{(x,x'),\gamma} \mathbf{1}_{\gamma' \subseteq \gamma} q^{-w(\gamma)} (q-1)^{w(\gamma')} SP_{\gamma',(y,y')}.
\end{aligned}$$

Defining the $2^N \times 2^N$ matrix $\overline{\mathsf{F}}$ by $\overline{\mathsf{F}}_{\gamma,\gamma'} = \mathbf{1}_{\gamma' \subseteq \gamma} q^{-w(\gamma)} (q-1)^{w(\gamma')}$ we obtain:

$$\mathsf{F} = PS \times \overline{\mathsf{F}} \times SP. \tag{2}$$

Similarly, for the S layer we have:

$$\mathsf{S}_{(x,x'),(y,y')} = \mathbf{1}_{\textsc{supp}(x \oplus x') = \textsc{supp}(y \oplus y')} q^{-N} (q-1)^{-w(x \oplus x')} = \sum_{\gamma} PS_{(x,x'),\gamma} SP_{\gamma,(y,y')}$$

and thus,

$$\mathsf{S} = PS \times SP. \tag{3}$$

5.3 Shrinking L, the Second Step

Given the structure of F_{KFC}, each linear layer L is surrounded by S or F layers. From equations (2) and (3), this means that each matrix L is surrounded by the conversion matrices PS and SP. Denoting $\overline{\mathsf{L}} = SP \times \mathsf{L} \times PS$ we obtain:

$$\begin{aligned}
\overline{\mathsf{L}}_{\gamma,\gamma'} &= \sum_{(x,x')} \sum_{(y,y')} SP_{\gamma,(x,x')} \mathsf{L}_{(x,x'),(y,y')} PS_{(y,y'),\gamma'} \\
&= q^{-N} (q-1)^{-w(\gamma)} \sum_{(x,x')} \mathbf{1}_{\gamma = \textsc{supp}(x \oplus x')} \mathbf{1}_{\gamma' = \textsc{supp}(\mathsf{L}(x \oplus x'))} \\
&= (q-1)^{-w(\gamma)} \sum_{x} \mathbf{1}_{\gamma = \textsc{supp}(x)} \mathbf{1}_{\gamma' = \textsc{supp}(\mathsf{L}(x))}.
\end{aligned}$$

The sum in this equation is the number of texts of a given support γ that are mapped by the MDS linear layer L on a text of support γ'. The number of codewords with given supports can be explicitly computed for any MDS code (see Theorem 3 in [12]) and, amazingly, only depends on the weights of the supports γ and γ'. We obtain the following formula:

$$\overline{\mathsf{L}}_{\gamma,\gamma'} = (q-1)^{-w(\gamma)} \frac{E(w(\gamma)+w(\gamma'))}{\binom{2N}{w(\gamma)+w(\gamma')}},$$

where $E(i) = \binom{2N}{i} \sum_{j=N+1}^{i} \binom{i}{j} (-1)^{i-j} (q^{j-N}-1)$ for $i > N$, $E(0) = 1$, and $E(i) = 0$ for $0 < i \leq N$. As the previous equation only depends on the weights of γ and γ', we can shrink L even more:

$$\begin{aligned}\overline{\mathsf{L}}_{\gamma,\gamma'} &= \sum_{w,w'} \mathbf{1}_{w(\gamma)=w} \mathbf{1}_{w(\gamma')=w'} (q-1)^{-w} \frac{E(w+w')}{\binom{2N}{w+w'}} \\ &= \sum_{w,w'} SW_{\gamma,w} \binom{N}{w'} (q-1)^{-w} \frac{E(w+w')}{\binom{2N}{w+w'}} WS_{w',\gamma'}.\end{aligned}$$

Defining the $(N+1)\times(N+1)$ matrix $\overline{\overline{\mathsf{L}}}$ by $\overline{\overline{\mathsf{L}}}_{w,w'} = \binom{N}{w'}(q-1)^{-w} \frac{E(w+w')}{\binom{2N}{w+w'}}$,

$$\overline{\mathsf{L}} = SW \times \overline{\overline{\mathsf{L}}} \times WS. \tag{4}$$

A Brief Summary of the Situation. We started from $[F_{\mathsf{KFC}}]^2 = (\mathsf{S}\times\mathsf{L})^{r_1} \times (\mathsf{F}\times\mathsf{L})^{r_2} \times \mathsf{S}$. To makes things clearer, we consider the case where $r_1 = 1$ and $r_2 = 2$. Using equations (2), (3), and (4) we obtain:

$$\begin{aligned}[F_{\mathsf{KFC}}]^2 &= \mathsf{S}\times\mathsf{L}\times\mathsf{F}\times\mathsf{L}\times\mathsf{F}\times\mathsf{L}\times\mathsf{S} \\ &= PS\times SP\times\mathsf{L}\times PS\times\overline{\mathsf{F}}\times SP\times\mathsf{L}\times PS\times\overline{\mathsf{F}}\times SP\times\mathsf{L}\times PS\times SP \\ &= PS\times SW\times\overline{\overline{\mathsf{L}}}\times WS\times\overline{\mathsf{F}}\times SW\times\overline{\overline{\mathsf{L}}}\times WS\times\overline{\mathsf{F}}\times SW\times\overline{\overline{\mathsf{L}}}\times WS\times SP \\ &= PW\times\overline{\overline{\mathsf{L}}}\times WS\times\overline{\mathsf{F}}\times SW\times\overline{\overline{\mathsf{L}}}\times WS\times\overline{\mathsf{F}}\times SW\times\overline{\overline{\mathsf{L}}}\times WP.\end{aligned}$$

Now we focus on the simplification of $WS\times\overline{\mathsf{F}}$.

5.4 Shrinking $WS \times \overline{\mathsf{F}}$, the Third (and Last) Step

We have:

$$\begin{aligned}(WS\times\overline{\mathsf{F}})_{w,\gamma'} &= \sum_{\gamma} WS_{w,\gamma}\overline{\mathsf{F}}_{\gamma,\gamma'} = \binom{N}{w}^{-1} q^{-w}(q-1)^{w(\gamma')} \sum_{\gamma} \mathbf{1}_{w(\gamma)=w}\mathbf{1}_{\gamma'\subseteq\gamma} \\ &= \binom{N}{w}^{-1} q^{-w}(q-1)^{w(\gamma')} \mathbf{1}_{w\geq w(\gamma')} \binom{N-w(\gamma')}{N-w},\end{aligned}$$

so that $(WS\times\overline{\mathsf{F}})_{w,\gamma'}$ only depends on w and on the *weight* of γ'. Consequently, letting $\overline{\overline{\mathsf{F}}}$ be the $(N+1)\times(N+1)$ matrix defined by $\overline{\overline{\mathsf{F}}}_{w,w'} = q^{-w}(q-1)^{w'}\mathbf{1}_{w\geq w'}\binom{w}{w'}$, we obtain:

$$WS\times\overline{\mathsf{F}} = \overline{\overline{\mathsf{F}}}\times WS.$$

Final Summary of the Situation. From the previous summary and the last shrinking step, we finally obtain that:

$$\begin{aligned}[F_{\mathsf{KFC}}]^2 &= PW \times \overline{\overline{\mathsf{L}}} \times \overline{\overline{\mathsf{F}}} \times WS \times SW \times \overline{\overline{\mathsf{L}}} \times \overline{\overline{\mathsf{F}}} \times WS \times SW \times \overline{\overline{\mathsf{L}}} \times WP \\ &= PW \times \overline{\overline{\mathsf{L}}} \times \overline{\overline{\mathsf{F}}} \times \overline{\overline{\mathsf{L}}} \times \overline{\overline{\mathsf{F}}} \times \overline{\overline{\mathsf{L}}} \times WP.\end{aligned}$$

In the general case, this means that $[F_{\mathsf{KFC}}]^2 = PW \times (\overline{\overline{\mathsf{L}}})^{r_1} \times (\overline{\overline{\mathsf{F}}} \times \overline{\overline{\mathsf{L}}})^{r_2} \times WP$.

5.5 Practical Computation of the Advantage

The expression we just obtained for $[F_{\mathsf{KFC}}]^2$ leads to a simple practical expression for $\|[F_{\mathsf{KFC}}]^2 - [F^*]^2\|_a$. Noting that an adversary cannot increase his advantage asking twice the same query, we have:

$$\|[F_{\mathsf{KFC}}]^2 - [F^*]^2\|_a = \max_x \sum_y \max_{x' \neq x} \sum_{y'} \left| [F_{\mathsf{KFC}}]^2_{(x,x'),(y,y')} - q^{-2N} \right|.$$

Let U be the $(N+1) \times (N+1)$ matrix defined by $U_{w,w'} = q^{-N}(q-1)^{w'}\binom{N}{w'}$, so that for all x, x', y, y' we have $(PW \times U \times WP)_{(x,x'),(y,y')} = q^{-2N}$. Consequently, $\|[F_{\mathsf{KFC}}]^2 - [F^*]^2\|_a$ is equal to:

$$\max_x \sum_y \max_{x' \neq x} \sum_{y'} \left| \left(PW \times \left((\overline{\overline{\mathsf{L}}})^{r_1} \times (\overline{\overline{\mathsf{F}}} \times \overline{\overline{\mathsf{L}}})^{r_2} - U \right) \times WP \right)_{(x,x'),(y,y')} \right|.$$

As the inner matrix only depends on $w(x \oplus x')$ and of $w(y \oplus y')$, we get

$$\|[F_{\mathsf{KFC}}]^2 - [F^*]^2\|_a = \max_{w \neq 0} \sum_{w'} \left| \left((\overline{\overline{\mathsf{L}}})^{r_1} \times (\overline{\overline{\mathsf{F}}} \times \overline{\overline{\mathsf{L}}})^{r_2} - U \right)_{w,w'} \right|$$

Similar computations show that $|||[F_{\mathsf{KFC}}]^2 - [F^*]^2|||_\infty = \|[F_{\mathsf{KFC}}]^2 - [F^*]^2\|_a$.

Theorem 6. *Let $\overline{\overline{\mathsf{L}}}$, $\overline{\overline{\mathsf{F}}}$, and U be $(N+1) \times (N+1)$ matrices defined as above. The advantage of the best 2-limited distinguisher $\mathcal{A}$ (whether adaptive or not) against $F_{\mathsf{KFC}} = \mathsf{S} \circ (\mathsf{L} \circ \mathsf{F})^{r_2} \circ (\mathsf{L} \circ \mathsf{S})^{r_1}$ is given by:*

$$\mathrm{Adv}_{\mathcal{A}}(F_{\mathsf{KFC}}, F^*) = \frac{1}{2} \max_{w \neq 0} \sum_{w'} \left| \left((\overline{\overline{\mathsf{L}}})^{r_1} \times (\overline{\overline{\mathsf{F}}} \times \overline{\overline{\mathsf{L}}})^{r_2} - U \right)_{w,w'} \right|.$$

Explicit values of this advantage for some typical values of N, q, r_1 and r_2 are given in Table 1. We note that $r_1 = 3$ is enough (at least for these parameters). Moreover, the advantage increases with the value of r_2. The reason is that the more F layers there is, the higher is the probability of an internal collision.

Table 1. Advantage of the best 2-limited distinguisher against F_{KFC}.

	$N=8$ and $q=2^8$				$N=8$ and $q=2^{16}$				$N=16$ and $q=2^8$			
r_1 \ r_2	0	1	10	100	0	1	10	100	0	1	10	100
0	1	2^{-5}	2^{-8}	2^{-8}	1	2^{-13}	2^{-16}	2^{-16}	1	2^{-4}	2^{-8}	2^{-8}
1	2^{-5}	2^{-50}	2^{-52}	2^{-49}	2^{-13}	2^{-114}	2^{-116}	2^{-113}	2^{-4}	2^{-95}	2^{-104}	2^{-103}
2	2^{-46}	2^{-53}	2^{-52}	2^{-49}	2^{-110}	2^{-117}	2^{-116}	2^{-113}	2^{-87}	2^{-104}	2^{-104}	2^{-103}
3	2^{-62}	2^{-53}	2^{-52}	2^{-49}	2^{-128}	2^{-117}	2^{-116}	2^{-113}	2^{-120}	2^{-104}	2^{-104}	2^{-103}

6 Bounding the Advantage of the Best d-limited Distinguisher against F_{KFC} for $d > 2$

6.1 Replacing F by $\mathsf{F} \circ \mathsf{S}$

To simplify the proofs, we will replace each F layer of F_{KFC} by $\mathsf{F} \circ \mathsf{S}$. Both constructions are completely equivalent in the sense that any decorrelation result holding for the latter also holds for the original construction, the reason being that $[\mathsf{F} \circ \mathsf{S}]^d = [\mathsf{F}]^d$ (see Appendix B for a proof). From now on, we thus study the following equivalent construction:

$$F_{\mathsf{KFC}} = F_{\mathsf{KFC}[r_1,r_2]} = \mathsf{S} \circ (\mathsf{L} \circ \mathsf{F} \circ \mathsf{S})^{r_2} \circ (\mathsf{L} \circ \mathsf{S})^{r_1}.$$

Assumption 7. *For $r_1 > 2$, any $i \in \{0, \dots, r_2\}$ and any 2-limited distinguisher $\mathcal{A}_2$, we have* $\mathrm{Adv}_{\mathcal{A}_2}(F_{\mathsf{KFC}[r_1,r_2]}, F^*) \geq \mathrm{Adv}_{\mathcal{A}_2}(F_{\mathsf{KFC}[r_1,i]}, F^*)$.

This assumption seems natural from Table 1, although it might prove wrong in the general case (in particular, the threshold for r_1 might be different for other values of N and q). However, we experimentally verified it for all values of the parameters we consider in the rest of this paper.

In practice, Assumption 7 means that, when the advantage of the best 2-limited distinguisher against F_{KFC} is negligible, this is also the case before any F layer. The inputs of any F layer can thus be considered as *pairwise independent*.

6.2 Taking Advantage of the Pairwise Independence

Let $i \in \{0, \dots, r_2\}$. Referring to Section 4, we denote α_{i-1} the event α and let $F_1 = F_{\mathsf{KFC}[r_1,i-1]}$, $F_2 = \mathsf{F}$, and $S_3 = \mathsf{S} \circ \mathsf{L}$. We these notations, $F_{\mathsf{KFC}[r_1,i]} = S_3 \circ F_2 \circ F_1$, so that equation (1) gives

$$\mathrm{Adv}_{\mathcal{A}_d}(F_{\mathsf{KFC}[r_1,i]}, F^*) \leq \mathrm{Adv}_{\mathcal{A}_{d-1}}(F_{\mathsf{KFC}[r_1,i]}, F^*) + \Pr[\overline{\alpha}_{i-1}].$$

Bounding $\Pr[\overline{\alpha}_{i-1}]$ for all i allows to recursively bound $\mathrm{Adv}_{\mathcal{A}_d}(F_{\mathsf{KFC}[r_1,i]}, F^*)$. As in Section 4, we denote the output of F_1 by $X_1, \dots, X_d$ where, for $k = 1, \dots, d$, we have $X_k = (X_{k,1}, \dots, X_{k,N})$. Let $0 \leq \lambda \leq d$ be the number of X_k's different from all other texts on all N blocks. We have:

$$\lambda = \sum_{k=1}^{d} \prod_{b=1}^{N} \prod_{\substack{j=1 \\ j \neq k}}^{d} \mathbf{1}_{X_{k,b} \neq X_{j,b}}.$$

Using the linearity of the mean and the mutual independence of the N blocks, we obtain $E(\lambda) = d \cdot \left(\Pr[X_{1,1} \notin \{X_{2,1}, \ldots, X_{d,1}\}] \right)^N$.

Property 8. *For $d > 0$ we have $\mathcal{P}_d = \Pr[X_{1,1} \notin \{X_{2,1}, \ldots, X_{d,1}\}] \geq 1 - \frac{d-1}{q}$ and thus, $E(\lambda) \geq d \cdot \left(1 - \frac{d-1}{q}\right)^N$.*

Proof. The proof is done by induction on d. For $d = 1$ the result is trivial. Assume $\mathcal{P}_d \geq 1 - (d-1)/q$ for an arbitrary d. As stated in Section 6.1, we can assume that the X_i's are pairwise independent and thus:

$$\begin{aligned} \mathcal{P}_{d+1} &= \mathcal{P}_d - \Pr[X_{1,1} \notin \{X_{2,1}, \ldots, X_{d,1}\}\,,\, X_{1,1} = X_{d+1,1}] \\ &\geq \mathcal{P}_d - \Pr[X_{1,1} = X_{d+1,1}] = \mathcal{P}_d - \tfrac{1}{q}. \end{aligned}$$

The expression we obtained for $E(\lambda)$ leads to the final result. □

Using this result, we can easily bound $\Pr[\overline{\alpha}_i]$ as $E(\lambda) = \sum_{k=1}^{d} k \Pr[\lambda = k] \leq d \Pr[\lambda \neq 0] = d \Pr[\alpha_i]$, so that, for all $i \in \{0, \ldots, r_2\}$,

$$\Pr[\overline{\alpha}_i] \leq 1 - \tfrac{E(\lambda)}{d} \leq 1 - \left(1 - \tfrac{d-1}{q}\right)^N. \tag{5}$$

6.3 Piling-up the Rounds

Obviously, the bound on $\Pr[\overline{\alpha}_i]$ we just obtained cannot be used directly to obtain a meaningful bound on the advantage of high order distinguishers. Consequently, we will consider t successive α_i events and give an upper bound on the probability that *none* of them occurs. We have $\Pr[\overline{\alpha}_1, \ldots, \overline{\alpha}_t] = \Pr[\overline{\alpha}_t | \overline{\alpha}_1, \ldots, \overline{\alpha}_{t-1}] \cdot \Pr[\overline{\alpha}_1, \ldots, \overline{\alpha}_{t-1}]$. As the bound on $E(\lambda)$ only relies on the pairwise independence of the inputs of the i-th round, the bound given by equation (5) can also be proven for $\Pr[\overline{\alpha}_t | \overline{\alpha}_1, \ldots, \overline{\alpha}_{t-1}]$. By induction, we finally obtain that:

$$\Pr[\overline{\alpha}_1, \ldots, \overline{\alpha}_t] \leq \left(1 - \left(1 - \tfrac{d-1}{q}\right)^N\right)^t.$$

Theorem 9. *Assume that the advantage of the best 2-limited distinguisher on $F_{\mathsf{KFC}[r_1,r_2]}$ is bounded by ϵ. For any d and set of integers $\{t_3, \ldots, t_d\}$ such that $\sum_{i=3}^{d} t_i \leq r_2$, the advantage of the best d-limited distinguisher $\mathcal{A}_d$ on $\mathsf{F}_{\mathsf{KFC}[r_1,r_2]}$ is such that:*

$$\mathrm{Adv}_{\mathcal{A}_d}(\mathsf{F}_{\mathsf{KFC}[r_1,r_2]}, F^*) \leq \epsilon + \sum_{i=3}^{d} \left(1 - \left(1 - \tfrac{i-1}{q}\right)^N\right)^{t_i}.$$

Fixing $r_1 = 3$, the previous theorem bounds, for any value of d, the advantage of the best d-limited distinguisher against a given number of rounds r_2 of F_{KFC}. In Table 2 we give the best bounds we obtain for various values of r_2, d, N, and q. If one aims at a specific value of d and wants to select r_2 in order to bound the advantage of the best d-limited distinguisher, the best choice is probably to select the t_i's such that $\Pr[\overline{\alpha}_1, \ldots, \overline{\alpha}_{t_i}] < \epsilon$, which bounds the advantage by $d \cdot \epsilon$. The following theorem generalizes this idea.

Table 2. Bounds on $\mathrm{Adv}_{\mathcal{A}_d}$ for $r_1 = 3$ and various parameters.

	$N=8$ and $q=2^8$							$N=8$ and $q=2^{16}$						
r_2 \ d	2	3	4	8	16	32	64	2	3	4	8	16	32	64
10	2^{-52}	2^{-40}	2^{-17}	2^{-2}	1	1	1	2^{-116}	2^{-116}	2^{-57}	2^{-11}	1	1	1
100	2^{-49}	2^{-49}	2^{-49}	2^{-46}	2^{-11}	1	1	2^{-113}	2^{-113}	2^{-113}	2^{-113}	2^{-66}	2^{-23}	2^{-5}
250	2^{-48}	2^{-48}	2^{-48}	2^{-48}	2^{-33}	2^{-5}	1	2^{-112}	2^{-112}	2^{-112}	2^{-112}	2^{-112}	2^{-69}	2^{-25}
1000	2^{-46}	2^{-46}	2^{-46}	2^{-46}	2^{-46}	2^{-35}	2^{-2}	2^{-110}	2^{-110}	2^{-110}	2^{-110}	2^{-110}	2^{-110}	2^{-110}

	$N=16$ and $q=2^8$						
r_2 \ d	2	3	4	8	16	32	64
10	2^{-104}	2^{-31}	2^{-12}	1	1	1	1
100	2^{-103}	2^{-103}	2^{-103}	2^{-31}	2^{-5}	1	1
250	2^{-103}	2^{-103}	2^{-103}	2^{-81}	2^{-18}	1	1
1000	2^{-102}	2^{-102}	2^{-102}	2^{-102}	2^{-82}	2^{-12}	1

Theorem 10. *Assume that the advantage of the best 2-limited distinguisher against $F_{\mathsf{KFC}[r_1,r_2]}$ is bounded by ϵ. Let:*

$$t_d(\beta) = \min_t \{\Pr[\overline{\alpha}_1, \ldots, \overline{\alpha}_t] < \beta \cdot \epsilon\} = \left\lceil \frac{\log(\beta \cdot \epsilon)}{\log\left(1 - \left(1 - \frac{d-1}{q}\right)^N\right)} \right\rceil.$$

For any d such that $\sum_{i=3}^{d} t_i(\beta) \leq r_2$, the advantage of the best d-limited distinguisher $\mathcal{A}_d$ against $\mathsf{F}_{\mathsf{KFC}[r_1,r_2]}$ is such that:

$$\mathrm{Adv}_{\mathcal{A}_d}(\mathsf{F}_{\mathsf{KFC}[r_1,r_2]}, F^*) \leq \epsilon + \sum_{i=3}^{d} \left(1 - \left(1 - \frac{i-1}{q}\right)^N\right)^{t_i(\beta)} \leq \epsilon \cdot (1 + (d-2) \cdot \beta).$$

7 Conclusion

We introduced KFC, a block cipher based on a three round Feistel scheme. Each of the three round functions has an SPN-like structure for which we can either compute or bound the advantage of the best d-limited adaptive adversary, for any value of d. Using results from the Decorrelation Theory, we extend these results to the whole KFC construction. At this time, no key schedule has been specified for KFC. We suggest to use the same trick as in [1], i.e., use a key schedule based on a cryptographically secure pseudo-random generator (for example the good old BBS [10] or a faster generator like QUAD [3, 4]). This way, all the results we have proven assuming the mutual independence of the random functions and permutations remain valid when implementing KFC in practice with a 128 bit secret key. We propose two sets of parameters:

Regular KFC: $N = 8$, $q = 2^8$, $r_1 = 3$, $r_2 = 100$. These parameters lead to provable security against 8-limited adaptive distinguishers. Consequently, Regular KFC is resistant to iterated attacks of order 4, which include linear and differential cryptanalysis, the boomerang attack and others. Based on existing implementation results on C, we estimate the encryption speed of Regular KFC to 15-25 Mbits/s on a Pentium IV 2GHz. The key schedule needs to generate approximatively 2^{22} cryptographically secure pseudo-random bits.

Extra Crispy KFC: $N = 8$, $q = 2^{16}$, $r_1 = 3$, $r_2 = 1000$. Using these quite *extreme* parameters, we manage to obtain provable security against 70-limited adaptive adversaries, but encryption rate could probably never reach more than 1 Mbit/s. Also, the key schedule should produce 2^{35} pseudo random bits, which means that Extra Crispy KFC requires at least 4 GB of memory.

To the best of our knowledge, KFC is the first practical block cipher to propose tight security proofs of resistance against large classes of attacks, including most classical cryptanalysis (such as linear and differential cryptanalysis, taking hull effect in consideration in both cases, higher order differential cryptanalysis, the boomerang attack, differential-linear cryptanalysis, or the rectangle attack).

References

1. T. Baignères and M. Finiasz. Dial C for Cipher. In Biham and Youssef [9]. To appear.
2. T. Baignères and S. Vaudenay. Proving the security of AES substitution-permutation network. In B. Preneel and S.E. Tavares, editors, *Selected Areas in Cryptography,* SAC *05*, volume 3897 of *LNCS*, pages 65–81. Springer-Verlag, 2006.
3. C. Berbain, O. Billet, and H. Gilbert. Efficient implementations of multivariate quadratic systems. In Biham and Youssef [9]. To appear.
4. C. Berbain, H. Gilbert, and J. Patarin. QUAD: a practical stream cipher with provable security. In S. Vaudenay, editor, *Advances in Cryptology* - Eurocrypt *'06*, volume 4004 of *LNCS*, pages 109–128. Springer-Verlag, 2006.
5. E. Biham, A. Biryukov, and A. Shamir. Cryptanalysis of Skipjack reduced to 31 rounds using impossible differentials. In Stern [22], pages 12–23.
6. E. Biham, O. Dunkelman, and N. Keller. The rectangle attack - rectangling the Serpent. In B. Pfitzmann, editor, *Advances in Cryptology* - Eurocrypt *'01*, volume 2045 of *LNCS*, pages 340–357. Springer-Verlag, 2001.
7. E. Biham and A. Shamir. Differential cryptanalysis of DES-like cryptosystems. *Journal of Cryptology*, 4:3–72, 1991.
8. E. Biham and A. Shamir. Differential cryptanalysis of DES-like cryptosystems (extended abstract). In A. Menezes and S. Vanstone, editors, *Advances in Cryptology* - Crypto *'90*, volume 537 of *LNCS*, pages 2–21. Springer-Verlag, 1991.
9. E. Biham and A.M. Youssef, editors. *Proceedings of Selected Areas in Cryptography,* SAC *06*, LNCS. Springer-Verlag, 2006. To appear.
10. L. Blum, M. Blum, and M. Shub. Comparison of two pseudo-random number generators. In D. Chaum, R.L. Rivest, and A. Sherman, editors, *Advances in Cryptology* - Crypto *'82*, pages 61–78. Plemum, 1983.
11. J. Daemen and V. Rijmen. *The Design of Rijndael.* Information Security and Cryptography. Springer-Verlag, 2002.
12. M. El-Khamy and R. McEliece. The partition weight enumerator of MDS codes and its applications. In *IEEE International Symposium on Information Theory, ISIT 2005.* IEEE, 2005. Available on `http://arxiv.org/pdf/cs.IT/0505054`.

13. H. Feistel. Cryptography and computer privacy. *Scientific American*, 228:15–23, 1973.
14. L. Granboulan, P. Nguyen, F. Noilhan, and S. Vaudenay. DFCv2. In D.R. Stinson and S.E. Tavares, editors, *Selected Areas in Cryptography,* SAC *'00*, volume 2012 of *LNCS*, pages 57–71. Springer-Verlag, 2001.
15. L. Knudsen. Truncated and higher order differentials. In B. Preneel, editor, *Fast Software Encryption* - FSE *'95*, volume 1008 of *LNCS*, pages 196–211. Springer-Verlag, 1995.
16. X. Lai. Higher order derivatives and differential cryptanalysis. In Kluwer Academic Publishers, editor, *Symposium on Communication, Coding and Cryptography*, pages 227–233, 1994.
17. X. Lai, J. Massey, and S. Murphy. Markov ciphers and differential cryptanalysis. In D.W. Davies, editor, *Advances in Cryptology* - EUROCRYPT *'91*, volume 547 of *LNCS*, pages 17–38. Springer-Verlag, 1991.
18. M. Luby and C. Rackoff. How to construct pseudorandom permutations from pseudorandom functions. *SIAM Journal on Computing*, 17(2):373–386, 1988.
19. M. Matsui. The first experimental cryptanalysis of the Data Encryption Standard. In Y.G. Desmedt, editor, *Advances in Cryptology* - CRYPTO *'94*, volume 839 of *LNCS*, pages 1–11. Springer-Verlag, 1994.
20. M. Matsui. Linear cryptanalysis method for DES cipher. In T. Helleseth, editor, *Advances in Cryptology* - EUROCRYPT *'93*, volume 765 of *LNCS*, pages 386–397. Springer-Verlag, 1994.
21. K. Nyberg. Linear approximation of block ciphers. In A. De Santis, editor, *Advances in Cryptology* - EUROCRYPT *'94*, volume 950 of *LNCS*, pages 439–444. Springer-Verlag, 1995.
22. J. Stern, editor. *Proceedings of Advances in Cryptology* - EUROCRYPT *'99*, volume 1592 of *LNCS*. Springer-Verlag, 1999.
23. A. Tardy-Corfdir and H. Gilbert. A known plaintext attack of FEAL-4 and FEAL-6. In J. Feigenbaum, editor, *Advances in Cryptology* - CRYPTO *'91*, volume 576 of *LNCS*, pages 172–182. Springer-Verlag, 1992.
24. S. Vaudenay. Provable security for block ciphers by decorrelation. In *STACS'98*, volume 1373 of *LNCS*, pages 249–275. Springer-Verlag, 1998.
25. S. Vaudenay. Resistance against general iterated attacks. In Stern [22], pages 255–271.
26. S. Vaudenay. Decorrelation: a theory for block cipher security. *Journal of Cryptology*, 16(4):249–286, 2003.
27. D. Wagner. The boomerang attack. In L. Knudsen, editor, *Fast Software Encryption* - FSE *'99*, volume 1636 of *LNCS*, pages 156–170. Springer-Verlag, 1999.

A Proof of $|1 - 2 \cdot \Pr[\mathcal{A}_d(F) = 1 \mid \alpha]| = \mathrm{Adv}_{\mathcal{A}_{d-1}}(F, F^*)$

Without loss of generality, we can assume that the adversary does not make the same query twice (as this would not increase its advantage) and that the event α is true for the dth query. In this case, we know that $(F_2 \circ F_1)(x_d)$ is a uniformly distributed random variable independent of $(F_2 \circ F_1)(x_i)$ for all $i < d$. As S_3 is a permutation, this property is still true for $(S_3 \circ F_2 \circ F_1)(x_d) = F(x_d)$. Denoting by Y this random variable we have:

$$\begin{aligned}\Pr[F(x_1) = y_1, \ldots, F(x_d) = y_d|\alpha] &= \Pr[F(x_1) = y_1 \ldots F(x_{d-1}) = y_{d-1}, Y = y_d] \\ &= 2^{-n} \Pr[F(x_1) = y_1 \ldots F(x_{d-1}) = y_{d-1}].\end{aligned}$$

Let $A = |1 - 2 \cdot \Pr[\mathcal{A}_d(F) = 1 \mid \alpha]|$. Similarly to the proof of Theorem 10 in [26] we know that:

$$A = \frac{1}{2} \max_{x_1} \sum_{y_1} \cdots \max_{x_d} \sum_{y_d} \left| \Pr[F(x_1) = y_1, \ldots, F(x_d) = y_d | \alpha] - 2^{-d \cdot n} \right| .$$

From the two previous equations we obtain that:

$$\begin{aligned}
A &= \frac{1}{2} \max_{x_1} \sum_{y_1} \cdots \max_{x_d} \sum_{y_d} 2^{-n} \left| \Pr[F(x_1) = y_1 \ldots F(x_{d-1}) = y_{d-1}] - 2^{-(d-1) \cdot n} \right| \\
&= \frac{1}{2} \max_{x_1} \sum_{y_1} \cdots \max_{x_{d-1}} \sum_{y_{d-1}} \left| \Pr[F(x_1) = y_1 \ldots F(x_{d-1}) = y_{d-1}] - 2^{-(d-1) \cdot n} \right| \\
&= \mathrm{Adv}_{\mathcal{A}_{d-1}}(F, F^*).
\end{aligned}$$

B Proof that $[\mathsf{F} \circ \mathsf{S}]^d = [\mathsf{F}]^d$

For any $x = (x_1, \ldots, x_d), y = (y_1, \ldots, y_d) \in \{0,1\}^{nd}$ we have:

$$\begin{aligned}
[\mathsf{F} \circ \mathsf{S}]^d_{(x,y)} &= \Pr[(x_1, \ldots, x_d) \xrightarrow{\mathsf{F} \circ \mathsf{S}} (y_1, \ldots, y_d)] \\
&= \prod_{i=1}^{d} \Pr[(x_{1,i}, \ldots, x_{d,i}) \xrightarrow{F^* \circ C^*} (y_{1,i}, \ldots, y_{d,i})] \\
&= \prod_{i=1}^{d} \frac{1}{2^\ell !} \sum_{c} \Pr[(c(x_{1,i}), \ldots, c(x_{d,i})) \xrightarrow{F^*} (y_{1,i}, \ldots, y_{d,i})] \\
&= \prod_{i=1}^{d} \frac{1}{2^\ell !} \sum_{c} \Pr[(x_{1,i}, \ldots, x_{d,i}) \xrightarrow{F^*} (c^{-1}(y_{1,i}), \ldots, c^{-1}(y_{d,i}))] \\
&= \prod_{i=1}^{d} \Pr[(x_{1,i}, \ldots, x_{d,i}) \xrightarrow{F^*} (y_{1,i}, \ldots, y_{d,i})] \\
&= \Pr[(x_1, \ldots, x_d) \xrightarrow{\mathsf{F}} (y_1, \ldots, y_d)] \\
&= [\mathsf{F}]^d_{(x,y)}
\end{aligned}$$