Feistel Ciphers with L_2-Decorrelation

Serge Vaudenay

Ecole Normale Supérieure – CNRS
e-mail: Serge.Vaudenay@ens.fr

Abstract. Recently, we showed how to strengthen block ciphers by decorrelation techniques. In particular, we proposed two practical block ciphers, one based on the $\mathrm{GF}(2^n)$-arithmetics, the other based on the $x \bmod p \bmod 2^n$ primitive with a prime $p = 2^n(1+\delta)$. In this paper we show how to achieve similar decorrelation with a prime $p = 2^n(1-\delta)$. For this we have to change the choice of the norm in the decorrelation theory and replace the L_∞ norm by the L_2 norm. We propose a new practical block cipher which is provably resistant against differential and linear cryptanalysis.

At the STACS'98 conference, the author of the present paper presented the technique of decorrelation which enables to strengthen block ciphers in order to make them provably resistant against the basic differential and linear cryptanalysis [13].[1] So far, this analysis which is based on Carter and Wegman's paradigm of universal functions [3, 17], has been used with the L_∞-associated matrix norm in order to propose two new practical block cipher families which are provably resistant against those cryptanalysis: COCONUT98 and PEANUT98. This technique has been shown to enable to propose real-life encryption algorithms as shown by the Advanced Encryption Standard submission [5] and related implementation evaluations on smart cards [9]. In this paper we present some earlier results based on the L_2 norm in order to make a new practical block cipher PEANUT97.[2]

1 Basic Definitions

We briefly recall the basic definitions used in the decorrelation theory. Firstly, let us recall the notion of d-wise distribution matrix associated to a random function.

[1] A full paper version [14] is available on the web site [15].

[2] The decorrelation technique with the L_2 norm happens to be somewhat less easy than the L_∞ norm, which is why is has not been published so far.

Definition 1 ([13]). *Given a random function F from a given set $\mathcal{A}$ to a given set $\mathcal{B}$ and an integer d, we define the "d-wise distribution matrix" $[F]^d$ of F as a $\mathcal{A}^d \times \mathcal{B}^d$-matrix where the (x, y)-entry of $[F]^d$ corresponding to the multi-points $x = (x_1, \ldots, x_d) \in \mathcal{A}^d$ and $y = (y_1, \ldots, y_d) \in \mathcal{B}^d$ is defined as the probability that we have $F(x_i) = y_i$ for $i = 1, \ldots, d$.*

Secondly, we recall the definition of two matrix norms: the L_∞-associated norm denoted $|||.|||_\infty$, and the L_2-norm denoted $||.||_2$.

Definition 2. *Given a matrix A, we define*

$$||A||_2 = \sqrt{\sum_{x,y} (A_{x,y})^2} \tag{1}$$

$$|||A|||_\infty = \max_x \sum_y |A_{x,y}| \tag{2}$$

where the sums run over all the (x, y)-entries of the matrix A.[3]

Finally, here is the definition of the general d-wise decorrelation distance between two random functions.

Definition 3 ([13]). *Given two random functions F and G from a given set $\mathcal{A}$ to a given set $\mathcal{B}$, an integer d and a matrix norm $||.||$ over the vector space $\mathbf{R}^{\mathcal{A}^d \times \mathcal{B}^d}$, we call $||[F]^d - [G]^d||$ the "d-wise decorrelation $||.||$-distance" between F and G. In addition, we call "d-wise decorrelation $||.||$-bias" of a random* function *(resp.* permutation*) F its d-wise decorrelation $||.||$-distance to a random function (resp. permutation) with a uniform distribution.*[4]

We consider block ciphers on a message-block space $\mathcal{M}$ with a key represented by a random variable K as a random permutation C_K defined by K over $\mathcal{M}$. Since the subscript K is useless in our context we omit it and consider the random variable C as a random permutation with a given distribution. Ideally, we consider the Perfect Cipher

[3] The strange $|||.|||_\infty$ notation used in [13] comes from the fact that this norm is associated to the usual $||.||_\infty$ norm over the vectors defined by $||V||_\infty = \max_x |V_x|$ by

$$|||A|||_\infty = \max_{||V||_\infty = 1} ||AV||_\infty.$$

[4] It is thus important to outline that we are considering a function or a permutation.

C^* for which the distribution of C^* is uniform over the set of the permutations over $\mathcal{M}$. Hence for any multi-point $x = (x_1, \ldots, x_d)$ with pairwise x_is and any multi-point $y = (y_1, \ldots, y_d)$ with pairwise y_is we have

$$[C^*]^d_{x,y} = \Pr[C^*(x_i) = y_i; i = 1, \ldots, d] = \frac{1}{\#\mathcal{M} \ldots (\#\mathcal{M} - d + 1)}.$$

We are interested in the decorrelation bias $||[C]^d - [C^*]^d||$ of a practical cipher C.

We recall that $|||.|||_\infty$ and $||.||_2$ are matrix norms (*i.e.* that the norm of any matrix-product $A \times B$ is at most the product of the norms of A and B) which makes the decorrelation bias a friendly measurement as shown by the following Lemma.

Lemma 4. *Let $||.||$ be a norm such that $||A \times B|| \leq ||A||.||B||$ for any matrix A and B. For any independent random ciphers denoted C_1, C_2, C_3, C_4, C^* (where C^* is perfect), the following properties hold.*

$$||[C_1 \circ C_2]^d - [C^*]^d|| \leq ||[C_1]^d - [C^*]^d||.||[C_2]^d - [C^*]^d|| \quad (3)$$
$$||[C_1 \circ C_2]^d - [C_1 \circ C_3]^d|| \leq ||[C_1]^d - [C^*]^d||.||[C_2]^d - [C_3]^d|| \quad (4)$$
$$\begin{aligned}||[C_1 \circ C_2]^d - [C_3 \circ C_4]^d|| \leq\ & ||[C_1]^d - [C^*]^d||.||[C_2]^d - [C_4]^d|| \\ & + ||[C_1]^d - [C_3]^d||.||[C_4]^d - [C^*]^d|| \quad (5)\end{aligned}$$

Those properties come from the easy facts $[C_1 \circ C_2]^d = [C_2]^d \times [C_1]^d$ and $[C^*]^d \times [C_1]^d = [C^*]^d$.

Feistel Ciphers are defined over $\mathcal{M} = \mathcal{M}_0^2$ for a given group $\mathcal{M}_0$ (*e.g.* $\mathcal{M}_0 = \mathbf{Z}_2{}^{\frac{m}{2}}$) by round functions $F_1, \ldots, F_r$ on $\mathcal{M}_0$. We let $C = \Psi(F_1, \ldots, F_r)$ denote the cipher defined by $C(x^l, x^r) = (y^l, y^r)$ where we iteratively compute a sequence (x^l_i, x^r_i) such that

$$\begin{aligned} x^l_0 = x^l &\text{ and } x^r_0 = x^r \\ x^l_i = x^r_{i-1} &\text{ and } x^r_i = x^l_{i-1} + F_i(x^r_{i-1}) \\ y^l = x^r_r &\text{ and } y^r = x^l_r \end{aligned}$$

(see Feistel [4]).

To illustrate the problem, we stress out that perfect decorrelation (*i.e.* decorrelation bias of zero) is achievable on a finite field (no matter which norm we take). For instance, a random $(d-1)$-degreeed polynomial with a uniform distribution is a perfectly d-wise decorrelated function. A random affine permutation with a uniform distribution is a perfectly pairwise decorrelated permutation. (Perfect decorrelation of higher degree is much more complicated.) Finite field arithmetic is however cumbersome in software for the traditional characteristic two. This is why we studied decorrelation biases.

2 Previous Security Results

Decorrelation enables to quantify the security of imperfectly decorrelated ciphers. Here we consider the security in the Luby-Rackoff model [6]. We consider opponents as Turing machines which have a limited access to an encryption oracle device and whose aim is to distinguish whether the device implements a given practical cipher $C_1 = C$ or a given cipher C_2 which is usually $C_2 = C^*$. When fed with an oracle c, the Turing machine $\mathcal{T}^c$ returns either 0 or 1. If we want to distinguish a random cipher C from C^*, we let p (resp. p^*) denote $\Pr[\mathcal{T}^C = 1]$ (resp. $\Pr[\mathcal{T}^{C^*} = 1]$) where the probability is over the distribution of the random tape of $\mathcal{T}$ and the distribution of the cipher. We say the attack is successful if $|p - p^*|$ is large. On the other hand, we say that the cipher C resists against the attack if we have $|p - p^*| \leq \epsilon$ for some small ϵ. This model is quite powerful, because if we prove that a cipher C cannot be distinguished from the Perfect Cipher C^*, then any attempt to decrypt a ciphertext provided by C will also be applicable to the cipher C^* for which we know the security. (For more motivation on this security model, see Luby-Rackoff [6].)

Inspired by Biham and Shamir's attack [2] we call *differential distinguisher* with the (fixed) characteristic (a, b) and complexity n the following algorithm:

Input: a cipher c, a complexity n, a characteristic (a, b)

1. for i from 1 to n do
   (a) pick uniformly a random X and query for $c(X)$ and $c(X \oplus a)$
   (b) if $c(X \oplus a) = c(X) \oplus b$, stop and output 1
2. output 0

Similarly, inspired by Matsui's attack [7] we call *linear distinguisher* with the characteristic (a,b) and complexity n the following algorithm:[5]

Input: a cipher c, a complexity n, a characteristic (a,b), a set A

1. initialize the counter value t to zero
2. for i from 1 to n do
   (a) pick a random X with a uniform distribution and query for $c(X)$
   (b) if $X \cdot a = c(X) \cdot b$, increment the counter t
3. if $t \in A$, output 1, otherwise output 0

Both linear and differentiel distinguishers are particular cases of iterative distinguisher attacks (see [14]).

Theorem 5 ([14]). *Let C be a cipher on the space $\mathcal{M} = \mathbf{Z}_2^m$, let C^* be the Perfect Cipher, and let $\epsilon = |||[C]^2 - [C^*]^2|||_\infty$. For any differential distinguisher between C and the Perfect Cipher C^* with complexity n, the advantage $|p - p^*|$ is at most $\frac{n}{2^m-1} + n\epsilon$. Similarly, for any linear distinguisher, the advantage is such that*

$$\lim_{n\to+\infty} \frac{|p-p^*|}{n^{\frac{1}{3}}} \leq 9.3\left(\frac{1}{2^m-1} + 2\epsilon\right)^{\frac{1}{3}}.$$

This theorem means that C is immune against any differential or linear distinguisher if $|||[C]^2 - [C^*]^2|||_\infty \approx 2^{-m}$. In this paper, we show we can obtain similar results with the L_2-decorrelation and that we can use them for an efficient real-life cipher.

3 Security by L_2-Decorrelation

It is well known that differential and linear cryptanalysis with characteristic (a,b) respectively depend on the following measurements. If C is a random cipher on $\mathbf{Z}_2^m$ where $\oplus$ denotes the group operation (the bitwise XOR) and $\cdot$ denotes the dot product (the parity of the bitwise *and*), we denote

$$\mathrm{EDP}^C(a,b) = \mathop{E}_{C}\left(\Pr_X[C(X\oplus a) = C(X)\oplus b]\right) = 2^{-m} \sum_{\substack{x_1\oplus x_2 = a\\ y_1\oplus y_2 = b}} [C]^2_{x,y}$$

$$\mathrm{ELP}^C(a,b) = \mathop{E}_{C}\left(\left(2\Pr_X[X\cdot a = C(X)\cdot b] - 1\right)^2\right) = 1 - 2^{2-2m} \sum_{\substack{x_1\cdot a = y_1\cdot b\\ x_2\cdot a \neq y_2\cdot b\\ x_1\neq x_2, y_1\neq y_2}} [C]^2_{x,y}$$

[5] For differential and linear cryptanalysis, we assume that the message space $\mathcal{M}$ is $\mathbf{Z}_2^m$ so that the addition $+$ is the bitwise exclusive *or* and the dot product $\cdot$ is the parity of the bitwise *and*.

where $x = (x_1, x_2)$ and $y = (y_1, y_2)$, and X is uniformly distributed.[6] In [14], Theorem 5 comes from the upper bounds

$$|\mathrm{EDP}^C(a,b) - \mathrm{EDP}^{C^*}(a,b)| \le |||[C]^2 - [C^*]^2|||_\infty$$
$$|\mathrm{ELP}^C(a,b) - \mathrm{ELP}^{C^*}(a,b)| \le 2|||[C]^2 - [C^*]^2|||_\infty.$$

The same inequalities hold with the L_2 norm. (These are the consequence of Cauchy-Schwarz Inequality.) We can thus adapt Theorem 5 with the L_2 bounds without any more argument.

Theorem 6. *Theorem 5 remains valid if we replace $|||.|||_\infty$ norm by the $||.||_2$ norm.*

This means that if $\epsilon = ||[C]^2 - [C^*]^2||_2$ is small (*i.e.* if $\epsilon < 2^{-m}$), the complexity of any basic differential or linear cryptanalysis is close to 2^m, thus no more efficient than exhaustive search.

In the following sections we show how to construct a practical cipher with a relatively small $||[C]^2 - [C^*]^2||_2$. For this we first study how to bound the decorrelation L_2-bias of a Feistel Cipher from the decorrelation of its round functions. Then we construct round functions with relatively small decorrelation L_2-bias and a corresponding dedicated cipher.

4 L_2-Decorrelation of Feistel Ciphers

Here we show how to measure the decorrelation L_2-bias of a Feistel cipher from the decorrelation of its round functions. We first study the case of a 2-round Feistel Cipher.

Lemma 7. *Let $\mathcal{M}_0$ be a group and let $\mathcal{M} = \mathcal{M}_0^2$. Let F_1, F_2, F_1^* and F_2^* be four independent random functions on $\mathcal{M}_0$ where F_1^* and F_2^* have a uniform distribution. If we have $||[F_i]^d - [F_i^*]^d||_2 \le \epsilon$ then we have*

$$||[\Psi(F_1, F_2)]^d - [\Psi(F_1^*, F_2^*)]^d||_2 \le \epsilon\sqrt{\epsilon^2 + 2P_d}$$

where P_d is the number of partitions of $\{1, \ldots, d\}$.

[6] Those notations are inspired from Matsui's [8]. Actually, Matsui defined DP and LP and we use here their expected values over the distribution of the cipher in order to measure the average complexity of the attacks.

Proof. Let $x = (x_1, \ldots, x_d)$ (resp. $y = (y_1, \ldots, y_d)$)be a multi-point with $x_i = (x_i^l, x_i^r)$ (resp. $y_i = (y_i^l, y_i^r)$). We recall that the relation $y_i = \Psi(g_1, g_2)(x_i)$ means that $y_i^r = x_i^l + g_1(x_i^r)$ and $y_i^l = x_i^r + g_2(y_i^r)$. We thus have

$$\begin{aligned}&\Pr[\Psi(G_1, G_2)(x_i) = y_i; i] \\ &\quad = \Pr[G_1(x_i^r) = y_i^r - x_i^l; i] \Pr[G_2(y_i^r) = y_i^l - x_i^r; i].\end{aligned}$$

The 1–1 relation between (x^l, x^r, y^l, y^r) and $(x^r, y^r - x^l, y^r, y^l - x^r)$ is an important point. In the following, we let (t, u, v, w) denote this family. Let us write the previous equation

$$\Pr_{G_1 G_2}[x \mapsto y] = \Pr_{G_1}[t \mapsto u] \Pr_{G_2}[v \mapsto w].$$

Let $\Delta \Pr$ denotes $\Pr_F - \Pr_{F^*}$ with obvious notations. We have

$$\begin{aligned}\Delta \Pr_{12}[x \mapsto y] = \Delta \Pr_1[t \mapsto u] \Delta \Pr_2[v \mapsto w] + \Pr_{F_1^*}[t \mapsto u] \Delta \Pr_2[v \mapsto w] \\ + \Pr_{F_2^*}[v \mapsto w] \Delta \Pr_1[t \mapsto u].\end{aligned}$$

Now we have

$$||[\Psi(F_1, F_2)]^d - [\Psi(F_1^*, F_2^*)]^d||_2^2 = \sum_{x,y} \left(\Delta \Pr_{12}[x \mapsto y] \right)^2.$$

We note that

$$\sum_{t,u} \Pr_{F_1^*}[t \mapsto u] \Delta \Pr_1[t \mapsto u] = 0$$

(and a similar property for $\Pr_2$), thus we have

$$\begin{aligned}||[\Psi(F_1, F_2)]^d - [\Psi(F_1^*, F_2^*)]^d||_2^2 = \epsilon_1^2 \epsilon_2^2 + \epsilon_1^2 \sum_{v,w} \left(\Pr_{F_2^*}[v \mapsto w] \right)^2 \\ + \epsilon_2^2 \sum_{t,u} \left(\Pr_{F_1^*}[t \mapsto u] \right)^2\end{aligned}$$

where $\epsilon_j = ||[F_j]^d - [F_j^*]^d||_2^2$. Hence

$$||[\Psi(F_1, F_2)]^d - [\Psi(F_1^*, F_2^*)]^d||_2^2 \leq \epsilon^4 + 2\epsilon^2 \sum_{t,u} \left(\Pr_{F_1^*}[t \mapsto u] \right)^2.$$

For any partition $\mathcal{P} = \{O_1, \ldots, O_k\}$ of $\{1, \ldots, d\}$ into k parts, let

$$\mathcal{M}_{\mathcal{P}} = \{t; \forall i, j \quad (t_i = t_j \Leftrightarrow \exists k \quad i, j \in O_k)\}.$$

We have

$$\sum_{t,u} \left(\Pr_{F_1^*}[t \mapsto u] \right)^2 = \sum_{\substack{\mathcal{P} \text{ into} \\ k \text{ parts}}} \sum_{t \in \mathcal{M}_{\mathcal{P}}} \sum_{u} \left(\Pr_{F_1^*}[t \mapsto u] \right)^2.$$

We have M^k u-terms for which the probability is not zero. Namely it is $1/M^k$. The number of t-terms which correspond to this partition is $M(M-1)\ldots(M-k+1)$ thus

$$\sum_{t,u} \left(\Pr_{F_1^*}[t \mapsto u] \right)^2 = \sum_{\substack{\mathcal{P} \text{ into} \\ k \text{ parts}}} \frac{M(M-1)\ldots(M-k+1)}{M^k}$$

which is less than P_d. $\square$

In order to measure the decorrelation distance between a 2-round Feistel Cipher and the Perfect Cipher, we thus have to study the case of a truly random 2-round Feistel Cipher.

Lemma 8. *Let $\mathcal{M}_0$ be a group and let $\mathcal{M} = \mathcal{M}_0^2$. Let F_1^* and F_2^* be two independent random functions on $\mathcal{M}_0$ with a uniform distribution and let C^* be the Perfect Cipher on $\mathcal{M}$. We have*

$$||[\Psi(F_1^*, F_2^*)]^d - [C^*]^d||_2 \leq \sqrt{P_d(P_d - 1)}$$

where P_d is the number of partitions of $\{1, \ldots, d\}$.

Proof. With obvious notations we have

$$||[\Psi(F_1^*, F_2^*)]^d - [C^*]^d||_2^2 = \sum_{x,y} \left(\Pr_{\Psi(F_1^*, F_2^*)} - \Pr_{C^*} \right)^2 [x \mapsto y].$$

The sums $\sum \Pr_{C^*}^2[x \mapsto y]$ and $\sum \Pr_{\Psi(F_1^*, F_2^*)} \Pr_{C^*}[x \mapsto y]$ are equal to P_d. (We observe it by fixing the partition associated to x and making the sum over all ys.) For the remaining sum, we use same ideas as in the previous proof:

$$\sum_{x,y} \left(\Pr_{\Psi(F_1^*, F_2^*)}[x \mapsto y] \right)^2 = \sum_{t,u,v,w} \left(\Pr_{F_1^*}[t \mapsto u] \Pr_{F_2^*}[v \mapsto w] \right)^2$$

which is less than P_d^2. $\square$

Lemma 8 may look useless because the decorrelation bias of is greater than one (so we cannot consider product cipher and get efficient bounds). We can however use it to study the case of a 4-round Feistel Cipher. From Lemma 7 and Lemma 8 and from Equation (5) we obtain the following Lemma in a straightforward way.

Lemma 9. *Let $\mathcal{M}_0$ be a group and let $\mathcal{M} = \mathcal{M}_0^2$. Let $F_1, \ldots, F_4$, $F_1^*, \ldots, F_4^*$ be eight independent random functions on $\mathcal{M}_0$ where the F_i^*s have a uniform distribution. If we have $||[F_i]^d - [F_i^*]^d||_2 \leq \epsilon \leq \sqrt{2}$ then we have*

$$||[\Psi(F_1, F_2, F_3, F_4)]^d - [\Psi(F_1^*, F_2^*, F_3^*, F_4^*)]^d||_2 \leq 2\sqrt{2}(P_d)^{\frac{3}{2}}\epsilon$$

where P_d is the number of partitions of $\{1, \ldots, d\}$.

It thus remains to study the decorrelation distance between a truly random 4-round Feistel Cipher and the Perfect Cipher: once we know that

$$||[\Psi(F_1, F_2, F_3, F_4)]^d - [\Psi(F_1^*, F_2^*, F_3^*, F_4^*)]^d||_2 \leq u_d$$

we obtain from Equation (3) that

$$||[\Psi(F_1, \ldots, F_{4r})]^d - [\Psi(F_1^*, \ldots, F_{4r}^*)]^d||_2 \leq \left(2\sqrt{2}(P_d)^{\frac{3}{2}}\epsilon + u_d\right)^r$$

where $\epsilon = \max_i ||[F_i]^d - [F_i^*]^d||_2 \leq \sqrt{2}$. Unfortunately, the problem of obtaining a general result on the d-wise decorrelation of a truly random 4-round Feistel Cipher is still open.[7] In the next section we propose a construction in the $d = 2$ case for which we can evaluate the decorrelation.

5 A Dedicated Construction

In a general finite field $\mathrm{GF}(q)$, an obvious way to construct pairwise decorrelated functions (resp. permutations) consists of taking

$$F(x) = a.x + b$$

[7] This problem has been solved in [13, 14] with the $|||.|||_\infty$ norm. This is why the L_2 norm looks less friendly.

where (a, b) is a random pair uniformly distributed in $\mathrm{GF}(q)^2$ (resp. $\mathrm{GF}(q)^* \times \mathrm{GF}(q)$). Unfortunately, the traditional message space $\mathbf{Z}_2^m$ requires that we use finite fields of characteristic two. If we aim to implement a cipher in software on a modern microprocessor, it looks cumbersome to implement a poor characteristic-two multiplication since there already is a built-in integer multiplication. For this reason we can think of the

$$F(x) = ((ax + b) \bmod p) \bmod 2^{\frac{m}{2}}$$

imperfectly decorrelated function to be inserted at the input of each round function of a Feistel Cipher, where p is a prime close to $2^{\frac{m}{2}}$.

In [14] the (m, r, d, p)-PEANUT Cipher Family is defined to be the set of all r-round Feistel Ciphers over $\mathbf{Z}_2^m$ in which all round functions can be written

$$F(x) = g\left(\sum_{i=1}^{d} k_i.x^{d-i} \bmod p \bmod 2^{\frac{m}{2}}\right)$$

where $(k_1, \ldots, k_d)$ is an (independent) round key which is uniformly distributed in $\{0, \ldots, 2^{\frac{m}{2}} - 1\}^d$, p is a prime, and g is a permutation. For $p > 2^{\frac{m}{2}}$, F has a friendly d-wise decorrelation $|||.|||_\infty$-bias which is roughly $2d\delta$ when $p = (1 + \delta)2^{\frac{m}{2}}$. For $p < 2^{\frac{m}{2}}$, the $|||.|||_\infty$-decorrelation is poor for $d \geq 2$. For instance, in the case $d = 2$, for $x = (0, p)$ we have

$$\sum_{y=(y_1,y_2)} \left| \Pr\begin{bmatrix} g(k_2 \bmod p) = y_1 \\ g(k_2 \bmod p) = y_2 \end{bmatrix} - \Pr\begin{bmatrix} F^*(0) = y_1 \\ F^*(p) = y_2 \end{bmatrix} \right| = 2 - 2^{1-\frac{m}{2}} + \delta.$$

Hence $|||[F]^2 - [F^*]^2|||_\infty \approx 2$. The $p < 2^{\frac{m}{2}}$ case can however be studied with the L_2 norm. In the following, we consider a PEANUT Cipher construction with $d = 2$ and $p < 2^{\frac{m}{2}}$.

Lemma 10. *Let A and B be two independent random variables with a uniform distribution over $\{0, \ldots, 2^{\frac{m}{2}} - 1\}$. We let $F(x) = Ax + B \bmod p$ where $p = (1 - \delta)2^{\frac{m}{2}}$ is a prime for $1/14 \geq \delta \geq 0$. Let F^* be a random function uniformly distributed over the same set. We have*

$$||[F]^2 - [F^*]^2||_2 \leq 2\sqrt{2\delta}.$$

Proof. We let $N = 2^{\frac{m}{2}}$. We want to upper bound the sum

$$\sum_{\substack{x=(x_1,x_2)\\y=(y_1,y_2)}} \left([F]^2_{x,y} - [F^*]^2_{x,y}\right)^2.$$

Table 1 shows all the possible cases for x and y, the number of times they occur and an upper bound for the probability difference.

For instance if $x_1 = x_2 \not\equiv 0 \pmod p$ and $y_1 = y_2 < p$, we have

$$[F]^2_{(x_1,x_2),(y_1,y_2)} = \Pr[Ax_1 + B \bmod p = y_1]$$

and $[F^*]^2_{(x_1,x_2),(y_1,y_2)} = N^{-1}$. We let a (resp. b, c, d) be the number of $(A \bmod p, B \bmod p)$ pairs such that $Ax_1 + B \bmod p = y_1$ and

- $A \bmod p < \delta N$ and $B \bmod p < \delta N$ (resp.
- $A \bmod p < \delta N$ and $B \bmod p \geq \delta N$,
- $A \bmod p \geq \delta N$ and $B \bmod p < \delta N$,
- $A \bmod p \geq \delta N$ and $B \bmod p \geq \delta N$).

We have $a + b = \delta N$, $a + c = \delta N$ and $a + b + c + d = p$. Hence

$$[F]^2_{(x_1,x_2),(y_1,y_2)} = \frac{4a + 2b + 2c + d}{N^2} = \frac{N + \delta N + a}{N^2}.$$

Since we have $0 \leq a \leq \delta N$, we have

$$(1 + \delta)N^{-1} \leq [F]^2_{(x_1,x_2),(y_1,y_2)} \leq (1 + 2\delta)N^{-1}.$$

The $x_1 \not\equiv x_2$ case is split into four cases which depend on x_1 and x_2. The last case is $y_1 \geq p$ or $y_2 \geq p$ for which $[F]^2_{x,y} = 0$. The three other cases correspond to cases on $(A \bmod p, B \bmod p)$ with $y_i = Ax_i + B \bmod p$.

Case 1: $A \bmod p < \delta N$, $B \bmod p < \delta N$. We have $[F]^2_{x,y} = 4N^{-1}$.
Case 3: $A \bmod p \geq \delta N$, $B \bmod p \geq \delta N$. We have $[F]^2_{x,y} = N^{-1}$.
Case 2: other values. We have $[F]^2_{x,y} = 2N^{-1}$.

We can now upper bound the whole sum. We obtain that the decorrelation bias $||[F]^2 - [F^*]^2||_2^2$ is less than

$$7\delta + 14\delta^2 - 6\frac{\delta}{N} - 4\delta^3 - 24\frac{\delta^2}{N} + 4\frac{\delta}{N^2} - 8\delta^4 + 16\frac{\delta^3}{N} - 8\frac{\delta^2}{N^2}$$

which is less than 8δ when $\delta \leq 1/14$. □

case x	case y	num. x	num. y	$\|[F]^2_{x,y} - [F^*]^2_{x,y}\| \leq$
	$y_1 = y_2 < \delta N$		δN	N^{-1}
$x_1 = x_2 \equiv 0$	$y_1 = y_2 \geq (1-\delta)N$	2	δN	N^{-1}
	other cases		$N^2 - 2\delta N$	0
	$y_1 = y_2 \geq (1-\delta)N$		δN	N^{-1}
$x_1 = x_2 \not\equiv 0$	$y_1 = y_2 < (1-\delta)N$	$N-2$	$(1-\delta)N$	$2\delta N^{-1}$
	$y_1 \neq y_2$		$N^2 - N$	0
	$y_1 = y_2 < \delta N$		δN	$2N^{-1} - N^{-2}$
$x_1 \neq x_2$, $x_1 \equiv x_2 \equiv 0$	$y_1 = y_2 \geq (1-\delta)N$	2	δN	N^{-2}
	other $y_1 = y_2$		$(1-2\delta)N$	$N^{-1} - N^{-2}$
	$y_1 \neq y_2$		$N^2 - N$	N^{-2}
$x_1 \neq x_2$, $x_1 \equiv x_2 \not\equiv 0$	$y_1 = y_2 < (1-\delta)N$	$2\delta N - 2$	$(1-\delta)N$	$(1+2\delta)N^{-1} - N^{-2}$
	other cases		$N^2 - (1-\delta)N$	N^{-2}
	case 1		$\delta^2 N^2$	$3N^{-2}$
$x_1 \not\equiv x_2$	case 2	$N^2 - (1+2\delta)N$	$2(1-2\delta)\delta N^2$	N^{-2}
	case 3		$(1-2\delta)^2 N^2$	0
	y_1 or $y_2 \geq (1-\delta)N$		$(2-\delta)\delta N^2$	N^{-2}

Table 1. Decorrelation of $A.x + B \bmod (1-\delta)N$

Lemma 11. *Let $\mathcal{M} = \mathbf{Z}_2^m$. Let $F_1^*, \ldots, F_4^*$ be four independent random functions on $\mathbf{Z}_2^{\frac{m}{2}}$ with a uniform distribution and let C^* be the Perfect Cipher on $\mathcal{M}$. We have*

$$||[\Psi(F_1^*, F_2^*, F_3^*, F_4^*)]^2 - [C^*]^2||_2 \leq \sqrt{2.2^{-m} + 4.2^{-\frac{3m}{2}}}.$$

Proof. For each input pair $x = (x_1, x_2)$ we have $x_i = (x_i^l, x_i^r)$. Similarly, for each output pair $y = (y_1, y_2)$ we have $y_i = (y_i^l, y_i^r)$. All (x, y) pairs can be split into 10 cases:

1. $y_1^r \neq y_2^r$, $x_1^r \neq x_2^r$
2. $y_1^r \neq y_2^r$, $x_1^r = x_2^r$, $x_1^l \oplus x_2^l \not\in \{0, y_1^r \oplus y_2^r\}$
3. $y_1^r \neq y_2^r$, $x_1^r = x_2^r$, $x_1^l \oplus x_2^l = y_1^r \oplus y_2^r$
4. $y_1^r \neq y_2^r$, $x_1 = x_2$
5. $y_1^r = y_2^r$, $y_1^l \neq y_2^l$, $x_1^r \oplus x_2^r \not\in \{0, y_1^l \oplus y_2^l\}$
6. $y_1^r = y_2^r$, $y_1^l \neq y_2^l$, $x_1^r \oplus x_2^r = y_1^l \oplus y_2^l$
7. $y_1^r = y_2^r$, $y_1^l \neq y_2^l$, $x_1^r = x_2^r$, $x_1^l \neq x_2^l$
8. $y_1^r = y_2^r$, $y_1^l \neq y_2^l$, $x_1 = x_2$
9. $y_1^r = y_2^r$, $y_1^l = y_2^l$, $x_1 = x_2$
10. $y_1 = y_2$, $x_1 \neq x_2$

Each case requires a dedicated study.

We consider a truly random $2r$-round Feistel cipher for $r \geq 1$ denoted $C = \Psi(F_1^*, \ldots, F_{2r}^*)$. We have

$$[C]^2_{x,y} = \begin{cases} A_r^1 = \frac{1}{N^2(N^2-1)}\left(1 - \frac{1}{N^{2r}}\right) & \text{if case 1} \\ A_r^2 = \frac{1}{N^2(N^2-1)}\left(1 - \frac{1}{N^{r-1}} - \frac{1}{N^r} + \frac{1}{N^{2r-1}}\right) & \text{if case 2} \\ A_r^3 = \frac{1}{N^2(N^2-1)}\left(1 + \frac{1}{N^{r-2}} - \frac{1}{N^{r-1}} - \frac{2}{N^r} + \frac{1}{N^{2r-1}}\right) & \text{if case 3} \\ 0 & \text{if case 4} \\ A_r^5 = \frac{1}{N^2(N^2-1)}\left(1 - \frac{1}{N^{r-1}} - \frac{1}{N^r} + \frac{1}{N^{2r-1}}\right) & \text{if case 5} \\ A_r^6 = \frac{1}{N^2(N^2-1)}\left(1 + \frac{1}{N^{r-2}} - \frac{1}{N^{r-1}} - \frac{2}{N^r} + \frac{1}{N^{2r-1}}\right) & \text{if case 6} \\ A_r^7 = \frac{1}{N^2(N^2-1)}\left(1 - \frac{1}{N^{2r-2}}\right) & \text{if case 7} \\ 0 & \text{if case 8} \\ N^{-2} & \text{if case 9} \\ 0 & \text{if case 10} \end{cases}$$

where $N = 2^{\frac{m}{2}}$. We prove this by an easy induction. Namely we show that

$$\begin{pmatrix} A_r^1 \\ A_r^2 \\ A_r^3 \end{pmatrix} = \begin{pmatrix} \frac{(N-1)^2}{N^2} + \frac{1}{N} & \frac{N-2}{N^2} & \frac{1}{N^2} \\ 1 - \frac{1}{N} & \frac{1}{N} & 0 \\ 1 - \frac{1}{N} & 0 & \frac{1}{N} \end{pmatrix}^{r-1} \begin{pmatrix} \frac{1}{N^4} \\ 0 \\ \frac{1}{N^3} \end{pmatrix}$$

and

$$\begin{pmatrix} A_r^5 \\ A_r^6 \\ A_r^7 \end{pmatrix} = \begin{pmatrix} \frac{(N-1)(N-2)}{N^2} + \frac{1}{N} & \frac{N-1}{N^2} & \frac{N-1}{N^2} \\ \frac{(N-1)(N-2)}{N^2} & \frac{N-1}{N^2} + \frac{1}{N} & \frac{N-1}{N^2} \\ 1 - \frac{2}{N} & \frac{1}{N} & \frac{1}{N} \end{pmatrix}^{r-1} \begin{pmatrix} 0 \\ \frac{1}{N^3} \\ 0 \end{pmatrix}$$

For instance, if $r = 1$ and $y_1^r \neq y_2^r$, $x_1^r = x_2^r$, $x_1^l \oplus x_2^l = y_1^r \oplus y_2^r$, the probability corresponds to the fact that $F_1^*(x_1^r)$ XORs the good value on both x_1^l and x_2^l (with probability $1/N$) and that both $F_2^*(y_1^r)$ and $F_2^*(y_2^r)$ XOR the good values on x_1^r and x_2^r respectively (with probability $1/N^2$).

To prove the matrix relations, we let x denote the input of C, y denote the output of the first two rounds and z denote the output. We have

$$\begin{aligned} z &= \Psi(F_1^*, F_2^*, F_3^*, \ldots, F_{2r}^*)(x) = \Psi(F_3^*, \ldots, F_{2r}^*)(y) \\ (y^r, y^l) &= \Psi(F_1^*, F_2^*)(x). \end{aligned}$$

For instance, transition from case 2 ($z_1^r \neq z_2^r$, $y_1^r = y_2^r$, $y_1^l \oplus y_2^l \notin \{0, z_1^r \oplus z_2^r\}$) to case 1 ($z_1^r \neq z_2^r$, $x_1^r \neq x_2^r$) corresponds to the $N(N-2)$ possibilities for y_1^l and y_2^l (all but for $y_1^l = y_2^l$ or $y_1^l \oplus y_2^l = z_1^r \oplus z_2^r$), all with probability $1/N^2$ (since $F_1^*(x_1^r)$ and $F_1^*(x_2^r)$ are independent), mixed with the N possibilities for $y_1^r = y_2^r$, all with probability $1/N^2$, which gives $(N-2)/N^2$. This means A_r^1 includes a term $\frac{N-2}{N^2} A_{r-1}^2$ which represents all possible ys coming from case 2.

With this result we can compute the pairwise decorrelation bias of C. We have

$$\begin{aligned} \|[C]^2 - [C^*]^2\|_2^2 = & n_1(\Delta A_r^1)^2 + n_2(\Delta A_r^2)^2 + n_3(\Delta A_r^3)^2 \\ & + n_5(\Delta A_r^5)^2 + n_6(\Delta A_r^6)^2 + n_7(\Delta A_r^7)^2 \end{aligned}$$

where $\Delta A_r^i = A_r^i - \frac{1}{N^2(N^2-1)}$ and n_i is the number of (x, y) pairs in case i. We obtain

$$\frac{2N^{4-2r} - 6N^{2-2r} - 4N^{1-2r} + N^{4-4r} + 2N^{3-4r} + N^{2-4r}}{(N-1)^2}.$$

For $r = 2$ (four rounds), this is less than $2N^{-2} + 4N^{-3}$. □

We can now define the PEANUT97 Cipher construction. It consists of a $(m, 4r, 2, p)$-PEANUT Cipher, *i.e.* a $4r$-round Feistel Cipher on m-bit message blocks which is characterized by some prime $p \leq 2^{\frac{m}{2}}$. Each round function of the cipher must be with the form

$$F_i(x) = g_i(K_{2i-1}x + K_{2i} \bmod p)$$

where $(K_1, \ldots, K_{8r})$ is uniformly distributed in $\mathbf{Z}_2^{4mr}$ and g_i is a (possibly independently keyed) permutation on the $\frac{m}{2}$-bit strings. The lemmata 9, 10 and 11 proves the following theorem.

Theorem 12. *Let C be a $(m, 4r, 2, p)$-PEANUT97 Cipher such that $p = (1-\delta)2^{\frac{m}{2}}$ with $0 \leq \delta \leq \frac{1}{14}$. Let C^* be the Perfect Cipher. We have*

$$\|[C]^2 - [C^*]^2\|_2 \leq \left(16\sqrt{2\delta} + \sqrt{2.2^{-m} + 4.2^{-\frac{3m}{2}}}\right)^r.$$

For instance, with $m = 64$ and $p = 2^{32} - 5$, we obtain $\|[C]^2 - [C^*]^2\|_2 \leq 2^{-10r}$. Thus for $r = 7$ we have $\|[C]^2 - [C^*]^2\|_2 \leq 2^{-70}$.

Theorem 6 thus shows that $|p - p^*| \leq 0.1$ for any differential distinguisher with complexity $n \leq 2^{60}$ and any linear distinguisher with complexity $n \leq 2^{44}$.

This PEANUT97 construction has been tested on a Pentium in assembly code. A 28-round 64-bit encryption required less than 790 clock cycles, which yields an encryption rate of 23Mbps working at 300MHz. The table below compares it with the PEANUT98 construction, for which the $|||.|||_\infty$-decorrelation theory enables to decrease the number of rounds (see [13]) and the DFC AES candidate which is a PEANUT98 128-bit block cipher (see [5]). All ciphers have similar security against differential and linear cryptanalysis. We remark that one PEANUT97 is much faster than the other rounds, so PEANUT97 may be faster than PEANUT98 if we can get tighter bounds in order to decrease the number of rounds.

cipher	PEANUT97	PEANUT98	DFC
block length	64	64	128
number of rounds	28	9	8
cycles/encryption	788	396	754
cycles/round	28	44	94
enc. rate at 300MHz	23Mbps	46Mbps	49Mbps
pairwise decorrelation	2^{-70} (L_2)	2^{-76} ($|||.|||_\infty$)	2^{-112} ($|||.|||_\infty$)
reference	[12], here	[13, 14]	[5, 9]

6 Conclusion

We have shown how to use the $ax + b \bmod p$ pairwise decorrelation primitive for $p \leq 2^{\frac{m}{2}}$. It requires that we use the L_2 norm in the decorrelation technique, which leads to more complicated computations than for the $|||.|||_\infty$ norm.

When used at the input of Feistel Ciphers, this primitive enables to protect it against differential and linear cryptanalysis. For 64-bit message block, it however requires at least 28 rounds.

Some extensions of the $|||.|||_\infty$-decorrelation results to the L_2-decorrelation is still open: it is not clear how to state results with higher degrees of decorrelation ($d > 2$) and how to prove the security of decorrelated ciphers against general iterated attacks as in [14].

References

1. E. Biham. A Fast New DES Implementation in Software. In *Fast Software Encryption*, Haifa, Israel, Lectures Notes in Computer Science 1267, pp. 260–272, Springer-Verlag, 1997.
2. E. Biham, A. Shamir. *Differential Cryptanalysis of the Data Encryption Standard*, Springer-Verlag, 1993.
3. L. Carter, M. Wegman. Universal Classes of Hash Functions. *Journal of Computer and System Sciences*, vol. 18, pp. 143–154, 1979.
4. H. Feistel. Cryptography and Computer Privacy. *Scientific American*, vol. 228, pp. 15–23, 1973.
5. H. Gilbert, M. Girault, P. Hoogvorst, F. Noilhan, T. Pornin, G. Poupard, J. Stern, S. Vaudenay. Decorrelated Fast Cipher: an AES Candidate. Advanced Encryption Standard Submissions, US Department of Commerce, 1998.
6. M. Luby, C. Rackoff. How to Construct Pseudorandom Permutations from Pseudorandom Functions. *SIAM Journal on Computing*, vol. 17, pp. 373–386, 1988.
7. M. Matsui. The First Experimental Cryptanalysis of the Data Encryption Standard. In *Advances in Cryptology CRYPTO'94*, Santa Barbara, California, U.S.A., Lectures Notes in Computer Science 839, pp. 1–11, Springer-Verlag, 1994.
8. M. Matsui. New Structure of Block Ciphers with Provable Security Against Differential and Linear Cryptanalysis. In *Fast Software Encryption*, Cambridge, United Kingdom, Lectures Notes in Computer Science 1039, pp. 205–218, Springer-Verlag, 1996.
9. G. Poupard, S. Vaudenay. Decorrelated Fast Cipher: an AES Candidate Well Suited for Low Cost Smart Cards Applications. Submitted to CARDIS'98.
10. C. E. Shannon. Communication Theory of Secrecy Systems. *Bell system technical journal*, vol. 28, pp. 656–715, 1949.
11. A. Shamir. How to Photofinish a Cryptosystem? Presented at the Rump Session of Crypto'97.
12. S. Vaudenay. A Cheap Paradigm for Block Cipher Security Strengthening. Technical Report LIENS-97-3, Ecole Normale Supérieure, 1997.
13. S. Vaudenay. Provable Security for Block Ciphers by Decorrelation. In *STACS 98*, Paris, France, Lectures Notes in Computer Science 1373, pp. 249–275, Springer-Verlag, 1998.
14. S. Vaudenay. Provable Security for Block Ciphers by Decorrelation. (Journal Version.) Submitted.
15. S. Vaudenay. The Decorrelation Technique Home-Page. URL:`http://www.dmi.ens.fr/~vaudenay/decorrelation.html`
16. G. S. Vernam. Cipher Printing Telegraph Systems for Secret Wire and Radio Telegraphic communications. *Journal of the American Institute of Electrical Engineers*, vol. 45, pp. 109–115, 1926.
17. M. N. Wegman, J. L. Carter. New Hash Functions and their Use in Authentication and Set Equality. *Journal of Computer and System Sciences*, vol. 22, pp. 265–279, 1981.