Decorrelation: A Theory for Block Cipher Security

Serge Vaudenay

Swiss Federal Institute of Technology (EPFL)
Serge.Vaudenay@epfl.ch

Abstract. Pseudorandomness is a classical model for the security of block ciphers. In this paper we propose convenient tools in order to study it in connection with the Shannon Theory, the Carter-Wegman universal hash functions paradigm, and the Luby-Rackoff approach. This enables the construction of new ciphers with security proofs under specific models. We show how to ensure security against basic differential and linear cryptanalysis and even more general attacks. We propose practical construction schemes.

1 Introduction

Conventional encryption is used in order to enforce confidentiality of communications in a network. Following the Kerckhoffs principles [34], schemes are defined by three *public* algorithms: a key generation scheme, an encryption scheme, and a decryption scheme. Two parties willing to communicate confidentially can generate a private key which is used as a parameter for encryption and decryption. Here encryption and decryption are formalized as functions C and D, respectively, such that $D(C(x)) = x$ for any message x.

In 1949, Shannon formalized the notion of secrecy [59]. He formally proved the unconditional security (in his security model) of the Vernam Cipher which had been published in 1926 [71]. Unfortunately, this scheme happens to be quite expensive to implement for networking because the sender and the receiver need to be synchronized, and they need quite cumbersome huge keys. Shannon's result also proves that unconditional security cannot be achieved in a better (*i.e.* cheaper) way. For this reason, empirical security seemed to be the only efficient alternative, and all secret key block ciphers which have been publicly developed were considered to be secure until some researcher published a dedicated attack on it. Therefore research mostly advanced like a tennis game between designers and analysts.

In the 70s the U.S. Government used to be far ahead of academic research on cryptography. By releasing the *Data Encryption Standard* (DES) [2] without development rationales, this paradoxically boosted research on block ciphers as researchers were trying to reverse engineer or attack the design of DES. Real advances on the attack strategies on block ciphers were made in the early 90s when Biham and Shamir invented *differential cryptanalysis* and applied it against DES [7,8,9,10]. The best version of this attack can recover a secret key with a simple 2^{47}-chosen plaintext attack.[1] Although this attack is heuristic, experiments confirmed the results. Biham and Shamir's attack was based on statistical cryptanalysis ideas which were later used by Gilbert and Chassé against another cipher [15,16]. Those ideas inspired Matsui who developed a *linear cryptanalysis* on DES [41,42]. This heuristic attack, which has been implemented, can recover the key with a 2^{43}-known plaintext attack. Since then, many researchers tried to generalize and improve these attacks (see, for instance, [22,27,29,31,35,36,37,48,61,62]), but the underlying ideas were quite the same.

The basic idea of differential cryptanalysis is to use properties like "if x and x' are two plaintext blocks such that $x' = x \oplus a$, then it is likely that $C(x') = C(x) \oplus b$".[2] Then the attack is an iterated two-chosen plaintexts attack which consists in getting the encrypted values of two random plaintexts which verify $x' = x \oplus a$ until the special event $C(x') = C(x) \oplus b$ occurs. Similarly, linear cryptanalysis consists in using the probability $\Pr[C(x) \in H_2 / x \in H_1]$ for two given hyperplanes H_1 and H_2. With the $\mathrm{GF}(2)$-vector space structure, hyperplanes are half-spaces, and this probability should be close to $1/2$. Linear cryptanalysis exploits the distance between this probability and $1/2$ when it is large enough. More precisely, linear cryptanalysis is an incremental one-known plaintext attack where we simply measure the correlation between the events $[x \in H_1]$ and $[C(x) \in H_2]$.

Cryptanalysis is not restricted to destructive purposes. It also has a positive side on which the analyst tries to prove the security of cryptographic schemes. Unlike the negative aspects which can be purely intuitive (there is no need

[1] Previously, the best known attack was an improvement of exhaustive search which required on average 2^{54} DES computations by using the complementation property.

[2] Here $\oplus$ denotes the bitwise exclusive OR function. However, this technique can be extended to any other group law.

for proving that an attack works if we can experiment it successfully), the positive aspects require more formal and systematic results.

Instead of breaking or proposing new encryption functions, Nyberg first formalized the notion of strength against differential cryptanalysis [50]. Similarly, Chabaud and Vaudenay formalized the notion of strength against linear cryptanalysis [12]. With this approach, we can study how to make internal computation boxes resistant against both attacks. This can be used in a heuristic way by usual active s-boxes counting tricks (*e.g.* see [22,23]). This has also been used to construct the PURE cipher for which we can prove the security against both attacks (see Nyberg and Knudsen [52]), but in an unsatisfactory way which introduces some algebraic properties which lead to other attacks as shown by Jakobsen and Knudsen [26]. The Nyberg-Knudsen approach was later used by Matsui in practical block ciphers including MISTY and KASUMI [1,43,44].

Another approach in order to study the security of block ciphers was introduced by Luby and Rackoff in 1988[3] [40]. They have shown how to formalize security by pseudo-randomness and how to prove the security of the underlying DES construction — the Feistel scheme [14] — provided that round functions are totally random. As for the Shannon result, this suffers from the expensive cost of random bits, and basically requires having an enormous private key. We can still use derandomization techniques, like the Carter-Wegman method [11,73] for sampling pairwise independent numbers. This leads us to the notion of decorrelation which enables measuring the pseudo-randomness with small keys and studying how it protects against attacks.

Inspired by Carter and Wegman, we use simple primitives which we call NUT (for "n-Universal Transformation") since they are so cheap to implement. We propose construction methods for block ciphers that we call COCONUT (for "Cipher Organized with Cute Operations and NUT"), PEANUT (for "Pretty Encryption Algorithm with NUT"), and WALNUT (for "Wonderful Algorithm with Light NUT"). Our construction is based on a theory which mixes all previous results and happens to offer new ways of investigation for research on block ciphers.

1.1 Related Work

Several researchers concentrated on the positive side of cryptanalysis: security arguments. Usually block cipher designers try to upper bound the probability of the best differential or linear characteristics in ad-hoc ways. Some results apply to multi-path characteristics like Nyberg-Knudsen [51,52], Aoki-Ohta [3], Keliher *et al.* [32,33], and Park *et al.* [53,54].

In another approach, Luby-Rackoff [39] and Maurer-Massey [45] studied the security of product ciphers.

One of our purpose is to quantify the security against ciphers when a limited number d of samples are available, starting from the seminal work Luby-Rackoff [40] related to Feistel schemes [14]. Some extensions investigated the security with higher values of d, *e.g.* Patarin [56], and Maurer-Pietrzak [46]. Many other researchers have applied the same techniques to other schemes. (See, for instance, [19,24,25,30,47,49].)

Our work studies provable security against specific models of attacks. We addressed the basic differential and linear cryptanalysis and the more general model of iterated attacks which are based on a (low) specific number d of plaintext/ciphertext samples. Our work was further extended by Junod [28] with techniques using statistics.

Some papers related to the theory presented in this article are collected on [66].

1.2 Structure of this Article

The paper is organized as follows. First we give some definitions on decorrelation (Section 2) and basic constructions for NUTs (Section 3). Then we investigate connections to Shannon's perfect secrecy notion (Section 4). We show how to express security results in the Luby-Rackoff security model (Section 5). We prove how pairwise decorrelation can protect a cipher against basic differential and linear cryptanalysis (Sections 6.1 and 6.2). We generalize those results with the notion of "iterated attacks of order d" (Section 6.3). Then we apply decorrelation upper bounds to practical constructions such as Feistel Ciphers (Section 7). Finally, we define the COCONUT, PEANUT and WALNUT families (Sections 8.1 and 8.2).

1.3 Notations

In what follows we use the following notations:

[3] An earlier version was presented at the CRYPTO '85 conference.

$\circ$: composition of two functions: $f \circ g$ is a function which maps x onto $f(g(x))$,
$\mathbf{R}$: set of real numbers,
$\mathrm{GF}(q)$: finite field with q elements,
$\mathcal{M}^d$: set of all sequences which consist of d elements of a set $\mathcal{M}$,
$\mathrm{Adv}_{\mathcal{A}}$: advantage of a distinguisher $\mathcal{A}$ (see Section 5),
1_P: variable which is set to 1 if the predicate P is satisfied or to 0 otherwise,
$\mathrm{DP}^c(a,b)$: differential probability of a function c with characteristic (a,b) (see Section 6.1),
$\mathrm{LP}^c(a,b)$: linear probability of a function c with characteristic (a,b) (see Section 6.2).

We represent all random variables by capital letters. They are associated to a probability distribution which will be clear from the context. For instance, X may denote a random variable and $\Pr[X = x]$ may represent the probability that it takes a given value x.

Given finite sets $\mathcal{I}$ and $\mathcal{J}$, a real matrix A of type $\mathcal{I} \times \mathcal{J}$ is defined by an array of real numbers whose row indices and column indices run in $\mathcal{I}$ and $\mathcal{J}$ respectively. We let $\mathbf{R}^{\mathcal{I}\times\mathcal{J}}$ denote the set of all these matrices. The term in row i and column j is denoted $A_{i,j}$. In Section 2.4 four norms $||A||_2$, $N_\infty(A)$, $|||A|||_\infty$, and $||A||_a$ of the matrix A will be defined.

Random functions or permutations will be considered. They will be represented by random variables, *e.g.* F or C. Section 2.1 defines the matrix $[F]^d$ or $[C]^d$ for any positive integer d. Random functions or permutations with "ideal" distributions will be denoted with a star superscript as F^* or C^*.

2 Decorrelation

2.1 Block Ciphers, Random Functions, Distribution Matrices

In what follows, we consider ciphers as random permutations C on a message-block space $\mathcal{M}$. Since we are considering block ciphers, and for simplicity reasons, messages are considered as elements of $\mathcal{M}$ which is assumed to be a finite set. In most of practical cases, we have $\mathcal{M} = \{0,1\}^m$. We emphasize on C being a *random* permutation. Here the randomness comes from the random choice of the secret key. In particular, for any (fixed) permutation c over $\mathcal{M}$, there is a probability $\Pr[C = c]$ that the C instance is equal to c.

Definition 1. *Given a random function F from a given set $\mathcal{M}_1$ to a given set $\mathcal{M}_2$ and an integer d, we define the* d-wise distribution matrix *$[F]^d$ of F as a $\mathcal{M}_1^d \times \mathcal{M}_2^d$-matrix where the (x,y)-entry of $[F]^d$ corresponding to the multi-points $x = (x_1,\ldots,x_d) \in \mathcal{M}_1^d$ and $y = (y_1,\ldots,y_d) \in \mathcal{M}_2^d$ is defined as the probability that we simultaneously have $F(x_i) = y_i$ for $i = 1,\ldots,d$. We denote it $[F]^d_{x,y} = \Pr[x \overset{F}{\mapsto} y]$.*

Basically, each row of the d-wise distribution matrix corresponds to the distribution of the d-tuple $(F(x_1),\ldots,F(x_d))$ where $(x_1,\ldots,x_d)$ corresponds to the index of the row. Intuitively, every experiment (or attack) on C with d samples will provide some information on some simultaneous equations $C(x_i) = y_i$. The experiment probability will thus correspond to a cell in the $[C]^d$ matrix.

2.2 Perfect Decorrelation

The d-wise distribution matrix of a random function intuitively defines its *d-wise decorrelation*. There is no precise definition of decorrelation, only ways to *compare* some, and models for *perfect decorrelation*. Two random functions have the same d-wise decorrelation if, and only if their d-wise distribution matrices are equal.

A random function (or a random permutation) will be compared with an ideal version of it which will have to be specified. Then, we will be able to compare the decorrelations of the function (or permutation) with its ideal version. For example, a block cipher C over $\mathcal{M}$ is compared with the ideal block cipher C^* over $\mathcal{M}$ which is defined to be a random permutation over $\mathcal{M}$ with uniform distribution. Note that for $\mathcal{M} = \{0,1\}^m$, we need $\log_2(2^m!) \approx m2^m$ bits in order to specify fully an instance of C^*, which is enormous. If C and C^* have the same decorrelation to the order d, we say that the d-wise decorrelation of the cipher C is *perfect*.

Similarly a random function F from $\mathcal{M}_1$ to $\mathcal{M}_2$ is compared with a uniformly distributed random function F^* from $\mathcal{M}_1$ to $\mathcal{M}_2$. We say that the d-wise decorrelation of the random function F is *perfect* if F and F^* have the same d-wise decorrelation.

Let F be a random function from $\mathcal{M}_1$ to $\mathcal{M}_2$. Saying that the *function* F has a perfect 1-wise decorrelation means that for any x_1 the distribution of $F(x_1)$ is uniform.

Saying that the *function* F has a perfect 2-wise decorrelation means that for any $x_1 \neq x_2$ the random variables $F(x_1)$ and $F(x_2)$ are uniformly distributed and independent. This is exactly the notion of strongly universal$_2$ function as defined by Carter and Wegman [73].

Saying that a *cipher* C on $\mathcal{M}$ has a perfect 2-wise decorrelation means that for any $x_1 \neq x_2$, the random variable $(C(x_1), C(x_2))$ is uniformly distributed among all the (y_1, y_2) pairs such that $y_1 \neq y_2$. This is exactly the notion of pairwise independent permutation as defined by Carter and Wegman [73].

2.3 Decorrelation Distance

The previous section provides a *qualitative* way to compare decorrelations. Here we introduce a *quantitative* way to do the same.

Definition 2. *Given two random functions F and G from a given set $\mathcal{M}_1$ to a given set $\mathcal{M}_2$, an integer d and a distance D over the matrix space $\mathbf{R}^{\mathcal{M}_1^d \times \mathcal{M}_2^d}$, we call $D([F]^d, [G]^d)$ the* d-wise decorrelation D-distance *between F and G. When G is the ideal version of F and is clear from the context we call $D([F]^d, [G]^d)$ the* d-wise decorrelation D-bias *of F.*

A decorrelation distance of zero means that for any multi-point $x = (x_1, \ldots, x_d)$ the multi-points $(F(x_1), \ldots, F(x_d))$ and $(G(x_1), \ldots, G(x_d))$ have the same distribution, so that F and G have the same *decorrelation*.

2.4 Classical Distances

For the purpose of our treatment, we define the L_2 norm, the infinity weighted pseudo-norm N_∞, the L_∞-associated matrix norm $|||.|||_\infty$, and the $||.||_a$-norm[4] on $\mathbf{R}^{\mathcal{M}_1^d \times \mathcal{M}_2^d}$ by:

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

$$N_\infty(A) = \max_{x,y} \frac{|A_{x,y}|}{\Pr[x \overset{C^*}{\mapsto} y]} \tag{2}$$

$$|||A|||_\infty = \max_x \sum_y |A_{x,y}| \tag{3}$$

$$||A||_a = \max_{x_1} \sum_{y_1} \ldots \max_{x_d} \sum_{y_d} |A_{x,y}| \tag{4}$$

where C^* is the Perfect Cipher, $x = (x_1, \ldots, x_d) \in \mathcal{M}_1^d$, and $y = (y_1, \ldots, y_d) \in \mathcal{M}_2^d$. For Equation (2) we use the convention that $0/0 = 0$ and $c/0$ is undefined for $c \neq 0$. Thus the N_∞ is not always defined. We can check that it is always defined when $A = [C_1]^d - [C_2]^d$ for any random permutations C_1 and C_2. Hence N_∞ still defines a distance in order to compare decorrelation of permutations.

We recall properties of matrix norms. First, $||.||_2$, $|||.|||_\infty$, and $||.||_a$ are norms, which means that

1. $||A|| = 0$ if and only if A is the identically zero matrix,
2. $||u.A|| = |u|.||A||$ for any real number u,
3. $||A + B|| \leq ||A|| + ||B||$.

The latter property is the "triangular inequality". We easily check these properties for $||.||_2$, $|||.|||_\infty$, and $||.||_a$. In addition, these are *matrix* norms, which means that we have the extra property (called "multiplicativity")

4. $||A \times B|| \leq ||A||.||B||$ whenever we can make the matrix product $A \times B$.

This property is quite well known for $||.||_2$ and $|||.|||_\infty$.[5] We prove it for the $||.||_a$ norm.

[4] This norm was first introduced in [69].

[5] It comes from the Cauchy-Schwarz property for $||.||_2$ and from the link $|||A||| = \max_{v \neq 0} \frac{||Av||}{||v||}$ with a vector norm for $|||.|||_\infty$.

Lemma 3. *Given two matrices $A \in \mathbf{R}^{\mathcal{M}_1^d \times \mathcal{M}_2^d}$ and $B \in \mathbf{R}^{\mathcal{M}_2^d \times \mathcal{M}_3^d}$, we have $||A \times B||_a \leq ||A||_a.||B||_a$.*

Proof. We prove it by induction on d. For $d = 1$ this is simply the result on the $|||.|||_\infty$ norm.

Given $x_1 \in \mathcal{M}_1$ and $y_1 \in \mathcal{M}_2$, we define $\pi_{x_1,y_1}(A) \in \mathbf{R}^{\mathcal{M}_1^{d-1} \times \mathcal{M}_2^{d-1}}$ by

$$(\pi_{x_1,y_1}(A))_{(x_2,\ldots,x_d),(y_2,\ldots,y_d)} = A_{(x_1,\ldots,x_d),(y_1,\ldots,y_d)}.$$

We notice that

$$||A||_a = \max_{x_1} \sum_{y_1} ||\pi_{x_1,y_1}(A)||_a.$$

We similarly define $\pi_{y_1,z_1}(B)$ and $\pi_{x_1,z_1}(A \times B)$. We have similar observations for $||B||_a$ and $||A \times B||_a$. Obviously we have

$$\pi_{x_1,z_1}(A \times B) = \sum_{y_1} \pi_{x_1,y_1}(A) \times \pi_{y_1,z_1}(B).$$

Using the triangular inequality and the induction hypothesis, we have

$$||A \times B||_a \leq \max_{x_1} \sum_{z_1} \sum_{y_1} ||\pi_{x_1,y_1}(A)||_a.||\pi_{y_1,z_1}(B)||_a.$$

By considering matrices in $\mathbf{R}^{\mathcal{M}_1 \times \mathcal{M}_2}$ and $\mathbf{R}^{\mathcal{M}_2 \times \mathcal{M}_3}$ whose terms are the $||\pi_{x_1,y_1}(A)||_a$ and $||\pi_{y_1,z_1}(B)||_a$ values we notice that this expression is yet another $|||.|||_\infty$ norm of a matrix product. Hence

$$||A \times B||_a \leq \left(\max_{x_1} \sum_{y_1} ||\pi_{x_1,y_1}(A)||_a \right) . \left(\max_{y_1} \sum_{z_1} ||\pi_{y_1,z_1}(B)||_a \right)$$

which is nothing but $||A \times B||_a \leq ||A||_a.||B||_a$. □

We also recall properties of distances. A distance D is such that

1. $D(A, B) = 0$ if and only if $A = B$,
2. $D(A, B) = D(B, A)$,
3. $D(A, C) \leq D(A, B) + D(B, C)$.

Matrix norms define distances by $D(A, B) = ||A - B||$. We easily check that N_∞ defines a distance on distribution matrices of ciphers as well.

In [69] the $||.||_s$ norm is introduced. It is used in order to study super-pseudorandomness whereas $||.||_a$ is used in order to study randomness. For simplicity we omit it in the paper, but we put discussion of it in Appendix B.

2.5 Multiplicativity of Decorrelation Distances

Theorem 4. *Let $C_1, \ldots, C_r$ be independent ciphers over $\mathcal{M}$. We consider $C = C_r \circ \ldots \circ C_1$ the product cipher. We let C^* be the perfect cipher over $\mathcal{M}$. For the distance D defined by either $||.||_2$, $|||.|||_\infty$, $||.||_a$, or N_∞ we have*

$$D([C]^d, [C^*]^d) \leq \prod_{i=1}^{r} D([C_i]^d, [C^*]^d).$$

It will be shown that the distance D characterizes the weakness of a cipher. Hence this theorem means that the weakness is multiplicative in a product cipher. This property makes the decorrelation bias of ciphers a multiplicative combinatorial measurement for those distances. It is quite convenient to prove the amplification phenomenon in product ciphers.

Proof. By induction we only need to prove it for $r = 2$. Let C_1 and C_2 be two independent random permutations over $\mathcal{M}$. We notice that $[C_2 \circ C_1]^d = [C_1]^d \times [C_2]^d$.

For any i we notice that $C_i \circ C^*$, $C^* \circ C_i$, and C^* have the same distribution. Hence we have $[C_i \circ C^*]^d = [C^* \circ C_i]^d = [C^*]^d$.

From those observations we notice that

$$[C_2 \circ C_1]^d - [C^*]^d = ([C_1]^d - [C^*]^d) \times ([C_2]^d - [C^*]^d).$$

We recall that the $||.||_2$, $|||.|||_\infty$, and $||.||_a$ norms are matrix norms, *i.e.* $||A \times B|| \leq ||A||.||B||$. Therefore

$$||[C_2 \circ C_1]^d - [C^*]^d|| \leq ||[C_1]^d - [C^*]^d||.||[C_2]^d - [C^*]^d||$$

for those norms. We easily check that we have a similar property for the N_∞ distance. □

3 Decorrelation Modules

The aim of this section is to provide cheap and efficient decorrelated random functions or permutations. We call them NUT, for *n-Universal Transformations* in order to remind us of the Carter-Wegman notion of universal function and to emphasize their low cost.

3.1 NUT-0: Perfect 1-Wise Decorrelated Permutations over a Group

Perfect 1-wise decorrelation is easy to achieve with permutations when the message-block space $\mathcal{M}$ is given a group structure. We let $+$ denote the group law in $\mathcal{M}$. We can use $C(x) = x + K$ where K is a uniformly distributed random key on $\mathcal{M}$, which is exactly the Vernam Cipher [71]. This primitive plays an important role in the construction of block ciphers, *e.g.* in order to construct Markov ciphers (see Lai-Massey-Murphy [37]) or in the Nyberg-Knudsen construction [52].

3.2 NUT-I: Perfect Decorrelated Functions over a Finite Field

Perfect decorrelated *functions* are easy to construct when $\mathcal{M}$ is given a finite field structure. We can take $F(x) = K_1 + K_2 x + K_3 x^2 + \ldots + K_d x^{d-1}$ where $K = (K_1, \ldots, K_d)$ is a uniformly distributed random key on $\mathcal{M}^d$. This random function has perfect d-wise decorrelation due to the Lagrange interpolation principle. This builds perfect decorrelation *functions* to arbitrary orders. Perfect decorrelated *permutations* to arbitrary orders are much harder to construct.

3.3 NUT-II: Perfect Pairwise Decorrelated Permutations over a Finite Field

We can construct perfect pairwise decorrelated *ciphers* on a field structure $\mathcal{M}$ as well by $C(x) = K_1 + K_2.x$ where $K = (K_1, K_2)$ is uniform in $\mathcal{M} \times \mathcal{M}^*$.[6]

3.4 NUT-III: Modulo p -Based Pairwise Decorrelated Functions for the L_2 Norm

On the standard space $\mathcal{M} = \{0,1\}^m$, our previous construction requires implementing arithmetic on the finite field $\mathrm{GF}(2^m)$, which may lead to a poor encryption rate on software for large m. We can take advantage of built-in integer multiplication by approximating the previous construction. The decorrelation is no longer perfect though.

Theorem 5. *Let $F(x) = K_1 + K_2.x \bmod p$ for $p = (1-\delta)2^m$ prime and K_1, K_2 independent uniformly distributed random variables in $\mathcal{M} = \{0, \ldots, 2^m - 1\}$. We assume that $1/14 \geq \delta \geq 0$. F is a random function from $\mathcal{M}$ to $\mathcal{M}$. Let F^* be a uniformly distributed random function from $\mathcal{M}$ to $\mathcal{M}$. We have $|||[F]^2 - [F^*]^2||_2 \leq 2\sqrt{2\delta}$.*

The proof is given in [65].

Note that the pairwise $|||.|||_\infty$-decorrelation of this primitive is pretty bad since the distribution of $(F(x_1), F(x_2))$ is odd when $x_1 = x_2 + p$: we always have $F(x_1) = F(x_2)$. We can however use this primitive in order to amplify the decorrelation of random cipher in the sense of the L_2 norm and still obtain provable security bounds.

[6] Here $\mathcal{M}^*$ denotes the set of all non-zero field elements.

3.5 NUT-IV: Modulo p -Based Decorrelated Functions for the $||.||_a$ Norm

Here instead of taking p smaller than 2^m as in the previous construction, we take p larger than 2^m.

Theorem 6. *Let $F(x) = (K_1 + K_2 x + K_3 x^2 + \ldots + K_d x^{d-1} \bmod p) \bmod 2^m$ for $p = (1+\delta)2^m$ prime, $\delta \geq 0$, and $K_1, \ldots, K_d$ independent uniformly distributed random variables in $\mathcal{M} = \{0, \ldots, 2^m - 1\}$. F is a random function from $\mathcal{M}$ to $\mathcal{M}$. Let F^* be a uniformly distributed random function from $\mathcal{M}$ to $\mathcal{M}$. We have $||[F]^d - [F^*]^d||_a \leq 2((1+\delta)^d - 1)$.*

This theorem generalizes to any finite field $\mathrm{GF}(p)$ with p not necessarily prime, and any $\mathcal{M}$ when using any injective representation from $\mathcal{M}$ to $\mathrm{GF}(p)$ for x and the K_i's and using any surjective mapping from $\mathrm{GF}(p)$ to $\mathcal{M}$ instead of the modulo 2^m reduction. (See Theorem 7 of [69].)

The proof of this theorem requires materials from Section 5. We provide it in Appendix A.

Note that a similar construction has been previously used by Halevi and Krawczyk for authentication in the MMH algorithm [21].

3.6 NUT-V: 3-Wise Decorrelated Permutations over a Finite Field

A similar way to construct (almost) perfect 3-wise decorrelated *permutation* on a field structure $\mathcal{M}$ is by $C(x) = a + b/(x+c)$ where $K = (a, b, c)$ with $b \neq 0$. (By convention we set $1/0 = 0$.) We can prove that $||[C]^3 - [C^*]^3||_a \leq \frac{6}{q}$ where q is the field cardinality, with the same techniques as for Theorem 6. (See [4].)

4 Links to the Shannon Secrecy Theory

4.1 Perfect Secrecy and Decorrelation

Shannon defines security by the notion of perfect secrecy [59]. Perfect secrecy is a property of a cipher and a random plaintext source. We say that C provides perfect secrecy for a given distribution of X if $H(X/C(X)) = H(X)$ where H denotes the Shannon entropy,[7] or equivalently if X and $C(X)$ have independent distributions. We can also consider ciphers C which provide perfect secrecy for *any* distribution of X. This means that the distribution of $C(x)$ does not depend on x.

In Shannon's formalism, X denotes the full stream of plaintext that we want to encrypt whereas X denotes one plaintext block in our approach. We usually bring the two approaches together by considering C as a one-time cipher which encrypts a single plaintext (big) block.

Obviously, if C is a perfect 1-wise decorrelated cipher, then C provides perfect secrecy for any plaintext source since $C(x)$ is uniformly distributed for any x. The Vernam cipher (see Section 3.1) is an example.

We easily capture the notion of a chosen plaintext or ciphertext attack with the following generalization.

Theorem 7. *Let C be a cipher with a perfect d-wise decorrelation. For any $x_1, \ldots, x_{d-1}$, if X is a random variable such that $X \neq x_i$, then*

$$H(X/C(x_1), \ldots, C(x_{d-1}), C(X)) = H(X).$$

This means that if an adversary knows $d-1$ pairs $(x_i, C(x_i))$ (either by a chosen plaintext or ciphertext attack), for any y_d which is different from all $C(x_i)$'s, his knowledge of $C^{-1}(y_d)$ is nothing more than knowing that it is different from all x_i's.

Proof. From the definitions, straightforward computations show that for any random variable X we have

$$H(X/C(x_1), \ldots, C(x_{d-1}), C(X)) = H(X) + p \log_2 p,$$

where $p = \Pr[X \neq x_i; i = 1, \ldots, d-1]$. Since we know that our X is different from all x_i we have $p = 0$. □

[7] We recall that by definition $H(X) = -\sum_x \Pr[X = x] \log_2 \Pr[X = x]$ with the convention that $0 \log_2 0 = 0$, and that $H(X/Y) = H(X,Y) - H(Y)$ where $H(X,Y)$ is the entropy of the joint variable $Z = (X,Y)$.

4.2 Key Length Lower Bound

The Shannon approach enables proving a lower bound on the private key length for ciphers which achieve perfect secrecy for any plaintext source. More precisely, the Shannon Theorem proves that if C provides perfect secrecy for any distribution of the plaintexts over $\mathcal{M}$, then $H(C) \geq \log_2 \#\mathcal{M}$. This means that the key parameter in C needs to have at least $\log_2 \#\mathcal{M}$ bits to be at least as long as the plaintext. The Vernam cipher achieves the equality case.

We can prove a similar result for perfect decorrelation.

Theorem 8. *If F is a random function from $\mathcal{M}_1$ to $\mathcal{M}_2$ with perfect d-wise decorrelation (for $d \leq \#\mathcal{M}_1$), then $H(F) \geq d.\log_2 \#\mathcal{M}_2$. If C is a cipher over $\mathcal{M}$ with perfect d-wise decorrelation, then $H(C) \geq d.\log_2 \#\mathcal{M} - d^2/\#\mathcal{M} - o(d^2/\#\mathcal{M})$ as $d/\#\mathcal{M}$ decreases toward zero.*

Proof. Let $x_1, \ldots, x_d$ be d pairwise different points in $\mathcal{M}_1$. Since F has a perfect d-wise decorrelation, then $Y = (F(x_1), \ldots, F(x_d))$ is uniformly distributed in $\mathcal{M}_2^d$. Hence we have $H(Y) = d.\log_2 \#\mathcal{M}_2$. Due to the property of joint entropy we have $H(F,Y) \geq H(Y)$. However, F fully determines Y hence $H(F,Y) = H(F)$, thus $H(F) \geq d.\log_2 \#\mathcal{M}_2$.

For ciphers we do the same. Y happens to be uniformly distributed among all multi-points with pairwise different entries. We have $N = \#\mathcal{M}(\#\mathcal{M}-1)\ldots(\#\mathcal{M}-d+1)$ values which is greater than $(\#\mathcal{M}-d)^d$. We obtain $H(C) \geq \log_2 N \geq d.\log_2(\#\mathcal{M}-d)$. Hence

$$H(C) \geq d.\log_2 \#\mathcal{M} + d.\log_2\left(1 - \frac{d}{\#\mathcal{M}}\right).$$

The result then comes from $\log(1-\varepsilon) = -\varepsilon - o(\varepsilon)$. □

5 Security against Distinguishers with Limited Oracle Accesses

In the Luby-Rackoff model [40], an attacker is an infinitely powerful Turing machine $\mathcal{A}^{\mathcal{O}}$ which has access to an oracle $\mathcal{O}$. Her aim is to distinguish a cipher C from the Perfect Cipher C^* by querying the oracle with a limited number d of inputs. The oracle $\mathcal{O}$ implements either C or C^*. The attacker must finally answer 0 ("reject") or 1 ("accept"). We measure the ability to distinguish C from C^* by the advantage $\mathrm{Adv}_{\mathcal{A}} = |p - p^*|$ where p (resp. p^*) is the probability of accepting C (resp C^*), *i.e.* the probability of answering 1 if $\mathcal{O}$ implements C (resp. C^*).

5.1 d-Limited Distinguishers and N_∞-Decorrelation

Since we put no upper bound on the computational capability of the distinguisher (the only limitation is on the number of queries to the oracle), we can assume without loss of generality that the best one is fully deterministic. Hence it can be defined by functions $f_1, \ldots, f_d$ and an acceptance set $\mathcal{A}$ as illustrated in Fig. 1.

Parameters: functions $f_1, \ldots, f_d$, a set $\mathcal{A}$
Oracle: a permutation c
1: select a fixed message $X_1 = f_1()$ and get $Y_1 = c(X_1)$
2: calculate a message $X_2 = f_2(Y_1)$ and get $Y_2 = c(X_2)$
3: ...
4: calculate a message $X_d = f_d(Y_1, \ldots, Y_{d-1})$ and get $Y_d = c(X_d)$
5: if $Y = (Y_1, \ldots, Y_d) \in \mathcal{A}$, output 1, otherwise output 0

Fig. 1. A General d-Limited Distinguisher.

Theorem 9. *Let d be an integer, and let C be a cipher. For any distinguisher between C and the perfect cipher C^* which is limited to d queries (as depicted in Fig. 1), we have*

$$\mathrm{Adv}_{\mathrm{Fig.\ 1}} \leq N_\infty([C]^d - [C^*]^d)$$

where the N_∞ norm is defined by Equation (2).

In particular, we have unconditional security when the decorrelation is perfect and we still have a proven quantified security when the decorrelation is small.

Proof. Obviously we have

$$p = \sum_{y \in \mathcal{A}} \Pr[x \overset{C}{\mapsto} y]$$

where $y = (y_1, \ldots, y_d)$ and $x = (x_1, \ldots, x_d)$ with $x_i = f_i(y_1, \ldots, y_{i-1})$ in the sum. Since $\Pr[x \overset{C}{\mapsto} y] \leq (1 + \varepsilon) \Pr[x \overset{C^*}{\mapsto} y]$ with $\varepsilon = N_\infty([C]^d - [C^*]^d)$, we have

$$p \leq (1 + \varepsilon) \sum_{y \in \mathcal{A}} \Pr[x \overset{C^*}{\mapsto} y] = (1 + \varepsilon) p^*$$

so we have $p - p^* \leq \varepsilon$ for any attacker. We can apply this result to the attacker which produces the opposite output to show that $|p - p^*| \leq \varepsilon$. □

5.2 Best Non-Adaptive Distinguisher and $|||.|||_\infty$-Decorrelation

Here is a more precise theorem in the non-adaptive case. We call a distinguisher "non adaptive" if no X_i queried to the oracle depends on some previous answers Y_j (see Fig. 2).

Theorem 10. *Let d be an integer and C be a cipher. The best d-limited non-adaptive distinguisher (as depicted in Fig. 2) for C is such that*

$$\mathrm{Adv}_{\text{Fig. 2}} = \frac{1}{2} |||[C]^d - [C^*]^d|||_\infty$$

where the $|||.|||_\infty$ norm is defined by Equation (3) and C^* is the perfect cipher.

Parameters: values $X_1, \ldots, X_d$, a set $\mathcal{A}$
Oracle: a permutation c
1: select some fixed messages $X = (X_1, \ldots, X_d)$
2: get $Y = (c(X_1), \ldots, c(X_d))$
3: if $Y \in \mathcal{A}$, output 1, otherwise, output 0

Fig. 2. A d-Limited Non-Adaptive Distinguisher.

Proof. The best attack is fully characterized by $x = (x_1, \ldots, x_d)$ and $\mathcal{A}$. With the notations of Theorem 9, we have

$$p = \sum_y 1_{y \in \mathcal{A}} \Pr\left[x \overset{C}{\mapsto} y\right],$$

thus, we have

$$\mathrm{Adv} = \left| \sum_y 1_{y \in \mathcal{A}} \left(\Pr\left[x \overset{C}{\mapsto} y\right] - \Pr\left[x \overset{C^*}{\mapsto} y\right] \right) \right|.$$

Looking for the best distinguisher thus consists of maximizing this expression over all possible choices for x and $\mathcal{A}$. We can easily see that this maximum is obtained when $\mathcal{A}$ consists of all y's such that $\Pr\left[x \overset{C}{\mapsto} y\right] - \Pr\left[x \overset{C^*}{\mapsto} y\right]$ have the same sign. Since the full sum for all y is zero, the sum of all positive terms is equal to the sum of negative terms, hence half of the sum of all absolute values. Hence for the best distinguisher

$$\mathrm{Adv} = \max_x \frac{1}{2} \sum_y \left| \Pr\left[x \overset{C}{\mapsto} y\right] - \Pr\left[x \overset{C^*}{\mapsto} y\right] \right|.$$

We can recognize here the $|||.|||_\infty$ distance between $[C]^d$ and $[C^*]^d$. □

5.3 Best Adaptive Distinguisher and $||.||_a$-Decorrelation

We can extend Theorem 10 and get a more precise result than Theorem 9.

Theorem 11. *Let d be an integer and let C be a cipher. The best d-limited distinguisher (as depicted in Fig. 1) for C is such that*

$$\mathrm{Adv}_{\mathrm{Fig.\ 1}} = \frac{1}{2}||[C]^d - [C^*]^d||_a$$

where the $||.||_a$ norm is defined by Equation (4) and C^* is the perfect cipher.

This motivates the introduction of the $||.||_a$ norm.

Proof. The best attack is fully characterized by $f_1, \ldots, f_d$ and $\mathcal{A}$. As for Theorem 10, for the optimal distinguisher we have

$$\mathrm{Adv} = \frac{1}{2}\sum_y \left|[C]^d_{x,y} - [C^*]^d_{x,y}\right|$$

with $x_i = f_i(y_1, \ldots, y_{i-1})$. By maximizing this expression in terms of $f_1, \ldots, f_d$ we obtain $\mathrm{Adv} = \frac{1}{2}||[C]^d - [C^*]^d||_a$. □

6 Resistance against Iterated Attacks

Since resisting against general d-limited distinguishers for d large costs too many bits of randomness in the private keys (as Theorem 8 says), we can wonder how useful this theory is for practical ciphers. In this section we investigate some particular class of distinguishers which capture many of the existing attack methods. We show that decorrelation with low degree is enough to resist them.

6.1 Differential Cryptanalysis

In this section we assume that $\mathcal{M}$ is given a group structure of order M. (Typically we consider $\mathcal{M} = \{0,1\}^m$ and the XOR group law.) We study the security of pairwise decorrelated ciphers against basic differential cryptanalysis.

Let C be a cipher on $\mathcal{M}$ and let C^* be the Perfect Cipher.

Most differential cryptanalysis of r-round block ciphers based on the Biham and Shamir attack (see [9,10]) use a simple distinguisher between $r - i$ rounds (for $i = 1$, 2, or 3) of the cipher and the perfect cipher. This distinguisher uses a fixed pair $(a, b) \in \mathcal{M}^2$ with $a \neq 0$ and is depicted in Fig. 3.

Parameters: a complexity n, a characteristic (a, b)
Oracle: a permutation c
1: **for** i from 1 to n **do**
2: pick uniformly a random X and query for $c(X)$ and $c(X + a)$
3: if $c(X + a) = c(X) + b$, output 1 and stop
4: **end for**
5: output 0

Fig. 3. Differential Distinguisher.

We define

$$\mathrm{DP}^C(a, b) = \Pr_X[C(X + a) = C(X) + b],$$

where X has a uniform distribution. It is well known that differential cryptanalysis depends on this quantity (see, for instance, [50]). This quantity depends on the choice of the cipher (*i.e.* on the key). Here we focus on average complexities of attacks with no prior information on the key.[8] For this we concentrate on the average value $E\left(\mathrm{DP}^C(a, b)\right)$

[8] The problem of successful attacks for sets of *weak keys* is not our purpose here.

over the distribution of C. We first mention that $E\left(\mathrm{DP}^C(a,b)\right)$ has an interesting linear expression with respect to the pairwise distribution matrix of C. Namely, straightforward computation shows that

$$E(\mathrm{DP}^C(a,b)) = \frac{1}{M}\sum_{\substack{x_1,x_2\\y_1,y_2}} 1_{\substack{x_2=x_1+a\\y_2=y_1+b}} \Pr\left[(x_1,x_2) \overset{C}{\mapsto} (y_1,y_2)\right]. \tag{5}$$

Lemma 12. *For the distinguisher of Fig. 3 between C and the perfect cipher C^* over the group $\mathcal{M}$ of order M we have*

$$\mathrm{Adv}_{\text{Fig. 3}} \le n.\max\left(\frac{1}{M-1}, E\left(\mathrm{DP}^C(a,b)\right)\right).$$

Proof. It is straightforward to see that the probability p^c, for some fixed oracle c, that the attack accepts c is

$$p^c = 1 - (1 - \mathrm{DP}^c(a,b))^n,$$

which is less than $n.\mathrm{DP}^c(a,b)$. The probability that it accepts C is $p = E(p^C) \le n.E\left(\mathrm{DP}^C(a,b)\right)$. Since from Equation (5) we have $E\left(\mathrm{DP}^{C^*}(a,b)\right) = \frac{1}{M-1}$, we obtain the result. □

Theorem 13. *Let C be a cipher on a group $\mathcal{M}$ of order M and let C^* be the perfect cipher. For any basic differential distinguisher between C and C^* (depicted in Fig. 3) of complexity n, we have*

$$\mathrm{Adv}_{\text{Fig. 3}} \le \frac{n}{M-1} + \frac{n}{2}|||[C]^2 - [C^*]^2|||_\infty.$$

Note that this result holds for differential cryptanalysis with *any* group law and captures the notion of *multi-path differential*.

Proof. We first consider the distinguisher with $n = 1$. It is a non adaptive distinguisher limited to two queries. Due to Theorem 10, this is less than $\frac{1}{2}|||[C]^2 - [C^*]|||_\infty$. As in the proof of Lemma 12, we obtain that it is further equal to

$$\left|E\left(\mathrm{DP}^C(a,b)\right) - \frac{1}{M-1}\right|.$$

Thus

$$E\left(\mathrm{DP}^C(a,b)\right) \le \frac{1}{M-1} + \frac{1}{2}|||[C]^2 - [C^*]|||_\infty.$$

We conclude by using Lemma 12. □

So, if the pairwise decorrelation bias has the order of $1/M$, basic differential cryptanalysis cannot work against C unless its complexity reaches the order of magnitude of M.

6.2 Linear Cryptanalysis

Linear cryptanalysis has been invented by Matsui [41,42] based on the notion of statistical attacks which are due to Gilbert *et al.* [15,16,60]. As for differential cryptanalysis, we study here the underlying distinguisher against $r - i$ rounds for small i.

In this section we assume that $\mathcal{M} = \{0,1\}^m$. The inner dot product $a \cdot b$ in $\{0,1\}^m$ is the parity of the bitwise AND of a and b.

Let C be a cipher on $\mathcal{M}$ and let C^* be the Perfect Cipher.

As in Section 6.1, we similarly call the *basic linear distinguisher* the distinguisher characterized by a pair $(a,b) \in \mathcal{M}^2$ with $b \neq 0$ which is depicted in Fig. 4. We notice here that the attack depends on the way it accepts or rejects based on the final counter u value.

As pointed out by Chabaud and Vaudenay [12], linear cryptanalysis is based on the quantity

$$\mathrm{LP}^C(a,b) = \left(2\Pr_X[X \cdot a = C(X) \cdot b] - 1\right)^2.$$

(Here we use Matsui's notations taken from [43].) As for differential cryptanalysis, we focus on $E\left(\mathrm{LP}^C(a,b)\right)$, and there is a linear expression of this mean value in terms of the pairwise distribution matrix $[C]^2$ which comes from straightforward computations as shown by the following lemma.

Parameters: a complexity n, a characteristic (a, b), a set $\mathcal{A}$
Oracle: a permutation c
1: initialize the counter value u to zero
2: **for** i from 1 to n **do**
3: pick a random X with a uniform distribution and query for $c(X)$
4: if $X \cdot a = c(X) \cdot b$, increment the counter u
5: **end for**
6: if $u \in \mathcal{A}$, output 1, otherwise output 0

Fig. 4. Linear Distinguisher.

Lemma 14. *Given a random permutation C over $\{0,1\}^m$, for any a and b, we have*

$$E(\mathrm{LP}^C(a,b)) = 2^{-2m} \sum_{\substack{x_1,x_2\\y_1,y_2}} (-1)^{(x_1\oplus x_2)\cdot a+(y_1\oplus y_2)\cdot b} \Pr\left[(x_1,x_2) \overset{C}{\mapsto} (y_1,y_2)\right]$$
$$= 1 - 2^{2-2m} \sum_{\substack{x_1\neq x_2\\y_1\neq y_2}} 1_{\substack{x_1\cdot a=y_1\cdot b\\x_2\cdot a\neq y_2\cdot b}} \Pr\left[(x_1,x_2) \overset{C}{\mapsto} (y_1,y_2)\right].$$

If C has a uniform distribution, $a \neq 0$, and $b \neq 0$, we have $E(\mathrm{LP}^C(a,b)) = \frac{1}{2^m-1}$. Note that $E(\mathrm{LP}^C(0,b)) = 0$ for $b \neq 0$.

Proof. In order to prove it, we first notice that $2\Pr_X[X \cdot a = C(X) \cdot b] - 1 = E\left((-1)^{X\cdot a+C(X)\cdot b)}\right)$, and we express $\mathrm{LP}^C(a,b)$ as

$$\mathrm{LP}^C(a,b) = E\left((-1)^{(X_1\oplus X_2)\cdot a+(C(X_1)\oplus C(X_2))\cdot b}\right)$$

where X_1 and X_2 are independent uniformly distributed random variables. We have

$$E(\mathrm{LP}^C(a,b)) = 2^{-2m} \sum_{\substack{x_1,x_2\\y_1,y_2}} (-1)^{(x_1\oplus x_2)\cdot a+(y_1\oplus y_2)\cdot b} \Pr\left[(x_1,x_2) \overset{C}{\mapsto} (y_1,y_2)\right].$$

The contribution of terms for which $x_1 = x_2$ is equal to 2^{-m}. Considering that C is a permutation we can concentrate on $x_1 \neq x_2$ and $y_1 \neq y_2$. Then we split the remaining sum into four groups depending on the two bits $(x_1 \cdot a \oplus y_1 \cdot b, x_2 \cdot a \oplus y_2 \cdot b)$. Let Σ_{b_1,b_2} be the sum of all probabilities for which the two bits are (b_1, b_2), $x_1 \neq x_2$, and $y_1 \neq y_2$. We have

$$E(\mathrm{LP}^C(a,b)) = 2^{-m} + 2^{-2m}\Sigma_{0,0} - 2^{-2m}\Sigma_{0,1} - 2^{-2m}\Sigma_{1,0} + 2^{-2m}\Sigma_{1,1}.$$

Due to symmetry we have $\Sigma_{0,1} = \Sigma_{1,0}$. Furthermore, the sum of the four sums is $2^m(2^m - 1)$. Hence

$$E(\mathrm{LP}^C(a,b)) = 2^{-m} + 2^{-2m} \times 2^m(2^m - 1) - 4 \times 2^{-2m}\Sigma_{0,1}$$

which leads to our second result. Computations when C is uniformly distributed are straightforward. □

Lemma 15. *For the distinguisher of Fig. 4 we let p^c be the probability that the output is 1 given an oracle c. We let p_0 be the probability that it outputs 1 when the counter is incremented with probability $\frac{1}{2}$ in each iteration instead of querying the oracle. We have*

$$|p^c - p_0| \leq 2\sqrt{n.\mathrm{LP}^c(a,b)}.$$

Furthermore, the maximum for $|p^c - p_0|$ is asymptotically equivalent to $\frac{1}{\sqrt{2\pi}}\sqrt{n.\mathrm{LP}^c(a,b)}$ when n increases and $\mathrm{LP}^c(a,b) = o(\frac{1}{n})$.

Proof. We first express the probability p^c that the distinguisher accepts c. Let N_i be the random variable defined as being 1 or 0 depending on whether or not we have $X \cdot a = c(X) \cdot b$ in the ith iteration. All N_i's are independent and

with the same 0-or-1 distribution. Let z be the probability that $N_i = 1$. We also define $\theta = 2z - 1 = \sqrt{\mathrm{LP}^c(a,b)}$. We thus want to prove that $|p^c - p_0| \leq 2\theta\sqrt{n}$. We have

$$p^c = \sum_{u \in \mathcal{A}} \binom{n}{u} z^u (1-z)^{n-u}$$

thus

$$p^c - p_0 = \sum_{u \in \mathcal{A}} \binom{n}{u} \left(z^u (1-z)^{n-u} - \frac{1}{2^n} \right).$$

We would like to upper bound $|p^c - p_0|$ over all possible $\mathcal{A}$ depending on z. Since z and $1-z$ play a symmetric role we assume without loss of generality that $z \geq \frac{1}{2}$. For $z = \frac{1}{2}$, the result is trivially true, so from now on we assume that $z > \frac{1}{2}$. Since $z^u(1-z)^{n-u}$ is an increasing function in terms of u we have

$$\max_{\mathcal{A}} |p^c - p_0| = \sum_{u=k}^{n} \binom{n}{u} \left(z^u (1-z)^{n-u} - \frac{1}{2^n} \right)$$

where k is the least integer u such that the difference in parentheses is nonnegative, *i.e.*

$$k = 1 + \left\lfloor n \frac{\log \frac{1}{2} - \log(1-z)}{\log z - \log(1-z)} \right\rfloor.$$

Replacing u by $\frac{n}{2}$ in the same expression in parentheses we obtain a negative difference. Hence $k \geq \frac{n+1}{2}$. Similarly, replacing u by $n.z$, the expression in parentheses turns out to be an increasing function in terms of z which is 0 for $z = \frac{1}{2}$. Since $z > \frac{1}{2}$ we obtain that $k \leq \lceil n.z \rceil$. Therefore $\frac{n-1}{2} \leq k - 1 \leq (n-1)z + z$.

If $n = 1$, we have $k = 1$ thus $\max_{\mathcal{A}} |p^c - p_0| = z - \frac{1}{2}$ so the result holds. If $n = 2$, we have $k \geq \frac{3}{2}$ thus $k = 2$ and

$$\max_{\mathcal{A}} |p^c - p_0| = \left(z - \frac{1}{2} \right) \left(z + \frac{1}{2} \right) \leq \frac{3}{2} \left(z - \frac{1}{2} \right)$$

so the result holds as well. We now concentrate on $n \geq 3$.

We use the following identity taken from [58].[9]

$$\sum_{u=k}^{n} \binom{n}{u} z^u (1-z)^{n-u} = k \binom{n}{k} \int_0^z t^{k-1} (1-t)^{n-k} dt. \tag{6}$$

We obtain

$$\max_{\mathcal{A}} |p^c - p_0| = k \binom{n}{k} \int_{\frac{1}{2}}^z t^{k-1} (1-t)^{n-k} dt \tag{7}$$

thus

$$|p^c - p_0| \leq k \binom{n}{k} \left(z - \frac{1}{2} \right) \max_{t \in [0,1]} \left(t^{k-1} (1-t)^{n-k} \right).$$

The maximum is obtained for $t = \frac{k-1}{n-1}$ hence

$$|p^c - p_0| \leq k \binom{n}{k} \left(z - \frac{1}{2} \right) \frac{(k-1)^{k-1} (n-k)^{n-k}}{(n-1)^{n-1}}.$$

Let $x = 2\frac{k-1}{n-1} - 1$. We have $k - 1 = \frac{n-1}{2}(1+x)$ and $n - k = \frac{n-1}{2}(1-x)$. We have $0 \leq x \leq 1$ and

$$|p^c - p_0| \leq k \binom{n}{k} \left(z - \frac{1}{2} \right) \frac{1}{2^{n-1}} \left((1+x)^{1+x} (1-x)^{1-x} \right)^{\frac{n-1}{2}}.$$

By using $k \binom{n}{k} = n \binom{n-1}{k-1}$ and the Stirling approximation we obtain that this bound is asymptotically equal to $\frac{\theta\sqrt{n}}{\sqrt{2\pi}}$ so the bound we want to prove is not so loose.

[9] We can easily prove it by derivating it in terms of z.

We can easily prove that $(1+x)^{1+x}(1-x)^{1-x} \leq 2^{2x^2}$. Hence

$$|p^c - p_0| \leq k\binom{n}{k}\left(z - \frac{1}{2}\right)\frac{1}{2^{n-1}}2^{(n-1)x^2}.$$

Since $k - 1 \leq (n-1)z + z$ we have $x \leq \theta + \frac{\theta}{n-1} + \frac{1}{n-1} = \frac{n\theta+1}{n-1}$. Thus

$$|p^c - p_0| \leq \theta \times \left[k\binom{n}{k}\frac{1}{2^n}\right] \times 2^{\frac{(n\theta+1)^2}{n-1}}.$$

For $n = 3$ we have $k\binom{n}{k}\frac{1}{2^n} \leq \frac{3}{4}$ thus

$$|p^c - p_0| \leq 2\theta\sqrt{n} \times \frac{1}{2\sqrt{3}} \times \frac{3}{4} \times 2^{\frac{(3\theta+1)^2}{n-1}}.$$

For $\theta \leq \frac{1}{2\sqrt{3}}$ we obtain $|p^c - p_0| \leq 2\theta\sqrt{n}$ and this remains true even for $\theta > \frac{1}{2\sqrt{3}}$. We now concentrate on $n \geq 4$.

The $\binom{n}{k}$ term is upper bounded by $\binom{n}{r}$ with $r = \lceil \frac{n}{2} \rceil$. Furthermore, we have

$$\binom{n}{r}\frac{1}{2^n} \leq \prod_{i=1}^{r}\left(1 - \frac{1}{2i}\right)$$

with equality when n is even. Then

$$\begin{aligned}
\log\left(\binom{n}{r}\frac{1}{2^n}\right) &\leq \sum_{i=1}^{r}\log\left(1 - \frac{1}{2i}\right) \\
&\leq -\frac{1}{2}\sum_{i=1}^{r}\frac{1}{i} \\
&\leq -\frac{1}{2}\int_1^{r+1}\frac{dt}{t} \\
&\leq -\frac{1}{2}\log(r+1) \\
&\leq -\frac{1}{2}\log\frac{n}{2} + 1
\end{aligned}$$

therefore

$$\binom{n}{k}\frac{1}{2^n} \leq \sqrt{\frac{2}{n+2}}.$$

Now we have

$$k\binom{n}{k}\frac{1}{2^n} = n\binom{n-1}{k-1}\frac{1}{2^n} \leq \frac{n}{2}\sqrt{\frac{2}{n+1}} \leq \sqrt{\frac{n}{2}}.$$

We deduce

$$|p^c - p_0| \leq 2\theta\sqrt{n} \times 2^{\frac{(n\theta+1)^2}{n-1} - \frac{3}{2}}.$$

When $\theta\sqrt{n} < \frac{1}{2}$ and $n \geq 4$ we have $\frac{(n\theta+1)^2}{n-1} - \frac{3}{2} < 0$ so we obtain $|p^c - p_0| \leq 2\theta\sqrt{n}$. When $\theta\sqrt{n} \geq \frac{1}{2}$ this also holds since the right-hand side of the inequality is greater than 1 and the left-hand side is a difference between two probabilities. This proves the upper bound.

By definition of k we have $z^{k-1}(1-z)^{n-k} \geq \frac{1}{z2^n}$, so we have $t^{k-1}(1-t)^{n-k} \geq \frac{1}{2^{n-1}(1+\theta)}$ for any $t \in [\frac{1}{2}, z]$. From Equation (7) we deduce

$$\max_{\mathcal{A}}|p^c - p_0| \geq \frac{\theta}{1+\theta} \times \left[k\binom{n}{k}\frac{1}{2^n}\right].$$

If $\theta = o(\frac{1}{\sqrt{n}})$, we have $k = \frac{n}{2} + o(\sqrt{n})$ thus $\binom{n}{k} \sim \frac{2^{n+1}}{\sqrt{2\pi n}}$ from the Stirling Formula. Hence $\max_{\mathcal{A}}|p^c - p_0|$ is asymptotically larger than $\frac{\theta\sqrt{n}}{\sqrt{2\pi}}$. Since it is also smaller, this is indeed an equivalent. □

Lemma 16. *Let C be a cipher on $\mathcal{M} = \{0,1\}^m$. For any linear distinguisher (as depicted in Fig. 4) between C and the ideal cipher C^* we have*

$$\mathrm{Adv}_{\text{Fig. 4}} \leq 3\sqrt[3]{n.E\left(\mathrm{LP}^C(a,b)\right)} + 3\sqrt[3]{\frac{n}{2^m-1}}.$$

Proof. We first notice that the advantage is zero when $a = 0$ or $b = 0$, so the bound holds. Let us now assume that $a \neq 0$ and $b \neq 0$. We now take a random permutation C with the corresponding Z and p^C as in the previous lemma. Let $\delta = E((2Z-1)^2)$. (Note that $\delta = E(\mathrm{LC}^C(a,b))$.) When $|2Z-1| \leq \alpha$ we have

$$|p^C - p_0| \leq 2 \times \alpha\sqrt{n}.$$

Since $(2Z-1)^2$ is positive, the probability that $|2Z-1|$ is greater than α is less than $\frac{\delta}{\alpha^2}$. Hence

$$|p - p_0| \leq 2 \times \alpha\sqrt{n} + \frac{\delta}{\alpha^2}$$

for any α.

We now fix $\alpha = \left(\frac{\delta}{\sqrt{n}}\right)^{\frac{1}{3}}$. We obtain $|p - p_0| \leq 3 \times \sqrt[3]{\delta n}$.

We recall that $\delta = E\left(\mathrm{LP}^C(a,b)\right)$. Since $a \neq 0$ and $b \neq 0$, we note that $E\left(\mathrm{LP}^{C^*}(a,b)\right) = \frac{1}{2^m-1}$ from Lemma 14 so we can have

$$|p^* - p_0| \leq 3\sqrt[3]{\frac{n}{2^m-1}}.$$

We finally use that $|p - p^*| \leq |p - p_0| + |p^* - p_0|$. □

Theorem 17. *Let C be a cipher on $\mathcal{M} = \{0,1\}^m$. For any linear distinguisher (as depicted in Fig. 4) between C and the ideal cipher C^* of complexity n we have*

$$\mathrm{Adv}_{\text{Fig. 4}} \leq 3\sqrt[3]{n.|||[C]^2 - [C^*]^2|||_\infty + \frac{n}{2^m-1}} + 3\sqrt[3]{\frac{n}{2^m-1}}.$$

Proof. Actually we have $E\left(\mathrm{LP}^{C^*}(a,b)\right) = \frac{1}{2^m-1}$ and

$$\left|E\left(\mathrm{LP}^C(a,b)\right) - \frac{1}{2^m-1}\right| \leq |||[C]^2 - [C^*]^2|||_\infty$$

from Lemma 14. We conclude by using the previous lemma. □

So, if the pairwise decorrelation bias has the order of 2^{-m}, linear distinguishers do not work against C unless its complexity reaches the order of magnitude of 2^m.

6.3 Non-Adaptive Iterated Attacks of Order d

Theorems 13 and 17 suggest that we try to generalize them to distinguishers in the model depicted in Fig. 5 as proposed in [67]. In this model, we iterate a d-limited non-adaptive distinguisher $\mathcal{T}$. We assume that this distinguisher obtains a sample (X,Y) with $X = (X_1,\ldots,X_d)$ and $Y = (Y_1,\ldots,Y_d)$ such that $y_i = c(X_i)$ for a given distribution of X. Thus, we can think of a known plaintext attack where X has a plaintext source distribution (*e.g.* a uniform distribution) or of a chosen plaintext attack where X has a given distribution (*e.g.* in differential cryptanalysis, $X = (X_1, X_1 + a)$ where X_1 has a uniform distribution). The result of the attack depends on the result of all iterated ones in a way characterized by a set $\mathcal{A}$. For instance, if $\mathcal{A} = \{0,1\}^n \setminus \{(0,\ldots,0)\}$ we can define the differential cryptanalysis (thus of order $d = 2$). If $\mathcal{A}$ is the set of all $(t_1,\ldots,t_n)$ with an acceptable sum we can define the linear cryptanalysis (of order $d = 1$).

One may believe that a cipher is resistant to this model of distinguisher once it has a small d-wise decorrelation bias. This is wrong as the following example shows. Let C be a cipher with a perfect d-wise decorrelation. We assume that an instance c of C is totally defined by d points (x_i, y_i) so that C is uniformly distributed in a set of

Parameters: a complexity n, a distribution on X, a test $\mathcal{T}$, a set $\mathcal{A}$
Oracle: a permutation c
1: **for** i from 1 to n **do**
2: pick $X = (X_1, \ldots, X_d)$ at random
3: get $Y = (c(X_1), \ldots, c(X_d))$
4: set $T_i = 0$ or 1 with an expected value $\mathcal{T}(X, Y)$
5: **end for**
6: if $(T_1, \ldots, T_n) \in \mathcal{A}$ output 1 otherwise output 0

Fig. 5. Non-Adaptive Iterated Attack of Order d.

$k = M(M-1)\ldots(M-d+1)$ permutations denoted $c_1, \ldots, c_k$. From $X = (X_1, \ldots, X_d)$ and $Y = (Y_1, \ldots, Y_d)$ we can define $I(X, Y)$ as the unique index j such that $c_j(X_i) = Y_i$ for $i = 1, \ldots, d$. We let

$$\mathcal{T}(X, Y) = \begin{cases} 1 \text{ if } I(X, Y) \equiv 0 \bmod \mu \\ 0 \text{ otherwise} \end{cases}$$

for a given modulus $\mu = n/a$ and

$$\mathcal{A} = \{0,1\}^n \backslash \{(0, \ldots, 0)\}.$$

If we feed this attack with C or C^*, we have

$$p \approx \frac{1}{\mu} = \frac{a}{n} \text{ or } p^* \approx 1 - \left(1 - \frac{1}{\mu}\right)^n \approx 1 - e^{-a}$$

for $a \ll n$ respectively. Thus Adv can be large even with a relatively large n. This problem actually comes from the fact that the tests $\mathcal{T}$ provide a same expected result for C and C^* but a totally different standard deviation.

As a more concrete counterexample we can consider C as the NUT-II decorrelation module over $\mathcal{M} = \mathrm{GF}(2^m)$ which achieves perfect decorrelation to the order $d = 2$. We can consider a kind of differential-linear attack as an iterated attack of order $d = 2$ which queries random pairs (X_1, X_2) with a fixed difference $X_1 \oplus X_2 = a$ and take T_i equal to one bit $b \cdot (c(X_1) \oplus c(X_2))$. Then we take $\mathcal{A} = \{(0, \ldots, 0), (1, \ldots, 1)\}$. Due to the NUT-II structure, T_i is a constant bit thus $p = 1$, but $p^* \approx 2.2^{-m}$ so the advantage of the distinguisher is close to 1. This simple example extends into a real attack due to Biham *et al.* [6] against the COCONUT98 cipher [64].

We can however prove the security when the cipher has a good decorrelation to the order $2d$ and an extra assumption about the distribution of X in every iteration.

Theorem 18. *Let C be a cipher on a message space of size M such that $|||[C]^{2d} - [C^*]^{2d}|||_\infty \leq \varepsilon$ for some given $d \leq M/2$ where C^* is the perfect cipher. Let us consider a non-adaptive iterated distinguisher (as depicted in Fig. 5) of order d between C and C^* of complexity n. We assume that the distinguisher generates sets of d plaintexts of independent and equal distribution in all iterations. We have*

$$\mathrm{Adv}_{\text{Fig. 5}} \leq 5 \sqrt[3]{\left(2\delta + \frac{5d^2}{2M} + \frac{3\varepsilon}{2}\right) n^2} + n\varepsilon,$$

where δ is the probability that any two different iterations send at least one query in common.

Note that this extra assumption on δ makes sense when considering either known plaintext attacks or chosen plaintext attacks with a sufficiently large sample space. For instance, if the distribution of X is uniform, we have $\delta \leq \frac{d^2}{M}$. If $X = (X_1, X_1 + a)$ with X_1 uniformly distributed, we have $\delta \leq \frac{3}{M}$.

Proof. Let Z (resp. Z^*) be the probability that the test accepts $(X, C(X))$ (resp. $(X, C^*(X))$), *i.e.*

$$Z = E_X(\mathcal{T}(X, C(X))).$$

Let p (resp. p^*) be the probability that the attack accepts, *i.e.*

$$p = \Pr_C[(T_1, \ldots, T_n) \in \mathcal{A}].$$

Since the T_i are independent and with the same expected value Z which only depends on C, we have

$$p = E_C\left(\sum_{(t_1,\ldots,t_n)\in\mathcal{A}} Z^{t_1+\ldots+t_n}(1-Z)^{n-(t_1+\ldots+t_n)}\right).$$

This can be written

$$p = \sum_{i=0}^{n} a_i E_C\left(Z^i(1-Z)^{n-i}\right)$$

for some integers a_i such that $0 \leq a_i \leq \binom{n}{i}$. Obviously the advantage $p - p^*$ is maximal when all a_i are either 0 or $\binom{n}{i}$ depending on the distributions of Z and Z^*. This proves that we can assume an iterated attack to have an acceptance set $\mathcal{A}$ of the form $\mathcal{A} = \{(t_1,\ldots,t_n); t_1+\ldots+t_n \in B\}$. Let

$$f(x) = \sum_{i\in B}\binom{n}{i}x^i(1-x)^{n-i}.$$

We have $p = E(f(Z))$. We have

$$f'(x) = \sum_{i\in B}\binom{n}{i}\frac{i-nx}{x(1-x)}x^i(1-x)^{n-i}.$$

The full sum over $i = 0,\ldots,n$ is the derivative of the binomial expansion of $(x+(1-x))^n$ which is 1. Hence the full sum is zero. We deduce that

$$|f'(x)| \leq \sum_{nx\leq i\leq n}\binom{n}{i}\frac{i-nx}{x(1-x)}x^i(1-x)^{n-i}.$$

Since $i \leq n$ we have

$$|f'(x)| \leq \frac{n}{x}\sum_{nx\leq i\leq n}\binom{n}{i}x^i(1-x)^{n-i}.$$

For $x \geq \frac{1}{2}$ we deduce that $|f'(x)| \leq 2n$. Since x and $1-x$ play a symmetric role we have $|f'(x)| \leq 2n$ for any x. Thus we have $|f(Z) - f(Z^*)| \leq 2n|Z - Z^*|$.

The crucial point in the proof is in proving that $|Z - Z^*|$ is small within a high probability. For this, we need $|E(Z) - E(Z^*)|$ and $|V(Z) - V(Z^*)|$ both to be small.

From Theorem 10 we know that $|E(Z) - E(Z^*)| \leq \frac{\varepsilon}{2}$. We note that Z^2 corresponds to another test but with $2d$ entries, namely,

$$\mathcal{T}((X_1,\ldots,X_d),(C(X_1),\ldots,C(X_d))) \times \mathcal{T}((X_{d+1},\ldots,X_{2d}),(C(X_{d+1}),\ldots,C(X_{2d})))$$

hence we have $|E(Z^2) - E((Z^*)^2)| \leq \frac{\varepsilon}{2}$. Hence $|V(Z) - V(Z^*)| \leq \frac{3}{2}\varepsilon$. Now from the Tchebichev's Inequality we have

$$\Pr[|Z - E(Z)| > \lambda] \leq \frac{V(Z)}{\lambda^2}$$

for any $\lambda > 0$. Since $|p - p^*| \leq E(|f(Z) - f(Z^*)|)$ we have

$$|p - p^*| \leq \sum_{z,z^*}\Pr[Z = z]\Pr[Z^* = z^*]|f(z) - f(z^*)|.$$

We separate values of z and z^* for which we have $|z - E(Z)| \leq \lambda$ and $|z^* - E(Z^*)| \leq \lambda$ from others and we get

$$|p - p^*| \leq \frac{V(Z)}{\lambda^2} + \frac{V(Z^*)}{\lambda^2} + 2n(|E(Z) - E(Z^*)| + 2\lambda).$$

We deduce

$$|p - p^*| \leq \frac{2V(Z^*) + \frac{3}{2}\varepsilon}{\lambda^2} + 2n\left(\frac{\varepsilon}{2} + 2\lambda\right)$$

so, with $\lambda = \left(\frac{2V(Z^*)+\frac{3}{2}\varepsilon}{n}\right)^{\frac{1}{3}}$ we have

$$|p - p^*| \leq 5\left(\left(2V(Z^*) + \frac{3\varepsilon}{2}\right)n^2\right)^{\frac{1}{3}} + n\varepsilon.$$

Now we have

$$\begin{aligned} V(Z^*) &= \sum_{\substack{(x,y)\in A \\ (x',y')\in A}} \Pr_X[x]\Pr_X[x']\left(\Pr_{C^*}\left[(x,x') \overset{C^*}{\mapsto} (y,y')\right] - \Pr_{C^*}[x \overset{C^*}{\mapsto} y]\Pr_{C^*}[x' \overset{C^*}{\mapsto} y']\right) \\ &\leq \frac{1}{2}\sum_{\substack{x,y \\ x',y'}} \Pr_X[x]\Pr_X[x']\left|\Pr_{C^*}\left[(x,x') \overset{C^*}{\mapsto} (y,y')\right] - \Pr_{C^*}[x \overset{C^*}{\mapsto} y]\Pr_{C^*}[x' \overset{C^*}{\mapsto} y']\right|. \end{aligned}$$

The sum over all x and x' entries with colliding entries (*i.e.* with some $x_i = x'_j$) is less than δ. The sum over all y and y' entries with colliding entries and no colliding x and x' is less than $d^2/4M$. The sum over all no colliding x and x' and no colliding y and y' is equal to

$$\frac{1-\delta}{2}\left(1 - \frac{M(M-1)\ldots(M-2d+1)}{M^2(M-1)^2\ldots(M-d+1)^2}\right),$$

which is less than $\frac{d^2}{2(M-d)}$. Thus we have $V(Z^*) \leq \delta + \frac{d^2}{4M} + \frac{d^2}{2(M-d)}$ which is less than $\delta + \frac{5d^2}{4M}$ when $2d \leq M$. □

This theorem proves that we need $n = \Omega(1/\sqrt{\varepsilon})$ or $n = \Omega(\sqrt{M})$ to have a meaningful iterated attack. If we apply it to linear cryptanalysis, this result is thus weaker than Theorem 17. It is however much more general.

Note that Theorem 18 could be extended with no assumption on δ and with adaptive attacks as long as we upper bound $V(Z^*)$.

7 Block Cipher Constructions

In the previous sections we have seen that it is enough to achieve a good decorrelation of low degree in order to resist many practical attack models. Here we show how to construct practical ciphers with those properties.

7.1 Decorrelation of Feistel Ciphers

In this section, we assume that $\mathcal{M} = {\mathcal{M}_0}^2$ where $\mathcal{M}_0$ is a group. Thus we can consider Feistel Ciphers on $\mathcal{M}$. We recall the notation Ψ for Feistel schemes. Given functions $f_1, \ldots, f_r$ over $\mathcal{M}_0$ we define

$$\Psi(f_1, \ldots, f_r)(x_L, x_R) = (z_R + f_r(z_L), z_L)$$

where $(z_L, z_R) = \Psi(f_1, \ldots, f_{r-1})(x_L, x_R)$ and $\Psi()(x_L, x_R) = (x_R, x_L)$, as illustrated by Fig. 6.

Lemma 19 (Luby-Rackoff [40]). *Let F_1, F_2, F_3 be three independent uniformly distributed random functions on $\mathcal{M}_0$ and let d be an integer. We consider the cipher $C = \Psi(F_1, F_2, F_3)$ on $\mathcal{M} = {\mathcal{M}_0}^2$ and we compare it with the perfect cipher C^*. For any distinguisher $\mathcal{A}$ between $\Psi(F_1, F_2, F_3)$ and C^* which is limited to d queries, we have*

$$\mathrm{Adv}_{\mathcal{A}} \leq \frac{d^2}{\sqrt{\#\mathcal{M}}}.$$

Thus from Theorem 11 we have

$$||[\Psi(F_1, F_2, F_3)]^d - [C^*]^d||_a \leq 2\frac{d^2}{\sqrt{\#\mathcal{M}}}.$$

The decorrelation $||.||_a$-bias of Feistel Ciphers can be estimated with the following lemma.

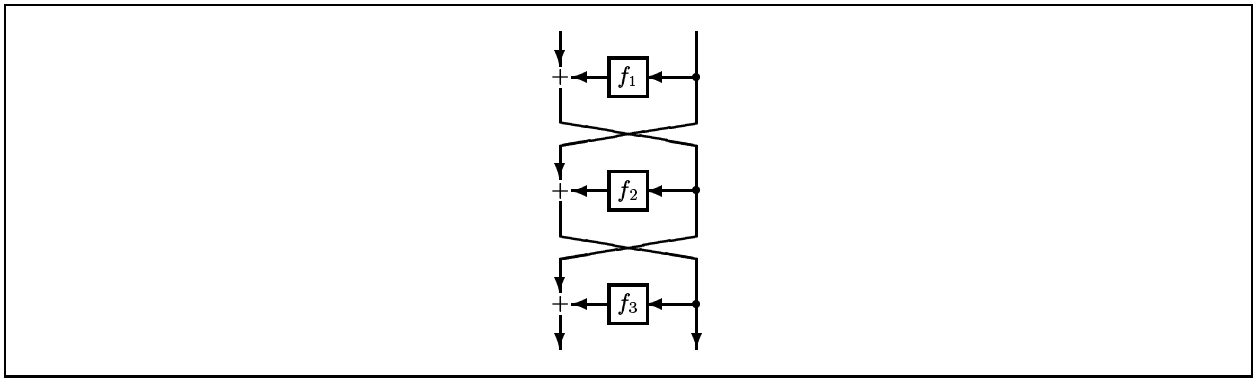

Fig. 6. Feistel Scheme $\Psi(f_1, f_2, f_3)$.

Lemma 20. *Let $F_1, \ldots, F_r$ (resp. $R_1, \ldots, R_r$) be r independent random functions on $\mathcal{M}_0$ such that $||[F_i]^d - [R_i]^d||_a \leq \varepsilon_i$ $(i = 1, \ldots, r)$. We have*

$$||[\Psi(F_1, \ldots, F_r)]^d - [\Psi(R_1, \ldots, R_r)]^d||_a \leq \varepsilon_1 + \ldots + \varepsilon_r.$$

Proof. Let $C_i = \Psi(R_1, \ldots, R_i, F_{i+1}, \ldots, F_r)$. We have $C_0 = \Psi(F_1, \ldots, F_r)$ and $C_r = \Psi(R_1, \ldots, R_r)$. Using the triangular inequality we have

$$||[\Psi(F_1, \ldots, F_r)]^d - [\Psi(R_1, \ldots, R_r)]^d||_a \leq \sum_{i=1}^{r} ||[C_{i-1}]^d - [C_i]^d||_a.$$

Let us prove that $||[C_{i-1}]^d - [C_i]^d||_a \leq ||[F_i]^d - [R_i]^d||_a$.

Using Theorem 11, we consider a distinguisher $\mathcal{A}$ between C_{i-1} and C_i with advantage $\frac{1}{2}||[C_{i-1}]^d - [C_i]^d||_a$. We can construct a distinguisher $\mathcal{B}$ between F_i and R_i as follows:

1. we simulate the random $R_1, \ldots, R_{i-1}, F_{i+1}, \ldots, F_r$
2. from an oracle $\mathcal{O}$ which implements a function on $\mathcal{M}_0$ we construct an oracle which implements the cipher $\Psi(R_1, \ldots, R_{i-1}, \mathcal{O}, F_{i+1}, \ldots, F_r)$
3. we run $\mathcal{A}$ on this oracle.

Obviously this distinguisher between F_i and R_i has the advantage $\frac{1}{2}||[C_{i-1}]^d - [C_i]^d||_a$. Since the best possible advantage is $\frac{1}{2}||[F_i]^d - [R_i]^d||_a$, we obtain the result. □

By using Theorem 10, this lemma, the Luby-Rackoff Lemma, and the multiplicativity of decorrelation distances, we obtain the following result.

Theorem 21. *Let $F_1, \ldots, F_r$ be r independent random functions on $\mathcal{M}_0$ such that $||[F_i]^d - [F^*]^d||_a \leq \varepsilon$ $(i = 1, \ldots, r)$ where F^* is a uniformly distributed random function on $\mathcal{M}_0$. We consider the cipher $C = \Psi(F_1, \ldots, F_r)$ on $\mathcal{M} = {\mathcal{M}_0}^2$ and we compare it with the perfect cipher C^*. Let $k \geq 3$ be an integer. For any distinguisher $\mathcal{A}$ between C and C^* which is limited to d queries, we have*

$$\mathrm{Adv}_{\mathcal{A}} \leq \frac{1}{2}\left(k\varepsilon + \frac{2d^2}{\sqrt{\#\mathcal{M}}}\right)^{\lfloor \frac{r}{k} \rfloor}.$$

We note that the result holds for practical Feistel schemes as long as rounds use independent subkeys and that we can measure the decorrelation biases of round functions.

Proof. By using the simulation technique as in the lemma, we notice that the $||.||_a$-decorrelation bias can only decrease with the number of rounds. Hence the decorrelation bias of a k-round Feistel scheme is at most $k\varepsilon + \frac{2d^2}{\sqrt{\#\mathcal{M}}}$. Next we use the multiplicativity of the $||.||_a$-decorrelation bias $\lfloor \frac{r}{k} \rfloor$ times. We may have a few extra rounds, but this can only make the decorrelation bias decrease. We finally use Theorem 11. □

We mention that there is a similar result for the $||.||_2$ norm in [65].

7.2 Generalization

The construction of decorrelated Feistel schemes based on the Luby-Rackoff Theorem generalizes to arbitrary structures. We provide here a useful lemma taken from [70] which was freely adapted from Patarin's "coefficient H techniques" [55].

Lemma 22. *Let d be an integer. Let F be a random function from a set $\mathcal{M}_1$ to a set $\mathcal{M}_2$. We let $\mathcal{X}$ be the subset of $\mathcal{M}_1^d$ of all $(x_1, \ldots, x_d)$ with pairwise different entries. We let F^* be a uniformly distributed random function from $\mathcal{M}_1$ to $\mathcal{M}_2$. We know that for all $x \in \mathcal{X}$ and $y \in \mathcal{M}_2^d$ the value $[F^*]_{x,y}^d$ is a constant $p_0 = (\#\mathcal{M}_2)^{-d}$. We assume there exist a subset $\mathcal{Y} \subseteq \mathcal{M}_2^d$ and two positive real values ε_1 and ε_2 such that*

- $|\mathcal{Y}|p_0 \geq 1 - \varepsilon_1$
- $\forall x \in \mathcal{X} \quad \forall y \in \mathcal{Y} \quad [F]_{x,y}^d \geq p_0(1 - \varepsilon_2)$.

Then we have $||[F]^d - [F^*]^d||_a \leq 2\varepsilon_1 + 2\varepsilon_2$.

This lemma intuitively means that if $[F]_{x,y}^d$ is close to $[F^*]_{x,y}^d$ for all x and almost all y, then the decorrelation bias of F is small.

Proof. We use the characterization of $||.||_a$-decorrelation bias in terms of best adaptive distinguisher by using Theorem 11. We let $\mathcal{A}$ be one d-limited distinguisher between F and F^* with maximum advantage. We can assume without loss of generality that $\mathcal{A}$ is deterministic and never sends the same query twice. The behavior of $\mathcal{A}$ is thus deterministically defined by the oracle responses $y = (y_1, \ldots, y_d)$. We let x_i denote the ith query defined by $y_1, \ldots, y_{i-1}$. We let $x = (x_1, \ldots, x_d)$ which is assumed to be in $\mathcal{X}$. We let A be the set of all rejected y_i, *i.e.* for which $\mathcal{A}$ outputs 0. It is straightforward that

$$\mathrm{Adv}_{\mathcal{A}} = -\sum_{y \in A} \left([F]_{x,y}^d - [F^*]_{x,y}^d \right).$$

Next we have

$$\mathrm{Adv}_{\mathcal{A}} \leq \sum_{\substack{y \in A \\ y \in \mathcal{Y}}} \varepsilon_2 [F^*]_{x,y}^d + \sum_{\substack{y \in A \\ y \notin \mathcal{Y}}} [F^*]_{x,y}^d.$$

The first sum is upper bounded by ε_2. For the second sum, we recall that all x_i's are pairwise different, so $[F^*]_{x,y}^d$ is always equal to p_0. This sum is thus less than ε_1. □

As a first application, here is a quite useful lemma.

Lemma 23. *For a random uniformly distributed function F^* and a random uniformly distributed permutation C^* defined over $\{0,1\}^m$, we have*

$$||[F^*]^d - [C^*]^d||_a \leq d(d-1)2^{-m}.$$

Proof. We use Lemma 22 with $F = C^*$. We let $\mathcal{Y}$ be equal to the set of all pairwise different outputs. We have

$$|\mathcal{Y}|p_0 \geq 1 - \frac{d(d-1)}{2} 2^{-m}$$

which gives ε_1. Since $[C^*]_{x,y}^d \geq p_0$, we can take $\varepsilon_2 = 0$ and apply Lemma 22. □

As an example of application we prove Lemma 19 in a few lines.

Proof (of Lemma 19). Following the Feistel scheme $C = \Psi(F_1, F_2, F_3)$, we let

$$\begin{aligned} x_i &= (z_i^0, z_i^1) \\ z_i^2 &= z_i^0 + F_1(z_i^1) \\ y_i &= (z_i^4, z_i^3) \end{aligned}$$

We let E be the event that $z_i^3 = z_i^1 + F_2(z_i^2)$ and $z_i^4 = z_i^2 + F_3(z_i^3)$ for all $i = 1, \ldots, d$. We have $[C]_{x,y}^d = \Pr[E]$. We now define

$$\mathcal{Y} = \left\{ (y_1, \ldots, y_d); \forall i < j \quad z_i^3 \neq z_j^3 \right\}.$$

(This is a set of non-pathological outputs when computing $[C]^d_{x,y}$.) We can easily check that $\mathcal{Y}$ fulfills the requirements of Lemma 22. Firstly we have

$$|\mathcal{Y}| \geq \left(1 - \frac{d(d-1)}{2} 2^{-\frac{m}{2}}\right) 2^{md},$$

thus we let $\varepsilon_1 = \frac{d(d-1)}{2} 2^{-\frac{m}{2}}$. Second, for $y \in \mathcal{Y}$ and any x (with pairwise different entries), we need to consider $[C]^d_{x,y}$. Let E^2 be the event that all z_i^2's are pairwise different over the distribution of F_1. We have

$$[C]^d_{x,y} \geq \Pr[E/E^2] \Pr[E^2].$$

For computing $\Pr[E/E^2]$ we know that z_i^3's are pairwise different, as for the z_i^2's. Hence $\Pr[E/E^2] = 2^{-md}$. It is then straightforward that $\Pr[E^2] \geq 1 - \frac{d(d-1)}{2} 2^{-\frac{m}{2}}$ which we define to be $1 - \varepsilon_2$. We thus obtain from Lemma 22 that $||[C]^d - [F^*]^d||_a \leq 2d(d-1)2^{-\frac{m}{2}}$. From this and Lemma 23 we thus obtain $||[C]^d - [C^*]^d||_a \leq 2d^2 2^{-\frac{m}{2}}$ for $d \leq 2^{1+\frac{m}{2}}$. Since $||.||_a$ is always less than 2, it also holds for larger d. □

This technique can be used for various applications. For instance, we can compare the decorrelation provided by top-level schemes of the candidates to the AES standardization process. This has been done in [47]. It was also applied to the Lai-Massey scheme (the construction of IDEA [36]) in [68]. This is used in Section 8.3.

8 Construction Examples

8.1 COCONUT: A Perfect Decorrelation Design

In this section we define the COCONUT Ciphers family which are perfectly decorrelated ciphers to the order 2. It uses the NUT-II decorrelation module.

The COCONUT Ciphers are characterized by some parameters (m, p) where m is the block length, and p is an irreducible polynomial of degree m in $\mathrm{GF}(2)$ (which defines a representation of the $\mathrm{GF}(2^m)$ Galois Field). A COCONUT Cipher of block length m is simply a product cipher $C_1 \circ C_2 \circ C_3$ where C_1 and C_3 are any (possibly weak) ciphers which can depend on each other, and C_2 is an independent cipher based on a $2m$-bit key which consists of two polynomials A and B of degree at most $m-1$ over $\mathrm{GF}(2)$ such that $A \neq 0$. For a given representation of polynomials into m-bit strings, we simply define

$$C_2(x) = A.x + B \bmod p.$$

C_2 is thus the NUT-II decorrelation module.

Since C_2 performs perfect decorrelation to the order 2 and since it is independent from C_1 and C_3, any COCONUT Cipher is obviously perfectly decorrelated to the order two. Therefore Theorems 13 and 17 show that COCONUT resists basic differential and linear cryptanalysis.

One can wonder what C_1 and C_3 are for. Actually, C_2 makes some classes of attacks provably impractical, but in a way which makes the cipher obviously weak against other attacks. (C_2 is actually a linear function, thus although we can prove it resists any attack with a parameter $d \leq 2$, it is fairly weak against some attacks with $d = 3$.) We believe that all real attacks on any real cipher have an intrinsic *order* d: that is, they use the d-wise correlation in the encryption of d messages. Attacks with a large d on real ciphers are naturally impractical, because the d-wise decorrelation can hardly be analyzed since it depends on too many factors. Therefore, the COCONUT approach consists in making the cipher provably resistant against attacks of order at most 2 such as differential or linear cryptanalysis, and heuristically secure against attacks of higher order by real life ciphers as C_1 and C_3.

Example 24. The COCONUT98 Cipher has been proposed in [64] with parameters $m = 64$ and $p = x^{64} + x^{11} + x^2 + x + 1$. Interestingly, this motivated Wagner to invent the "boomerang attack" [72] in order to break it. This attack is an iterated attack of order 4 which uses pretty bad differential properties of C_1 and C_3. Another attack was found by Biham *et al.* [6] based on a non-adaptive iterated attack of order 2 (namely a differential-linear attack, see Section 6.3). This shows that despite the COCONUT98 Cipher provably resisting any differential distinguisher as depicted in Fig. 3, one must not neglect the intrinsic strength of C_1 and C_3. The existence of attacks when using stronger C_1 and C_3 is still an open problem.

For completeness, we mention that an extension of the COCONUT construction (called DONUT for "Double Operations with NUT") was proposed by Cheon *et al.* [13].

8.2 PEANUT: A Partial Decorrelation Design

In this section we define the PEANUT Ciphers family, which achieve an example of partial decorrelation. This family is based on the NUT-IV decorrelation module.

The PEANUT Ciphers are characterized by some parameters (m, r, d, p). They are Feistel Ciphers with a block length of m bits (m even) and r rounds. The parameter d is the order of partial decorrelation that the cipher performs, and p must be a prime number greater than $2^{\frac{m}{2}}$.

The cipher is defined by a key of $\frac{mrd}{2}$ bits which consists of a sequence of r lists of d $\frac{m}{2}$-bit numbers, one for each round. In each round, the F function has the form

$$F(x) = g(k_1.x^{d-1} + k_2.x^{d-2} + \ldots + k_{d-1}.x + k_d \bmod p \bmod 2^{\frac{m}{2}})$$

where g is any permutation on the set of all $\frac{m}{2}$-bit numbers.

From Theorem 21 with $k = 3$ we thus obtain the following theorem.

Theorem 25. *Let C be a cipher in the PEANUT family with parameters (m, r, d, p). We have*

$$||[C]^d - [C^*]^d||_a \leq \left(\left(1 + 2\left(p^d 2^{-\frac{md}{2}} - 1\right)\right)^3 - 1 + \frac{2d^2}{2^{\frac{m}{2}}}\right)^{\lfloor \frac{r}{3} \rfloor}$$

where C^* is the perfect cipher.

When $p \approx 2^{\frac{m}{2}}$, the upper bound for $||[C]^d - [C^*]^d||_a$ can be approximated by

$$\left(\frac{6d\left(p - 2^{\frac{m}{2}}\right) + 2d^2}{2^{\frac{m}{2}}}\right)^{\lfloor \frac{r}{3} \rfloor}.$$

Example 26. We can use the parameters $m = 64$, $r = 9$, $d = 2$ and $p = 2^{32} + 15$. We obtain that $||[C]^2 - [C^*]^2||_a \leq 2^{-76}$. Therefore from Theorems 13 and 17 no differential or linear distinguisher can be efficient. The PEANUT98 Cipher has been proposed with these parameters in [64].

Example 27. In an earlier version of this work [63], we proposed a similar construction (called PEANUT97) based on the NUT-III decorrelation module which uses prime numbers smaller than $2^{\frac{m}{2}}$. However, the result above does not hold with the $||.||_a$ norm, but rather with the $||.||_2$ one. The drawback is that this norm has less friendly theorems for constructing Feistel ciphers, and in particular we need more rounds to make the cipher provably secure. (See [65].)

Example 28. The AES candidate DFC was proposed based on the PEANUT construction (see [17,18,20]). Nominal parameters are $m = 128$, $r = 8$, $d = 2$, and $p = 2^{64} + 13$, so we have $||[C]^2 - [C^*]^2||_a \leq 2^{-115}$.

8.3 WALNUT: An Alternate Design

The Feistel cipher is based on a round mapping defined by

$$(x_L, x_R) \mapsto (x_R, x_L \oplus F(x_R)).$$

The Feistel scheme benefits the Luby-Rackoff lemma which enables building a PEANUT cipher with a provably low decorrelation bias. Instead, we can use the Lai-Massey scheme on which IDEA relies and which is based on the round mapping

$$(x_L, x_R) \mapsto (O(x_L + F(x_L - x_R)), x_R + F(x_L - x_R))$$

as illustrated by Fig. 7 where $+$ is any group addition law, $-$ is the corresponding subtraction, and O is an *orthomorphism* for the group, *i.e.* a permutation such that $x \mapsto O(x) - x$ is also a permutation. As shown in [68], the Luby-Rackoff lemma holds for this scheme as well with the same bound and same number of rounds. We can thus construct the WALNUT cipher is the same way as the PEANUT cipher, but with the Lai-Massey scheme instead of the Feistel one.

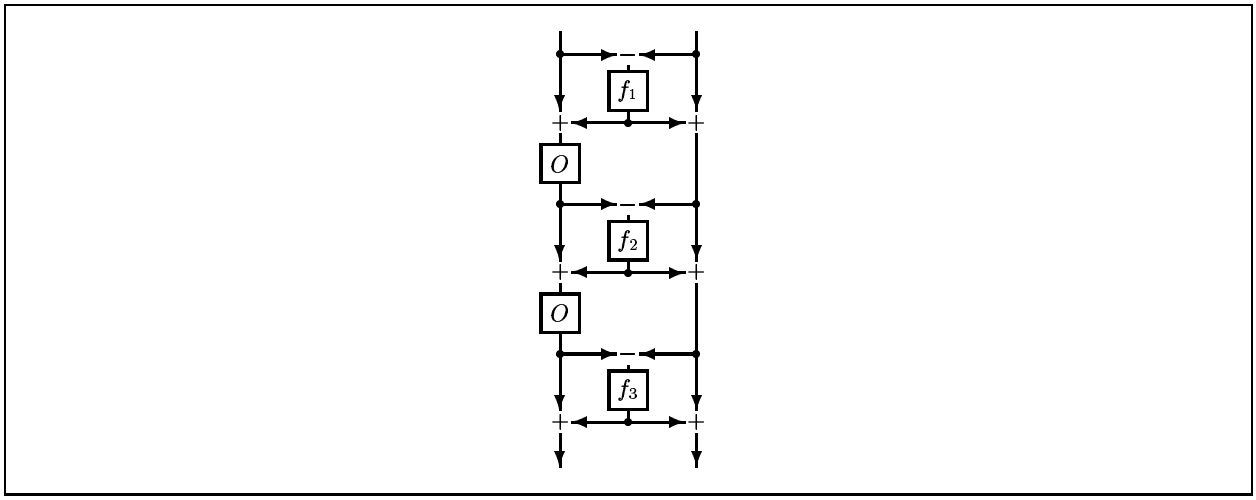

Fig. 7. Lai-Massey Scheme.

9 Conclusion and Further Work

Decorrelation modules are cheap and friendly tools which can strengthen the security of block ciphers. Actually, we can quantify their security against a class of cryptanalysis which includes differential and linear cryptanalysis. To illustrate this paradigm, we proposed prototype ciphers PEANUT97 [63], COCONUT98 and PEANUT98 [64], and DFCv2 [20].

One problem with the COCONUT, PEANUT, or WALNUT constructions is that they require a long key (in order to make the internal random functions independent). In real-life examples, we can generate this long key by using a pseudorandom generator fed with a short key, but the results on the security based on decorrelation are no longer valid. However, provided that the pseudorandom generator produces outputs which are indistinguishable from truly random sequences, we can still prove the security. This approach has been developed in [17,18,20] with the submission of DFC to the *Advanced Encryption Standard* process.

Security against some other generic models of attacks is still open. In particular we may investigate security against the Boomerang attack [72], the rectangle attack [5], or the linear-differential attack [6,38]. Although we can directly use results from Section 6.3 with a high order of decorrelation it is not quite clear at this time what the minimal order of decorrelation required is. Extensions of Theorem 18 to adaptive attacks is also open. It is further not quite sure that $2d$-decorrelation is necessary for getting provable security against iterated attacks of order d, although we have proven it is sufficient and that d-decorrelation is not.

It is further problematic to estimate the decorrelation bias of concrete ciphers like DES or AES candidates unless we approximate them to an ideal model [47].

10 Acknowledgments

I thank the anonymous referees, Eli Biham, Ueli Maurer, David Wagner, and Thomas Baignères for helpful comments, as well as Pascal Junod for interesting discussions and a pointer to Equation (6). My gratitude also goes to my co-authors from [17,18,20,57], Henri Gilbert, Marc Girault, Louis Granboulan, Philippe Hoogvorst, Phong Nguyen, Fabrice Noilhan, Thomas Pornin, Guillaume Poupard, Jacques Stern, and from [4,47], Kazumaro Aoki and Shiho Moriai.

References

1. ETSI. Universal Mobile Telecommunications System (UMTS); Specification of the 3GPP confidentiality and integrity algorithms. Document 2: Kasumi algorithm specification (3GPP TS 35.202 version 3.1.2 Release 1999). `http://www.etsi.org/`

2. Data Encryption Standard. *Federal Information Processing Standard Publication 46*, U. S. National Bureau of Standards, 1977.
3. K. Aoki, K. Ohta. Strict Evaluation of the Maximum Average of Differential Probability and the Maximum Average of Linear Probability. *IEICE Transactions on Fundamentals*, vol. E80-A, pp. 1–8, 1997.
4. K. Aoki, S. Vaudenay. On the Use of GF-Inversion as a Cryptographic Primitive. To appear in the Proceedings of SAC'03, LNCS, Springer-Verlag.
5. E. Biham, O. Dunkelman, N. Keller. The Rectangle Attack — Rectangling the Serpent. In *Advances in Cryptology EUROCRYPT'01*, Innsbruck, Austria, Lecture Notes in Computer Science 2045, pp. 340–357, Springer-Verlag, 2001.
6. E. Biham, O. Dunkelman, N. Keller. Enhancing Differential-Linear Cryptanalysis. In *Advances in Cryptology ASIACRYPT'02*, Queenstown, New Zeland, Lecture Notes in Computer Science 2501, pp. 254–266, Springer-Verlag, 2002.
7. E. Biham, A. Shamir. Differential Cryptanalysis of DES-like Cryptosystems. In *Advances in Cryptology CRYPTO'90*, Santa Barbara, California, U.S.A., Lecture Notes in Computer Science 537, pp. 2–21, Springer-Verlag, 1991.
8. E. Biham, A. Shamir. Differential Cryptanalysis of DES-like Cryptosystems. *Journal of Cryptology*, vol. 4, pp. 3–72, 1991.
9. E. Biham, A. Shamir. Differential Cryptanalysis of the Full 16-Round DES. In *Advances in Cryptology CRYPTO'92*, Santa Barbara, California, U.S.A., Lecture Notes in Computer Science 740, pp. 487–496, Springer-Verlag, 1993.
10. E. Biham, A. Shamir. *Differential Cryptanalysis of the Data Encryption Standard*, Springer-Verlag, 1993.
11. J. L. Carter, M. N. Wegman. Universal Classes of Hash Functions. *Journal of Computer and System Sciences*, vol. 18, pp. 143–154, 1979.
12. F. Chabaud, S. Vaudenay. Links between Differential and Linear Cryptanalysis. In *Advances in Cryptology EUROCRYPT'94*, Perugia, Italy, Lecture Notes in Computer Science 950, pp. 356–365, Springer-Verlag, 1995.
13. D. H. Cheon, S. J. Lee, J. I. Lim, S. J. Lee. New Block Cipher DONUT Using Pairwise Perfect Decorrelation. In *Progress in Cryptology INDOCRYPT'00*, Calcutta, India, Lecture Notes in Computer Science 1977, pp. 262–270, Springer-Verlag, 2000.
14. H. Feistel. Cryptography and Computer Privacy. *Scientific American*, vol. 228, pp. 15–23, 1973.
15. H. Gilbert. *Cryptanalyse Statistique des Algorithmes de Chiffrement et Sécurité des Schémas d'Authentification*, Thèse de Doctorat de l'Université de Paris 11, 1997.
16. H. Gilbert, G. Chassé. A Statistical Attack of the FEAL-8 Cryptosystem. In *Advances in Cryptology CRYPTO'90*, Santa Barbara, California, U.S.A., Lecture Notes in Computer Science 537, pp. 22–33, Springer-Verlag, 1991.
17. H. Gilbert, M. Girault, P. Hoogvorst, F. Noilhan, T. Pornin, G. Poupard, J. Stern, S. Vaudenay. Decorrelated Fast Cipher: an AES Candidate. (Extended Abstract.) In *Proceedings from the First Advanced Encryption Standard Candidate Conference*, National Institute of Standards and Technology (NIST), Ventura, California, U.S.A., August 1998.
18. H. Gilbert, M. Girault, P. Hoogvorst, F. Noilhan, T. Pornin, G. Poupard, J. Stern, S. Vaudenay. Decorrelated Fast Cipher: an AES Candidate. Submitted to the Advanced Encryption Standard process. In *CD-ROM "AES CD-1: Documentation"*, National Institute of Standards and Technology (NIST), August 1998.
19. H. Gilbert, M. Minier. New Results on the Pseudorandomness of Some Blockcipher Constructions. In *Fast Software Encryption'01*, Yokohama, Japan, Lecture Notes in Computer Science 2355, pp. 248–266, Springer-Verlag, 2002.
20. L. Granboulan, P. Nguyen, F. Noilhan, S. Vaudenay. DFCv2. In *Selected Areas in Cryptography'00*, Waterloo, Ontario, Canada, Lecture Notes in Computer Science 2012, pp. 57–71, Springer-Verlag, 2001.
21. S. Halevi, H. Krawczyk. MMH: Software Message Authentication in the Gbit/second Rates. In *Fast Software Encryption'97*, Haifa, Israel, Lecture Notes in Computer Science 1267, pp. 172–189, Springer-Verlag, 1997.
22. H. M. Heys. *The Design of Substitution-Permutation Network Ciphers Resistant to Cryptanalysis*, Ph.D. Thesis of Queen's University, Kingston, Ontario, Canada, 1994.
23. H. M. Heys, S. E. Tavares. Substitution-Permutation Networks Resistant to Differential and Linear Cryptanalysis. *Journal of Cryptology*, vol. 9, pp. 1–19, 1996.
24. T. Iwata, K. Kurosawa. On the Pseudorandomness of the AES Finalists — RC6 and Serpent. In *Fast Software Encryption'00*, New York, NY, USA, Lecture Notes in Computer Science 1978, pp. 231–243, Springer-Verlag, 2001.
25. T. Iwata, T. Yoshino, T. Yuasa, K. Kurosawa. Round Security and Super-Pseudorandomness of MISTY Type Structure. In *Fast Software Encryption'01*, Yokohama, Japan, Lecture Notes in Computer Science 2355, pp. 233–247, Springer-Verlag, 2002.
26. T. Jakobsen, L. R. Knudsen. The Interpolation Attack on Block Ciphers. In *Fast Software Encryption'97*, Haifa, Israel, Lecture Notes in Computer Science 1267, pp. 28–40, Springer-Verlag, 1997.
27. P. Junod. On the Complexity of Matsui's Attack. In *Selected Areas in Cryptography'01*, Toronto, Ontario, Canada, Lecture Notes in Computer Science 2259, pp. 199-211, Springer-Verlag, 2001.
28. P. Junod. On the Optimality of Linear, Differential and Sequential Distinguishers. In *Advances in Cryptology EUROCRYPT'03*, Warsaw, Poland, Lecture Notes in Computer Science 2656, pp. 17–32, Springer-Verlag, 2003.
29. P. Junod, S. Vaudenay. Optimal Key Ranking Procedures in a Statistical Cryptanalysis. To appear in FSE'03.
30. J.-S. Kang, S.-U. Shin, D. Hong, O. Yi. Provable Security of KASUMI and 3GPP Encryption Mode f8. In *Advances in Cryptology ASIACRYPT'00*, Brisbane, Australia, Lecture Notes in Computer Science 2248, pp. 255–271, Springer-Verlag, 2001.
31. B. R. Kaliski Jr., M. J. B. Robshaw. Linear Cryptanalysis using Multiple Approximations. In *Advances in Cryptology CRYPTO'94*, Santa Barbara, California, U.S.A., Lecture Notes in Computer Science 839, pp. 26–39, Springer-Verlag, 1994.
32. L. Keliher, H. Meijer, S. Tavares. New Method for Upper Bounding the Maximum Average Linear Hull Probability for SPNs. In *Advances in Cryptology EUROCRYPT'01*, Innsbruck, Austria, Lecture Notes in Computer Science 2045, pp. 420–436, Springer-Verlag, 2001.

33. L. Keliher, H. Meijer, S. Tavares. Improving the Upper Bound on the Maximum Average Linear Hull Probability for Rijndael. In *Selected Areas in Cryptography'01*, Toronto, Ontario, Canada, Lecture Notes in Computer Science 2259, pp. 112–128, Springer-Verlag, 2001.
34. A. Kerckhoffs. *La Cryptographie Militaire*, Librairie militaire de L. Baudouin & Cie., Paris, 1883.
35. L. R. Knudsen. *Block Ciphers — Analysis, Design and Applications*, Aarhus University, 1994.
36. X. Lai. *On the Design and Security of Block Ciphers*, ETH Series in Information Processing, vol. 1, Hartung-Gorre Verlag Konstanz, 1992.
37. X. Lai, J. L. Massey, S. Murphy. Markov Ciphers and Differential Cryptanalysis. In *Advances in Cryptology EUROCRYPT'91*, Brighton, United Kingdom, Lecture Notes in Computer Science 547, pp. 17–38, Springer-Verlag, 1991.
38. S. K. Langford, M. E. Hellman Differential-linear Cryptanalysis. In *Advances in Cryptology CRYPTO'94*, Santa Barbara, California, U.S.A., Lecture Notes in Computer Science 839, pp. 17–25, Springer-Verlag, 1994.
39. M. Luby, C. Rackoff. Pseudo-random Permutation Generators and Cryptographic Composition. In *Proceedings of the 17th ACM Symposium on Theory of Computing*, Providence, Rhode Island, U.S.A., pp. 363–365, ACM Press, 1985.
40. M. Luby, C. Rackoff. How to Construct Pseudorandom Permutations from Pseudorandom Functions. *SIAM Journal on Computing*, vol. 17, pp. 373–386, 1988.
41. M. Matsui. Linear Cryptanalysis Methods for DES Cipher. In *Advances in Cryptology EUROCRYPT'93*, Lofthus, Norway, Lecture Notes in Computer Science 765, pp. 386–397, Springer-Verlag, 1994.
42. M. Matsui. The First Experimental Cryptanalysis of the Data Encryption Standard. In *Advances in Cryptology CRYPTO'94*, Santa Barbara, California, U.S.A., Lecture Notes in Computer Science 839, pp. 1–11, Springer-Verlag, 1994.
43. M. Matsui. New Structure of Block Ciphers with Provable Security against Differential and Linear Cryptanalysis. In *Fast Software Encryption'96*, Cambridge, United Kingdom, Lecture Notes in Computer Science 1039, pp. 205–218, Springer-Verlag, 1996.
44. M. Matsui. New Block Encryption Algorithm MISTY. In *Fast Software Encryption'97*, Haifa, Israel, Lecture Notes in Computer Science 1267, pp. 54–68, Springer-Verlag, 1997.
45. U. M. Maurer, J. L. Massey. Cascade Ciphers: The Importance of Being First. *Journal of Cryptology*, vol. 6, pp. 55–61, 1993.
46. U. Maurer, K. Pietrzak. The Security of Many-Round Luby-Rackoff Pseudo-Random Permutations. In *Advances in Cryptology EUROCRYPT'03*, Warsaw, Poland, Lecture Notes in Computer Science 2656, pp. 544–561, Springer-Verlag, 2003.
47. S. Moriai, S. Vaudenay On the Pseudorandomness of Top-Level Schemes of Block Ciphers. In *Advances in Cryptology ASIACRYPT'00*, Kyoto, Japan, Lecture Notes in Computer Science 1976, pp. 289–302, Springer-Verlag, 2000.
48. S. Murphy, F. Piper, M. Walker, P. Wild. Likelihood Estimation for Block Cipher Keys. Unpublished.
49. M. Naor, O. Reingold. On the Construction of Pseudorandom Permutations: Luby-Rackoff Revisited. *Journal of Cryptology*, vol. 12, pp. 29–66, 1999.
50. K. Nyberg. Perfect Nonlinear S-Boxes. In *Advances in Cryptology EUROCRYPT'91*, Brighton, United Kingdom, Lecture Notes in Computer Science 547, pp. 378–385, Springer-Verlag, 1991.
51. K. Nyberg, L. R. Knudsen. Provable Security against a Differential Cryptanalysis. In *Advances in Cryptology CRYPTO'94*, Santa Barbara, California, U.S.A., Lecture Notes in Computer Science 839, pp. 566–574, Springer-Verlag, 1994.
52. K. Nyberg, L. R. Knudsen. Provable Security against a Differential Cryptanalysis. *Journal of Cryptology*, vol. 8, pp. 27–37, 1995.
53. S. Park, S. H. Sung, S. Chee, E-J. Yoon, J. Lim On the Security of Rijndael-Like Structures against Differential and Linear Cryptanalysis. In *Advances in Cryptology ASIACRYPT'02*, Queenstown, New Zeland, Lecture Notes in Computer Science 2501, pp. 176–191, Springer-Verlag, 2002.
54. S. Park, S. H. Sung, S. Lee, J. Lim Improving the Upper Bound on the Maximum Differential and Maximum Linear Hull Probability for SPN Structures and AES. To appear in FSE'03.
55. J. Patarin. *Etude des Générateurs de Permutations Basés sur le Schéma du D.E.S.*, Thèse de Doctorat de l'Université de Paris 6, 1991.
56. J. Patarin. About Feistel Schemes with Six (or More) Rounds. In *Fast Software Encryption'98*, Paris, France, Lecture Notes in Computer Science 1372, pp. 103–121, Springer-Verlag, 1998.
57. G. Poupard, S. Vaudenay. Decorrelated Fast Cipher: an AES Candidate well suited for Low Cost Smart Cards Applications. In *CARDIS'98*, Louvain-la-Neuve, Belgium, Lecture Notes in Computer Science 1820, pp. 254–264, Springer-Verlag, 2000.
58. A. Rényi. *Probability Theory*, Elsevier, 1970.
59. C. E. Shannon. Communication Theory of Secrecy Systems. *Bell system technical journal*, vol. 28, pp. 656–715, 1949.
60. A. Tardy-Corfdir, H. Gilbert. A Known Plaintext Attack of FEAL-4 and FEAL-6. In *Advances in Cryptology CRYPTO'91*, Santa Barbara, California, U.S.A., Lecture Notes in Computer Science 576, pp. 172–181, Springer-Verlag, 1992.
61. S. Vaudenay. *La Sécurité des Primitives Cryptographiques*, Thèse de Doctorat de l'Université de Paris 7, Technical Report LIENS-95-10 of the Laboratoire d'Informatique de l'Ecole Normale Supérieure, 1995.
62. S. Vaudenay. An Experiment on DES — Statistical Cryptanalysis. In *3rd ACM Conference on Computer and Communications Security*, New Delhi, India, pp. 139–147, ACM Press, 1996.
63. S. Vaudenay. A cheap Paradigm for Block Cipher Security Strengthening. Technical Report LIENS-97-3, 1997.

64. S. Vaudenay. Provable Security for Block Ciphers by Decorrelation. In *STACS'98*, Paris, France, Lecture Notes in Computer Science 1373, pp. 249–275, Springer-Verlag, 1998.
65. S. Vaudenay. Feistel Ciphers with L_2-Decorrelation. In *Selected Areas in Cryptography'98*, Kingston, Ontario, Canada, Lecture Notes in Computer Science 1556, pp. 1–14, Springer-Verlag, 1999.
66. S. Vaudenay. The Decorrelation Technique Home-Page.
URL:`http://lasecwww.epfl.ch/decorrelation.shtml`
67. S. Vaudenay. Resistance Against General Iterated Attacks. In *Advances in Cryptology EUROCRYPT'99*, Prague, Czech Republic, Lecture Notes in Computer Science 1592, pp. 255–271, Springer-Verlag, 1999.
68. S. Vaudenay. On the Lai-Massey Scheme. In *Advances in Cryptology ASIACRYPT'99*, Singapore, Lecture Notes in Computer Science 1716, pp. 8–19, Springer-Verlag, 2000.
69. S. Vaudenay. Adaptive-Attack Norm for Decorrelation and Super-Pseudorandomness. In *Selected Areas in Cryptography'99*, Kingston, Ontario, Canada, Lecture Notes in Computer Science 1758, pp. 49–61, Springer-Verlag, 2000.
70. S. Vaudenay. On Provable Security for Conventional Cryptography. Invited talk. In *Information Security and Cryptology ICISC'99*, Seoul, Korea, Lecture Notes in Computer Science 1787, pp. 1–16, Springer-Verlag, 1999.
71. 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.
72. D. Wagner. The Boomerang Attack. In *Fast Software Encryption'99*, Roma, Italy, Lecture Notes in Computer Science 1636, pp. 156–170, Springer-Verlag, 1999.
73. 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.

A Proof of Theorem 6

Proof. We consider the best adaptive distinguisher $\mathcal{A}$ between F and F^*. Without loss of generality we can assume that it is deterministic and it never asks the same query twice. Let $x_i = f_i(y_1, \ldots, y_{i-1})$, $x = (x_1, \ldots, x_d)$, and $y = (y_1, \ldots, y_d)$. Let $\mathcal{A}$ be the set of accepted y by the distinguisher. The decorrelation distance is simply twice its advantage, hence

$$||[F]^d - [F^*]^d||_a = 2 \sum_{y \in \mathcal{A}} \left([F]^d_{x,y} - [F^*]^d_{x,y} \right) .$$

Obviously this sum is maximal when $\mathcal{A}$ consists of all y for which we have $[F]^d_{x,y} > [F^*]^d_{x,y}$. Since the full sum is zero, we have

$$||[F]^d - [F^*]^d||_a = \sum_y \left| [F]^d_{x,y} - [F^*]^d_{x,y} \right| .$$

The sample space for F has cardinality 2^{md}. Hence for any multi-point y the probability $[F]^d_{x,y}$ can be written $j.2^{-md}$ for some integer j which is at most 2^{md}. Let x be a multi-point. Let N_j be the number of y multi-points such that $[F]^d_{x,y} = j.2^{-md}$. Note that $[F^*]^d_{x,y} = 2^{-md}$ since all x_i are pairwise different. We have

$$||[F]^d - [F^*]^d||_a = \sum_{j=0}^{2^{md}} N_j |j - 1| 2^{-md}.$$

Since we have $\sum_j N_j = 2^{md}$ and

$$\sum_j N_j . j 2^{-md} = \sum_y [F]^d_{x,y} = 1$$

we obtain $\sum_j N_j |j - 1| 2^{-md} = 2N_0 2^{-md}$. Hence

$$||[F]^d - [F^*]^d||_a = 2N_0 2^{-md}.$$

So we only need to count the number of y so that $[F]^d_{x,y} = 0$, *i.e.* the number of unreached multi-points y by the $(K_1, \ldots, K_d) \mapsto y$ mapping.

Let $z_i = K_1 + K_2 x_i + \ldots + K_d x_i^{d-1} \bmod p$ and $z = (z_1, \ldots, z_d)$. We have $y_i = z_i \bmod 2^m$ so z is a lift of y. If y is unreached, it means that none of its lift are reached. We can thus map at least one (unreached) $z(y)$ in an injective way to any unreached y multi-point. Hence we have at least N_0 unreached multi-points z by the $(K_1, \ldots, K_d) \mapsto z$

mapping. We notice that this mapping has no collision: if $K = (K_1, \ldots, K_d)$ and $K' = (K'_1, \ldots, K'_d)$ lead to the same z, they lead to the same y, therefore to the same x, so we must have $K = K'$ due to interpolation reasons. Hence at least 2^{md} multi-points z are reached, so we have $2^{md} + N_0 \leq p^d$. So N_0 is less than $p^d - 2^{md}$ and we have $||[F]^d - [F^*]^d||_a \leq 2((1+\delta)^d - 1)$. □

B On Super-Pseudorandomness

Super-pseudorandomness addresses the distinguishability of random *permutations* with distinguishers which can submit inputs or outputs to the oracle and get the corresponding output or input in return. Fig. 8 depicts a general distinguisher limited to d queries which are either chosen inputs or chosen outputs. For completeness we state here the results without a proof.

Parameters: functions $f_1, \ldots, f_d$, a set $\mathcal{A}$
Oracle: permutations c and c^{-1}
1: select a fixed direction and message $(B_1, Z_1^0) = f_1()$ and get $Z_1^1 = c(Z_1^0)$ if $B_1 = 0$ or $Z_1^1 = c^{-1}(Z_1^0)$ otherwise
2: calculate a direction and a message $(B_2, Z_2^0) = f_2(Z_1^1)$ and get $Z_2^1 = c(Z_2^0)$ if $B_2 = 0$ or $Z_2^1 = c^{-1}(Z_2^0)$ otherwise
3: ...
4: calculate a direction and a message $(B_d, Z_d^0) = f_d(Z_1^1, \ldots, Z_{d-1}^1)$ and get $Z_d^1 = c(Z_d^0)$ if $B_d = 0$ or $Z_d^1 = c^{-1}(Z_d^0)$ otherwise
5: if $(Z_1^1, \ldots, Z_d^1) \in \mathcal{A}$, output 1, otherwise output 0

Fig. 8. A General d-Limited Distinguisher with Chosen Inputs or Outputs.

In the same way that we define the $||.||_a$ norm with Equation (4), we define a $||.||_s$ norm on $\mathbf{R}^{\mathcal{M}^d \times \mathcal{M}^d}$ by

$$||A||_s = \max_{b_1 \in \{0,1\}} \max_{z_1^0} \sum_{z_1^1} \cdots \max_{b_d \in \{0,1\}} \max_{z_d^0} \sum_{z_d^1} |A_{(z_1^{b_1}, \ldots, z_d^{b_d}),(z_1^{1-b_1}, \ldots, z_d^{1-b_d})}|.$$

We can transform the matrix A into a matrix $\bar{A}$ on $\mathbf{R}^{\mathcal{M}_1^d \times \mathcal{M}^d}$ where $\mathcal{M}_1 = \{0,1\} \times \mathcal{M}$ by

$$\bar{A}_{((b_1,z_1^0),\ldots,(b_d,z_d^0)),(z_1^1,\ldots,z_d^1)} = A_{(z_1^{b_1},\ldots,z_d^{b_d}),(z_1^{1-b_1},\ldots,z_d^{1-b_d})}.$$

Then we have $||A||_s = ||\bar{A}||_a$. We easily deduce that

- $||.||_s$ is actually a matrix norm,
- we have $|||A|||_\infty \leq ||A||_a \leq ||A||_s$ for any A,
- the following equivalent of Theorem 9 holds.

Theorem 29 ([69]). *Let d be an integer and let C be a cipher. The best d-limited distinguisher (as depicted in Fig. 8) for C is such that*

$$\mathrm{Adv}_{\text{Fig. 8}} = \frac{1}{2}||[C]^d - [C^*]^d||_s$$

where C^* is the perfect cipher.

The equivalent of Lemma 22 is as follows.

Lemma 30 ([70]). *Let d be an integer. Let C be a random permutation on a set $\mathcal{M}$. We let $\mathcal{X}$ be the subset of $\mathcal{M}^d$ of all $(x_1, \ldots, x_d)$ with pairwise different entries. We let C^* be a uniformly distributed random permutation on $\mathcal{M}$. We know that for all $x, y \in \mathcal{X}$, the value $[C^*]^d_{x,y}$ is a constant p_0. If there exists a positive real value ε such that*

$$\forall x, y \in \mathcal{X} \quad [C]^d_{x,y} \geq p_0(1-\varepsilon)$$

then we have $||[C]^d - [C^*]^d||_s \leq 2\varepsilon$. *Similarly, if there exists a positive real value* ε *such that*

$$\forall x, y \in \mathcal{X} \quad [C]^d_{x,y} \geq \frac{1-\varepsilon}{\#\mathcal{M}^d}$$

then we have $||[C]^d - [C^*]^d||_s \leq 2\varepsilon + \frac{d^2}{\#\mathcal{M}}$.

Lemma 19 can also be stated in terms of super-pseudorandomness.

Lemma 31 (Luby-Rackoff [40]). *Let* F_1, F_2, F_3, F_4 *be four independent uniformly distributed random functions on* $\mathcal{M}_0$ *and let* d *be an integer. We consider the cipher* $C = \Psi(F_1, F_2, F_3, F_4)$ *on* $\mathcal{M} = {\mathcal{M}_0}^2$ *and we compare it with the perfect cipher* C^*. *For any distinguisher* $\mathcal{A}$ *between* $\Psi(F_1, F_2, F_3, F_4)$ *and* C^* *which is limited to* d *chosen inputs or outputs, we have*

$$\mathrm{Adv}_{\mathcal{A}} \leq \frac{d^2}{\sqrt{\#\mathcal{M}}}.$$

The consequence is the following equivalent of Theorem 21.

Theorem 32 ([69]). *Let* $F_1, \ldots, F_r$ *be* r *independent random functions on* $\mathcal{M}_0$ *such that* $||[F_i]^d - [F^*]^d||_a \leq \varepsilon$ *(*$i = 1, \ldots, r$*) where* F^* *is a uniformly distributed random function on* $\mathcal{M}_0$. *We consider the cipher* $C = \Psi(F_1, \ldots, F_r)$ *on* $\mathcal{M} = {\mathcal{M}_0}^2$ *and we compare it with the perfect cipher* C^*. *Let* $k \geq 4$ *be an integer. For any distinguisher* $\mathcal{A}$ *between* C *and* C^* *which is limited to* d *chosen inputs or outputs, we have*

$$\mathrm{Adv}_{\mathcal{A}} \leq \frac{1}{2}\left(k\varepsilon + \frac{2d^2}{\sqrt{\#\mathcal{M}}}\right)^{\lfloor \frac{r}{k} \rfloor}.$$

Finally, Theorem 25 extends as follows.

Theorem 33. *Let* C *be a cipher in the PEANUT family with parameters* (m, r, d, p). *We have*

$$||[C]^d - [C^*]^d||_s \leq \left(\left(1 + 2\left(p^d 2^{-\frac{md}{2}} - 1\right)\right)^4 - 1 + \frac{2d^2}{2^{\frac{m}{2}}}\right)^{\lfloor \frac{r}{4} \rfloor},$$

where C^* *is the perfect cipher.*

The parameters of PEANUT98 lead to $||[C]^2 - [C^*]^2||_s \leq 2^{-48}$. The parameters of DFC lead to $||[C]^2 - [C^*]^2||_s \leq 2^{-112}$.