Pseudorandom permutation families over abelian groups

Louis Granboulan, Éric Levieil, and Gilles Piret**

École Normale Supérieure
Louis.Granboulan@ens.fr,Eric.Levieil@ens.fr,
Gilles.Piret@ens.fr

Abstract. We propose a general framework for differential and linear cryptanalysis of block ciphers when the block is not a bitstring. We prove piling-up lemmas for the generalized differential probability and the linear potential, and we study their lower bounds and average value, in particular in the case of permutations of $\mathbb{F}_p$. Using this framework, we describe a toy cipher, that operates on blocks of 32 decimal digits, and study its security against common attacks.
Keywords: block cipher, arbitrary domain, differential and linear cryptanalysis.

1 Introduction

1.1 Motivations

While all well-known block ciphers are pseudo-random permutation families of some set $\{0,1\}^n$ where $n = 64$ or 128, there exists some applications where a pseudo-random permutation of an arbitrary set is needed. For example, if one wants to add an encryption layer within a system that stores its data in decimal value, this encryption layer should encrypt decimal numbers without any expansion, and this is not possible if a binary encoding of these numbers is encrypted by a standard block cipher. Another example appears in some public-key cryptography protocols, where one assumes the existence of some ideal permutation or of some ideal cipher, that permutes elements of a set of cardinality other than 2^n, for example the set of points of an elliptic curve.

Moreover, while there are many studies on the cryptographic properties of boolean functions, and some studies on the cryptographic properties of addition modulo 2^n, no published results really looks into the generalization of these binary properties to the case where the characteristic is $\ell > 2$. A general framework for differential and linear cryptanalysis for arbitrary characteristic may bring a new insight into the understanding of these attacks.

** This work is supported in part by the French government through X-Crypt, in part by the European Commission through ECRYPT.

1.2 Previous work

Black and Rogaway [3] have described how to design block ciphers that permute arbitrary domains. Hence our problem already has a solution. However, their techniques are modes of use of conventional ciphers, and we prefer to study the feasibility of ad hoc designs.

The generalization of differential cryptanalysis to any abelian group is classical, and this generalization appears in the study of ciphers using addition modulo 2^n [10, 16, 11] but also more exotic operations like the $\otimes$ in IDEA [10] or multiplication in Multiswap [4]. Nyberg [13] wrote one of the few papers that study S-boxes over $\mathbb{F}_p$ with respect to differential cryptanalysis[1]. The generalization of linear cryptanalysis is a new result, the only similar work being the $\mathbb{Z}_4$-linear cryptanalysis by Parker and Raddum [15].

Our toy cipher we describe is based on a straightforward adaptation of Rijndael [6], which is a typical example of key-alternating cipher [8].

1.3 Our setting

To study the differential cryptanalysis of a function $f : G \to G'$, we need to provide both G and G' with a structure of abelian group. The number of elements in these groups will be denoted q and q'. The minimal integer ℓ such that all elements of these groups are of ℓ-torsion (i.e. $\ell.G = \ell.G' = \{0\}$) will be called the characteristic of f and be a key parameter for linear cryptanalysis. We will investigate more deeply the prime case, where $G = G' = \mathbb{F}_p$.

1.4 Outline of the paper

Sections 2 and 3 explain how differential and linear cryptanalysis are generalized. We give definitions and basic properties, then we show that *piling-up* theorems exist, and therefore these techniques can be used to evaluate the security of a whole cipher, based on the study of its non-linear components. We also show that in the prime case, optimality is equivalent to f being a degree 2 polynomial. Finally, we give an estimation of the non-linearity of a random function, with respect to differential cryptanalysis.

In section 4 we describe our cipher TOY100. We explain the design criteria. Because our toy cipher has non-prime characteristic $\ell = 100$, some technical difficulties appear in the study of the linear part of the cipher. We solve the problem for our specific example. Section 5 is a security analysis of TOY100.

2 Differential cryptanalysis

2.1 Definition

Introduced by Biham and Shamir in [2], differential cryptanalysis is one of the most useful techniques in modern cryptanalysis. The idea is to encrypt pairs of

[1] But there is a small mistake in its proposition 7.

plaintexts having a fixed difference, and to observe the differences between the pairs of ciphertexts.

We recall that in our setting f is a function from G to G', abelian groups of cardinality q and q'. Let us define a $q \times q'$ matrix Δ that describes the action of f over differences by $\Delta(f)_{a,b} = \#\{x | f(x+a) - f(x) = b\}$.

The complexity of an attack by differential cryptanalysis is of order $1/DP(f)$ where the differential probability $DP(f) = D(f)/q$ is defined by:

$$D(f) = \max_{(a,b) \in G \times G' \setminus \{(0,0)\}} \Delta(f)_{a,b}.$$

The exact value of $D(f)$ being too expensive to compute, one usually computes the exact values for the elementary functions used in f and combine these values using piling-up theorems.

2.2 Properties valid for any group

Differential probability for the inverse. If $f : G \to G$ is bijective, then $\Delta(f^{-1}) = {}^t\Delta(f)$ and therefore $D(f^{-1}) = D(f)$.

Proof. If $f(x+a) - f(x) = b$, then $f^{-1}(y) + a = f^{-1}(y+b)$ with $y = f(x)$. □

Parallel execution. If f is the parallel execution of functions f_1 and f_2, then its differential properties are easily deduced from the differential properties of f_1 and f_2. More precisely, if we define f over $G_1 \times G_2$ by $f(x,y) = (f_1(x), f_2(y))$, then $\Delta(f)_{(a_1,a_2),(b_1,b_2)} = \Delta(f_1)_{a_1,b_1} \Delta(f_2)_{a_2,b_2}$ and $D(f) = \max(q_2 D(f_1), q_1 D(f_2))$.

Sequential execution. If f is the sequential execution of two functions, the differential properties cannot be directly combined, because the image of a uniform distribution of input pairs with fixed difference does not necessarily have uniform distribution for all output pairs with given difference. The distribution can be made uniform by adding a random key, and ciphers using this design are named Markov ciphers [10]. In this setting, we compose the function $f : G \to G'$, the translation in G' that we name ADD_KEY, and the function $g : G' \to G''$ to obtain $h_K = g \circ \text{ADD\_KEY}(K) \circ f$.

Theorem 1.

$$DP(h_K) \approx DP(f) DP(g)$$

if the following hypothesis hold:

- *Stochastic equivalence.* $\Delta(h_K)$ *does not depend heavily on* K*;*
- *Dominant characteristic for* (a,c)*.* $\sum_b \Delta(f)_{a,b} \Delta(g)_{b,c} \approx \max_b \Delta(f)_{a,b} \Delta(g)_{b,c}$.
- *Independence.* $\max_b \Delta(f)_{a,b} \Delta(g)_{b,c} \approx \max_{b,b'} \Delta(f)_{a,b} \Delta(g)_{b',c}$.

Proof. Let $\phi_a(x,K) = (x,\ f(x)+K,\ f(x+a)-f(x))$. The restriction of ϕ_a to the set of solutions (x,K) of the equation $h_K(x+a)-h_K(x)=c$ is a one-to-one mapping to the set of solutions (x,y,b) of the pair of equations $g(y+b)-g(y)=c$ and $f(x+a)-f(x)=b$. Therefore the following formula over Δ matrices holds:

$$\sum_{K\in G'} \Delta(h_K) = \Delta(f)\Delta(g).$$

Under the first hypothesis, $D(h_K) \approx \frac{1}{q'} \max_{(a,b)\neq(0,0)} \big(\Delta(f)\Delta(g)\big)_{a,b}$. Now we apply the second hypothesis to a pair (a,b) for which $D(h_K) \approx \frac{1}{q'}\big(\Delta(f)\Delta(g)\big)_{a,b}$, and then we apply the third hypothesis. □

Lower bound. If $D(f)=1$ then f is not bijective.

Proof. For any non-zero a, $\sum_{b\in G}\Delta(f)_{a,b}=q$, therefore all elements of the a-th row of $\Delta(f)$ are equal to 1 and in particular $\Delta(f)_{a,0}$. □

2.3 The case of $\mathbb{F}_p$

All functions in $\mathbb{F}_p$ can be interpolated by a polynomial which is unique if its degree is less than p. The degree of f has some impact on $D(f)$.

Proposition 1. *(i) $D(f)=p$ is equivalent to f linear or constant.*
(ii) $D(f)=p-1$ is impossible.
(iii) If f has degree $d\geq 2$, then $D(f)\leq d-1$. In particular, if f is of degree 2, then $D(f)=1$.
(iv) For all d between 2 and $p-1$, there are polynomials of degree d, such as $D(f)=d-1$.

Proof. (i): Let $a\neq 0$ and b be such that $\Delta(f)_{a,b}=p$. Then $f(x)=a^{-1}bx+f(0)$.
(ii): Let $a\neq 0$ and b be such that $\Delta(f)_{a,b}=p-1$. There exists $b'\neq b$ such that $\Delta(f)_{a,b'}=1$. But $0=\sum_{x\in G} f(x+a)-f(x)=(p-1)b+b'=b'-b$.
(iii): $f(x+a)-f(x)-b$ is a polynomial of degree $d-1$, so it has at most $d-1$ roots.
(iv): We want to find f such that $f(x+1)-f(x)$ is a polynomial with $d-1$ distinct roots. First, we choose any polynomial with $d-1$ distinct roots then we write the equality between the coefficients. We obtain a triangular system with a non-zero diagonal, which implies it is invertible. □

Conjecture for the lower bound.

Conjecture 1. If $D(f)=1$, then the degree of f is 2.

If we define the differential $df_a(x)=f(x+a)-f(x)$, it has the property of being a zero-sum function i.e. $\sum_{x\in G} df_a(x)=0$. The hypothesis $D(f)=1$ of our conjecture is equivalent to $\forall a\neq 0,\ df_a$ is bijective.

In spite of this simple formulation, and a computer-aided verification that it is true for $p\leq 19$, we could not prove this conjecture. However, if the following lemma holds, then this conjecture is true, as shown in appendix A.5.

Lemma 1 (Key lemma). *If $\varphi : \mathbb{F}_p \to \mathbb{Z}$ satisfies $\sum_{y\in\mathbb{F}_p} \varphi(y)^2 = p-1$, and $\forall x \neq 0,\ \sum_{y\in\mathbb{F}_p} \varphi(y)\varphi(x+y) = -1$ then $\forall x,\ \varphi(x) \in \{0, \pm 1\}$.*

Average value. To find functions with high degree but low $D(f)$, we can try random functions. The following theorem evaluates the average value of $D(f)$ and its proof (in appendix A.1) contains upper bounds on the number of functions with low or high $D(f)$.

Theorem 2. *Let us define $z(p) = \lfloor \Gamma^{-1}(p/(6\log p)) \rfloor - 1$ where as usual $\Gamma(z+1) = z!$, then*

$$\lim_{p\to\infty} Pr[z(p) \leq D(f) \leq 3z(p)] = 1$$

It is possible to decrease the constant 3 to 2 (the proof will be in the full version of the paper). There is no reason that prevents this result to be applied to $\mathbb{Z}/q\mathbb{Z}$, except perhaps the human's lack of taste for lengthy computations. However, it is impossible to have really precise results on this subject, unless one can explicit the dependence between the differentials of a function. Assuming independence is the usual way to deal with this problem (see for example [8]), but it is not true for small p.

2.4 The case of $\mathbb{Z}/q\mathbb{Z}$

The case where G is isomorphic to $\mathbb{Z}/q\mathbb{Z}$ cannot be seen as a generalization of the prime case for two reasons:

- there exist many functions that cannot be interpolated by polynomials
- even when this interpolation exists, the form of canonical interpolations is tricky to define

The following theorem, proven in appendix A.2, shows that polynomials are a negligible fraction of the functions over $\mathbb{Z}/q\mathbb{Z}$. For example, over $\mathbb{Z}/100\mathbb{Z}$ there are $2 \cdot 10^{12}$ polynomials and 10^{200} functions.

Theorem 3. *(i) Let $q = p^2$ with p prime. Then the number of distinct polynomials over $\mathbb{Z}/q\mathbb{Z}$ is equal to p^{3p}.*
(ii) Let $q = q_1q_2$, with q_1, q_2 coprime. Then the number of distinct polymials over $\mathbb{Z}/q\mathbb{Z}$ is the product of this number over $\mathbb{Z}/q_1\mathbb{Z}$ and $\mathbb{Z}/q_2\mathbb{Z}$.

If $q = q_1q_2$, with q_1, q_2 coprime, and if f is a polynomial, then its differential properties need only to be studied over $\mathbb{Z}/q_1\mathbb{Z}$ and over $\mathbb{Z}/q_2\mathbb{Z}$, as proved in the following theorem. If it is not a polynomial, such a decomposition is not possible.

Theorem 4. *Let $f \in \mathbb{Z}/q\mathbb{Z}[X]$ and for $i = 1, 2$ $f_i \in \mathbb{Z}/q_i\mathbb{Z}$ defined by $f_i(x) = f(x) \pmod{q_i}$. Then $D(f) = D(f_1)D(f_2)$.*

Proof. $z \to (z \bmod q_1, z \bmod q_2)$ is an isomorphism. □

3 Linear cryptanalysis

3.1 Definition

Linear cryptanalysis is a known-plaintext attack that was discovered just after differential cryptanalysis [18, 17, 12]. It is based on the study of linear approximations of the cipher. Linear cryptanalysis has been defined for boolean functions: a linear approximation of a function f is described by two masks (a, b) which select respectively bits of the input and of the output. If we denote by $\langle a|x\rangle$ the dot product of a and x, then linear approximations are given by comparing $\langle a|x\rangle - \langle b|f(x)\rangle$ for random x and $\langle a|x\rangle - \langle b|y\rangle$ for random x and y.

Linear cryptanalysis can be generalized to the study of the functions $f : G \to G'$, if there is some integer q such that all elements of both groups are of ℓ-torsion. This condition implies that both G and G' are isomorphic to a product of cyclic groups of order dividing ℓ. Under this condition and using this isomorphism, we can define scalar products over G and G' with output in $\mathbb{Z}/\ell\mathbb{Z}$, denoted $\langle\cdot|\cdot\rangle$. And finally we define the scalar product on $G \times G'$ by $\langle a, b|x, y\rangle = \langle a|x\rangle - \langle b|y\rangle$. The generalization of linear cryptanalysis can be done using two approaches.

Bias from random behavior. For any pair $(a, b) \in (G, G')$, let us define the distribution vector $\Lambda_0(f)_{a,b} = (\#\{x \in G \mid \langle a, b|x, f(x)\rangle = u\})_{u\in\mathbb{Z}/\ell\mathbb{Z}}$. The random behavior is given by $S_{a,b;u} = \frac{1}{q'}\#\{(x, y) \in G \times G' \mid \langle a, b|x, y\rangle = u\}$.

Therefore, if we define the bias $\Lambda_S(f)_{a,b;u} = \Lambda_0(f)_{a,b;u} - S_{a,b;u}$, then all elements of this matrix sum up to zero $\sum_u \Lambda_S(f)_{a,b;u} = 0$ and its greatest term is a measure of non-linearity.

$$L(f) = \max_{a,b\neq 0,u} (\Lambda_S(f)_{a,b;u})^2$$

The complexity of the attack is expected to be of order $1/LP(f)$, where the linear potential $LP(f) = L(f)/q^2$.

Dual of differential cryptanalysis. The other approach generalizes the duality between differential and linear cryptanalysis, as it has been done for example by Chabaud and Vaudenay [5]. First, we need to define the characteristic function of f, which is $\theta_f : G \times G' \to \{0, 1\}$ such that $\theta_f(x, y) = 1$ iff $y = f(x)$. We also define the convolutional product of two functions by $(f*g)(a) = \sum_x f(x)g(a+x)$. As in Chabaud-Vaudenay, we can prove that $(\theta_f * \theta_f)(a, b) = \Delta_{a,b}$.

Let us choose a ℓ-th root[2] of unity $\xi \in \mathbb{C}$ and define the transform of $\phi : X \to Y$ by $\hat{\phi}(a) = \sum_x \phi(x)\xi^{\langle a|x\rangle}$. Note that $\hat{\phi}(-a)$ and $\hat{\phi}(a)$ are complex conjugates, that $\hat{\hat{\phi}}(x) = \#Y.\phi(-x)$, and also that $\widehat{(\phi * \phi)} = |\hat{\phi}|^2$ and therefore is real-valued. By duality, we define $\lambda(f)_{a,b} = (\widehat{\theta_f * \theta_f})(a, b)$ and $\lambda(f) = \max_{(a,b)\neq(0,0)} \lambda(f)_{a,b}$.

[2] Replacing -1 by a ℓ-th root of unity is not a new idea. For example, it appeared as footnote 4 of [1]. The fact that it is a different approach than computing the bias was probably not noticed.

Links between both approaches. In the binary case (i.e. $\ell = 2$) we have $\xi = -1$ and $L(f)_{a,b;1} = -L(f)_{a,b;0}$ therefore $\hat{\theta}_f(a,b) = 2L(f)_{a,b;0}$ and $\lambda(f) = 4L(f)$. When $\ell \neq 2$, no such simple relation exists. For example, in $\mathbb{Z}/7\mathbb{Z}$, let us take $f(x) = x^6 + x^3$ and $g(x) = x^6 + x^3 + x^2$. Then $L(f) = L(g) = 9$ but $\lambda(f) = 39.96\cdots$ while $\lambda(g) = 26.19\cdots$. The list of all possible values for $\mathbb{Z}/5\mathbb{Z}$ and $\mathbb{Z}/7\mathbb{Z}$ is in appendix A.4.

Both approaches give some insight into the security of a cipher. However, in the following, we mainly consider the measure of bias, which is easier to implement as a concrete cryptanalysis.

3.2 Properties valid for any group

Main properties of $S_{a,b;u}$. When a, b are fixed, $S_{a,b;u}$ is either 0 or another fixed value denoted $S_{a,b}$. The set $T_{a,b} = \{u | S_{a,b;u} \neq 0\}$ is a subgroup of $\mathbb{Z}/\ell\mathbb{Z}$, and $q' = S_{a,b}\ \#T_{a,b}$.

Proof. If $\langle a, b | x_0, y_0\rangle = u_0$ and $\langle a, b | x_1, y_1\rangle = u_1$, then $\langle a, b | x_0 + x_1, y_0 + y_1\rangle = u_0 + u_1$. Therefore, the sets of solutions of the equations $\langle a, b | x, y\rangle = u$ can be translated one to another. $\square$

The inverse. If $f : G \to G$ is bijective, then $\Lambda_S(f^{-1})_{a,b;u} = \Lambda_S(f)_{a,b;-u}$ and therefore $L(f^{-1}) = L(f)$.

Parallel execution. If f is the parallel execution of functions f_1 and f_2 of same characteristic, then bias matrices are combined by convolution. More precisely, if we define f over $G_1 \times G_2$ by $f(x, y) = (f_1(x), f_2(y))$, then $\Lambda_0(f)_{(a_1,a_2),(b_1,b_2)} = \Lambda_0(f_1)_{a_1,b_1} * \Lambda_0(f_2)_{a_2,b_2}$. If the sets T_{a_1,b_1} and T_{a_2,b_2} are equal, then this formula also applies to Λ_S.

Proof. Note that the hypothesis on the sets T_{a_i,b_i} is mandatory. A simple counterexample is $G_1 = G_2 = \mathbb{Z}/100\mathbb{Z}$, $f_1(x) = 2x$, $f_2(x) = x$, $a_1 = 5$, $a_2 = 10$, $b_1 = b_2 = 0$ and $\langle .|.\rangle$ is the usual multiplication over $\mathbb{Z}/100\mathbb{Z}$.

To prove the formula for Λ_0, we decompose $\langle a, b | x, f(x)\rangle = u$ into its components $\langle a_1, b_1 | x_1, f_1(x_1)\rangle = v$ and $\langle a_2, b_2 | x_2, f_2(x_2)\rangle = u - v$. Then we use the following facts: $S^{G_1 \times G_2}_{a,b;u} = \sum_v S^{G_1}_{a_1,b_1;u-v} S^{G_2}_{a_2,b_2;v}$ and $\sum_{v \in T_{a,b}} \Lambda_G(f)_{a,b;v} = 0$. $\square$

Sequential execution. As for the differential cryptanalysis, we suppose we have a Markov cipher. In this case, the following theorem, proven in appendix A.3 allows us to approximate the value of $LP(h_K)$, for $h_K = g \circ \text{ADD\_KEY}(K) \circ f$:

Theorem 5 (Piling-up for LP).

$$LP(h_K) \approx LP(f)LP(g)$$

if the following hypothesis hold:

– *Stochastic equivalence.* $\Lambda_S(h_K)_{a,c;u+\langle b|K\rangle}$ *does not depend heavily on* K*;*
– *Dominant trail and independence.*

$$\max_{\substack{a,b,c,u\\T_{a,b}=T_{b,c}}}\left|\sum_v \Lambda_S(f)_{a,b;u-v}\Lambda_S(g)_{b,c;v}\right| \approx \max_{a,b_f,b_g,c,u,v}\left|\Lambda_S(f)_{a,b_f;u-v}\Lambda_S(g)_{b_g,c;v}\right|$$

Piling-up $\boldsymbol{\lambda(f)}$. This other approach also has composition results. For example, we prove in appendix A.3 a piling-up lemma that shows that under some appropriate hypothesis, $\lambda(h_K)/q^2 \approx \lambda(f)/q^2\lambda(g)/q'^2$.

3.3 The case of $\mathbb{F}_p$

Functions over $\mathbb{F}_p$ that have optimal resistance against linear cryptanalysis have degree 2.

Theorem 6. *Let G be a group of cardinality p, with p prime. If $L(f) = 1$, then f can be interpolated by a polynomial of degree* 2.

Proof. Let us work in $PG(2, p)$, the projective plane over $\mathbb{F}_p$. Let

$$E(f) = \{x, f(x), 1|x \in \mathbb{F}_p\} \cup (0, 1, 0).$$

$E(f)$ is a $p + 1$-arc, i.e. a set of $p + 1$ points, no three of which are collinear. According to the corollary of theorem 10.4.1, p.236, of [9], a $p+1$-arc in $PG(2, p)$ with p odd, is a conic. So

$$E(f) = \{(x_0, x_1, x_2)|a_{00}x_0^2 + a_{11}x_1^2 + a_{22}x_2^2 + a_{01}x_0x_1 + a_{02}x_0x_2 + a_{12}x_1x_2 = 0\}$$

But $(0, 1, 0) \in E(f)$, therefore $a_{11} = 0$.
And $(a_{01}, -a_{00}, 0) \in E(f)$, therefore $(a_{01}, -a_{00}, 0) \equiv (0, 1, 0)$. Therefore $a_{01} = 0$.
If $a_{12} = 0$, $(0, f(0), 1) \in E(f)$ implies $\{(0, y, 1)|y \in \mathbb{F}_p\} \subset E(f)$. Therefore a_{12} is not null and f is described by a degree 2 polynomial. □

3.4 Relation with the linear cryptanalysis over $\mathbb{Z}/4\mathbb{Z}$

Matthew Parker and Haavard Raddum have suggested a generalization of linear cryptanalysis over $\mathbb{Z}/4\mathbb{Z}$ in [15]. Their method allows better approximations of the S-boxes but the combination of those approximations is less efficient than in classical linear cryptanalysis. Their method is a very particular case of ours, where a $2n$-bit string is seen as an element of $(\mathbb{Z}/4\mathbb{Z})^n$.

4 A Toy Cipher: TOY100

4.1 High-level description

In this section we aim at showing that it is possible to design a secure and efficient block cipher that does not use words of n bits as a block.

The structure of the cipher is quite similar to Rijndael [6, 7]. It works on blocks of 32 decimal digits, with keys of the same size. It is composed of 11 identical rounds, followed by a slightly different final round. A block A is divided in 16 subblocks, each subblock being a number between 0 and 99. A block is represented as a 4×4 matrix $A = (a_{i,j})_{i,j \in \{0,\ldots,3\}}$, of which each element is a subblock. Round r ($r = 0 \ldots 10$) is made out of the application of a key addition layer $\sigma[K^r]$ which adds modulo 100 a subkey to each subblock, followed by the parallel application, denoted γ, of a certain S-box to each subblock, and finally a linear function θ that mixes the subblocks. The last round has a final key addition instead of the linear layer, so it is written as $\sigma[K^{11}] \circ \gamma \circ \sigma[K^{12}]$.

4.2 Our Choice of Components

The S-Box. The S-box was chosen to satisfy $D(f) \leq 5$ and $L(f) \leq 5^2$. An iteration of RC4-100 consists, being given an array of 100 numbers, and two pointers i, j, to increment i, add $t[i]$ to j (modulo 100) then exchange $t[i]$ and $t[j]$. Starting from the permutation identity, $i = 1$, $j = 0$, we checked the permutation every 100 iterations until we find a permutation satisfying the criteria on $D(f)$ and $L(f)$. The permutation found is the 3 409 672th, after 340 967 200 iterations of RC4-100.

This function has $D(f) = 5$, $L(f) = 5^2$ and $\lambda(f) = 734.122\cdots$

	0	1	2	3	4	5	6	7	8	9
0	0	67	12	32	30	53	34	37	71	38
10	42	94	58	95	78	35	6	22	36	81
20	61	93	43	72	25	27	15	69	90	47
30	1	91	84	86	24	79	66	40	10	33
40	59	8	11	48	28	76	73	82	39	51
50	45	13	97	74	9	7	52	88	62	96
60	23	29	3	4	75	56	5	64	17	49
70	68	77	80	55	85	92	44	21	98	50
80	20	31	65	83	19	57	41	70	18	99
90	89	60	46	26	63	14	87	16	54	2

The Diffusion Function The diffusion function θ is composed of two similar parts, `MixColumns` and `MixRows`.
First, we define a function Mix that takes 4 subblocks as an input:

$$Mix \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{pmatrix} = \begin{pmatrix} a_4 + a_1 + a_2 \\ a_1 + a_2 + a_3 \\ a_2 + a_3 + a_4 \\ a_3 + a_4 + a_1 \end{pmatrix}$$

Mix is bijective and its inverse is:

$$Mix^{-1}\begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{pmatrix} = \begin{pmatrix} S - a_3 \\ S - a_4 \\ S - a_1 \\ S - a_2 \end{pmatrix}$$

with $S = (a_1 + a_2 + a_3 + a_4)/3$.

`MixColumns` (resp. `MixRows`) consists in applying Mix to each column (resp. row). Note that `MixColumns` and `MixRows` commute.

We define the subblock weight of a block B as the number of non-zero subblocks, and we denote it as $SW(B)$. The *branch number* is a measure of the efficiency of a diffusion layer.

Definition 1. *The branch number of a diffusion function f, $BN(f)$ is defined as:*

$$BN(f) = \min_{B \neq 0}(SW(B) + SW(f(B))$$

Proposition 2.

$$BN(\theta) = 6$$

Proof. The first step of the proof enumerates the cases where there are one or two non-zero subblocks in the input B, and show that there will be at least six non-zero subblocks in $\theta(B)$; it is the same for $\theta^{-1}(B)$. We conclude by observing that if $b_{21} = b_{22} = b_{23} = 50$ and the other subblocks of B are 0, then $C = \theta(B)$ is such that $c_{12} = c_{22} = c_{32} = 50$ and the other subblocks are 0. $\square$

The Key Schedule The key expansion is very similar to the one of AES. As always, additions are modulo 100. The first round key K^0 is the key itself. For the following rounds, we iterate as follows:

$$\begin{aligned} k^{r+1}_{0,j} &= k^r_{0,j} + S(k^r_{3,(j+1) \bmod 4}) + 3^r && (j \in \{0,1,2,3\}) \\ k^{r+1}_{i,j} &= k^r_{i,j} + k^{r+1}_{i-1,j} && (i \in \{1,2,3\}, j \in \{0,1,2,3\}) \end{aligned}$$

5 Security Analysis of TOY100

5.1 Differential Cryptanalysis

The best differentials we found rely on the following property of the linear layer:

$$\begin{bmatrix} \delta & -\delta & 0 & 0 \\ -\delta & \delta & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \xrightarrow{\theta} \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \delta & -\delta \\ 0 & 0 & -\delta & \delta \end{bmatrix} \tag{1}$$

We estimated the probability of these differentials for $n = 2, 3, 4, 5, 6$. That is to say, we computed

$$\max_{\Delta_0,\Delta_n \in \{1,\dots,99\}} \sum_{\Delta_1,\dots,\Delta_{n-1}\in\{1,\dots,99\}} \frac{\Pi(\Delta_0 \to \Delta_1)^2 \cdot \ldots \cdot \Pi(\Delta_{n-1} \to \Delta_n)^2}{10^{8n}}, \quad (2)$$

where $\Pi(\Delta_i \to \Delta_j) := \Delta(f)_{\Delta_i,\Delta_j} \cdot \Delta(f)_{-\Delta_i,-\Delta_j}$. Remark that our choice of the linear transform makes the "modified difference distribution table" Π particularly important. There is always some "interaction" between the linear transform and the S-box regarding resistance against differential (and linear) cryptanalysis, but it is rarely so explicit.

Our results are given in Table 1. Note that the probabilities given are only lower bounds, as other characteristics exist for the same differential; however they have more active S-boxes, so we expect their contribution to the overall probability to be small. Such n-round differential can be used in an attack on $n+1$ rounds. This way we can attack up to 6 (and maybe 7) rounds. Details are given in appendix B.1.

Table 1. Estimated probability for the best n-round differential

# Rounds n	Best Probability
2	$4.05 \cdot 10^{-11}$
3	$2.83 \cdot 10^{-16}$
4	$2.61 \cdot 10^{-21}$
5	$2.72 \cdot 10^{-26}$
6	$3.47 \cdot 10^{-31}$

5.2 Linear Cryptanalysis

The best linear characteristic we found relies on the same type of observation as the one used for differential cryptanalysis.

Namely, θ transforms mask $\begin{bmatrix} \alpha & -\alpha & 0 & 0 \\ -\alpha & \alpha & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$ into mask $\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \alpha & -\alpha \\ 0 & 0 & -\alpha & \alpha \end{bmatrix}$.

The piling-up lemma can be iterated, so for an $(n+1)$-round characteristic we have

$$\begin{aligned} &\sum_{K^1,\dots,K^n \in\ (\mathbb{Z}/100\mathbb{Z})^{16}} \Lambda_S(h_{K^1,\dots,K^n})_{a,c;u+b_1K^1+\ldots+b_nK^n} \\ &= \sum_{v_1,\dots,v_n\in\mathbb{Z}/100\mathbb{Z}} \Lambda_S(\rho_{n+1})_{a,b_n;u-v_1-\ldots-v_n} \cdot \\ &\qquad \Lambda_S(\rho_n)_{b_n,b_{n-1};v_n} \cdot \ldots \cdot \Lambda_S(\rho_1)_{b_1,c;v_1} \end{aligned} \quad (3)$$

where $\rho_i = \gamma \cdot \theta$ $(i \neq n+1)$ and $\rho_{n+1} = \gamma$.
This equation holds under the hypothesis that

$$S_{a,b_n} = S_{b_n,b_{n-1}} = ... = S_{b_1,c}. \tag{4}$$

Informally, equation (3) gives the average bias taken over all n-tuples of round keys (the first and last round keys are not considered here; they only contribute to the linear equation by a constant, which is unknown). Note that for the equation to be useful for linear cryptanalysis, it is required that the characteristic roughly equally holds for all keys. This hypothesis is common; it is known as *hypothesis of stochastic equivalence* [10, 14]. We computed the maximum of (3) over all possible $(n+3)$-uples $(a, b_1, ..., b_n, c; u)$, for various numbers of rounds. The maxima we found correspond to $a = b_1 = ... = b_n = c = 10$ and $u = 0$. We note that condition (4) is satisfied. Detailed figures are given in Appendix B.2. Taking the first and last round keys into consideration, the corresponding linear approximation for $n+1$ rounds of the cipher is

$$\begin{aligned} &10 \cdot (c_{33} + c_{44} - c_{34} - c_{43}) - 10 \cdot (p_{11} + p_{22} - p_{12} - p_{21}) \\ &= \sum_{i=0}^{\lfloor \frac{n+1}{2} \rfloor} 10(k_{11}^{2i} + k_{22}^{2i} - k_{12}^{2i} - k_{21}^{2i}) + \sum_{i=1}^{\lceil \frac{n+1}{2} \rceil} 10(k_{33}^{2i-1} + k_{44}^{2i-1} - k_{34}^{2i-1} - k_{43}^{2i-1}), \end{aligned} \tag{5}$$

if r is odd; $c_{33} + c_{44} - c_{34} - c_{43}$ must be replaced by $c_{11} + c_{22} - c_{12} - c_{21}$ if it is even.

5.3 Structural Attacks

The diffusion layer of our cipher operates on well-aligned blocks, which could make it vulnerable to structural attacks. We explored truncated differential, impossible differential, and square attacks. The best such attack we found is a square-like attack, which can be used for a practical cryptanalysis of up to 4 rounds of TOY100. Details are given in appendix B.3.

6 Conclusion

In this paper we extended usual block cipher theory over $\mathbb{Z}_2^n$ to a more general framework in which the input and output spaces are arbitrary abelian groups. We studied quite extensively how differential and linear cryptanalysis apply in this context. We observe that many concepts, such as differential and linear parameters of a function or piling-up lemmas, can be generalized. Moreover, constructing a cipher by using the classical key-alternating paradigm still seems to be appropriate.

However several problems remain unsolved. The link between the differential parameter $D(f)$ and linear parameters $L(f)$ and $\lambda(f)$ should be investigated. Constructing functions with good such parameters, without using some kind of random search, is an open problem as well. A formalization of the "special role" of elements of small characteristic is also a goal for further research. Finally, our toy cipher would deserve a more consequent cryptanalytic effort.

Acknowledgement

The idea of the proof of theorem 6 was found by Mathieu Dutour and David Madore. We also thank David Madore for the proof of theorem 3.

References

1. T. Baignères, P. Junod, and S. Vaudenay. How Far Can We Go Beyond Linear Cryptanalysis? *Advances in Cryptology - Asiacrypt 2004*, LNCS 3329, Springer-Verlag, 2004. `http://lasecwww.epfl.ch/php_code/publications/search.php?ref=BJV04`
2. E. Biham and A. Shamir. Differential Cryptanalysis of DES-like cryptosystems. *Advances in Cryptology, CRYPTO '90*, Springer-Verlag, pp. 2-21.
3. John Black and Phillip Rogaway. Ciphers with Arbitrary Finite Domains. *RSA Data Security Conference, Cryptographer's Track (RSA CT '02)*, LNCS, vol. 2271, pp. 114-130, Springer, 2002.
4. Nikita Borisov, Monica Chew, Rob Johnson, and David Wagner. Cryptanalysis of Multiswap. 2001. `http://www.cs.berkeley.edu/~rtjohnso/multiswap/`
5. F. Chabaud and S. Vaudenay. Links between differential and linear cryptalysis. *Advances in Cryptology, Proceedings Eurocrypt'94*, LNCS 950, Springer-Verlag, 1995, pp.356-365.
6. J. Daemen and V. Rijmen. The Design of Rijndael: AES- the Advanced Encryption Standard. Springer-Verlag, 2002.
7. J. Daemen and V. Rijmen. AES proposal: Rijndael. *First Advanced Encryption Standard (AES) Conference*, Ventura, Canada National Institute of Standards and Technology, 1998.
8. J. Daemen et V. Rijmen. Statistics of Correlation and Differentials in Block Ciphers. Cryptology ePrint Archive, Report 2005/212, 2005 `http://eprint.iacr.org/2005/212`
9. J.W.P. Hirschfeld. Projective Geometries Over Finite Fields. *Oxford University Press, Oxford.* 1979.
10. X. Lai, J.L. Massey, and S. Murphy. Markov ciphers and differential cryptanalysis. *Advances in Cryptology, Proceedings Eurocrypt'91*, LNCS 547, D.W. Davies, Ed., Springer-Verlag, 1991, pp. 17-38.
11. Helger Lipmaa, Johan Wallén and Philippe Dumas. On the Additive Differential Probability of Exclusive-Or. In Bimal Roy and Willi Meier, editors, *Fast Software Encryption 2004*, volume 3017 of Lecture Notes in Computer Science, pages 317–331, Delhi, India, February 5–7, 2004. Springer-Verlag.
12. M. Matsui. Linear cryptanalysis method for DES cipher. *Advances in Cryptology, Proceedings Eurocrypt'93*, LNCS 765, T. Helleseth, Ed., Springer-Verlag, 1994, pp. 386-397.
13. K. Nyberg. Differentially uniform mappings for cryptography. *Advances in Cryptology, Proceedings Eurocrypt'93*, LNCS 765, T. Helleseth, Ed., Springer-Verlag, 1994, pp. 55-64.

14. K. Nyberg. Linear Approximation of Block Ciphers. *Advances in Cryptology, Proceedings Eurocrypt'94*, LNCS 950, pages 439–444. Springer-Verlag, 1995.
15. M.G.Parker and H.Raddum. $\mathbb{Z}_4$-Linear Cryptanalysis. NESSIE Internal Report, 27/06/2002: NES/DOC/UIB/WP5/018/1
16. B. Schneier, J. Kelsey, D. Whiting, D. Wagner, C. Hall, and N. Ferguson. New Results on the Twofish Encryption Algorithm. *Second AES Candidate Conference*, April 1999.
17. M. Matsui, A. Yamagishi. A New Method for Known Plaintext Attack of FEAL Cipher. *Advances in Cryptology, Proceedings Eurocrypt'92*, pages 81–91.
18. A. Tardy-Corfdir, H. Gilbert. A Known Plaintext Attack of FEAL-4 and FEAL-6. *Advances in Cryptology, CRYPTO 1991*, pages 172–181.

A Proofs

A.1 Average value of $D(f)$ in the prime case

We prove theorem 2.

We note $f^{<-1>}(y)$ the preimage of y and we define the function bp (biggest preimage) as $bp(f) = \max_y \# f^{<-1>}(y)$.

If $df_1 = dg_1$, then $f - g$ is a constant. Moreover, $bp(df_1) \leq D(f)$. Therefore, the function $f \to (df_1, f(0))$ is injective from the set of functions with $D(f) < k$ to the product of the set of functions with $bp < k$ by G.

We define $C_{k,p}$ as $\binom{p}{k}(p-1)^{p-k}$.

Using the precedent remark and lemma 2 just below, we deduce that:

$$Pr[D(f) < k] \leq p^{1-p} \#\{f \mid bp(f) < k\} \leq p(1 - \frac{C_{k,p}}{p^p})^p$$

$z(p)$ satisfies $z(p)! \leq \dfrac{p}{6 \log p} < (z(p)+1)!$

Then $C_{z(p),p} \sim \dfrac{p^p}{ez(p)!}$

For p big enough, we have $\dfrac{C_{z(p),p}}{p^p} \geq 2 \log p / p$ and therefore

$$\lim_{p \to \infty} Pr[D(f) < z(p)] = 0$$

If $D(f) > 3z(p)$ there is an x such that $bp(df_x) > 3z(p)$. But knowing x, df_x, and $f(0)$ determines uniquely f.

Therefore $\Pr[D(f) > 3z(p)] \leq p^{2-p} \#\{f \mid bp(f) > 3z(p)\}$.

Using lemma 2 just below, we obtain that $\Pr[D(f) > 3z(p)] \leq p^{3-p} C_{3z(p),p}$.

Using Stirling's formula, we deduce:

$$\lim_{p \to \infty} Pr[D(f) > 3z(p)] = 0.$$

Lemma 2. *$\#\{f \mid bp(f) = k\} \leq pC_{k,p}$ and $\#\{f \mid bp(f) < k\} \leq p^p(1 - \frac{C_{k,p}}{p^p})^p$.*

Proof. First, we remark that $\#\{f \mid \text{the cardinality of the preimage of } i = k\} = C_{k,p}$

(i) If $bp(f) = k$ then it exists $i \in G$ such as $\#f^{<-1>}(i) = k$. We conclude using the above remark and the fact that the cardinality of an union is upper-bounded by the sum of cardinalities.

(ii) If $bp(f) < k$ then for all y we have $\#\{x \mid f(x) = y\} < k$, and also $\#\{f|\#f^{<-1>}(y) < k\} \leq p^p - \#\{f|\#f^{<-1>}(y) = k\} \leq p^p(1 - \frac{C_{k,p}}{p^p})$.

Those events are anti-correlated, i.e. if an element has a small preimage, then the probability that the other elements have also a small preimage is smaller. So we can bound the global probablity by the product of probabilities.

Therefore

$$\#\{f \mid bp(f) < k\} \leq p^p \left(1 - \frac{C_{k,p}}{p^p}\right)^p$$

□

A.2 Counting polynomial over $\mathbb{Z}/q\mathbb{Z}$

We prove theorem 3.

(i): Let P be a polynomial over $\mathbb{Z}/q\mathbb{Z}$. We can write P in the form: $P(X) = A(X)(X^p - X)^2 + B(X)p(X^p - X) + C(X)(X^p - X) + pD(X) + E(X)$ with A, B, C, D, E polynomials such as B, C, D, E have degree at most $p - 1$ and coefficients between 0 and $p - 1$.

We want to prove that

$$\forall x \in \mathbb{Z}/q\mathbb{Z} \quad P(x) = 0 \pmod{p^2} \Leftrightarrow C = D = E = 0$$

Only the direct sense is difficult. Clearly, $E = 0$ because the equation is also true modulo p. We remark that $Q(xp + y) = Q(y) \pmod{p}$.
Then, we have $P(p + y) = C(y)(y^p - y) - pC(y) + pD(y) \pmod{p^2}$. And $0 = P(P + y) - P(y) = -pC(y) \pmod{p^2}$. Therefore $C = 0$ and $D = 0$.

(ii) We define the function ϕ from $\mathbb{Z}/q\mathbb{Z}[X]$ to $\mathbb{Z}/q_1\mathbb{Z}[X] \times \mathbb{Z}/q_2\mathbb{Z}[X]$ as $\phi(P) = (P_1, P_2)$ such that $P_i(x) = P(x) \pmod{q_i}$. The function ϕ is well-defined and bijective.

A.3 Piling-up for linear cryptanalysis

Piling-up for bias-based approach. We prove theorem 5.

Let $\phi_{a,c}(x, K) = (x, f(x) + K, \langle a, c|x, h_K(x)\rangle)$. For any b, $\phi_{a,c}$ is a bijection from the set of solutions (x, K) of the equation $\langle a, c|x, h_K(x)\rangle = u + \langle b|K\rangle$ onto the set of solutions (x, y, v) of the equations $\langle a, b|x, f(x)\rangle = u - v$ and $\langle b, c|y, g(y)\rangle = v$. Therefore, for any b, we have a sort of generalized matrix product $\sum_K \Lambda_0(h_K) = \Lambda_0(f)\Lambda_0(g)$ where the elements of these matrix are multiplied by convolution with respect to u. More precisely,

$$\sum_{K \in G'} \Lambda_0(h_K)_{a,c;u+\langle b|K\rangle} = \sum_v \Lambda_0(f)_{a,b;u-v} \Lambda_0(g)_{b,c;v}.$$

We will prove now that the formula remains true translated to Λ_S, for any b such that $T_{a,b} = T_{b,c}$. Both sides are zero if $u \notin T_{a,b}$, therefore we suppose that $u \in T_{a,b}$.

First, we recall that $\sum_{v \in T_{a,b}} \Lambda_S(f)_{a,b;v} = 0$. On one hand, we compute $\sum_v \Lambda_0(f)_{a,b;u-v} S_{b,c;v} = \sum_{v \in T_{b,c}} \Lambda_0(f)_{a,b;u-v} S_{b,c} = q'\ S_{b,c}$, and also, because $T_{a,b}$ is a group and $\langle c|y\rangle \in T_{a,b}$, we compute $\sum_K S_{a,c;u+\langle b|K\rangle} = \sum_v S_{a,b;v} \#\{y \mid \langle c|y\rangle = u - v\} = S_{a,b} \sum_{v \in T_{a,b}} \#\{y \mid \langle c|y\rangle = u - v\} = q'\ S_{a,b}$.

Therefore

$$\sum_{K \in G'} \Lambda_S(h_K)_{a,c;u+\langle b|K\rangle} = \sum_v \Lambda_S(f)_{a,b;u-v} \Lambda_S(g)_{b,c;v}.$$

We need the additional hypothesis that there exists some b such that $T_{a,b} = T_{b,c}$. This is true if $G = G' = G''$, because $\langle a, a+c|x, y\rangle = \langle a+c, c|x-y, x\rangle$ and therefore $T_{a,a+c} = T_{a+c,c}$. It follows that:

$$LP(h_K) = \max_{a,c,u} \left(\frac{\Lambda_S(h_K)_{a,c;u}}{q} \right)^2 \approx \max_{\substack{a,b,c,u \\ T_{a,b}=T_{b,c}}} \left(\frac{\sum_v \Lambda_S(f)_{a,b;u-v} \Lambda_S(g)_{b,c;v}}{qq'} \right)^2$$
$$\approx \max_{a,b_f,b_g,c,u,v} \left(\frac{\Lambda_S(f)_{a,b_f;u-v}}{q} \frac{\Lambda_S(g)_{b_g,c;v}}{q'} \right)^2 = LP(f)LP(g)$$

Piling-up for duality-based approach. $\lambda(h_K)_{a,c} = \sum_{x,z} \Delta(h_K)_{x,z} \xi^{\langle a,c|x,z\rangle}$. If $\Delta(h_K)$ does not depend heavily on K, then $\Delta(h_K)_{x,z} \approx \frac{1}{q'} \sum_y \Delta(f)_{x,y} \Delta(g)_{y,z}$. If $\sum\limits_y \Delta(f)_{x,y} \Delta(g)_{y,z} \xi^{\langle a,b|x,y\rangle} \xi^{\langle b,c|y,z\rangle} \approx \frac{1}{q'} \sum\limits_{y_f,y_g} \Delta(f)_{x,y_f} \Delta(g)_{y_g,z} \xi^{\langle a,b|x,y_f\rangle} \xi^{\langle b,c|y_g,z\rangle}$, then $\lambda(h_K)_{a,c} \approx \sum\limits_{x,z} \frac{1}{q'} \frac{1}{q'} \sum_{y_f,y_g} \Delta(f)_{x,y_f} \Delta(g)_{y_g,z} \xi^{\langle a,b|x,y_f\rangle} \xi^{\langle b,c|y_g,z\rangle}$ which means that $\lambda(h_K)_{a,c} \approx \frac{1}{q'^2} \lambda(f)_{a,b} \lambda(g)_{b,c}$ and therefore

$$\frac{\lambda(h_K)}{q^2} \approx \frac{\lambda(f)}{q'^2} \frac{\lambda(g)}{q^2}$$

A.4 A list of all triples $D(f), L(f), \lambda(f)$ for small values of p.

This is a table of all possible values for non affine functions and for $p = 5$ and 7:

p	D	L	λ	example
5	1	1	5	x^2
	2	4	$9.472\cdots$	$x^4 + 2x^2$
	2	4	$13.090\cdots$	x^3
	3	4	$16.708\cdots$	$x^4 + x^2$
	3	9	$19.472\cdots$	x^4
7	1	1	7	x^2
	2	4	$13.097\cdots$	$x^6 + x^4 + 6x^2$
	2	9	14	x^4
	2	4	$14.185\cdots$	$x^6 + 6x^4 + x^2$
	2	9	$14.454\cdots$	$x^5 + x^2$
	2	4	$14.603\cdots$	$x^6 + x^4$
	3	4	$14.603\cdots$	$x^6 + 3x^4$
	2	4	$15.207\cdots$	$x^6 + x^5 + 4x^2$
	2	4	$16.899\cdots$	$x^6 + x^2$
	3	4	$16.899\cdots$	$x^6 + 5x^5 + x^4 + x^3$
	2	9	$17.048\cdots$	$x^6 + x^5 + x^3 + 5x^2$
	3	9	$17.048\cdots$	$x^6 + 3x^4 + 2x^3$
	2	9	$17.234\cdots$	$x^6 + x^4 + x^3 + 3x^2$
	3	4	$17.234\cdots$	$x^6 + 5x^2$
	3	4	$18.256\cdots$	$x^6 + 3x^3 + x^2$
	3	9	$18.256\cdots$	$x^6 + x^4 + 3x^3$
	2	4	$18.591\cdots$	$x^5 + x^4$
	3	4	$18.591\cdots$	$x^5 + 2x^2$
	3	9	$18.591\cdots$	$x^5 + x^4 + 2x^3$
	2	9	$19.076\cdots$	$x^6 + 5x^3 + 5x^2$
	2	4	$19.195\cdots$	$x^6 + 2x^4 + 5x^2$
	3	4	$19.195\cdots$	$x^6 + 2x^5 + x^4 + x^3$
	3	9	$19.195\cdots$	$x^6 + 2x^5 + 4x^3$
	2	4	$21.640\cdots$	$x^6 + 2x^5 + x^3 + x^2$
	3	4	$21.640\cdots$	$x^6 + 2x^3 + x^2$
7	2	4	$22.476\cdots$	x^3
	3	9	$22.878\cdots$	$x^6 + 4x^5 + x^3 + x^2$
	4	9	$22.878\cdots$	$x^6 + +2x^4 + 2x^3$
	2	4	$23.481\cdots$	$x^6 + 3x^3$
	2	4	$23.481\cdots$	$x^6 + 2x^5 + x^4 + 2x^2$
	2	4	$24.689\cdots$	$x^6 + 2x^5 + 3x^4$
	3	4	$24.689\cdots$	$x^6 + 2x^4 + x^3 + x^2$
	4	4	$24.689\cdots$	$x^6 + 3x^5 + x^3$
	3	9	$24.921\cdots$	$x^6 + 3x^4 + 6x^2$
	4	9	$24.921\cdots$	$x^6 + 3x^3 + 3x^2$
	3	9	$25.591\cdots$	$x^4 + x^3$
	4	9	$25.591\cdots$	$x^5 + 3x^3 + x^2$
	3	9	$26.195\cdots$	$x^6 + x^4 + x^3$
	3	4	$26.799\cdots$	$x^6 + 2x^5$
	4	16	$29.207\cdots$	$x^6 + 3x^4 + x^3$
	3	4	$30.183\cdots$	$x^4 + 3x^2$
	4	4	$30.183\cdots$	x^5
	3	9	$31.689\cdots$	$x^6 + 2x^4 + x^2$
	4	9	$31.689\cdots$	$x^6 + 2x^5 + x^4 + 4x^2$
	4	16	$32.256\cdots$	$x^6 + x^4 + 2x^3 + 5x^2$
	3	4	$32.628\cdots$	$x^6 + 2x^3$
	4	4	$32.628\cdots$	$x^6 + 3x^5 + x^2$
	3	9	$35.073\cdots$	$x^6 + x^4 + x^3 + x^2$
	4	9	$35.073\cdots$	$x^6 + 2x^5 + 2x^3$
	4	16	$39.024\cdots$	$x^5 + x^3$
	3	9	$39.963\cdots$	$x^6 + x^3$
	5	9	$39.963\cdots$	$x^6 + x^5 + 5x^3$
	5	16	$41.169\cdots$	$x^6 + x^4 + x^2$
	5	25	$44.481\cdots$	x^6

A.5 In the prime case, $D(f) = 1 \Rightarrow L(f) = 1$

We will use the following lemma, for which we did not find a proof.

Lemma 1 (Key lemma). *If* $\varphi : \mathbb{F}_p \to \mathbb{Z}$ *satisfies* $\forall x \neq 0,\ (\varphi * \varphi)(x) = -1$, *and* $(\varphi * \varphi)(0) = p - 1$ *then* $\forall x,\ \varphi(x) \in \{0, \pm 1\}$.

Let us fix f, a, and b. We denote $\eta(u) = \Lambda_S(f)_{a,b;u}$ and $\sigma = \eta * \eta$. Note that $\sum_u S_{a,b;u}\xi^u = 0$ and therefore $\hat{\theta}_f(a,b) = \sum_u \eta(u)\xi^u$ and also $\lambda(f)_{a,b} = \sum_{u,v} \eta(u)\eta(v)\xi^{u-v} = \sum_v \eta(v)^2 + \sum_{u=1}^{(\ell+1)/2}(\xi^u + \xi^{-u})\sigma(u)$ which is a real number.

In general, the lower bound for $D(f)$ is q/q'; if this lower bound is reached, then the matrix $\Delta(f)$ if fully known: $\Delta(f)_{a\neq 0,b} = q/q'$, $\Delta(f)_{0,b\neq 0} = 0$, and $\Delta(f)_{0,0} = q$, and therefore we can completely compute its transform: $\lambda(f)_{a,b\neq 0} = q$, $\lambda(f)_{a\neq 0,0} = 0$, and $\lambda(f)_{0,0} = q^2$.

Now, let us look at the case where $G = G' = \mathbb{F}_p$. If f is a polynomial of degree 2, we can check that $D(f) = L(f) = 1$. We want to prove that $D(f) = 1 \Rightarrow L(f) = 1$.

Let us suppose that $D(f) = 1$, then the duality implies that $\lambda(f)_{a,b\neq 0} = p$. However, $\lambda(f)_{a,b} = \sum_v \eta(v)^2 + \sum_{u=1}^{\ell-1} \sigma(u)\xi^u$. Since $\eta(v)$ is an integer, the second sum is also an integer. Because the $(\xi^u)_{u=1\ldots\ell-2}$ are linearly independent over $\mathbb{Q}$, the fact that $\sum_{u\neq 0} \sigma(u)\xi^u$ is an integer implies that all $\sigma(u)$ are equal to some common value σ. Therefore $\sum_v \eta(v)^2 = p + \sigma$.

We also know that $\sum_v \eta(v) = 0$ and therefore $0 = (\sum_v \eta(v))^2 = \sum_v \eta(v)^2 + 2\sum_u \sigma(u) = p(\sigma + 1)$ and we proved that $\sigma = -1$ and $\sigma(0) = p - 1$.

We apply the key lemma to the function η, which means that $\Lambda_S(f)_{a,b;u} = \{0, \pm 1\}$, and therefore $L(f) = 1$.

B Security Analysis of TOY100

B.1 Differential Cryptanalysis

A Key Recovery Attack The attack uses the differential described in Section 5.1, followed by one round of key guess. More precisely, the differential is followed by $\sigma[K^n] \cdot \gamma \cdot \sigma[K^{n+1}]$. The attack goes as follows:

(i) Encrypt N plaintext pairs $(P, P + \Delta_0)$.
(ii) The corresponding ciphertext pairs that actually follow the differential are equal on 12 words (of which the position is fixed). Consider only the pairs satisfying this condition.
(iii) The key guess is performed on the 4 words of the last round key for which the difference is non zero. A counter is set for each candidate. It is incremented when the difference before the last S-box layer corresponding to the candidate is $\theta(\Delta_n)$.
(iv) After enough pairs have been considered, the most counted candidate is selected. The remaining key material is retrieved using a similar attack or by exhaustive key search.

Let $\mathbf{T}_0$ denote the event that 12 words of the output difference are 0, as specified in step 2 of the attack. Let $\mathbf{D}$ be the event that the differential is followed. We consider the 5-round differential, with $D := Pr[\mathbf{D}] \simeq 3 \cdot 10^{-26}$. Then we have

$$\begin{aligned} Pr[\mathbf{T}_0] &= Pr[\mathbf{T}_0|\mathbf{D}] \cdot Pr[\mathbf{D}] + Pr[\mathbf{T}_0|\neg\mathbf{D}] \cdot Pr[\neg\mathbf{D}] \\ &\simeq 1 \cdot 3 \cdot 10^{-26} + 10^{-24} \cdot (1 - 3 \cdot 10^{-26}) \\ &\simeq 10^{-24} \end{aligned} \tag{6}$$

The right 4-subblock subkey will be counted $N \cdot D \simeq 3 \cdot 10^{-26} \cdot N$ times. A wrong 4-subblock subkey will be counted $N \cdot P[\mathbf{T}_0] \cdot 100^{-4} \simeq 10^{-32} \cdot N$ times. Hence the SNR of the attack is $3 \cdot 10^6$, and the subkey can be recovered using less than $2/D = 2/3 \cdot 10^{26}$ pairs. The best way to retrieve the remaining part of the key is exhaustive search.

Applying the same attack for one more round is probably possible, but almost requires the whole codebook. To the best of our investigations, more complex variants of the attack do not significantly improve its efficiency.

Another Property of the Linear Layer The following property of the linear layer, which corresponds to the branch number bound, seems promising:

$$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 50 & 50 & 50 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \xrightarrow{\theta} \begin{bmatrix} 0 & 50 & 0 & 0 \\ 0 & 50 & 0 & 0 \\ 0 & 50 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \tag{7}$$

However this pattern can be used only if $\Delta(f)_{50,50}$ is big enough. For the function we selected it is 0. We note here the particular role played by $\Delta(f)_{50,50}$. The existence of such "specially important" elements in the matrix $\Delta(f)$ is related to the fact that we are working over a ring. Other elements of small characteristic can be important as well for the same kind of reason. In Table 2 we give elements of $\Delta(f)$ corresponding to input and output differences which are multiple of 25; we observe that all of them are small.

Table 2. Values of $\Delta(f)_{a,b}$ when a and b have small characteristic.

(a,b)	$\Delta(f)_{a,b}$
(25,25)	0
(25,50)	0
(25,75)	0
(50,25)	1
(50,50)	0
(50,75)	1
(75,25)	0
(75,50)	0
(75,75)	0

B.2 Linear Cryptanalysis

The following linear equation (equation 5 in section 5.2)

$$\begin{aligned} &10 \cdot (c_{33} + c_{44} - c_{34} - c_{43}) - 10 \cdot (p_{11} + p_{22} - p_{12} - p_{21}) \\ = &\sum_{i=0}^{\lfloor \frac{n+1}{2} \rfloor} 10(k_{11}^{2i} + k_{22}^{2i} - k_{12}^{2i} - k_{21}^{2i}) + \sum_{i=1}^{\lceil \frac{n+1}{2} \rceil} 10(k_{33}^{2i-1} + k_{44}^{2i-1} - k_{34}^{2i-1} - k_{43}^{2i-1}), \end{aligned}$$

holds with probability 1/10 for a random permutation, and with probability $1/10 + \epsilon$ for TOY100 parameterized by a random key, where $|\epsilon|$ is given in

Table 3. Therefore it can be used to build a distinguisher, by identifying the value of $10 \cdot (c_{33} + c_{44} - c_{34} - c_{43}) - 10 \cdot (p_{11} + p_{22} - p_{12} - p_{21})$ occurring the most often, and comparing its frequency of apparition to a certain threshold, in order to distinguish both probability distributions. The data complexity of the attack is $O(\epsilon^{-2})$. This distinguisher can be used in a key-recovery attack, by performing key guesses on the first and/or last round key. Up to 7 rounds of the cipher can be attacked this way, and we are close to an attack on 8 rounds. The data and time complexity are $O(\epsilon^{-2})$.

Finally, we note that relying on property (7) to build a characteristic is not

Table 3. Estimated bias for the best $(n+1)$-round linear characteristic

# Rounds $n+1$	Best Bias
2	$2.49 \cdot 10^{-6}$
3	$8.78 \cdot 10^{-9}$
4	$3.10 \cdot 10^{-11}$
5	$1.09 \cdot 10^{-13}$
6	$3.86 \cdot 10^{-16}$

possible, as our S-box satisfies $\Lambda_S(f)_{50,50;0} = \Lambda_S(f)_{50,50;50} = 0$.

B.3 A Square-like Attack

Our square-like attack aims at the cipher

$$P^{(i)} \xrightarrow{\sigma[K^1]\cdot\gamma\cdot\theta} A^{(i)} \xrightarrow{\sigma[K^2]\cdot\gamma\cdot\theta} B^{(i)} \xrightarrow{\sigma[K^3]\cdot\gamma} C^{(i)} \xrightarrow{\theta} D^{(i)} \xrightarrow{\sigma[K^4]} E^{(i)} \xrightarrow{\gamma\cdot\sigma[K^5]} F^{(i)}$$

It exploits batches of 100^4 plaintexts with the following structure:

$$P^{(i)} = \begin{bmatrix} p_{11}^{(i)} & p_{12}^{(i)} & p_{13}^{(i)} & p_{14}^{(i)} \\ p_{21}^{(i)} & p_{22}^{(i)} & p_{23}^{(i)} & p_{24}^{(i)} \\ p_{31}^{(i)} & p_{32}^{(i)} & p_{33}^{(i)} & p_{34}^{(i)} \\ p_{41}^{(i)} & p_{42}^{(i)} & p_{43}^{(i)} & p_{44}^{(i)} \end{bmatrix} = \begin{bmatrix} a^{(i)} & b^{(i)} & \kappa & \kappa \\ c^{(i)} & d^{(i)} & \kappa & \kappa \\ \kappa & \kappa & \kappa & \kappa \\ \kappa & \kappa & \kappa & \kappa \end{bmatrix},$$

where $(a^{(i)}, b^{(i)}, c^{(i)}, d^{(i)})$ takes every possible value. As the value of constants does not matter for our attack, all κ's denote constants that are *not* necessarily equal.

Let us define:

$$\begin{aligned}
&S_{rs}(x) := S(x + k^1_{rs}) \\
&m^{(i)} := S_{11}(a^{(i)}) + S_{12}(b^{(i)}) \\
&n^{(i)} := S_{21}(c^{(i)}) + S_{22}(d^{(i)}) \\
&o^{(i)} := S_{11}(a^{(i)}) + S_{21}(c^{(i)}) \\
&p^{(i)} := S_{12}(b^{(i)}) + S_{22}(d^{(i)}) \\
&x^{(i)} := m^{(i)} + n^{(i)} = o^{(i)} + p^{(i)}
\end{aligned}$$

After the first round $\sigma[k^1] \cdot \gamma \cdot \theta$ the data become:

$$\begin{bmatrix} x^{(i)} & x^{(i)} & p^{(i)} & o^{(i)} \\ x^{(i)} & x^{(i)} & p^{(i)} & o^{(i)} \\ n^{(i)} & n^{(i)} & S_{22}(d^{(i)}) & S_{21}(c^{(i)}) \\ m^{(i)} & m^{(i)} & S_{12}(b^{(i)}) & S_{11}(a^{(i)}) \end{bmatrix} + \begin{bmatrix} \kappa & \kappa & \kappa & \kappa \\ \kappa & \kappa & \kappa & \kappa \\ \kappa & \kappa & \kappa & \kappa \\ \kappa & \kappa & \kappa & \kappa \end{bmatrix}$$

It is then easy to see that the state $B^{(i)}$ after the second round $\sigma[k^2] \cdot \gamma \cdot \theta$ is such that $b^{(i)}_{11}, b^{(i)}_{12}, b^{(i)}_{21}, b^{(i)}_{22}$ are still active (i.e. take every value equally often). This property is preserved after passing through $\sigma[k^3] \cdot \gamma$. In order to push the distinguisher further, we use the following property of θ again:

$$D^{(i)} = \theta(C^{(i)}) \Rightarrow d^{(i)}_{33} + d^{(i)}_{44} - d^{(i)}_{34} - d^{(i)}_{43} = c^{(i)}_{11} + c^{(i)}_{22} - c^{(i)}_{12} - c^{(i)}_{21} \tag{8}$$

So we have

$$\begin{aligned}
& \sum_{1 \le i \le 100^4} e^{(i)}_{33} + e^{(i)}_{44} - e^{(i)}_{34} - e^{(i)}_{43} \\
= & \sum_{1 \le i \le 100^4} (d^{(i)}_{33} + k^4_{33}) + (d^{(i)}_{44} + k^4_{44}) - (d^{(i)}_{34} + k^4_{34}) - (d^{(i)}_{43} + k^4_{43}) \\
= & \sum_{1 \le i \le 100^4} d^{(i)}_{33} + d^{(i)}_{44} - d^{(i)}_{34} - d^{(i)}_{43} \\
= & \sum_{1 \le i \le 100^4} c^{(i)}_{11} + c^{(i)}_{22} - c^{(i)}_{12} - c^{(i)}_{21} \\
= & \sum_{1 \le i \le 100^4} c^{(i)}_{11} + \sum_{1 \le i \le 100^4} c^{(i)}_{22} - \sum_{1 \le i \le 100^4} c^{(i)}_{12} - \sum_{1 \le i \le 100^4} c^{(i)}_{21} = 0,
\end{aligned}$$

where the last equality results from the fact that $c^{(i)}_{11}, c^{(i)}_{12}, c^{(i)}_{21}$ and $c^{(i)}_{22}$ are active.

By guessing 4 words $k^5_{33}, k^5_{34}, k^5_{43}, k^5_{44}$ of the last round key we can check this property. The probability of a false alarm is 1/100, so about 4 batches of 100^4 plaintexts are necessary to retrieve this part of the key. Besides it is clear that our analysis holds for any "square of four words" of the plaintext. Hence we can retrieve the remaining 12 subblocks using the same method. The global complexity is about $16 \cdot 100^4$ chosen plaintexts. The offline work is of the same order of magnitude.