Some Properties of Modular Addition (Extended abstract)

Victoria Vysotskaya

JSC «InfoTeCS», Moscow, Russia
vysotskaya.victory@gmail.com

Abstract

In this paper we study a problem which emerged during an attempt to apply a differential cryptanalysis method to the «Magma» algorithm. We obtained a general formula of distribution in the difference distribution table of addition modulo 2^n and provided an efficient method for computing the distribution in a row with given index. Moreover, an exact formula that may be used to solve the task of counting all the distributions was obtained, and an asymptotically accurate approximation of number of distinct distributions was proved. Finally, we designed an algorithm to generate all distributions in $2^{O(\sqrt{(n)})}$ operations (whereas the corresponding brute-force method takes $2^{\Omega(n)}$).

Keywords: modular addition, partitions, differential cryptanalysis.

1 Introduction

The problem studied in the paper emerged during an attempt to estimate the applicability of differential cryptanalysis to the Russian government standard symmetric key block cipher (GOST 28147-89) rounds [1]. It is vital since the algorithm (called "Magma") is still present in the modern Russian GOST R 34.12-2015 of symmetric key block cipher [2].

During the research on the topic the following equation emerged:

$$\Delta f = \left[(x \oplus \Delta x) \boxplus_n y\right] \oplus (x \boxplus_n y). \tag{1}$$

Let us introduce the function $P_n(\Delta x, \Delta f)$:

$$P_n(\Delta x, \Delta f) = \Big|\big\{(x, y) : \Delta f = \left[(x \oplus \Delta x) \boxplus_n y\right] \oplus (x \boxplus_n y); \\ \Delta x, \Delta f \in \{0, \ldots, 2^n - 1\}\big\}\Big|$$

(it is analogous to a special case of the differential probability of addition modulo 2^n studied in [3]). Let us consider the table of values of this function $(P_n)_{\Delta x,\Delta f}$. In this table rows are indexed by Δx and columns by Δf. Such a table is usually called difference distribution table (DDT).

Let us introduce an equivalence relation on the rows of matrix P_n as follows: two rows are called *equivalent* if they coincide up to permutations of elements. Next, we study the set of equivalence classes into which matrix rows are divided. Let us call such equivalence classes the *distributions*.

Note. *Let us consider the calculation of number of different distributions or enumerating them, as algorithmic tasks. Then trivial (brute force) algorithm requires* $2^{\Omega(n)}$ *operations as one need to calculate the value of* Δf *for all* x, y, $\Delta x \in \{0, \ldots, 2^n - 1\}$. *At the same time the algorithm based on the results presented in our article requires polynomial number of operations for the first task and* $2^{O(\sqrt{n})}$ *operations for the second.*

2 Parametrization of distributions

Lemma 1. *Let matrix* P_n *have the form*

$$P_n = \begin{bmatrix} A & B \\ C & D \end{bmatrix}.$$

Then matrix P_{n+1} *has the form*

$$P_{n+1} = 2 \left[\begin{array}{cc|cc} 2A & B & 0 & B \\ C & D & C & D \\ \hline 0 & B & 2A & B \\ C & D & C & D \end{array}\right].$$

The proof of Lemma 1 is given in the Appendix A, since it is quite cumbersome.

This Lemma can be reworded: if

$$P_n = 2^{n+1} \begin{bmatrix} A_n & B_n \\ B_n & A_n \end{bmatrix},$$

then

$$A_n = \begin{bmatrix} 2A_{n-1} & B_{n-1} \\ B_{n-1} & A_{n-1} \end{bmatrix}, \quad B_n = \begin{bmatrix} 0 & B_{n-1} \\ B_{n-1} & A_{n-1} \end{bmatrix}.$$

Let us denote $(\alpha_{n-1}, \alpha_{n-2}, \ldots, \alpha_1, \alpha_0)$ the binary representation of number i. Then let us match each distribution located in some row of matrix P_n with a polynomial in the following way. A row p_i corresponds to polynomial

$\sum_{j=0}^{n+2} c_j x^j$, where c_i is the amount of numbers 2^i in p_i. Hence multiplication by 2 corresponds to multiplication by x and concatenation to addition of polynomials. For $a_n^i(x)$ and $b_n^i(x)$ corresponding to i-th rows of A_n and B_n respectively we have:

$$a_n^i(x) = \begin{cases} x a_{n-1}^i(x) + b_{n-1}^i(x), & \text{if } \alpha_{n-2} = 0, \\ a_{n-1}^i(x) + b_{n-1}^i(x), & \text{if } \alpha_{n-2} = 1; \end{cases}$$

$$b_n^i(x) = \begin{cases} b_{n-1}^i(x), & \text{if } \alpha_{n-2} = 0, \\ a_{n-1}^i(x) + b_{n-1}^i(x), & \text{if } \alpha_{n-2} = 1. \end{cases}$$

Thus,

$$\begin{bmatrix} a_n^i(x) \\ b_n^i(x) \end{bmatrix} = W_{\alpha_{n-2}} \begin{bmatrix} a_{n-1}^i(x) \\ b_{n-1}^i(x) \end{bmatrix},$$

where

$$W_0 = \begin{bmatrix} x & 1 \\ 0 & 1 \end{bmatrix}, \quad W_1 = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}.$$

Moreover,

$$A_1 = \begin{bmatrix} 1 \end{bmatrix}, \ B_1 = \begin{bmatrix} 0 \end{bmatrix}, \ a_1 = 1, \ b_1 = 0.$$

Repeating the same argument $n-2$ more times we finally get

$$a_n^i(x) + b_n^i(x) = \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} a_n^i(x) \\ b_n^i(x) \end{bmatrix} = \begin{bmatrix} 1 & 1 \end{bmatrix} W_{\alpha_{n-2}} W_{\alpha_{n-3}} \ \dots \ W_{\alpha_0} \begin{bmatrix} 1 \\ 0 \end{bmatrix}. \tag{2}$$

Let us denote i' the number with binary representation $(\alpha_{n-2}, \alpha_{n-3}, \dots, \alpha_0)$. This choice is based on the knowledge that the most significant bit does not affect the distribution. Let us separate groups of 0's and 1's in i'. We assume that the first one is a group of 1's, and the last one is a group of 0's (both can be empty). The number of 1's is $K = k_1 + k_2 + \cdots + k_s$, the number of 0's is $L = \ell_1 + \cdots + \ell_s$ and $L + K = n - 1$. Then

$$i' = \underbrace{11\dots1}_{k_1}\underbrace{0\dots0}_{\ell_1}\underbrace{1\dots1}_{k_2}\underbrace{0\dots0}_{\ell_2}\dots\underbrace{1\dots1}_{k_s}\underbrace{0\dots0}_{\ell_s}$$

and expression (2) becomes

$$a_n^i(x) + b_n^i(x) = \begin{bmatrix} 1 & 1 \end{bmatrix} W_1^{k_1} W_0^{\ell_1} \dots W_1^{k_s} W_0^{\ell_s} \begin{bmatrix} 1 \\ 0 \end{bmatrix}. \tag{3}$$

We will use the following statements, easily provable by induction:

$$W_1^k = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \cdots \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = 2^{k-1} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix},$$

$$W_0^\ell = \begin{bmatrix} x & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x & 1 \\ 0 & 1 \end{bmatrix} \cdots \begin{bmatrix} x & 1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} x^\ell & x^{\ell-1} + x^{\ell-2} + \cdots + 1 \\ 0 & 1 \end{bmatrix}.$$

Then (3) may be represented as:

$$a_n^i(x) + b_n^i(x) = \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} 2^{k_1-1} \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} x^{\ell_1} & x^{\ell_1-1} + \cdots + 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \cdots$$

$$\cdots \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} x^{\ell_s} & x^{\ell_s-1} + \cdots + 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

Note that

$$\begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} x^\ell & x^{\ell-1} + \cdots + 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = x^\ell + x^{\ell-1} + x^{\ell-2} + \cdots + 2.$$

Then

$$a_n^i(x) + b_n^i(x) = 2 \cdot 2^{K-s}(x^{\ell_1} + x^{\ell_1-1} + \cdots + 2) \ \cdots \ (x^{\ell_s} + x^{\ell_s-1} + \cdots + 2)x^{\ell_s}.$$

Hence

$$p_n^i(x) = 2^{K-s+1} \prod_{j=1}^{s-1} (x^{\ell_j} + x^{\ell_j-1} + \cdots + 2)x^{\ell_s}. \tag{4}$$

Let us denote Q_n the set of tuples $(s, L, \underline{\ell_s}, \widetilde{\ell})$, where $s \in \{1, \ldots, n-1\}$, $\ell_s \in \{0, \ldots, n-1\}$, $L \in \{0, \ldots, n-s\}$ and $\widetilde{\ell}$ is a multiset of $s-1$ positive integers summing up to $L - \ell_s$. We now want to prove that there is a one-to-one correspondence between the set of polynomials $p_n^i(x)$ and the set Q_n. It is obvious that there is a corresponding set $q_i \in Q_n$ to each polynomial $p_n^i(x)$ and vice versa. So it is enough to show that if two polynomials are equal then corresponding sets of parameters coincides.

Let us fix numbers d_1 and d_2 and then compare two expressions

$$p_n^{d_1}(x) = 2^{K'-s'+1} \prod_{j=1}^{s'-1} (x^{\ell'_j} + x^{\ell'_j-1} + \cdots + 2)x^{\ell'_{s'}},$$

$$p_n^{d_2}(x) = 2^{K''-s''+1} \prod_{j=1}^{s''-1} (x^{\ell''_j} + x^{\ell''_j-1} + \cdots + 2)x^{\ell''_{s''}}.$$

If polynomials are equal, then $2^{K'-s'+1}x^{L'} = 2^{K''-s''+1}x^{L''}$, hence $L' = L''$ and $K' - s' + 1 = = K'' - s'' + 1$. Since the counts of 0's are equal, the counts of 1's are equal too, so $s' = s''$. Besides, the lower powers of the polynomials must coincide, hence $\ell'_{s'} = \ell''_{s''}$. Now it remains to prove that under stated assumptions the equality of polynomials also mean the equality of parameters $\ell'_1, \ldots, \ell'_{s'}$ and $\ell''_1, \ldots, \ell''_{s''}$ up to a permutation.

Let us denote

$$\mathcal{G}_{\ell'_j}(x) = x^{\ell'_j} + x^{\ell'_j - 1} + \cdots + 2,$$
$$\mathcal{G}_{\ell''_j}(x) = x^{\ell''_j} + x^{\ell''_j - 1} + \cdots + 2.$$

We now show that if

$$\prod_{j'=1}^{s'-1} \mathcal{G}_{\ell'_{j'}}(x) = \prod_{j''=1}^{s''-1} \mathcal{G}_{\ell''_{j''}}(x), \tag{5}$$

then the multiset $\{\mathcal{G}_{\ell'_{j'}}(x)\}_{j'=1}^{s'-1}$ equals to the multiset $\{\mathcal{G}_{\ell''_{j''}}(x)\}_{j''=1}^{s''-1}$, or in other words the decomposition of such polynomials into factors of form $\mathcal{G}_j(x)$ is unique. For this purpose we prove that polynomials $\mathcal{G}_j(x)$ are pairwise coprime. Let us compute the greatest common divisor of $\mathcal{G}_u(x)$ and $\mathcal{G}_v(x)$ for $u > v$:

$$(\mathcal{G}_u(x), \mathcal{G}_v(x)) = (x^u + x^{u-1} + \cdots + 2, x^v + x^{v-1} + \cdots + 2) =$$
$$= (x^{u-v-1} + \cdots + 1, x^v + x^{v-1} + \cdots + 2) = \left(\frac{x^{u-v} - 1}{x - 1}, \frac{x^{v+1} - 1}{x - 1} + 1\right) =$$
$$= \frac{1}{x-1}\left(x^{u-v} - 1, x^{v+1} + x - 2\right).$$

The roots of the polynomial $f(x) = x^{u-v} - 1$ are all roots of unity of the degree $(u - v)$. Let us check which of these roots can be roots of polynomial $h(x) = x^{v+1} + x - 2$.

Let $\varepsilon = \cos\frac{2\pi}{u-v} + i \cdot \sin\frac{2\pi}{u-v}$ be a primitive root of unity of the degree $(u - v)$, then $\{\varepsilon^\ell\}_{\ell=0}^{u-v-1}$ is a set of all roots of unity of the degree $(u - v)$. So

$$\varepsilon^{\ell(v+1)} + \varepsilon^\ell - 2 = 0.$$

Therefore $\varepsilon^{\ell(v+1)} = \varepsilon^\ell = 1$, as $|\varepsilon^k| \leqslant 1$ for all k. Hence

$$\frac{1}{x-1}\left(x^{u-v} - 1, x^{v+1} + x - 2\right) = \frac{1}{x-1}(x - 1) = 1.$$

Now let us return to the case (5). We decompose polynomials of the left

and right sides into irreducible ones. Then we consider the first irreducible polynomial $f(x)$ on the left-hand side. In order for equality to hold, $f(x)$ also has to be present on the right-hand side. So there are some $\mathcal{G}_u$ on the left-hand side and $\mathcal{G}_v$ on the right-hand side divisible by $f(x)$. Hence, u must be equal to v. Divide both sides by $\mathcal{G}_u$ and continue in the same fashion, arriving at the conclusion that the decomposition is unique up to a permutation.

Thus, we proved the following

Theorem 1. *There is a one-to-one correspondence between the set of distributions of the rows of matrix P_n and the parameters set Q_n.*

Using Theorem 1 one can enumerate the distributions in time proportional to their number. More precisely, one can iterate over all distinct distribution and list them in time $O\left(|Q_n| \cdot \mathrm{poly}(n)\right)$, where $\mathrm{poly}(n)$ is a polynomial of n. The only tricky part is to enumerate all the multisets with given sum, but it can be done using one of various recursive algorithms in $O(1)$ amortised time per iteration (e. g. see [4]).

3 The number of distributions

Let $p(n,k)$ be the number of partitions of n into exactly k parts. Moreover, let $p(n,k) = 0$, if $k \leqslant 0$ or $n \leqslant 0$, but $p(0,0) = 1$. If we fix s, L and ℓ_s then the number of tuples from the set Q_n with these parameters is equal to $p(L-\ell_s, s-1)$. Obviously there are only n tuples with $s = 1$: $(1,1,\ldots,1,1), (1,1,\ldots,1,0), \ldots, (1,0,\ldots,0,0), (0,0,\ldots,0,0)$. We will consider this case separately and we will assume that $s \geqslant 2$. Finally, note that $\sum\limits_{L=\ell_s}^{n-s} p(L-\ell_s, s-1) = \sum\limits_{L=0}^{n-s-\ell_s} p(L, s-1)$. Then

$$|Q_n| = \left[\sum_{s=2}^{n-1}\sum_{\ell_s=0}^{n-1}\sum_{L=0}^{n-s-\ell_s} p(L, s-1)\right] + n. \tag{6}$$

We make one more note to be used later.

Lemma 2. $p(n,k) = p(n-1,k-1) + p(n-k,k)$.

Proof. Note that the partition of number n into k parts can either include some number of 1's or not include any. In the first case, there is a one-to-one correspondence between such partitions and (unconstrained) partitions of $n-1$ into $k-1$ parts (just put additional 1 to a partition) — there are $p(n-1,k-1)$ of them. In the second case, there is a correspondence between such partitions and (unconstrained) partitions of $n-k$ into k parts (just add 1 to each number in partition) — there are $p(n-k,k)$ of them. □

We now show that the expression (6) can be simplified.

Theorem 2. $|Q_n| = \sum_{j=1}^{n-1} p(j) + 1$, *where* $p(j) = \sum_{s=1}^{j} p(j,s)$, $n > 3$.

Proof (by induction). For $n = 4$ formula (6) gives $|Q_4| = 7$. At the same time $p(3) + p(2) + p(1) + 1 = 3 + 2 + 1 + 1 = 7$.

Let us show the induction step. In other words, let us prove that the following holds

$$\sum_{s=2}^{n}\sum_{\ell_s=0}^{n}\sum_{L=0}^{n-s+1-\ell_s} p(L,s-1) - \sum_{s=2}^{n-1}\sum_{\ell_s=0}^{n-1}\sum_{L=0}^{n-s-\ell_s} p(L,s-1) = p(n) - 1.$$

$$\sum_{s=2}^{n}\sum_{\ell_s=0}^{n}\sum_{L=0}^{n-s+1-\ell_s} p(L,s-1) - \sum_{s=2}^{n-1}\sum_{\ell_s=0}^{n-1}\sum_{L=0}^{n-s-\ell_s} p(L,s-1) =$$

$$= \sum_{s=2}^{n-1}\left[\sum_{\ell_s=0}^{n}\sum_{L=0}^{n-s+1-\ell_s} p(L,s-1) - \sum_{\ell_s=0}^{n-1}\sum_{L=0}^{n-s-\ell_s} p(L,s-1)\right] +$$

$$+ \underbrace{\sum_{\ell_s=0}^{n}\sum_{L=0}^{n-n+1-\ell_s} p(L,n-1)}_{=0} =$$

$$= \sum_{s=2}^{n-1}\left[\sum_{\ell_s=0}^{n-1}\left[\sum_{L=0}^{n-s+1-\ell_s} p(L,s-1) - \sum_{L=0}^{n-s-\ell_s} p(L,s-1)\right] + \underbrace{\sum_{L=0}^{n+1-s-n} p(L,s-1)}_{=0}\right] =$$

$$= \sum_{s=2}^{n-1}\sum_{\ell_s=0}^{n-1} p(n+1-s-\ell_s, s-1).$$

Now we will prove by induction that the latter is equal to $p(n) - 1$. For $n = 4$ both of them are equal to 4.

Induction step: let us check the validity of equation

$$p(n+1)-1=\sum_{s=2}^{n}\sum_{\ell_s=0}^{n}p(n+2-s-\ell_s,s-1)=$$

$$=\sum_{s=2}^{n}\sum_{\ell_s=-1}^{n-1}p(n+1-s-\ell_s,s-1)=\sum_{s=2}^{n-1}\sum_{\ell_s=0}^{n-1}p(n+1-s-\ell_s,s-1)+$$

$$+\underbrace{\sum_{\ell_s=-1}^{n-1}p(n+1-n-\ell_s,n-1)}_{=p(1-\ell_s,n-1)=0}+\sum_{s=2}^{n}p(n+1-s-(-1),s-1)=$$

$$=p(n)-1+\sum_{s=2}^{n}p(n+2-s,s-1).$$

Since

$$p(n)=\sum_{s=1}^{n}p(n,s),\ \ p(n+1)=\sum_{s=1}^{n+1}p(n+1,s),$$

the equation becomes

$$\sum_{s=1}^{n+1}p(n+1,s)=\sum_{s=1}^{n+1}p(n,s)+\sum_{s=1}^{n-1}p(n+1-s,s). \tag{7}$$

Let us continue transforming the expression (7):

$$\sum_{s=1}^{n}p(n+1,s)+\underbrace{p(n+1,n+1)}_{=1}=\sum_{s=2}^{n+1}p(n,s-1)+\sum_{s=1}^{n}p(n+1-s,s)-$$

$$-\underbrace{p(n+1-n,n)}_{=p(1,n)=0}=\sum_{s=1}^{n}p(n,s-1)+\underbrace{p(n,n)}_{=1}-\underbrace{p(n,0)}_{=0}+\sum_{s=1}^{n}p(n+1-s,s).$$

Eventually,

$$\sum_{s=1}^{n}p(n+1,s)=\sum_{s=1}^{n}p(n,s-1)+\sum_{s=1}^{n-1}p(n+1-s,s).$$

Lemma 2 ends the proof.

□

Theorem 2 makes it possible to solve the task of counting all the distributions. We just have to calculate values of $p(j,s)$ for $j\in\{1,\ldots,n-1\},\ s\in$

$\{1, \ldots, j\}$, then all the $p(j)$ and finally $|Q_n|$. The complexity of computing $p(j, s)$ dominates the other steps and according to Lemma 2 may be done in $O(n^2)$ additions of n-bit numbers. Thus we need $O(n^3)$ bit operations for the counting problem.

4 Asymptotical approximation

In [5] the following asymptotic formula for the number of partitions $p(n)$ was obtained:

$$p(n) \sim \frac{1}{4\sqrt{3}n} e^{\pi\sqrt{\frac{2n}{3}}}.$$

Hence

$$|Q_n| \sim \sum_{j=1}^{n-1} \frac{1}{4\sqrt{3}j} e^{\pi\sqrt{\frac{2j}{3}}} + 1 \text{ as } n \to \infty. \tag{8}$$

The following Lemma allows us to claim it.

Lemma 3. *Let $f(n) \sim g(n)$ as $n \to \infty$, $f(n) \geqslant 0$, $g(n) \geqslant 0$, $f(n)$ and $g(n)$ monotonically increase and are unbounded, $F(n) = \sum_{k=1}^{n} f(k)$, $G(n) = \sum_{k=1}^{n} g(k)$, then $F(n) \sim G(n)$, $n \to \infty$.*

You can find the proof of Lemma 3 in Appendix B.

Now we will prove an auxiliary Lemma.

Lemma 4.

$$\sum_{j=1}^{n-2\sqrt{n}\ln n} \frac{1}{4\sqrt{3}j} e^{\pi\sqrt{\frac{2j}{3}}} = o\left(\frac{1}{4\sqrt{3}n} e^{\pi\sqrt{\frac{2n}{3}}}\right) \text{ as } n \to \infty.$$

Proof. Let us show that

$$\lim_{n\to\infty} \sum_{j=1}^{n-2\sqrt{n}\ln n} \frac{n}{j} e^{\pi\sqrt{\frac{2}{3}}\left(\sqrt{j}-\sqrt{n}\right)} = 0.$$

Since

$$\frac{n}{j} < n, \; j < n - 2\sqrt{n}\ln n$$

and

$$n - 2\sqrt{n}\ln n < n,$$

it is sufficient to prove that

$$\lim_{n\to\infty} n^2 e^{\pi\sqrt{\frac{2}{3}}\left(\sqrt{n-2\sqrt{n}\ln n}-\sqrt{n}\right)} = 0.$$

From

$$\sqrt{n-2\sqrt{n}\ln n} = \sqrt{n}\sqrt{1-\frac{2\ln n}{\sqrt{n}}} = \sqrt{n}\left(1-\frac{2\ln n}{2\sqrt{n}}+o\left(\frac{\ln n}{\sqrt{n}}\right)\right) = \\ = \sqrt{n}-\ln n+o(\ln n)$$

it follows that

$$\lim_{n\to\infty} n^2 e^{\pi\sqrt{\frac{2}{3}}\left(\sqrt{n-2\sqrt{n}\ln n}-\sqrt{n}\right)} = \lim_{n\to\infty} n^2 e^{\pi\sqrt{\frac{2}{3}}(-\ln n+o(\ln n))} = \lim_{n\to\infty} n^{2-\pi\sqrt{\frac{2}{3}}+o(1)}.$$

Whereas the exponent is negative, Lemma is proved. □

Theorem 3.

$$\sum_{j=1}^{n} p(j) \sim \frac{e^{\pi\sqrt{\frac{2n}{3}}}}{2\sqrt{2}\pi\sqrt{n}} \ as\ n\to\infty.$$

Proof. It can be proved that there exists a number N_0 such that the function on the right-hand side monotonically increases on $[N_0;+\infty)$. We will estimate the sum from N_0 to n as first N_0-1 summands do not influence the asymptotic.

The following holds

$$\int_{N_0}^{n} \frac{e^{\pi\sqrt{\frac{2x}{3}}}dx}{4\sqrt{3}x} = \int_{N_0}^{n} \frac{1}{2\sqrt{2}\pi\sqrt{x}} de^{\pi\sqrt{\frac{2x}{3}}} = \frac{e^{\pi\sqrt{\frac{2x}{3}}}}{2\sqrt{2}\pi\sqrt{x}}\bigg|_{N_0}^{n} + \int_{N_0}^{n} \frac{e^{\pi\sqrt{\frac{2x}{3}}}}{2\sqrt{2}\pi x^{3/2}}dx = \\ = \frac{e^{\pi\sqrt{\frac{2n}{3}}}}{2\sqrt{2}\pi\sqrt{n}} - \frac{e^{\pi\sqrt{\frac{2N_0}{3}}}}{2\sqrt{2}\pi\sqrt{N_0}} + \int_{N_0}^{n} \frac{e^{\pi\sqrt{\frac{2x}{3}}}}{2\sqrt{2}\pi x^{3/2}}dx.$$

Now we will show that the last two summands here are $o(e^{\pi\sqrt{\frac{2n}{3}}}n^{-\frac{1}{2}})$. For the first of them it is obvious, so let us focus on the second. For this purpose we note that

$$\sum_{j=N_0}^{n-1} f(x) \leqslant \int_{N_0}^{n} f(x)dx \leqslant \sum_{j=N_0+1}^{n} f(x)$$

for non-decreasing function f. So by Lemma 4

$$\int_{N_0}^{n} \frac{e^{\pi\sqrt{\frac{2x}{3}}}}{2\sqrt{2}\pi x^{3/2}} \leqslant \sum_{N_0+1}^{n} \frac{e^{\pi\sqrt{\frac{2x}{3}}}}{2\sqrt{2}\pi x^{3/2}} \sim \sum_{j=n-2\sqrt{n}\ln n}^{n} \frac{e^{\pi\sqrt{\frac{2x}{3}}}}{2\sqrt{2}\pi x^{3/2}} \leqslant$$

$$\leqslant \frac{e^{\pi\sqrt{\frac{2n}{3}}}}{2\sqrt{2}\pi(n-2\sqrt{n}\ln n)^{\frac{3}{2}}} \cdot 2\sqrt{n}\ln n \sim \frac{e^{\pi\sqrt{\frac{2n}{3}}}}{\sqrt{2}\pi n^{\frac{3}{2}}}\sqrt{n}\ln n = \frac{e^{\pi\sqrt{\frac{2n}{3}}}}{\sqrt{2}\pi n}\ln n.$$

Finally for $n \to \infty$

$$\left(\frac{e^{\pi\sqrt{\frac{2n}{3}}}}{\sqrt{2}\pi n}\ln n\right)\left(\frac{e^{\pi\sqrt{\frac{2n}{3}}}}{\sqrt{n}}\right)^{-1} = \frac{\ln n}{\sqrt{2}\pi\sqrt{n}} \to 0.$$

In addition,

$$\sum_{j=N_0+1}^{n} f(x) - \sum_{j=N_0}^{n-1} f(x) = f(n) - f(N_0) = \frac{e^{\pi\sqrt{\frac{2n}{3}}}}{4\sqrt{3}n} - \frac{e^{\pi\sqrt{\frac{2N_0}{3}}}}{4\sqrt{3}N_0} = o\left(\frac{e^{\pi\sqrt{\frac{2n}{3}}}}{\sqrt{n}}\right)$$

and the following equality

$$\sum_{j=N_0}^{n} \frac{e^{\pi\sqrt{\frac{2j}{3}}}}{4\sqrt{3}j} \sim \int_{N_0}^{n} \frac{e^{\pi\sqrt{\frac{2x}{3}}}}{4\sqrt{3}x}dx.$$

concludes the proof of the Theorem. □

According to the above Theorem and the note after Theorem 1 we can enumerate all the distributions in time $2^{O(\sqrt{n})}$ that is obviously substantially better than brute force algorithm with complexity $2^{\Omega(n)}$.

5 Conclusion

We obtained a general form of distributions in DDT. Moreover, we provided an efficient method for computing the distribution in a row with given index. The obtained results imply a possibility to substantially accelerate the construction of all possible distributions. We showed that all the distributions now can be generated in time proportional to the amount of them. We have proved that the number of distinct distributions is $2^{O(\sqrt{n})}$, so the whole generating algorithm would take $2^{O(\sqrt{n})}$ operations. At the same time the brute force algorithm requires $2^{\Omega(n)}$ operations.

References

[1] *GOST 28147-89. National Standard of the USSR. Cryptographic Protection for Data Processing System*, (in Russian).

[2] *GOST R 34.12-2015. National Standard of the Russian Federation. Cryptographic Data Security. Block Ciphers*, (in Russian).

[3] Lipmaa H., Moriai S., "Efficient Algorithms for Computing Differential Properties of Addition", *Fast Software Encryption*, ed. Matsui M., 2002, 336–350.

[4] Kelleher J., O'Sullivan B., "Generating All Partitions: A Comparison Of Two Encodings", *CoRR*, **abs/0909.2331** (2009), arXiv: http://arxiv.org/abs/0909.2331.

[5] Hardy G. H., Ramanujan S., "Asymptotic Formulæ in Combinatory Analysis", *Proceedings of the London Mathematical Society*, **s2-17**:1, 75–115.

Appendix

A The proof of Lemma 1

Let us find a rule by which, knowing the form of matrix P_n for some n, we can construct such a matrix for P_{n+1}. Write down all variables, setting two most significant bits apart:

$$x = x_n \cdot 2^n + x_{n-1} \cdot 2^{n-1} + \widehat{x},$$

$$y = y_n \cdot 2^n + y_{n-1} \cdot 2^{n-1} + \widehat{y},$$

$$\Delta x = \Delta x_n \cdot 2^n + \Delta x_{n-1} \cdot 2^{n-1} + \Delta\widehat{x},$$

$$\Delta f = \Delta f_n \cdot 2^n + \Delta f_{n-1} \cdot 2^{n-1} + \Delta\widehat{f},$$

where

$$x,\ y,\ \Delta x,\ \Delta f \in \{0, \ldots, 2^{n+1} - 1\},$$

$$x_n,\ \Delta y_n,\ \Delta x_n,\ \Delta f_n \in \{0, 1\},$$

$$x_{n-1},\ \Delta y_{n-1},\ \Delta x_{n-1},\ \Delta f_{n-1} \in \{0, 1\},$$

$$\widehat{x},\ \widehat{y},\ \Delta\widehat{x},\ \Delta\widehat{f} \in \{0, \ldots, 2^{n-1} - 1\}.$$

Besides, denote

$$m_k(a, b, c) = \begin{cases} 0, & \text{if } a + b + c < 2^k, \\ 1, & \text{if } a + b + c \geqslant 2^k, \end{cases}$$

the function that returns a carry bit of addition $a + b + c$ modulo 2^k. We also denote $m_k(a, b) = m_k(a, b, 0)$. In addition let us note that $m_1(a, b) = a \,\&\, b$ and $m_1(a, b, c) = a \,\&\, b \vee a \,\&\, c \vee b \,\&\, c$ (the last is called the "majority" function).

Let us denote

$$c = m_{n-1}(\widehat{x}, \widehat{y}),$$
$$c_\Delta = m_{n-1}(\widehat{x} \oplus \Delta\widehat{x}, \widehat{y}).$$

So let us rewrite the first part of expression (1) in more detail:

$$\begin{aligned}(x + \Delta x) \boxplus_{n+1} y = \big[(\widehat{x} \oplus \Delta\widehat{x}) \boxplus_{n-1} \widehat{y}\big] + \\ + \big[c_\Delta \oplus \big((x_{n-1} \oplus \Delta x_{n-1}) + y_{n-1}\big)\big] \cdot 2^{n-1} + \big[(x_n \oplus \Delta x_n) + y_n\big] \cdot 2^n = \\ = \big[(\widehat{x} \oplus \Delta\widehat{x}) \boxplus_{n-1} \widehat{y}\big] + \big[c_\Delta \oplus \big((x_{n-1} \oplus \Delta x_{n-1}) \oplus y_{n-1}\big)\big] \cdot 2^{n-1} + \\ + \big[m_1(c_\Delta, x_{n-1} \oplus \Delta x_{n-1}, y_{n-1}) \oplus \big((x_n \oplus \Delta x_n) \oplus y_n\big)\big] \cdot 2^n.\end{aligned}$$

Similarly, we get

$$x \boxplus_{n+1} y = (\widehat{x} \boxplus_{n-1} \widehat{y}) + (c \oplus x_{n-1} \oplus y_{n-1}) \cdot 2^{n-1} + \big[m_1(c, x_{n-1}, y_{n-1}) \oplus (x_n \oplus y_n)\big] \cdot 2^n.$$

Then the equation (1) can be rewritten as

$$\begin{aligned}\Delta f = \big[(\widehat{x} \boxplus_{n-1} \widehat{y}) \oplus \big((\widehat{x} \oplus \Delta\widehat{x}) \boxplus_{n-1} \widehat{y}\big)\big] + (\Delta x_{n-1} \oplus c \oplus c_\Delta) \cdot 2^{n-1} + \\ + \big[\Delta x_n \oplus m_1(c_\Delta, x_{n-1} \oplus \Delta x_{n-1}, y_{n-1}) \oplus m_1(c, x_{n-1}, y_{n-1})\big] \cdot 2^n.\end{aligned}$$

Let us denote

$$\varphi(\widehat{x}, \widehat{y}, \Delta\widehat{x}) = \big[(\widehat{x} \boxplus_{n-1} \widehat{y}) \oplus \big((\widehat{x} \oplus \Delta\widehat{x}) \boxplus_{n-1} \widehat{y}\big)\big].$$

Hence, the equation (1) is equivalent to the following system:

$$\begin{cases} c = m_{n-1}(\widehat{x}, \widehat{y}), & (9) \\ c_\Delta = m_{n-1}(\widehat{x} \oplus \Delta\widehat{x}, \widehat{y}), & (10) \\ \varphi(\widehat{x}, \widehat{y}, \Delta\widehat{x}) = \Delta\widehat{f}, & (11) \\ c \oplus c_\Delta = \underbrace{\Delta f_{n-1} \oplus \Delta x_{n-1}}_{z_{n-1}}, & (12) \\ m_1(c_\Delta, x_{n-1} \oplus \Delta x_{n-1}, y_{n-1}) \oplus m_1(c, x_{n-1}, y_{n-1}) = \underbrace{\Delta f_n \oplus \Delta x_n}_{z_n}. & (13) \end{cases}$$

Let us fix Δx and Δf modulo 2^{n+1}. We denote the set of solutions (x, y) of the equation (1) modulo 2^k, where $x, y \in \{0, \ldots, 2^k - 1\}$ by M_k. Obviously, M_{n-1} is a set of solutions of equations (9)–(11), M_n — solutions of equations (9)–(12) and M_{n+1} — of equations (9)–(13). Additionally, we introduce two additional sets: $\widehat{U}_{n-1}$ is the set of solutions $(\widehat{x}, \widehat{y})$ of the system (9)–(12) and $\widehat{U}^{(c)}_{n-1}$ for $c \in \{0, 1\}$ is a subset of $\widehat{U}_{n-1}$ where $m_{n-1}(\widehat{x}, \widehat{y}) = c$.

We will try to find sets $(x_{n-1}, y_{n-1}, c, c_\Delta)$ satisfying conditions (12), (13).

But noting that $c_\Delta = c \oplus z_{n-1}$ we will search for solutions (x_{n-1}, y_{n-1}, c) of equation

$$m_1(c \oplus z_{n-1}, x_{n-1} \oplus \Delta x_{n-1}, y_{n-1}) \oplus m_1(c, x_{n-1}, y_{n-1}) = z_n. \qquad (14)$$

Depending on values of x_{n-1} and y_{n-1} this equation may be rewritten as

x_{n-1}	y_{n-1}	(14)
0	0	$m_1((c \oplus z_{n-1}), \Delta x_{n-1}, 0) \oplus m_1(0, 0, c) = z_n,$
0	1	$m_1((c \oplus z_{n-1}), \Delta x_{n-1}, 1) \oplus m_1(0, 1, c) = z_n,$
1	0	$m_1((c \oplus z_{n-1}), 1 \oplus \Delta x_{n-1}, 0) \oplus m_1(1, 0, c) = z_n,$
1	1	$m_1((c \oplus z_{n-1}), 1 \oplus \Delta x_{n-1}, 1) \oplus m_1(1, 1, c) = z_n.$

Or, equivalently,

x_{n-1}	y_{n-1}	(14)
0	0	$(c \oplus z_{n-1}) \cdot \Delta x_{n-1} = z_n,$
0	1	$\left[(c \oplus z_{n-1}) \cdot \Delta x_{n-1} \vee (c \oplus z_{n-1}) \vee \Delta x_{n-1}\right] \oplus c = z_n,$
1	0	$(c \oplus z_{n-1}) \cdot (1 \oplus \Delta x_{n-1}) \oplus c = z_n,$
1	1	$\left[(c \oplus z_{n-1}) \cdot (1 \oplus \Delta x_{n-1}) \vee (c \oplus z_{n-1}) \vee (1 \oplus \Delta x_{n-1})\right] \oplus 1 = z_n.$

Finally, simplifying, we obtain

x_{n-1}	y_{n-1}	(14)
0	0	$(c \oplus z_{n-1}) \cdot \Delta x_{n-1} = z_n,$
0	1	$(c \oplus z_{n-1}) \vee \Delta x_{n-1} = z_n \oplus c,$
1	0	$(c \oplus z_{n-1}) \cdot \Delta\overline{x_{n-1}} = z_n \oplus c,$
1	1	$(c \oplus z_{n-1}) \vee \Delta\overline{x_{n-1}} = \overline{z_n}.$

Let us consider two cases, $\Delta x_{n-1} = 0$ and $\Delta x_{n-1} = 1$, separately. In the first case we see:

x_{n-1}	y_{n-1}	(14)
0	0	$0 = z_n,$
0	1	$c \oplus z_{n-1} = z_n \oplus c,$
1	0	$c \oplus z_{n-1} = z_n \oplus c,$
1	1	$1 = \overline{z_n}.$

That is, in fact we have only two conditions:

$$z_n = 0, \;\; z_{n-1} = z_n.$$

Thus, depending on the values z_{n-1} and z_n (that is, on the values of Δx and Δf) the equation (14) may have a varying number of solutions $\sigma(z_{n-1}, z_n)$:

1. if $z_{n-1} = z_n = 0$, then $\sigma(z_{n-1}, z_n) = 4$,
2. if $z_{n-1} = 1, z_n = 0$, then $\sigma(z_{n-1}, z_n) = 2$,
3. if $z_{n-1} = z_n = 1$, then $\sigma(z_{n-1}, z_n) = 2$,
4. if $z_{n-1} = 0, z_n = 1$, then $\sigma(z_{n-1}, z_n) = 0$.

That is, the solution set is $\varepsilon \times \{0, 1\}$, where ε is a set of zero, two or four pairs (x_{n-1}, y_{n-1}).

In the case of $\Delta x_{n-1} = 1$ the equation (14) has more complex form:

x_{n-1}	y_{n-1}	(14)
0	0	$c \oplus z_{n-1} = z_n$,
0	1	$1 = z_n \oplus c$,
1	0	$0 = z_n \oplus c$,
1	1	$c \oplus z_{n-1} = \overline{z_n}$,

which is equivalent to:

$$c = z_{n-1} \oplus z_n,\ c = z_n \oplus 1,\ c = z_n,\ c = z_{n-1} \oplus z_n \oplus 1.$$

It is obvious that at any values $z_{n-1}, z_n \in \{0, 1\}$ exactly two of these conditions will be fulfilled, since up to the permutation they are equivalent to conditions

$$c = 0,\ c = 0,\ c = 1,\ c = 1.$$

So the solution set in this case is $(\varepsilon' \times \{0\}) \cup (\varepsilon'' \times \{1\})$, where ε', ε'' are sets of pairs (x_{n-1}, y_{n-1}), $|\varepsilon'| = |\varepsilon''| = 2$.

Let us introduce the function

$$S_k(T, \alpha, \beta) = \{(x + \alpha \cdot 2^k, y + \beta \cdot 2^k) \mid (x, y) \in T\},$$

where T is a set of pairs (x, y) for some $x, y \in \{0, \ldots, 2^k - 1\}$. It is easy to see that for each T holds $|S_k(T, \alpha, \beta)| = |T|$. Denote

$$\Psi = \Big\{(x_{n-1}, y_{n-1}, c) \;\Big|\; x_{n-1}, y_{n-1}, c \in \{0, 1\},\ (x_{n-1}, y_{n-1}, c) \text{ are solutions of the equation (14)}\Big\}.$$

We note that in the above notation

$$M_n = \bigsqcup_{x_{n-1}, y_{n-1} \in \{0,1\}} S_{n-1}(\widehat{U}_{n-1}, x_{n-1}, y_{n-1}),$$

$$M_{n+1} = \bigsqcup_{x_n,y_n\in\{0,1\}} \bigsqcup_{(x_{n-1},y_{n-1},c)\in\Psi} S_n(S_{n-1}(\widehat{U}_{n-1}^{(c)}, x_{n-1}, y_{n-1}), x_n, y_n).$$

Furthermore,

$$|M_n| = 4|\widehat{U}_{n-1}|,$$
$$|M_{n+1}| = 4 \sum_{(x_{n-1},y_{n-1},c)\in\Psi} |\widehat{U}_{n-1}^{(c)}|,$$
$$\widehat{U}_{n-1} = \widehat{U}_{n-1}^{(0)} \bigsqcup \widehat{U}_{n-1}^{(1)}.$$

That is,

$$\frac{|M_{n+1}|}{|M_n|} = \frac{\sum_{(x_{n-1},y_{n-1},c)\in\Psi} |\widehat{U}_{n-1}^{(c)}|}{|\widehat{U}_{n-1}|} = \begin{cases} \frac{\sum_{\Psi}|\widehat{U}_{n-1}|}{|\widehat{U}_{n-1}|}, & \text{if } \Delta x_{n-1} = 0, \\ \frac{2|\widehat{U}_{n-1}^{(0)}|+2|\widehat{U}_{n-1}^{(1)}|}{|\widehat{U}_{n-1}|}, & \text{if } \Delta x_{n-1} = 1, \end{cases}$$

$$= \begin{cases} |\Psi|, & \Delta x_{n-1} = 0, \\ \frac{2|\widehat{U}_{n-1}|}{|\widehat{U}_{n-1}|}, & \Delta x_{n-1} = 1, \end{cases} = \begin{cases} \sigma(z_{n-1}, z_n), & \Delta x_{n-1} = 0, \\ 2, & \Delta x_{n-1} = 1, \end{cases}$$

Lemma 1 is proved.

B The proof of Lemma 3

We will show that for any $\varepsilon > 0$ and $n \geqslant N$ for some number N, the inequality $\frac{F(n)}{G(n)} \leqslant 1 + \varepsilon$ holds. By the assumption of Lemma, $\frac{f(k)}{g(k)} \leqslant 1 + \frac{\varepsilon}{2}$ for all $k \geqslant N_1$ for some number N_1. Equivalently, for all $k \geqslant N_1$ holds $f(k) \leqslant (1 + \frac{\varepsilon}{2})g(k)$. Then

$$F(n) = \sum_{k=1}^{n} f(k) = \sum_{k=1}^{N_1-1} f(k) + \sum_{k=N_1}^{n} f(k) \leqslant \sum_{k=1}^{N_1-1} f(k) + \left(1 + \frac{\varepsilon}{2}\right) \sum_{k=N_1}^{n} g(k) =$$
$$= \sum_{k=1}^{N_1-1} f(k) - \left(1 + \frac{\varepsilon}{2}\right) \sum_{k=1}^{N_1-1} g(k) + \left(1 + \frac{\varepsilon}{2}\right) \sum_{k=1}^{n} g(k) =$$
$$= \sum_{k=1}^{N_1-1} \left(f(k) - \left(1 + \frac{\varepsilon}{2}\right) g(k)\right) + \left(1 + \frac{\varepsilon}{2}\right) G(n).$$

That is, for all $n \geqslant N_1$ for some number N_1

$$F(n) \leqslant \left(1 + \frac{\varepsilon}{2}\right) G(n) + c.$$

Since $g(k)$ is a monotonically increasing unbounded function, then for all $n \geqslant N_2$ for some number N_2 holds

$$g(n) \geqslant c \cdot \frac{2}{\varepsilon}.$$

Then

$$G(n) \geqslant c \cdot \frac{2}{\varepsilon},$$

hence

$$c \leqslant \frac{\varepsilon}{2} G(n).$$

And for all $n \geqslant \max\{N_1, N_2\}$

$$F(n) \leqslant \left(1 + \frac{\varepsilon}{2}\right) G(n) + c \leqslant \left(1 + \frac{\varepsilon}{2}\right) G(n) + \frac{\varepsilon}{2} G(n) = (1 + \varepsilon) G(n).$$

Similarly, for any $\varepsilon > 0$ starting with some n, the inequality $\frac{F(n)}{G(n)} \geqslant 1 - \varepsilon$ holds. So

$$\lim_{n \to \infty} \frac{F(n)}{G(n)} = 1 \Leftrightarrow F(n) \sim G(n).$$