Tatra Mt. Math. Publ. **29** (2004), 183–191

TATRA MOUNTAINS
Mathematical Publications

ON NON-POLYNOMIALITY OF XOR OVER $\mathbb{Z}_2^n$

Otokar Grošek — Mirka Miller — Joe Ryan

ABSTRACT. It turns out that Latin squares which are hard to approximate by a polynomial are suitable to be used as a part of block cipher algorithms. In this paper we state that XOR operation $\oplus$ over $\mathbb{Z}_2 \times \mathbb{Z}_2 \times \cdots \times \mathbb{Z}_2 = \mathbb{Z}_2^n$ as a coordinate-vise mod 2 group operation is possible to approximate by a polynomial over $\mathbb{Z}_{2^n}$ for $n = 1, 2, 3$. Moreover, such a polynomial does not exist for any $n \geq 4$.

1. Introduction

Let $\mathbb{Z}_{2^n}$ be the ring of integers taken $\bmod 2^n$. In this paper we use $\mathcal{F}(n)$ for the set of all polynomial functions $f\colon \mathbb{Z}_{2^n} \times \mathbb{Z}_{2^n} \to \mathbb{Z}_{2^n}$,

$$
\begin{aligned}
f(u, v) &= a_0 + a_{10}u + a_{01}v + a_{20}u^2 + a_{11}uv + a_{02}v^2 + \dots \\
&= a_0 + \sum_{i,j} a_{ij} u^i v^j \mod 2^n, \qquad (1)
\end{aligned}
$$

where $a_0, a_{ij} \in \mathbb{Z}_{2^n}$. The XOR operation $\oplus$ over $\mathbb{Z}_2 \times \mathbb{Z}_2 \times \cdots \times \mathbb{Z}_2 = \mathbb{Z}_2^n$ can be viewed as a coordinate-vise mod 2 group operation. Then there is a one to one natural mapping, say β, $\beta\colon \mathbb{Z}_2^n \to \mathbb{Z}_{2^n}$ such that for $x = (x_0, x_1, \dots, x_{n-1}) \in \mathbb{Z}_2^n$, $\beta(x) = \sum\limits_{i=0}^{n-1} 2^i x_i \in \mathbb{Z}_{2^n}$.

Let $L(n)$ be the Latin square of order 2^n with the symbol set $\{0, 1, \dots, 2^n - 1\}$ corresponding to the group $(\mathbb{Z}_2^n, \oplus)$, and where elements are denoted using β. Then for $n = 2, 3$ we have for $L(2)$ and $L(3)$, Tables 1 and 2 below.

Consider the following scenario: An attacker has access to outputs from a composition $x \oplus k$ of messages x and round keys k, both belonging to the group $(\mathbb{Z}_2^n, \oplus)$. Then his/her aim is to find a polynomial function $f \in \mathcal{F}(n)$ such that for all $\beta(x), \beta(y) \in \mathbb{Z}_{2^n}$, $\beta(x \oplus y) = f\bigl(\beta(x), \beta(y)\bigr)$. If this is the case then such an operation is easy to approximate by a polynomial in $\mathcal{F}(n)$.

2000 Mathematics Subject Classification: 68P30.
Keywords: latin squares, polynomial approximation of XOR, block ciphers.

Table 1. Cayley table for group $(Y_2^2, \oplus)$ when mapped by β.

$\oplus$	0	1	2	3
0	0	1	2	3
1	1	0	3	2
2	2	3	0	1
3	3	2	1	0

Table 2. Cayley table for group $(Y_2^3, \oplus)$ when mapped by β.

$\oplus$	0	1	2	3	4	5	6	7
0	0	1	2	3	4	5	6	7
1	1	0	3	2	5	4	7	6
2	2	3	0	1	6	7	4	5
3	3	2	1	0	7	6	5	4
4	4	5	6	7	0	1	2	3
5	5	4	7	6	1	0	3	2
6	6	7	4	5	2	3	0	1
7	7	6	5	4	3	2	1	0

The main goal of this paper is to show that for XOR operation there exist such polynomials for $n = 1, 2, 3$, and such polynomials do not exist for even $n \geq 4$.

We will understand the Cayley table of $(\mathbb{Z}_2^n, \oplus)$, $\beta(\mathbb{Z}_2^n) = \{0, 1, \dots, 2^n - 1\}$, as a Latin square $L = L(\ell_{ij})$ with $\ell_{ij} = \beta(i \oplus j)$. Therefore the notions of a group and a Latin square will be freely interchanged in the paper. Since it is more natural to express definitions and results concerning the topic in the language of Latin squares, the notion of a group will be used only sporadically.

If there is a polynomial function $f \in \mathcal{F}(n)$ such that for all $x, y \in \mathbb{Z}_2^n$, $\beta(x) = i$, $\beta(y) = j$, $\beta(x \oplus y) = \ell_{ij} = f(i, j)$, then the Latin square $L = L(\ell_{ij})$ (the group $(\mathbb{Z}_2^n, \oplus)$) is called polynomial, otherwise we call L non-polynomial ([2]). A simple example of polynomial Latin squares are the groups represented by $L(2)$ and $L(3)$ with the operation $\oplus$ defined above.

Example 1.1. Let f_2, f_3 be the respective polynomials. It was found, by an exhaustive computer search, that

$$\begin{aligned}
\beta(x \oplus y) = f_2\big(\beta(x), \beta(y)\big) &= \beta(x) + \beta(y) + 2\beta(x)\beta(y) \mod 4\,,\\
\beta(x \oplus y) = f_3\big(\beta(x), \beta(y)\big) &= \beta(x) + \beta(y) + 3\beta(x)\beta(y) + \beta(x)\beta(y)^2\\
&\quad + \beta(x)^2\beta(y) + \beta(x)^2\beta(y)^2 \mod 8\,.
\end{aligned}$$

2. XOR is polynomial over $\mathbb{Z}_{2^n}$, $n = 1, 2, 3$

Since from the context the difference between x and $\beta(x)$ is always clear, in the sequel we omit the mapping β. Then, e.g.,

$$\begin{aligned} f_1(x,y) &= x + y \mod 2\,, \\ f_2(x,y) &= x + y + 2xy \mod 4\,, \\ f_3(x,y) &= x + y + 3xy + xy^2 + x^2y + x^2y^2 \mod 8\,. \end{aligned} \tag{2}$$

From Tab. 1 and 2, one can easily derive the following general properties of f_n.

1. $x \oplus x = 0$ or
$$f_n(x,x) = 0 \mod 2^n, \tag{3}$$
which yields $a_0 = 0$ in (1);
2. Symmetry
$$f_n(x,y) = f_n(y,x) \mod 2^n, \tag{4}$$
which yields $a_{ij} = a_{ji}$ in (1);
3. Let λ be the Carmichael function. Then for $m = 2^n$ we have
$$\lambda(m) = \begin{cases} 1 & \text{if } n = 0,1\,, \\ 2 & \text{if } n = 2\,, \\ 2^{n-2} & \text{if } n > 2\,, \end{cases}$$
and due to the Global Euler–Fermat theorem (see [4]) for any $x \in \mathbb{Z}_{2^n}$ we have
$$x^{n+\lambda(2^n)} = x^n \mod 2^n, \tag{5}$$
which yields an upper bound for $i,j \le d$, $d = n + \lambda(2^n) - 1$ in (1). For example, for $n = 2, 3$ and 4, e.g., we have $i, j \le 3, 4$ and 7 respectively.
4. There are polynomials such that for any $x \in \mathbb{Z}_{2^n}$, $h(x) = 0 \bmod 2^n$. Thus polynomials f_n are not uniquely determined. For instance, it is easy to prove that for any $x \in \mathbb{Z}_{2^n}$,
$$h_n(x) = 2^{n-1}x(x+1) = 0 \mod 2^n, \tag{6}$$
and this polynomial is of the lowest order with such a property. There are some other polynomials which at first glance look different. Here are examples of such polynomials: for any $x, y \in \mathbb{Z}_8$
$$\begin{aligned} 4(x+y) + 4(x^2+y^2) &= 0 \mod 8\,; \\ 4xy(x+y) &= 0 \mod 8\,; \\ 2xy + 2(xy^2 + x^2y) + 2x^2y^2 &= 0 \mod 8\,. \end{aligned}$$
Hence, e.g., polynomial f_3 mentioned above is one of possible only.

Next we rewrite $f_n(x,y)$ in the following form which in fact allows us to prove our main theorem:

$$f_n(x,y) = \sum_{j=1}^{d} a_{0j}(x^j + y^j) + \sum_{i=1}^{d} x^i g_i(y) \mod 2^n, \tag{7}$$

where

$$g_i(y) = \sum_{j=1}^{d} a_{ij} y^j, \quad a_{ij} = a_{ji}\,, \tag{8}$$

which yields a matrix form

$$f_n(x,y) - \sum_{j=1}^{d} a_{0j}(x^j + y^j) = (x, x^2, \dots x^d) A (y, y^2, \dots y^d)' \mod 2^n \tag{9}$$

with a symmetric matrix $A = (a_{ij})$. Thus the number of unknown coefficients is $2d + (d^2 - d)/2 = d(d+3)/2$.

2.1 Case $n = 2$

In this case the powers of $x^i y^j$ are $d = 2 + \lambda(2^2) - 1 = 3$ at most. Let $u(x,y)$ be any polynomial of two variables over $\mathbb{Z}_4$, and h_2 is defined in (6). Then clearly the polynomial

$$f_2(x,y) + h_2(x)u(x,y)$$

is also an approximation polynomial. Hence we can list all the possible polynomials which approximate $L(2)$ over $\mathbb{Z}_4$. With respect to the symmetry (4) we have 9 unknown coefficients in (9), and thus $4^9 = 2^{18}$ possibilities. By an exhaustive computer search one can find 1023 different polynomials f_2.

2.2 Case $n = 3$

In this case the powers of $x^i y^j$ are $d = 3 + \lambda(2^3) - 1 = 4$ at most. Let again $u(x,y)$ be any polynomial of two variables over $\mathbb{Z}_8$, and h_3 is defined in (6). Similarly to the previous case, polynomial

$$f_3(x,y) + h_3(x)u(x,y)$$

is also an approximation polynomial. With respect to the symmetry (4) we have 14 unknown coefficients in (9), and thus $8^{14} = 2^{42}$ possibilities. Unfortunately, in this case an exhaustive computer search is outside of possibilities of standard computers.

2.3 Case $n = 4$

In this case the powers of $x^i y^j$ are $d = 4 + \lambda(2^4) - 1 = 7$ at most. Let $u(x, y)$ be any polynomial of two variables over $\mathbb{Z}_{16}$, and h_4 is defined in (6). Then again if such a polynomial exists then

$$f_4(x, y) + h_4(x) u(x, y)$$

is also an approximation polynomial. In this case it is not possible to find by exhaustive search all the polynomials which approximate $L(4)$ over $\mathbb{Z}_{16}$. Recall that matrix A is of size 7×7, and thus the number of unknowns is $14 + (7^2 - 7)/2 = 35$. Hence the computational complexity is at least $16^{35} = 2^{140}$ plus comparisons with $L(4)$.

Let us suppose that in (9) $a_{01} = 1$, $a_{0j} = 0$ for $j \geq 2$. Then it is straightforward to see that $f_n(x, 0) = x$. Moreover, using the binary expansion of x and y respectively, we get the difference between $\beta(x \oplus y)$ and $x + y \bmod 2^n$ on the left side of (9)

$$f_n(x, y) - (x + y) = -2 \sum_{i:\ x_i = y_i} 2^i x_i . \tag{10}$$

We observed this form in the cases $n = 1, 2, 3$. Unfortunately this is not the case now. An easy comparison of the left and right sides of (9) for $x = y = 4$ shows 0 on the right side and 8 on the left side. Thus we will search for the approximation polynomial in the general form of

$$f_4(x, y) - \sum_{j=1}^{7} a_{0j}(x^j + y^j) = xy(1, x, x^2, \dots x^6) A (1, y, y^2, \dots y^6)' \mod 2^4 \tag{11}$$

which requires for $y = 0$ a permutation polynomial (in fact identity)

$$f_4(x, 0) = x = \sum_{j=1}^{7} a_{0j} x^j \mod 2^4, \tag{12}$$

and for $x = y = 4$,

$$-\sum_{j=1}^{7} 2 a_{0j} 4^j = 8 a_{01} = 0 \mod 2^4. \tag{13}$$

We know[1] that

$$f_4(x, 0) = x + h_4(x) g(x) = x + 8x(x + 1) g(x) \tag{14}$$

[1] A standard procedure for $x = 1, 2, \dots, 7$ would lead to the Vandermond determinant which is equal to zero in this case.

where $g(x)$ is an arbitrary polynomial. But this is a contradiction with (13) since in this case a_{01} must be odd. Thus we conclude that such a polynomial does not exist.

Similar arguments are valid for all even values of n. The contradiction will be obtained for $x = y = 2^{n/2}$. Thus we proved

THEOREM 2.1. *Let $n \geq 4$ be an even number. Then there is no polynomial $f \in \mathcal{F}(n)$ such that it approximates the Latin square $L = L(\ell_{ij})$ for the group $(\mathbb{Z}_2^n, \oplus)$.*

2.4 Case $n \geq 5$

Clearly, any approximation polynomial $f(x,y)$ for a Latin square must be a permutation polynomial, when one variable is fixed. The following is known for polynomials over $\mathbb{Z}_{2^n}$, $n > 1$.

Let $r\colon \mathbb{Z}_{2^n} \to \mathbb{Z}_{2^n}$ be a polynomial. D u e b a l l ([1]) proved that the values of $r(x)$ are determined $\bmod 2^n$ for all values of x, if $r(x)$ is known for $x = 0, 1, \dots, 2(\alpha + 1) - 1$, where α is defined as follows: Let 2^{n_ν} be the highest power of 2 which divides $2^\nu \nu!$. The integers $n_0 = 1, n_1, \dots$ form an increasing sequence, and hence to every n there exists a unique integer $\alpha(n) = \alpha$ such that $n_\alpha < n \leq n_{\alpha+1}$. Table 3 illustrates this result where $e(\nu) = \max\{k\colon 2^k | \nu!\}$.

TABLE 3. Table for the sequence n_ν, and numbers $\alpha(n) + 1$.

ν	$\nu!$	$e(\nu)$	$n_\nu = \nu + e(\nu)$	n	$\alpha(n) + 1$
2	2	1	3	2	2
3	6	1	4	3	2
4	24	3	7	4	3
5	120	3	8	5	4
6	720	4	10	6	4
7	7!	4	11	7	4
8	8!	7	15	8	5
9	9!	7	16	9	6
10	10!	8	18	10	6

In the table there is a value $n_\nu = \nu + e(\nu)$. It is not difficult to see that

$$n_\nu = \nu + e(\nu) = \nu + \sum_{j=1}^{\lfloor \log_2 \nu \rfloor} \left\lfloor \frac{\nu}{2^j} \right\rfloor = \sum_{j=0}^{\lfloor \log_2 \nu \rfloor} \left\lfloor \frac{\nu}{2^j} \right\rfloor < 2\nu . \tag{15}$$

Thus for $n \le n_\nu < 2\nu$ we have

$$\nu = \alpha(n) + 1 \ge \lfloor n/2 \rfloor + 1 . \tag{16}$$

Next relation gives an exact value n_ν in a special case when $\nu = 2^\ell$. Then (15) yields

$$e(\nu) = \sum_{j=1}^{\ell} \frac{\nu}{2^j} = \sum_{j=1}^{\ell} 2^{\ell-j} = 2^\ell - 1 \tag{17}$$

or $n_\nu = \nu + e(\nu) = 2^{\ell+1} - 1$.

Let $n > 1$ be from the interval $n_{2^{\ell-1}} < n \le n_{2^\ell}$. Then $2^\ell - 1 < n \le 2^{\ell+1} - 1$, and

$$2^\ell - 1 < 2(\alpha(n) + 1) - 1 \le 2^{\ell+1} - 1 . \tag{18}$$

Thus from Dueball's result and (18) have

Lemma 2.2. *Let $r\colon \mathbb{Z}_{2^n} \to \mathbb{Z}_{2^n}$ be a polynomial, $2^\ell - 1 < n \le 2^{\ell+1} - 1$. Then the values of $r(x)$ are determined $\bmod 2^n$ for all values of x, if $r(x)$ is known for $x = 0, 1, \dots, 2\big(\alpha(n) + 1\big) - 1$ where $2(\alpha(n) + 1) - 1$ is bounded by (18).*

To illustrate the result, for $n = 5$ we have $2^2 - 1 < 5 \le 2^3 - 1$, and $2^2 - 1 < 2\big(\alpha(5) + 1\big) - 1 \le 2^3 - 1$. From Table 3 we have the exact value for $2\big(\alpha(5) + 1\big) - 1 = 7$. Thus the values of a polynomial $r\colon \mathbb{Z}_{2^5} \to \mathbb{Z}_{2^5}$ are determined $\bmod 2^5$ for all values of x, if $r(x)$ is known for eight values $x = 0, 1, \dots, 7$. Clearly, if $r(x)$ coincide with some values $\bmod 2^n$ then it coincides $\bmod 2^{n-t}$, $1 \le t < n$, too.

Let $f_n(x, y)$ be any approximation polynomial for the Latin square $L(n)$ of order 2^n. Then for any fixed $x = a$, $f(a, y)$ must be a permutation polynomial over the ring $\mathbb{Z}_{2^n}$. Due to Lemma 2.2 this polynomial is determined $\bmod 2^n$ by its values of x, if $f(a, y)$ is known at least for 2^ℓ values $y = 0, 1, \dots, 2^\ell - 1$. For $n = 4, 5, \dots, 7$ this yields $y = 0, 1, \dots, 7$. Similarly for $f(x, b)$. Moreover, by Lemma 2.2, e.g., $f_5 \bmod 2^3$ would be one of approximation polynomials for $L(3)$, $f_5 \bmod 2^3 \in \mathcal{F}(3)$.

Theorem 2.3. *There is no polynomial $f \in \mathcal{F}(5)$ such that it approximates Latin square $L = L(\ell_{ij})$ for the group $(\mathbb{Z}_2^5, \oplus)$.*

P r o o f. Let $f_5(x, y)$ be any approximation polynomial for the Latin square $L(5)$ of order 2^5. Then for any fixed $x = a$, $f_5(a, y)$ must be a permutation polynomial over the ring $\mathbb{Z}_{2^5}$. Especially, this is true for $0 \le a < 2^4 - 1$. By symmetry, the same is true for $f_5(x, b)$, $0 \le b < 2^4 - 1$. Thus if $f_5(x, y)$ would be any approximation polynomial for the Latin square $L(5)$ of order 2^5, then it should be an approximation polynomial for the Latin square $L(4)$ of order 2^4, a contradiction with Theorem 2.1. $\square$

COROLLARY 2.4. *Let $n \geq 4$. Then there is no polynomial $f \in \mathcal{F}(n)$ such that it approximates Latin square $L = L(\ell_{ij})$ for the group $(\mathbb{Z}_2^n, \oplus)$.*

3. Conclusions

It turns out that Latin squares which are hard to approximate by a polynomial are suitable to be used as a part of block cipher algorithms. In this paper we proved that XOR operation $\oplus$ over $\mathbb{Z}_2 \times \mathbb{Z}_2 \times \cdots \times \mathbb{Z}_2 = \mathbb{Z}_2^n$ as a coordinate-vise $\bmod 2$ group operation is possible to approximate by a polynomial over $\mathbb{Z}_{2^n}$ for $n = 1, 2, 3$. Moreover, such a polynomial does not exist for any $n \geq 4$.

We recall (see [2]) a similar question about polynomial approximation over the field $\mathrm{GF}(2^n)$. To any Latin square $L = L(\ell_{ij})$ of order 2^n there exists a unique polynomial f_n such that for all i, j, $f_n(i,j) = \ell_{ij}$. This polynomial is given by ($r = 2^n$)

$$f_n(x,y) = \sum_{u=0}^{r-1} \sum_{v=0}^{r-1} \left(1 - (x-u)^{n-1}\right)\left(1 - (y-v)^{n-1}\right) \ell_{uv}. \tag{19}$$

Thus any Latin square of size 2^n is a polynomial over the field $\mathrm{GF}(2^n)$. Unfortunately, this polynomial is of a very high order. Our results show that this is not the case over the ring $\mathbb{Z}_{2^n}$, $n > 3$. Moreover, we emphasize that block ciphering algorithms make use of this ring to operate on blocks of data, and thus the question about approximation over the ring $\mathbb{Z}_{2^n}$ is closer to practice.

Lai in [3] investigated relations between the groups $\left(\mathbb{Z}_2^n, \oplus\right)$ and $\left(\mathbb{Z}_2^n, +\right)$. Clearly, these groups are not isotopic as Latin squares because the latter one is the cyclic group while the first one is not. Moreover, he studied polynomial expressions over the ring $\mathbb{Z}_{2^n}$ for multiplication modulo $2^n + 1$, $n = 2, 4, 8, 16$. This research was motivated by his design of IDEA cipher.

REFERENCES

[1] DUEBALL, F.: *Bestimmung von Polynomen aus ihren Werten* mod p^n, Math. Nachr. **3** (1949), 71–76.

[2] GROŠEK, O.—HORÁK, P.—TRUNG VAN TRAN: *On non-polynomial Latin squares*, Des., Codes Cryptogr. **32** (2004), 217–226.

[3] XUEJIA, LAI: *On the Design and Security of Block Ciphers*, ETH Series in information processing, Vol. 1, Zürich, Hartung-Gorre Verlag, 1992.

[4] SCHWARZ, Š.: *The role of semigroups in the elementary theory of numbers*, Math. Slovaca **4** (1981), 369–395.

Received December 15, 2003

Otokar Grošek
Department of Mathematics
Slovak University of Technology
Ilkovičova 3
SK–812 19 Bratislava
SLOVAKIA
E-mail: grosek@kmat.elf.stuba.sk

Mirka Miller, Joe Ryan
School of ITMS
University of Ballarat
Ballarat VIC 3353
AUSTRALIA
E-mail: m.miller@ballarat.edu.au
joe.ryan@ballarat.edu.au