Distribution of 0 and 1 in the highest level of primitive sequences over $\mathbb{Z}/(2^e)$ *

QI Wenfeng (戚文峰) and ZHOU Jinjun (周锦君)

(Department of Applied Mathematics, Zhengzhou Information Engineering Institute, Zhengzhou 450002, China)

Received December 2, 1996

Abstract The distribution of 0 and 1 is studied in the highest level a_{e-1} of primitive sequences over $\mathbb{Z}/(2^e)$, and the upper and lower bounds on the ratio of the number of 0 to the number of 1 in one period of a_{e-1} are obtained. It is revealed that the larger e is, the closer to 1 the ratio will be.

Keywords: linear recurring sequence, primitive sequence, highest level sequence, distribution of 0 and 1.

Let $\mathbb{Z}$ be the ring of integers, and let $\mathbb{Z}/(2^e)$ be the residue ring of $\mathbb{Z}$ modulo 2^e. Let $f(x) = x^n + c_{n-1}x^{n-1} + \cdots + c_0$ be a monic polynomial with coefficients in $\mathbb{Z}/(2^e)$. We say that the sequence $a = (a_0, a_1, a_2, \cdots)$ over $\mathbb{Z}/(2^e)$ satisfying the linear recursion

$$a_{i+n} = -(c_0a_i + c_1a_{i+1} + \cdots + c_{n-1}a_{i+n-1}), i = 0, 1, 2, \cdots \quad (1)$$

is a linear recurring sequence generated by $f(x)$, and $f(x)$ is called a characteristic polynomial of a. $G(f(x))_e$ denotes the set of all sequences over $\mathbb{Z}/(2^e)$ generated by $f(x)$.

Remark. Recursion (1) is equivalent to $f(x)\ a = 0 = (0, 0, 0, \cdots)$, where x is the left-shift operator; that is, $xa = (a_1, a_2, a_3, \cdots)$.

For each element b in $\mathbb{Z}/(2^e)$, there exists a unique binary decomposition

$$b = b_0 + b_1 2 + \cdots + b_{e-1}2^{e-1},$$

where $b_i = 0$ or 1, and b_i is called the ith level bit of b.

Similarly, the sequence a over $\mathbb{Z}/(2^e)$ has a unique binary decomposition

$$a = a_0 + a_1 2 + \cdots + a_{e-1}2^{e-1},$$

where $a_i = (a_{i0}, a_{i1}, a_{i2}, \cdots)$ is binary sequences with $a_{ij} = 0$ or 1, a_i is called the ith level sequence of a, and a_{e-1} is called the highest level of a.

For a monic polynomial $f(x)$ over $\mathbb{Z}/(2^e)$, if $f(0)$ (i.e. c_0) is an invertible element, then there exists a positive integer T such that $f(x)$ divides $x^T - 1$ over $\mathbb{Z}/(2^e)$, and the smallest T is called the period of $f(x)$ over $\mathbb{Z}/(2^e)$, denoted by $\mathrm{per}(f(x))_e$. By ref. [1], $\mathrm{per}(f(x))_e \leqslant 2^{e-1}(2^n - 1)$, where $n = \deg f(x)$. If $\mathrm{per}(f(x)) = 2^{e-1}(2^n - 1)$, $f(x)$ is called a primitive polynomial over $\mathbb{Z}/(2^e)$ with degree n. Ref. [2] provides a coefficient criterion for primitiveness of polynomials over $\mathbb{Z}/(2^e)$. The sequences generated by a primitive polynomial are called primi-

* Project supported by the State Key Laboratory of Information Security, Graduate School of Chinese Academy of Sciences.

tive sequences over $\mathbb{Z}/(2^e)$. Ref. [3] has shown the following entropy-preservation theorem with significance of cryptography.

Let $f(x)$ be a primitive polynomial over $\mathbb{Z}/(2^e)$ and a, $b \in G(f(x))_e$. Then $a = b$ if and only if $a_{e-1} = b_{e-1}$.

Reference [4] presented the lower bounds on linear complexity of a_{e-1} and ref. [5] studied the minimal polynomial of a_{e-1}. These results have shown the prospects for application of the highest level sequences as cryptographic sequences.

For another problem of cryptographic sequences, we shall study the distribution of 0 and 1 in a_{e-1}. The results show that if e is sufficiently large, the ratio of the number of 0 to the number of 1 in one period of a_{e-1} is close to 1.

1 The distribution of 0 and 1 in a_{e-1}

We always let $f(x)$ be a primitive polynomial of degree n over $\mathbb{Z}/(2^e)$, $a \in G(f(x))_e$ and $a \not\equiv 0 \bmod 2$, i.e. $a_0 \neq 0$. By ref. [5], the period per (a_k) of kth level sequence a_k of a is $2^k T$, where $T = 2^n - 1$. By refs. [1, 5], for $1 \leqslant k \leqslant e-1$, over $\mathbb{Z}/(2^e)$ we have

$$x^{2^{k-1}T} - 1 \equiv 2^k h_k(x) \qquad \bmod f(x),$$

where $h_k(x)$ is a polynomial over $\mathbb{Z}/(2^e)$ with degree less than n and $h_k(x) \not\equiv 0 \bmod 2$.

Set $d = [e/2]$ and set $s \equiv h(x)a \pmod{2^d}$ to be a sequence over $\mathbb{Z}/(2^d)$, where

$$h(x) = \begin{cases} h_d(x), & e = 2d, \\ h_{d+1}(x), & e = 2d+1. \end{cases} \tag{2}$$

Remark. (i) $s \in G(f(x))_d$ and $s \not\equiv 0 \bmod 2$.

(ii) While $a \pmod{2^d}$ takes over all sequences with $a_0 \neq 0$ in $G(f(x))_d$, s takes over all sequences with $s_0 \neq 0$ in $G(f(x))_d$ too.

(iii) The period $\mathrm{per}(s)_d$ of s over $\mathbb{Z}/(2^d)$ is $2^{d-1}T$.

$\mathrm{N}(s,0)$ denotes the number of 0 in one period of s, $\mathrm{N}(a_{e-1},0)$ and $\mathrm{N}(a_{e-1},1)$ denote the numbers of 0 and 1 in a period of a_{e-1}. We obtain the following result of the distribution of 0 and 1.

Theorem 1. *Let $f(x)$ be a primitive polynomial of degree n over $\mathbb{Z}/(2^e)$, $d = [e/2]$, $T = 2^n - 1$, $a \in G(f(x))_e$ and $a_0 \neq 0$, $s \equiv h(x)a \bmod 2^d$, where $h(x)$ is defined by (2). Then*

$$\frac{2^{d-1}T - \mathrm{N}(s,0)}{2^{d-1}T + \mathrm{N}(s,0)} \leqslant \frac{\mathrm{N}(a_{e-1},0)}{\mathrm{N}(a_{e-1},1)} \leqslant \frac{2^{d-1}T + \mathrm{N}(s,0)}{2^{d-1}T - \mathrm{N}(s,0)}.$$

To prove Theorem 1, we first introduce the following lemma.

Lemma 1. *Let u, $v \in \mathbb{Z}/(2^d)$ and $v \neq 0$. Set*

$$S(i) = \{k \in \mathbb{Z}/(2^d) \mid \text{the } (d-1)\text{th level of } u + kv \text{ is } i\},$$

$i = 0, 1$. *Then* $|S(0)| = |S(1)| = 2^{d-1}$.

Proof. Since $v\neq 0$, we let the binary decomposition of v be

$$v = v_j 2^j + v_{j+1}2^{j+1} + \cdots + v_{d-1}2^{d-1},$$

where $0\leqslant j\leqslant d-1$, $v_j = 1$.

Set $v' = v_j + v_{j+1}2 + \cdots + v_{d-1}2^{d-j-1}$. Then $v = v'2^j$. Since $v_j = 1$, v' is an invertible element in $\mathbb{Z}/(2^d)$, i.e. there exists $s\in\mathbb{Z}/(2^d)$ such that $v's = 1$. We have $vs = 2^j$. Then for any $t\in\mathbb{Z}/(2^d)$ and $t\equiv 0 \bmod 2^j$, there exists $w\in\mathbb{Z}/(2^d)$ such that $t = vw$. And for any $w'\equiv w \bmod 2^{d-j}$, $vw' = vw = t$.

So while k takes over all elements in $\mathbb{Z}/(2^d)$, kv takes over all elements with the form $2^j t$ in $\mathbb{Z}/(2^d)$ and every such element occurs 2^j times. Let

$$u = u_0 + u_1 2 + \cdots + u_{j-1}2^{j-1} + u_j 2^j + \cdots + u_{d-1}2^{d-1},$$

and let $u' = u_0 + u_1 2 + \cdots + u_{j-1}2^{j-1}$. Then while k takes over all elements in $\mathbb{Z}/(2^d)$, $u + kv$ takes over all elements with form $u' + 2^j t$ in $\mathbb{Z}/(2^d)$ and every such element occurs 2^j times. It is easy to get $|S(0)| = |S(1)|$. Since

$$|S(0)| + |S(1)| = 2^d,$$

we have

$$|S(0)| = |S(1)| = 2^{d-1}.$$

Proof of Theorem 1. First let $e = 2d$. Since $x^{2^{d-1}T} - 1 \equiv 2^d h(x) \bmod f(x)$, over $\mathbb{Z}/(2^e)$ we have

$$(x^{2^{d-1}T} - 1)a = 2^d h(x)a.$$

By $a = a_0 + a_1 2 + \cdots + a_{e-1}2^{e-1}$ and per $(a_i) = 2^{i-1}T$, over $\mathbb{Z}/(2^e)$ we have

$$(x^{2^{d-1}T} - 1)(a_d + a_{d+1}2 + \cdots + a_{e-1}2^{d-1})2^d = h(x)(a_0 + a_1 2 + \cdots a_{d-1}2^{d-1})2^d.$$

So over $\mathbb{Z}/(2^d)$ we get

$$(x^{2^{d-1}T} - 1)(a_d + a_{d+1}2 + \cdots a_{e-1}2^{d-1}) = h(x)(a_0 + a_1 2 + \cdots + a_{d-1}2^{d-1}). \quad (3)$$

Set $t = a_d + a_{d+1}2 + \cdots + a_{e-1}2^{d-1} = (t_0, t_1, t_2, \cdots)$, $t_i\in\mathbb{Z}/(2^d)$, $i = 0,1,2,\cdots$. Then by (3)

$$(x^{2^{d-1}T} - 1)t = s. \quad (4)$$

Let $s = (s_0, s_1, s_2, \cdots)$, and set $R = 2^{d-1}T$. Then by (4), for any integer $i\geqslant 0$, $t_{i+R} = t_i + s_i$. So for any positive integer k,

$$t_{i+kR} = t_{i+(k-1)R} + s_{i+(k-1)R}.$$

Since per$(s)_d = R$; that is $s_{i+R} = s_i$, we have

$$t_{i+kR} = t_i + ks_i. \quad (5)$$

While i takes over 0 to $R-1$ and k takes over 0 to $2^d - 1$, t_{i+kR} exactly takes over first period of t.

For a fixed i, $0\leqslant i\leqslant R-1$, if $s_i\neq 0$, then while k takes over 0 to $2^d - 1$ and by Lemma 1, the $(d-1)$th level bit of t_{i+kR} takes 0 and 1 2^{d-1} times, respectively. If $s_i = 0$, then while k takes over 0 to $2^d - 1$, the $(d-1)$th level bit of t_{i+kR} always takes the $(d-1)$th level bit of t_i. If $\mathrm{N}(s,0)$ denotes the number of 0 in one period of s, then the number of nonzero in one period of s is $2^{d-1}T - \mathrm{N}(s,0)$. So the number $\mathrm{N}(t_{d-1},0)$ of 0 in one period of the $(d-1)$th level component of t satisfies

$$(2^{d-1}T - \mathrm{N}(s,0))2^{d-1} \leqslant \mathrm{N}(t_{d-1},0) \leqslant (2^{d-1}T - \mathrm{N}(s,0))2^{d-1} + \mathrm{N}(s,0)2^d,$$

Similarly

$$(2^{d-1}T - \mathrm{N}(s,0))2^{d-1} \leqslant \mathrm{N}(t_{d-1},1) \leqslant (2^{d-1}T - \mathrm{N}(s,0))2^{d-1} + \mathrm{N}(s,0)2^{d}.$$

Since $t_{d-1} = a_{e-1}$, we get

$$M \leqslant \frac{\mathrm{N}(a_{e-1},0)}{\mathrm{N}(a_{e-1},1)} \leqslant \frac{1}{M},$$

where

$$M = \frac{(2^{d-1}T - \mathrm{N}(s,0))2^{d-1}}{(2^{d-1}T - \mathrm{N}(s,0))2^{d-1} + \mathrm{N}(s,0)2^{d}} = \frac{2^{d-1}T - \mathrm{N}(s,0)}{2^{d-1}T + \mathrm{N}(s,0)}.$$

So when $e = 2d$, Theorem 1 holds.

Next suppose $e = 2d + 1$. Since $x^{2^dT} - 1 \equiv 2^{d+1}h(x) \bmod f(x)$, over $\mathbb{Z}/(2^e)$ we have

$$(x^{2^dT} - 1)a = 2^{d+1}h(x)a;$$

$$(x^{2^dT} - 1)(a_{d+1} + a_{d+2}2 + \cdots + a_{e-1}2^{d+1})2^{d+1} = h(x)(a_0 + a_12 + \cdots + a_{d-1}2^{d-1})2^{d+1}.$$

So over $\mathbb{Z}/(2^d)$

$$(x^{2^dT} - 1)(a_{d+1} + a_{d+2}2 + \cdots + a_{e-1}2^{d-1}) = h(x)(a_0 + a_12 + \cdots + a_{d-1}2^{d-1}).$$

Applying the proof method in case $e = 2d$, we can get

$$\frac{2^{d-1}T - \mathrm{N}(s,0)}{2^{d-1}T + \mathrm{N}(s,0)} \leqslant \frac{\mathrm{N}(a_{e-1},0)}{\mathrm{N}(a_{e-1},1)} \leqslant \frac{2^{d-1}T + \mathrm{N}(s,0)}{2^{d-1}T - \mathrm{N}(s,0)}.$$

Remark. Let s be a random variable taken from $G(f(x))_d$ with $s_0 \neq 0$. Then by the following Lemma 2, the average of $\mathrm{N}(s,0)$ is $2^{n-1} - 1$, where $n = \deg f(x)$. If the average could substitute for $\mathrm{N}(s,0)$ in Theorem 1, we could get estimates of $\frac{\mathrm{N}(a_{e-1},0)}{\mathrm{N}(a_{e-1},1)}$ in

$$\frac{2^d - 1}{2^d + 1} < \frac{\mathrm{N}(a_{e-1},0)}{\mathrm{N}(a_{e-1},1)} < \frac{2^d + 1}{2^d - 1}.$$

When e is sufficiently large, the ratio $\frac{\mathrm{N}(a_{e-1},0)}{\mathrm{N}(a_{e-1},1)}$ is close to 1. This would be a good distribution of 0 and 1. But now there are no good estimates of upper bound of $\mathrm{N}(s,0)$. It is not known if there exists a primitive sequence which contains a lot of zero. If such a sequence exists, then using Theorem 1 we cannot deduce whether the distribution is good or bad.

The upper bound of $\mathrm{N}(s,0)$ has not been solved. However, we shall show that when e is sufficiently large, there are few a of which $\frac{\mathrm{N}(a_{e-1},0)}{\mathrm{N}(a_{e-1},1)}$ is not close to 1.

Lemma 2. *Let $f(x)$ be a primitive polynomial of degree n over $\mathbb{Z}/(2^d)$. Set*

$$G'(f(x))_d = \{s \in G(f(x))_d \mid s_0 \neq 0\}.$$

For $s, t \in G'(f(x))_d$, if there exists a non-negative integer i such that $s = x^it$, then s is shift-equivalent to t. $G'(f(x))_d$ can be classified by shift-equivalence. Then

(i) *There are $2^{(n-1)(d-1)}$ shift-equivalent classes in $G'(f(x))_d$ and each class has $2^{d-1}T$ sequences, where $T = 2^n - 1$.*

(ii) *Let $s_{(1)}, s_{(2)}, \cdots, s_{(w)}$ be the representatives of all classes, where $w = 2^{(n-1)(d-1)}$. Then*

$$\sum_{i=1}^{w}\mathrm{N}(s_{(i)},0) = w(2^{n-1}-1) = 2^{(n-1)(d-1)}(2^{n-1}-1).$$

Proof. (i) For a state $u = (u_0, u_1, \cdots, u_{n-1}), u_i \in \mathbb{Z}/(2^d)$, if $u \not\equiv 0 = (0,\cdots,0) \bmod 2$, then u must be a state of one and only one sequence in $\{s_{(1)}, \cdots, s_{(w)}\}$. Conversely, each state u of some sequence in $\{s_{(1)}, \cdots, s_{(w)}\}$ must satisfy $u \not\equiv 0 \bmod 2$. Since the number of states over $\mathbb{Z}/(2^d)$ with $u \not\equiv 0 \bmod 2$ is

$$2^{nd} - 2^{n(d-1)} = 2^{n(d-1)}(2^n - 1) = 2^{n(d-1)}T,$$

and each $s_{(i)}$ has $2^{d-1}T$ states, the number of equivalent classes is $\frac{2^{n(d-1)}T}{2^{d-1}T} = 2^{(n-1)(d-1)}$.

(ii) By the process in the proof of (i), $\sum_{i=1}^{w}\mathrm{N}(s_{(i)},0)$ is the number $|U_0|$ of elements in the set:

$$U_0 = \{u = (0, u_1, \cdots, u_{n-1}) \mid u_i \in \mathbb{Z}/(2^d), \text{ and } u \not\equiv 0 \bmod 2\}.$$

Since

$$\begin{aligned}|U_0| &= 2^{d-1}2^{d(n-2)} + 2^{2(d-1)}2^{d(n-3)} + \cdots + 2^{(d-1)i}2^{d(n-i-1)} + \cdots + 2^{(d-1)(n-1)}\\ &= 2^{(d-1)(n-1)}(2^n - 1),\end{aligned}$$

where $2^{(d-1)i}2^{d(n-i-1)}$ is the number of $u = (0, u_1, \cdots, u_{n-1})$ which satisfies the condition that $u_1, \cdots, u_{i-1}$ are zero divisors and u_i is an invertible element in $\mathbb{Z}/(2^d)$, $1 \leqslant i \leqslant n-1$, we have

$$\sum_{i=1}^{w}\mathrm{N}(s_{(i)},0) = w(2^{n-1}-1) = 2^{(n-1)(d-1)}(2^{n-1}-1):$$

Lemma 3. *Let* $0 \leqslant k \leqslant d-1$. *Then the number of sequences in* $\Omega = \{s_{(1)}, \cdots, s_{(w)}\}$ *with* $\mathrm{N}(s_{(i)},0) \geqslant 2^k(2^{n-1}-1)$ *is* $2^{(d-1)(n-1)-k}$ *at most*.

Proof. Let S be the number of sequences $s_{(i)}$ in Ω with $\mathrm{N}(s_{(i)},0) \geqslant 2^k(2^{n-1}-1)$. Then

$$S2^k(2^{n-1}-1) \leqslant 2^{(d-1)(n-1)}(2^{n-1}-1).$$

So $S \leqslant 2^{(d-1)(n-1)-k}$.

Remark. By Lemma 3, the proportion of the number S to $|\Omega| = w$ is $1/2^k$ at most. So the proportion of sequences s with $\mathrm{N}(s,0) \geqslant 2^k(2^{n-1}-1)$ in $G'(f(x))_d$ is $1/2^k$ at most; that is, the proportion of sequences s with $\mathrm{N}(s,0) < 2^k(2^{n-1}-1)$ in $G'(f(x))_d$ is $(2^k-1)/2^k$, at least.

Theorem 2. *The condition is the same as that in Theorem* 1. *Then in* $G'(f(x))_e$ *the proportion of sequences with*

$$\frac{2^{d-k}-1}{2^{d-k}+1} < \frac{\mathrm{N}(a_{e-1},0)}{\mathrm{N}(a_{e-1},1)} < \frac{2^{d-k}+1}{2^{d-k}-1}$$

is $(2^k-1)/2^k$ *at least*.

Proof. Let $a \in G'(f(x))_e$, $s \equiv h(x)a \bmod 2^d$. If $\mathrm{N}(s,0) < 2^k(2^{n-1}-1)$, then by Theorem 1,

$$\frac{2^{d-1}(2^n-1) - 2^k(2^{n-1}-1)}{2^{d-1}(2^n-1) + 2^k(2^{n-1}-1)} \leqslant \frac{\mathrm{N}(a_{e-1},0)}{\mathrm{N}(a_{e-1},1)} \leqslant \frac{2^{d-1}(2^n-1) + 2^k(2^{n-1}-1)}{2^{d-1}(2^n-1) - 2^k(2^{n-1}-1)};$$

that is,

$$\frac{2^{d-k-1}(2^n-1)-(2^{n-1}-1)}{2^{d-k-1}(2^n-1)+(2^{n-1}-1)} \leqslant \frac{N(a_{e-1},0)}{N(a_{e-1},1)} \leqslant \frac{2^{d-k-1}(2^n-1)+(2^{n-1}-1)}{2^{d-k-1}(2^n-1)-(2^{n-1}-1)}.$$

So

$$\frac{2^{d-k}-1}{2^{d-k}+1} < \frac{N(a_{e-1},0)}{N(a_{e-1},1)} < \frac{2^{d-k}+1}{2^{d-k}-1}.$$

By Lemmas 2 and 3 and the above remark, the result is true.

Now we give examples for some e and examine the distribution of 0 and 1 in a_{e-1}.

(i) Set $e=32$, $d=16$, and take $k=8$. Then

$$\frac{2^{d-k}-1}{2^{d-k}+1} = \frac{2^8-1}{2^8+1} > 0.992\,2, \quad \frac{2^{d-k}+1}{2^{d-k}-1} = \frac{2^8+1}{2^8-1} < 1.007\,8,$$

$$\frac{2^k-1}{2^k} = \frac{2^8-1}{2^8} = 99.6\%.$$

So for any primitive polynomial of degree n over $\mathbb{Z}/(2^e)$, in $G'(f(x))_e$ the proportion of sequences with

$$0.992\,2 < \frac{N(a_{e-1},0)}{N(a_{e-1},1)} < 1.007\,8$$

is at least 99.6%.

(ii) Set $e=64$, $d=32$, and take $k=16$. Then

$$\frac{2^{d-k}-1}{2^{d-k}+1} = \frac{2^{16}-1}{2^{16}+1} > 0.999\,969\,48, \quad \frac{2^{d-k}+1}{2^{d-k}-1} = \frac{2^{16}+1}{2^{16}-1} < 1.000\,030\,52,$$

$$\frac{2^k-1}{2^k} = \frac{2^{16}-1}{2^{16}} > 99.998\,474\%.$$

So for any primitive polynomial of degree n over $\mathbb{Z}/(2^e)$, in $G'(f(x))_e$ the proportion of sequences with

$$0.999\,969\,48 < \frac{N(a_{e-1},0)}{N(a_{e-1},1)} < 1.000\,030\,52$$

is at least 99.998 474%.

So if e is sufficiently large and a is taken at random from $G'(f(x))_e$, then the distribution of 0 and 1 in a_{e-1} is very good.

References

1 Ward, M., The arithmetical theory of linear recurring sequences, *Trans. Amer. Math. Soc.*, 1933, 35(6):600.

2 Dai Zongduo, Huang Minqing, A criterion for primitiveness of polynomials over $\mathbb{Z}$ mod 2^d, *Chinese Science Bulletin*, 1990, 36(11):892.

3 Huang Minqiang, Dai Zongduo, Projective maps of linear recurring sequences with maximal p-adic periods, *Fibonacci Quart.*, 1992, 30(2):139.

4 Dai Zongduo, Beth, T., Gollman, D., Lower bounds for the linear complexity of sequences over residue rings, in *Advances in Cryptology, Eurocrypt' 90, Lncs*, Vol.473, Berlin: Springer-Verlag, 1991, 189—195.

5 Dai Zongduo, Binary sequences derived from ML-sequences over rings I: periods and minimal polynomials, *Journal of Cryptology*, 1990, 5(2), 193.