Linear Complexity of Sequences over Unknown Symbol Sets and Constructions of Sequences over $GF(p^k)$ whose Characteristic Polynomials are over $GF(p)$

ISIT
[redacted] 2002

Yun-Pyo HONG
Yu-chang EUN
Jeong-Heon Kim & Hong-Yeop Song
Department of Electrical and Electronics Engineering
Coding and Information Theory Lab
Yonsei University
Seoul, Korea

Content

- Introduction
  - point-to-point FH Communication (with interceptor)
  - large linear complexity ?
  - motivation
  - 2 problems

Part I

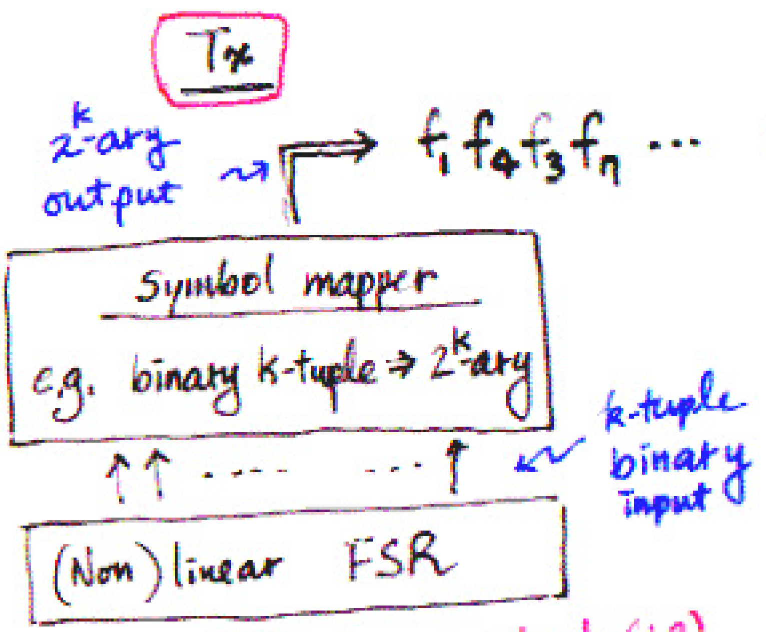
- Alternative definition of linear complexity

Part II

- Sufficient condition for a p^k-ary sequence over $GF(p^k)$ to have the minimal polynomial over $GF(p)$

- Concluding Remarks

Point-to-point FH Communication System (with interceptor)



Rx

- Interceptor tries to break the code using BM algorithm based on the observed copy of frequency slots.
- Need to choose an algebraic structure S' (e.g. finite field or integers mod n) with "addition" and "multiplication" as well as symbol correspondence between observed symbols and symbols of S'.

NEED: LARGE linear complexity (LC)

- of binary sequence ?
- of 2^k-ary sequence ?

FACT: "k-tuple reading" of a binary sequence (with $LC = \mathcal{L}$ over $GF(2)$) has in general $LC = \mathcal{L}' \le \mathcal{L}$, where $\mathcal{L}'$ over $GF(2^k)$.

What's even worse: $\mathcal{L}'$ cannot be uniquely determined (by BM) unless a basis of $GF(2^k)$ over $GF(2)$ is selected.

- What if the interceptor finds some $p > 2$ such that LC over $GF(p)$ is <u>much smaller</u> than the LC over $GF(2)$ that was originally designed?

- Do we need an alternative definition of "<u>linear complexity</u>" of a binary sequence?

- Sufficient condition that <u>LC</u> of k-tuple reading over $GF(2^k)$ does not decrease (is Not smaller than the <u>LC</u> of the original, designed, binary sequence over $GF(2)$)

Examples

◇ Example 1

- A sequence S with period 64:

$$0\,2\,2\,1\,2\,1\,2\,2\,0\,1\,1\,1\,1\,2\,2\,0\,0\,1\,2\,1\,1\,2\,0\,2\,2\,1\,1\,2\,0\,0\,0\,0$$
$$1\,0\,1\,0\,1\,1\,0\,1\,1\,1\,0\,2\,0\,0\,0\,1\,0\,0\,1\,0\,2\,2\,1\,2\,0\,0\,2\,2\,0\,0\,1\,2\ \ldots.$$

- The LC of S of Example 1 over various algebraic structures.

Over	$GF(3)$	$GF(4)$	$GF(5)$	Z_6	$GF(7)$	$\cdots$
LC	60	64	61	63	64	$\cdots$

◇ Example 2

- A sequence S with period 8: $0\ 1\ 3\ 7\ 6\ 5\ 2\ 4\ \ldots.$
- The distribution of the LC of S of Example 2.

LC over $GF(8)$	2	3	4	5	6	7	total
No. of sequences	0	0	0	2688	5376	32256	8!
LC over Z_8	2	3	4	5	6	7	total
No. of sequences	128	256	768	5888	14848	18432	8!

6

Arbitrary, not necessarily, by reading k-tuple from a binary sequence

◇ Example 3

- An 8-ary sequence S with period 63:

$$1\,3\,6\,4\,1\,4\,6\,6\,2\,0\,1\,1\,1\,3\,1\,3\,3\,6\,3\,4\,7\,4\,6\,1\,4\,6\,5\,4\,6\,7\,7$$
$$6\,3\,2\,5\,0\,3\,3\,3\,6\,7\,3\,2\,5\,1\,0\,5\,7\,5\,4\,3\,4\,6\,5\,5\,3\,3\,5\,1\,2\,4\,3\,6\,\ldots$$

0 ↔ 000
1 ↔ 001
2 ↔ 010
3 ↔ 011
4 ↔ 100
5 ↔ 101
6 ↔ 110
7 ↔ 111

⇔ symbols over GF(8) using $(\alpha^2, \alpha, 1)$ where $\alpha \sim x^3+x^2+1$ or $\alpha \sim x^3+x+1$

- Each s_n is represented as a binary 3-tuple as defined in *Eq. (1)*, and lifted up to $GF(8)$ using only the polynomial basis as given in *Eq. (2)* with two different primtive elements.

- It turned out that the LC with $x^3 + x^2 + 1$ is 59 and that with $x^3 + x + 1$ is 61.

◇ Example 4

For a sequence over two-symbol alphabet, the LC based on BM algorithm may be changed by ± 1 according to 2 different correspondences of the symbols with elements of $GF(2)$. Recall that the characteristic polynomial of the sequence would have (or not have) the factor $x + 1$ according to the interpretation of 0 and 1 as they are (or as switched, respectively).

Definition 5

The linear complexity (LC) of a sequence S (over an unknown symbol set) is the minimum LC over all possible algebraic sturctures and the symbol correspondences.

$$\underset{\sim}{S} \longrightarrow T(k, \underset{\sim}{S})$$

$$\{s_n\} \qquad\qquad \{t_n\}$$

$$t_n \triangleq (s_n, s_{n+1}, \cdots, s_{n+k-1})$$

p-ary sequence over GF(p) $\longrightarrow$ p-ary k-tuple sequence over GF(p)

$\Downarrow$

sequence over $GF(p^k)$

- need to fix a basis of $GF(p^k)$ over $GF(p)$
- Then, there is an LFSR over GF(p) that generates both $\underset{\sim}{S}$ and $T(k, \underset{\sim}{S})$

Prop. 6.

The (shortest) LFSR over GF(p) that generates $\underset{\sim}{S}$ will *also* generate $T(k, \underset{\sim}{S})$ for any k.

Question.

Is it also the *shortest* LFSR for $T(k, \underset{\sim}{S})$? (over $GF(p^k)$)

Example 8 *(a)* A binary sequence S_1 with period 63 is given by

$$110010000011111110101001001001101010111011011011101001111110010\ldots.$$

- LC of S_1 over $GF(2)$ is 62.
- LC of $T(3, S_1)$ over $GF(2^3)$ is 60 with respect to any polynomial basis as in *Eq. (2)*.

(b) A binary sequence S_2 with period 63 is given by

$$010111111100110000011011111101010011111100011001110100101001011\ldots.$$

- LC of $T(3, S_2)$ over $GF(2^3)$ is 55 with respect to the polynomial basis using $x^3 + x + 1$.
- LC of $T(3, S_2)$ over $GF(2^3)$ is 53 with respect to the polynomial basis using $x^3 + x^2 + 1$.

■

Proposition 9 The (minimum-degree) characteristic polynomial of a sequence over $GF(q)$ divides any connection polynomial of the LFSR that generates the sequence over $GF(q)$. Therefore, it is uniquely determined up to the multiplication by a constant.

A question at this point is the following: **is it possible that the shortest LFSR that generates S over $GF(p)$ is indeed the shortest LFSR that generates $T(k, S)$ over $GF(p^k)$ with respect to some basis of $GF(p^k)$ over $GF(p)$ for $k \geq 2$?** If it is possible to characterize such p-ary sequences S, then $T(k, S)$ over $GF(p^k)$ has the same characteristic polynomial as S and hence it is over $GF(p)$.

Theorem 10 (Main Theorem) :

- minimum-degree Characteristic polynomial $C(x)$ of $S = \{s_n\}$ over $GF(p)$ be given by $C(x) = \prod_{i \in I}(f_i(x))^{m_i}$,
- $f_i(x)$'s are some irreducible polynomials of degree d_i over $GF(p)$,
- m_i's are some positive integers, and
- I is an index set.
- Let $T(k, S)$ over $GF(p^k)$ be defined as in $t_n = (s_n, s_{n+1}, \ldots, s_{n+k-1})$ with respect to ANY but fixed basis for $k \geq 1$.
- k and d_i are relatively prime for all $i \in I$.

Then, we have the following:

- The shortest LFSR that generates S is also the shortest LFSR that generates $T(k, S)$ over $GF(p^k)$.
- Furthermore, it is also the shortest LFSR of $T(k, S)$ over $GF(p^m)$ for any $m \geq k$ such that m and d_i are relatively prime for all $i \in I$.

■

Corollary 12 (Main Corollary) . The linear complexity of $T(k,S)$ over $GF(p^k)$ as constructed in *Theorem 10* is **fixed regardless of the choice of basis** when symbols are represented as k-tuples over $GF(p)$. Furthermore, so is the LC of $T(k,S)$ over $GF(p^m)$ for $m \geq k$, if m and d_i are relatively prime for all $i \in I$.

i.e., lifting S over $GF(p)$ up to $GF(p^k)$ will NOT decrease Linear Complexity. (by reading successive k-tuples & by using any basis)

Corollary 13 For a p-ary m-sequence S of period $p^r - 1$ with p a prime, the shortest LFSR that generates S is also the shortest LFSR that generates $T(k,S)$ over $GF(p^k)$ as defined in *Eq. (3)* with respect to ANY basis if k is relatively prime to r. Furthermore, it is also the shortest LFSR of $T(k,S)$ over $GF(p^m)$ for any $m \geq k$ which is relatively prime to r.

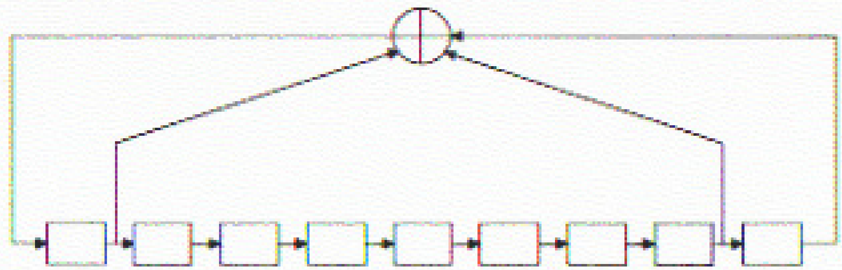
For binary sequences, besides the case of m-sequences, we would like to pick up one additional case to which Theorem 10 applies.

Corollary 14 If a binary sequence S has a period 2^r (for example, binary de Bruijn sequences), then the shortest LFSR that generates S is also the shortest LFSR that generates $T(k,S)$ over $GF(2^k)$ as defined in *Eq. (3)* for any positive integer k. Furthermore, it is also the shortest LFSR of $T(k,S)$ over $GF(2^m)$ for any $m \geq k$.

Example 15 A binary sequence S with period 16 is given by

$$0\ 0\ 0\ 0\ 1\ 0\ 1\ 1\ 1\ 1\ 1\ 1\ 0\ 1\ 0\ 0\ \ldots,$$

generated by the following LFSR over $GF(2)$:



An 8-ary sequence $T(3, S)$ with $k = 3$ over $GF(8)$ becomes

$$000\ \ 000\ \ 001\ \ 010\ \ 101\ \ 011\ \ 111\ \ 111\ \ 111\ \ 111\ \ 110\ \ 101\ \ 010\ \ \ldots.$$

An 8-ary sequence $T'(3, S)$ over $GF(16)$ becomes

$$0000\ \ 0000\ \ 0001\ \ 0010\ \ 0101\ \ 0011\ \ 0111\ \ 0111\ \ 0111\ \ 0111\ \ 0110\ \ 0101\ \ 0010\ \ \ldots.$$

Here, the symbol 0 is padded at the leftmost position of the every term of $T(3, S)$, and the resulting 4-tuples are regarded as the elements of $GF(16)$.
A 16-ary sequence $T(4, S)$ becomes

$$0000\ \ 0001\ \ 0010\ \ 0101\ \ 1011\ \ 0111\ \ 1111\ \ 1111\ \ 1111\ \ 1110\ \ 1101\ \ 1010\ \ 0100\ \ \ldots.$$

All these sequences have the same characteristic polynomial and the corresponding LFSR is shown above. ■

Example 16 A ternary sequence S in *Ex. 7* is indeed an m-sequence with the characteristic polynomial $x^3 + 2x + 1$ of degree 3. Therefore, the ternary 4-tuple sequence $T(4, S)$ in the example has the LFSR shown in *Fig. 1* as the shortest LFSR over $GF(3^4)$. *Theorem 10* implies that so does $T(k, S)$ over $GF(3^k)$ for any k not divisible by 3. ■

The converse of Theorem 1 is not generally true by the following example.

Example 11 A binary m-sequence S with characteristic polynomial $C(x) = x^4 + x + 1$, is given by

$$1\ 1\ 1\ 1\ 0\ 1\ 0\ 1\ 1\ 0\ 0\ 1\ 0\ 0\ 0\ \ldots.$$

The 4-ary sequence $T(2, S)$ has the same charateristic polynomial as S with respect to the polynomial basis in *Eq. (2)* even though the degree of $C(x)$ and $k = 2$ are not relatively prime. ■

Remark 17 Some interesting and related discussions are given by Gong and Xiao in 1994:

- They have given some algorigthm of constructing p^k-ary m-sequences using several p-ary m-sequences of the same period.
- We note that the resulting m-sequences over $GF(p^k)$ do not have the same characteristic polynomial as the component p-ary m-sequences.
- Their purpose is to construct m-sequences over $GF(p^k)$ very simply using m-sequences over $GF(p)$, and not on m-sequences with specified characteristic polynomial.
- In fact, they have considered the other case in which $C(x)$ factors over $GF(p^k)$ which is irreducible over $GF(p)$.
- That is, if the characteristic polynomial $C(x)$ of the component p-ary m-sequence over $GF(p)$ has degree kn, then the characteristic polynomial of resulting p^k-ary m-sequence over $GF(p^k)$ has degree n, and in fact, it must be a factor of $C(x)$ over $GF(p^k)$.

■

Now, let $U = \{u_n\}$, where $n = 1, 2, \ldots$, be a p-ary k-tuple sequence in general. In order to determine its characteristic polynomial of U over $GF(p^k)$, we need to fix one basis for BM algorithm. Following theorem charaterizes those U which do not need this.

Theorem 18 Let $U = \{u_n\}$, where $n = 1, 2, \ldots$, be a p-ary k-tuple sequence in general, where $u_n = (u_{n1}, u_{n2}, \ldots, u_{nk})$. Let a basis of $GF(p^k)$ over $GF(p)$ be fixed, and the characteristic polynomial $C(x)$ of U over $GF(p^k)$ using BM algorithm be determined to be of the form $\prod_{i \in I}(f_i(x))^{m_i}$, where $f_i(x)$ are irreducible polynomials of degree d_i over $GF(p)$, m_i are positive integers, and I is some index set. Then, $C(x)$ is a uniquely determined characteristic polynomial of U over $GF(p^k)$ regardless of the choice of basis, if k and d_i are relatively prime for all $i \in I$. Furthermore, $C(x)$ is the unique characteristic polynomial of $U(p, k)$ over $GF(p^m)$ for any $m \geq k$ using any basis such that m and d_i are relatively prime for all $i \in I$.

Concluding Remarks

An observed FH pattern by an interceptor must be a non-binary sequence over some unknown symbol set, and this causes a problem of determining the LC of the pattern since some specific operations of the LFSR must be provided. Therefore, it is reasonable that the interceptor will use such choice that leads to the least LC over all possiblities, and the system designer on the other hand must consider the LC of the FH pattern over various algebraic structures and the symbol correspondences including the true choice of the system.

In reality, however, we believe that a good choice would be the smallest size finite field of characteristic 2 that can just cover all the symbols of the sequence, because the computations over characteristic 2 are most efficiently implemented as hardware systems and the usual practice follows this idea.

We have tried several other options but failed to extract any further reasonable behavior of non-binary sequences over $GF(p^k)$ whose characteristic polynomial is uniquely determined regardless of the choice of basis other than those given in Theorem 10 of Section III. Theorem 18 is slightly more general in that the p-ary k-tuple sequences are not necessarily constructed as a k-tuple version of a p-ary sequence.

We note that Theorem 10 and its corollaries also apply equally well to $T(k,S)$ defined by

$$t_n = (s_{n+\sigma(1)}, s_{n+\sigma(2)}, \cdots, s_{n+\sigma(k)}), \tag{4}$$

where σ is any permutation on $\{1, 2, \ldots, k\}$. A further generalization is also possible by using any non-negative integers instead of $\sigma(i)$ for each i.