Linear Complexity of Finite Field Sequences over Different Fields

Andrew KLAPPER[†]

[†] Dept. of Computer Science, 779A Anderson Hall, University of Kentucky, Lexington, KY, 40506-0046, USA
E-mail: [†]klapper@cs.uky.edu

Abstract In this paper we study relationships between the linear complexities of a sequence when treated as a sequence over two distinct fields. We obtain bounds for one linear complexity in the form of a constant multiple of the other, where the constant depends only on the fields, not on the particular sequence.

key words: *Linear complexity, stream cipher, finite field, pseudorandom sequence.*

1. Introduction

Cryptographic stream ciphers use pseudorandom sequences to scramble messages. The security of a stream depends on the unpredictability of the sequence. In many cases there is a natural way in which we can think of the sequence entries as elements of an algebraic ring. For example, the entries in a binary sequence can be thought of as elements of the field with two elements $\mathbf{F}_2$. It is common that stream cipher designers demonstrate that their systems are secure against attacks that exploit such obvious algebraic interpretations. However they often ignore less natural algebraic interpretations that may lead to more successful attacks. Typically a binary sequence is shown to have high linear complexity over $\mathbf{F}_2$, which shows it is secure against the Berlekamp-Massey attack. However, it is possible that if we interpret the sequence as a sequence over some other finite field, even with a different characteristic, then it in fact has much lower linear complexity and can be attacked. For example, for the sequences called *geometric sequences* by Chan and Games [1], such a vulnerability was demonstrated by the author [2]. A geometric sequence is formed by applying a function f from a finite field $\mathbf{F}_q$ of odd characteristic p to $\mathbf{F}_2$. Chan and Games showed that geometric sequences have linear complexity close to their period. But we can also consider a binary sequence to be a sequence over $\mathbf{F}_p$ that just happens to have no entries other than 0 and 1. The author showed that these sequence in fact have small linear complexity over $\mathbf{F}_p$ (and in fact p can be found by the attacker with high probability).

Any binary sequence can be treated as a sequence over any ring and cryptanalyzed as such. Thus in principal if we are to believe in the security of a sequence we at least must believe it has high linear complexity with respect to every ring. In practice we need only be concerned with small rings, since the algebra in large enough rings would be complex enough to slow an attack. More generally, if S is a sequence over a ring R, we would like to understand the linear complexity of S when we treat it as a sequence over another ring via some mapping between the rings. In this paper we explore this question for certain rings.

$$\begin{tikzcd}
 & R \arrow[dl] \arrow[dr] & \\
R/I \arrow[rr, dashed] & & R/J
\end{tikzcd}$$

Fig. 1 Reinterpretation by lifting and reducing.

2. Basics

In this section we define terms and describe more precisely the problem we are concerned with. Throughout this section R denotes a commutative algebraic ring.

Definition 2.1: Let S be an infinite sequence with entries in R.

1. S is *linearly recurrent* if, for some integer t and coefficients $c_0, \cdots, c_{t-1} \in R$ we have
$$s_{n+t} = c_{t-1}s_{n+t-1} \cdots + c_0 s_n, \quad n = 0, \cdots.$$
Such an equation is called a *recurrence over R*. The integer t is the *length* of the recurrence.
2. The smallest length of a recurrence satisfied by S is called the *linear complexity of S over R*, denoted by $L_R(S)$.

See Lidl and Niederreiter's book on finite fields for background on linearly recurrent sequences [4]. See Lang's book on algebra for background on rings [3].

If I is an ideal in R, then we can reduce the elements of sequence S modulo I to obtain a sequence S_I over R/I. This reduced sequence is necessarily also linearly recurrent since every recurrence over R satisfied by S reduces to a recurrence over R/I satisfied by the reduced sequence. It follows that $L_{R/I}(S_I) \leq L_R(S)$. The purpose of this paper is to obtain lower bounds for $L_{R/I}(S_I)$ in terms of $L_R(S)$ for certain rings R. By transitivity we obtain lower bounds for $L_{R/I}(S_I)$ in terms of $L_{R/J}(S_J)$, where J is a second ideal in R.

We can also start with a sequence over R/I and attempt to reinterpret it as a sequence over R/J. The cryptanalyst hopes that this reinterpretation results in a much lower linear complexity, resulting in vulnerability of the stream cipher. If this is done by lifting the sequence to R, then reducing modulo J, our methods may give useful bounds on the extent to which the linear complexity decreases. See Figure 1.

The case when R is the ordinary integers was treated by

Shparlinski and Winterhof [5]. They proved the following theorem.

Theorem 2.2: Let $R = \mathbf{Z}$ be the integers. Let $m \geq 2$ and $h \geq 1$ be integers. Then for any sequence $S = s_0, s_1, \cdots$ of integers with $s_i \leq h$ for $i = 0, 1, \cdots$, we have

$$L_{\mathbf{Z}/(m)}(S_{(m)}) \geq \frac{1}{2\log_2(m)} \min\{L_{\mathbf{Z}}(S), m/h - 1\}.$$

Note that in many cases this is a very weak bound. For example, if a sequence is defined naturally as a sequence over $\mathbf{Z}/(m)$, then it is likely that $h = m - 1$. In this case the lower bound is less than 1, so the result says nothing.

In this paper we consider the case when R is an algebra over a finite field. We prove a similar bound, but one that is much sharper in the sense that the second term $m/h - 1$ is missing. The techniques we use are similar to Shparlinski and Winterhof's.

3. Main Theorem

Throughout this section let F be a finite field with q elements, and let R be a commutative F-algebra. If I is an ideal in R and $a, b \in R$, then we say a is congruent to b modulo I, or $a \equiv b \pmod{I}$, if $a - b \in I$. This is equivalent to saying that a and b have the same image in R/I. We refer to the cardinality of a basis for R/I over F as the *degree of* I, denoted $\deg(I)$.

Again let I be an ideal in R. Let B be a basis for R/I over F, and let B' be a subset of R that maps one to one and onto B under reduction by I. Let U be the F-linear span of B'. Then U is a complete set of representatives for R/I (that is, U that maps one to one and onto R/I under reduction by I). We say that U is a *linear set of representatives for* R/I.

Theorem 3.1: Let $I \subseteq R$ be an ideal with $\deg(I) = e < \infty$. Let $U \subseteq R$ be a linear set of representatives for R/I. Let $S = s_0, s_1, \cdots,$ be a sequence over R such that $s_i \in U$ for all i. Then

$$L_{R/I}(S_I) \geq \frac{L_R(S)}{e}.$$

Proof: Let

$$\begin{aligned} s_{n+t} \equiv\ & c_{t-1}s_{n+t-1} + \cdots + c_0 s_n \pmod{I},\\ & n = 0, 1, \cdots, \end{aligned}$$

be any recurrence of length $t = L_{R/I}(S_I)$ satisfied by S. We can iterate the recurrence arbitrarily many times. Thus for any integer $k \geq 0$ we have

$$s_{n+k} \equiv \sum_{j=0}^{t-1} c_{j,k} s_{n+j} \pmod{I}, \quad n = 0, 1, \cdots, \qquad (1)$$

for some $c_{j,k} \in R$, and we can take $c_{j,k} \in U$. Let $\mathbf{C}_k = (c_{0,k}, \cdots, c_{t-1,k}) \in R^t$.

Let $r = et + 1$ and consider the function Γ from F^r to R^t defined by

$$\Gamma(a_0, \cdots, a_{r-1}) = \sum_{k=0}^{r-1} a_k \mathbf{C}_k.$$

This is an $\mathbf{F}$-linear function from a vector space of dimension r to a vector space of dimension $et < r$, so it has a nontrivial kernel. That is, there is some $(a_0, \cdots, a_{r-1}) \neq (0, \cdots, 0)$ such that $\sum_{k=0}^{r-1} a_k \mathbf{C}_k = 0$. Let $u \leq r$ be the largest integer such that $a_u \neq 0$. Using equation (1) we find that

$$\begin{aligned} \sum_{k=0}^{u} a_k s_{n+k} &\equiv \sum_{k=0}^{u} a_k \sum_{j=0}^{r-1} c_{j,k} s_{n+j} \pmod{I}\\ &= \sum_{j=0}^{r-1} s_{n+j} \sum_{k=0}^{u} a_k c_{j,k}\\ &= 0, \end{aligned}$$

for every $n = 0, \cdots, N - u - 1$. That is,

$$\sum_{k=0}^{u} a_k s_{n+k} \in I.$$

However,

$$\sum_{k=0}^{u} a_k s_{n+k} \in U, \quad n = 0, 1, \cdots.$$

It follows that

$$\sum_{k=0}^{u} a_k s_{n+k} = 0, \quad n = 0, 1, \cdots.$$

Therefore

$$s_{n+u} = \sum_{k=0}^{u-1} (-a_u^{-1} a_k) s_{n+k} = 0, \quad n = 0, 1, \cdots.$$

Thus $L_R(S) \leq u \leq r - 1 = et$ so $L_R(S)/e \leq L_{R/I}(S_I)$, as claimed. □

We can interpret this theorem as saying that If we have a sequence T with entries in R/I, then if we lift the sequence to an arbitrary linear set of representatives for R/I in R, then the linear complexity of the lifted sequence will be no more than e times the linear complexity of T.

Suppose $U \subseteq V$ are a finite dimensional F-subspaces of R. Let I be an ideal such that U is a linear set of representatives modulo I, and let J be an ideal such that V is a linear set of representatives modulo J. If S is a sequence over R/I, then we can treat S as a sequence over R/J by lifting to $U \subseteq V$ and reducing modulo J. The following corollary is immediate.

Corollary 3.2: If I and J are as in the preceding paragraph, then for any sequence S over R/I we have

$$L_{R/I}(S_I) \geq \frac{L_{R/J}(S_J)}{\deg(I)}.$$

A simple general case occurs when $R = F[x]$, the polynomial ring in one variable. This ring is a principal ideal domain, so we can write $I = (f)$, where $\deg(I) = \deg(f)$ (in the usual sense). We can take for U the set of polynomials of degree less than e. For J we can take any other ideal $J = (g)$ where g has degree at least e. Theorem 3.1 then says the following.

Theorem 3.3: Let $f \in F[x]$ be a polynomial of degree $e > 0$. Let $S = s_0, s_1, \cdots,$ be a sequence over $F[x]$ such that $\deg(s_i) \leq e$ for all i. Then

$$L_{R/(f)}(S_{(f)}) \geq \frac{L_R(S)}{e}.$$

From this we obtain the specialization of Corollary 3.2.

Corollary 3.4: If f and g are polynomials in $F[x]$ and $\deg(g) \geq \deg(f)$, then for any sequence S of polynomials over F of degree less than $\deg(f)$ we have

$$L_{R/(f)}(S) \geq \frac{L_{R/(g)}(S)}{\deg(f)}.$$

The bound given by Theorem 3.1 is tight in the sense that there are sequences that meet this bound. For example, let $R = F[x]$ as above. The sequence with one period equal to

$$1, 1, x, 1, 0, x+1, x+1, 1, x+1, 0, x, x, x+1, x, 0$$

is an m-sequence over $\mathbf{F}_4 = \mathbf{F}_2[x]/(x^2+x+1)$, hence its linear complexity over $\mathbf{F}_4$ is 2. It is straightforward to see that its linear complexity over $\mathbf{F}_2[x]$ is 4. In fact the shortest recurrence it satisfies over $\mathbf{F}_2[x]$ is $s_{n+4} = s_{n+1} + s_n$.

More generally, let q be prime and f be an irreducible polynomial of degree e over F. Let S be an m-sequence with period $q^{ek} - 1$ with entries in $\mathbf{F}_{q^e} = F[x]/(f)$. Suppose we identify $\mathbf{F}_{q^e}$ with the set of polynomials of degree less than e over F. Using this identification, let $T : \mathbf{F}_{q^e} \to F$ be the function defined by evaluation at 0 (that is, $T(g)$ is the constant term of g). Then T is F-linear and it follows that the image of S under T is an m-sequence with entries in F and period $q^{ek} - 1$. Hence its linear complexity over F is ek.

Now interpret S as a sequence of elements of $R = F[x]$. Any linear recurrence with coefficients in R satisfied by S induces a linear recurrence of the same length satisfied by $T(S)$ since evaluation of polynomials at 0 is a ring homomorphism from R to F — it is just reduction modulo x. Hence $L_R(S) \geq ek$. That is, in this case

$$L_{R/(f)}(S_{(f)}) \leq \frac{L_R(S)}{e}.$$

It follows that we have equality.

One might wonder how common it is to find a reinterpretation of a sequence over one finite F-algebra as a sequence over another F-algebra as in Theorem 3.1. Let R_1 and R_2 be two finite F-algebras with $m = \dim(F_1) \leq \dim(F_2) = n$. Let $\tau : R_1 \to R_2$ be any injective F-linear function. Then any sequence S over R_1 can be interpreted as a sequence over R_2. We claim that this reinterpretation arises from the setting of Corollary 3.2, so that the given bound applies.

Let $\gamma_1, \cdots, \gamma_m$ be a basis for R_1 over F. Let $R = F[x_1, \cdots, n]$. Let $\delta_i = \tau(\gamma_i)$ for $1 \leq i \leq m$, and complete this to a basis $\delta_1, \cdots, \delta_n$ for R_2 over F. Then there are ring homomorphisms $\phi : R \to R_1$ and $\psi : R \to R_2$ defined by $\phi(x_i) = \gamma_i$ if $1 \leq i \leq m$, $\phi(x_i) = 0$ if $m < i \leq n$, and $\psi(x_i) = \delta_i$ if $1 \leq i \leq n$. The ideals $I = \ker(\phi)$ and $J = \ker(\psi)$ with the linear sets of representatives U, the F-linear span of $x_1, \cdots, x_m$, and V, the F-linear span of $x_1, \cdots, x_n$ satisfy the conditions of Corollary 3.2 as desired.

The situation when $F = \mathbf{F}_2$ is especially interesting. Let T be any finitely generated $\mathbf{F}_2$-algebra and $\psi : \mathbf{F}_2 \to T$ be any function (not necessarily linear). If S is a sequence over $\mathbf{F}_2$, we want to compare the linear complexity of S and $\psi(S)$. If we replace $\psi(x)$ by $\psi(x) - \psi(0)$, then we obtain a function that is necessarily $\mathbf{F}_2$-linear. The effect on the linear complexity of $\psi(S)$ is to increase or decrease it by at most 1. This is because addition of c amounts to adding $c/(1-x)$ to the generating function, so the degree of the denominator in the rational representation of the generating function changes by at most 1. Now we want to connect $\mathbf{F}_2$ and T by an $\mathbf{F}_2$-algebra R. Suppose that $T = \mathbf{F}_2[a_1, \cdots, a_n]/K$ for some ideal K. Let $R = \mathbf{F}_2[x_0, x_1, \cdots, x_n]$, $I = (x_0 - 1, x_1, \cdots, x_n)$, and $J = (x_0 - \psi(1)) + K$. Then $\mathbf{F}_2 = R/I$, $T = R/J$, and $U = \{1\}$ is a linear set of representatives for R/I. Let V be any linear set of representatives for R/J containing 1. It follows that the hypotheses of Theorem 3.1 hold with $e = 1$. Therefore, if $\psi(0) = 0$

$$L_T(\psi(S)) \leq L_{\mathbf{F}_2}(S).$$

For arbitrary ψ we obtain

$$L_T(\psi(S)) \leq L_{\mathbf{F}_2}(S) + 1.$$

4. Conclusion

We have given explicit bounds relating the linear complexities of a sequence over one finite dimensional F-algebra and its image in another finite dimensional F-algebra via an F-linear transformation τ. We showed that these linear complexities are within a constant factor (depending only on the rings, not the sequence) of each other.

We leave as an open problem the question of what happens when τ is non-linear.

References

[1] A. H. Chan and R. Games, "On the Linear Span of Binary Sequences from Finite Geometries, q Odd," in Advances in Crypology – Proceedings of Crypto 1986, Springer-Verlag Lecture Notes in Computer Science, pp. 405-417, 1986.

[2] A. Klapper, "The Vulnerability of Geometric Sequences Based on Fields of Odd Characteristic," *Journal of Cryptology* **7** (1994), 33-52.

[3] S. Lang, "Algebra, 2nd ed.," Addison-Wesley, Reading, MA, 1984.

[4] Lidl, R., and Niederreiter, H., "Finite Fields," in Encyclopedia of Mathematics Vol. 20, Cambridge University Press, Cambridge, 1983.

[5] I. Shparlinski and A. Winterhof, "On the Linear Complexity of Bounded Integer Sequences over Different Moduli," *Information Processing Letters* **96** (2005), 175-177.