Randomness of Elliptic Curve Sequences

Charles C.Y. Lam
Department of Combinatorics & Optimization
University of Waterloo
Waterloo, Ontario N2L 3G1, CANADA
cy3lam@cacr.uwaterloo.ca
Guang Gong
Department of Electrical and Computer Engineering
University of Waterloo
Waterloo, Ontario N2L 3G1, CANADA
ggong@cacr.uwaterloo.ca

June 20, 2002

Abstract

In this paper, we introduce linear feedback shift register sequences (LFSR) over the group of the elliptic curve points, and a way of generating binary sequences from these LFSR sequences. The former is called *LFSR-EC sequences*. Properties on representation, period, and linear span of these two types of sequences are discussed. A special case of this generation is then discussed in detail.
Index terms: Elliptic Curve, Sequences

1 Introduction and Preliminaries

Since Koblitz and Miller [6, 10] proposed to use elliptic curves over $GF(q)$ to construct public-key cryptosytems, elliptic curve public-key cryptosystems over $GF(2^l)$ or $GF(q)$ have become widely used in applications such as smart cards which provide limited space for implementation of modular computations. Currently, the operations of points on elliptic curves over $GF(2^l)$ or $GF(q)$ have a matured technology on both hardware and software implementations.

Linear recusive sequences (or equivalently, linear feedback shift register sequences) over $GF(p)$ have a significant influence in the development of communications and cryptography. In this work, we combine elliptic curves and linear feedback shift registers to generate pseudorandom sequences which provide strong potential applications in stream ciphers and pseudorandom number generators. In constrained systems or devices, for example, wireless devices which employ an elliptic public cryptosystem, then such a construction automatically provide a pseudorandom sequence (or number) generator, which can be used for generating random numbers in the elliptic curve digital signature algorithm and session key in their encryption phases.

Elliptic curve sequences with strong cryptographic properties have been studied by in the literature. Kaliski [7] used the randomness criteria based on the computational complexity of the discrete log problem over the elliptic curves to construct a class of elliptic curve pseudorandom sequences. Gong, Berson and Stinson [3] discussed the one-stage LFSR over supersingular elliptic curves over finite fields. Mahassni and Shparlinski [5] obtained a general bound for the $0-1$ distribution of the one-stage LFSR on elliptic curves over finite fields.

In this paper, we intend to generalise the construction introduced in [3].

The paper is organised as follows. In section 2, we give a generalised construction of LFSR-EC sequences, and discuss a few properties relating to the construction. We then specialise the generating polynomial to a primitive polynomial, and proceed to generate the binary sequences from the elliptic curve points. A special case of the generating polynomial is explained in section 3.

We conclude this section with the definition of elliptic curves over finite fields.

Elliptic Curves over Finite Fields

Let $\mathbb{F} = GF(q)$ where q is a power of a prime. If the characteristic of the field is 2, a elliptic curve E over $\mathbb{F}$ can be written into one of the following standard forms [9]

$$y^2 + c_3 y = x^3 + c_4 x + c_6, c_i \in \mathbb{F}$$

if E is supersingular, or

$$y^2 + xy = x^3 + c_2 x^2 + c_6, c_i \in \mathbb{F}$$

if E is non-supersingular. If the characteristic is greater than 3, E can be written in the following form

$$y^2 = x^3 + ax + b,\ a, b \in \mathbb{F}.$$

The points $P = (x, y)$, $x, y \in \mathbb{F}$ that satisfy the equation, together with a "point at infinity" denoted by $\mathcal{O}$, form an abelian group $(E, +, \mathcal{O})$ with identity element $\mathcal{O}$. (For the case where the characteristic is 3, the reader is referred to [9].)

2 Structure and randomness of elliptic curve sequences

Let $P = (x_P, y_P)$ be a point on a elliptic curve E with order r. The sequence of points produced is $\{P, 2P, 3P, \ldots\}$. Let $P_i = iP = (x_i, y_i)$. In this section we present the sequences through mappings specified by [3] on the different types of elliptic curves, and attempt to analyse several statistical properties relating to the sequences.

2.1 Structure

2.1.1 Sequences from Elliptic Curves over GF(p), $p \neq 2, 3$

Let p be a prime. The sequence is defined as

$$a_i = \begin{cases} 1, & y_i \geq \frac{p+1}{2} \\ 0, & \text{otherwise} \end{cases}$$

for $i = 1, \ldots, r-1$, and $a_0 = a_r = 1$.

The sequence $\{a_i\}$ can be represented as follows.

Lemma 1 *When $r \equiv 0 \pmod 2$, we can write the sequence as*

$$\underline{a} = (1, A, 0, \overleftarrow{A} + \underline{1}),$$

and when $r \equiv 1 \pmod 2$, we can write the sequence as

$$\underline{a} = (1, A, \overleftarrow{A} + \underline{1}),$$

where $A = (a_1, a_2, \ldots, a_m)$ for some $m \in \mathbb{Z}$, $\overleftarrow{A} = (a_m, a_{m-1}, \ldots, a_1)$, $\underline{1} = (11\ldots1)$, and $\overleftarrow{A} + \underline{1} = (a_m + 1, a_{m-1} + 1, \ldots, a_1 + 1)$.

Proof: When $r \equiv 0 \pmod 2$, we have $r = 2m$ for some $m \in \mathbb{Z}$. Hence we can write

$$\underline{a} = (a_0, a_1, \ldots, a_{m-1}, a_m, a_{m+1}, a_{m+2}, \ldots, a_{2m-1}).$$

Note that $(m+1)P$ is a point of order 2, hence $(m+1)P = (x_{m+1}, 0)$. Also, for $i = 1, \ldots, m$, $(m+1+i)P = (x_{m+1+i}, y_{m+1+i}) = (x_{m+1+i}, -y_{m+1-i})$. Hence $a_{m+1+i} = a_{m+1-i} + 1 \pmod 2$.

Therefore let $A = (a_1, a_2, \ldots, a_m)$. We can write $\underline{a} = (1, A, 0, \overleftarrow{A} + \underline{1})$.

When $r \equiv 1 \pmod 2$, we have $r = 2m+1$ for some $m \in \mathbb{Z}$. Hence we can write $\underline{a} = (1, A, \overleftarrow{A} + \underline{1})$, where we do not have a point of order 2. ■

2.1.2 Sequences from Elliptic Curves over GF(2^n)

Supersingular Curves over GF(2^n)

Let

$$E : y^2 + c_3 y = x^3 + c_4 x + c_6, \tag{1}$$

be a *supersingular* elliptic curve over GF(2^n). It is well known that the order of a supersingular elliptic curve over GF(2^n) must have order $\#E = 2^n + 1 - t$ where $t^2 = 0, 2^n, 2^{n+1}, 3 \cdot 2^n$ or 2^{n+2}.

The sequence is defined as

$$a_i = \mathrm{Tr}(y_i), i = 1, \ldots, r-1,$$

with $a_0 = a_r = 1$.

Note that the order of the point P is always odd as the curve has odd order.

The sequence above can be represented as follows.

Lemma 2 *The order r of the sequence satisfies $r \equiv 1 \pmod 2$. When* $\mathrm{Tr}(c_3) = 1$*. The sequence can be represented as*

$$\underline{a} = (1, A, \overleftarrow{A} + \underline{1}),$$

where $A = (a_1, a_2, \ldots, a_m)$. When $\mathrm{Tr}(c_3) = 0$*. The sequence can be written as*

$$\underline{a} = (1, A, \overleftarrow{A}),$$

Proof: For $i = 1, \ldots, m$, $(m+i)P = (x_{m+i}, y_{m+i}) = (x_{m+1-i}, y_{m+1-i} + c_3)$. Hence $a_{m+i} = a_{m+1-i} + \mathrm{Tr}(c_3) \pmod 2$. Hence result. ■

Non-Supersingular Curves over GF(2^n)

Similarly, let

$$E : y^2 + xy = x^3 + c_2 x^2 + c_6 \tag{2}$$

be a *non-supersingular* elliptic curve over $\mathrm{GF}(2^n)$. Note that the order of this curve is always even, and can be shown [14, Section A] that the order is divisible by 4 if and only if $\mathrm{Tr}(c_2) = 0$. Again, the sequence is defined as

$$a_i = \mathrm{Tr}(y_i), i = 1, \ldots, r-1,$$

with $a_0 = a_r = 1$.

Lemma 3 *The sequence above can appear in the following forms of representation:*

1. *When* $\mathrm{Tr}(c_2) = 0$ *and* $Tr(x_P) = 0$, *if* $r \equiv 0 \pmod 2$ *we have*

$$\underline{a} = (1, A, \mathrm{Tr}(c_6), \overleftarrow{A}),$$

and if $r \equiv 1 \pmod 2$ *we have*

$$\underline{a} = (1, A, \overleftarrow{A}).$$

2. *When* $\mathrm{Tr}(c_2) = 0$, $\mathrm{Tr}(x_P) = 1$, *then* $r \equiv 0, 2 \pmod 4$. *When* $r \equiv 0 \pmod 4$,

$$\underline{a} = (1, A, \mathrm{Tr}(c_6), \overleftarrow{A} + (1010 \cdots 101)),$$

and when $r \equiv 2 \pmod 4$,

$$\underline{a} = (1, A, \mathrm{Tr}(c_6), \overleftarrow{A} + (010 \cdots 101)).$$

3. *When* $\mathrm{Tr}(c_2) = 1$ *and* $Tr(x_P) = 0$, *we have*

$$\underline{a} = (1, A, \overleftarrow{A})$$

when $r \equiv 1 \pmod 2$ *and*

$$\underline{a} = (1, A, \mathrm{Tr}(c_6), \overleftarrow{A})$$

when $r \equiv 0 \pmod 2$.

4. *When* $\mathrm{Tr}(c_2) = 1$ *and* $r \equiv 1 \pmod 2$ *we have*

$$\underline{a} = (1, A, \overleftarrow{A} + \underline{1}).$$

If $r \equiv 0 \pmod 2$, *if* $\mathrm{Tr}(x_P) = 1$ *then*

$$\underline{a} = (1, A, \mathrm{Tr}(c_6), \overleftarrow{A} + \underline{1}),$$

and if $\mathrm{Tr}(x_P) = 0$, *then*

$$\underline{a} = (1, A, \mathrm{Tr}(c_6), \overleftarrow{A} + (010\ldots1010))$$

if $r \equiv 0 \pmod 4$ *and*

$$\underline{a} = (1, A, \mathrm{Tr}(c_6), \overleftarrow{A} + (10\ldots1010))$$

if $r \equiv 2 \pmod 4$.

To prove the lemma above, we need the following [14, Proposition 91]:

Proposition 1 *Let* (x_2, y_2) *be a point on curve (2). Then*

$$(x_2, y_2) = 2(x_1, y_1)$$

for some (x_1, y_1) *on (2) if and only if* $\mathrm{Tr}(x_2) = \mathrm{Tr}(c_2)$.

Now we proceed to prove lemma 3:

Proof: First note that for any point $(x, y) \in E$, $-(x, y) = (x, y+x)$. Let $iP = (x_i, y_i)$. Hence for any of the aforementioned sequences, for i from 1 to $r-1$, we have $a_i = \mathrm{Tr}(y_i)$ and $a_{r-i} = \mathrm{Tr}(y_{r-i}) = \mathrm{Tr}(y_i + x_i) = a_i + \mathrm{Tr}(x_i)$.

When $\mathrm{Tr}(x_P) = \mathrm{Tr}(x_P)$ then all points $P_i = iP$ are points that are double of another point, hence $\mathrm{Tr}(x_i) = \mathrm{Tr}(x_P)$ for all $1 \le i \le r-1$. Otherwise, the points are alternatively being a double of another point, hence $\mathrm{Tr}(x_i) = i + \mathrm{Tr}(x_P) \bmod 2$.

When $\mathrm{Tr}(c_2) \neq \mathrm{Tr}(x_P)$, the order r of P cannot be odd as otherwise P is the double of $\frac{r+1}{2}P$, leading to a contradiction.

When $r \equiv 0 \pmod 2$, there is a point of order 2. This point is characterised by $(x, y) = -(x, y) = (x, y+x)$. Hence $x = 0$. Solving for y gives $y = \sqrt{c_6}$.

The results follow from the observations. ■

2.2 Randomness of the Elliptic Curve Sequences

2.2.1 0-1 Distribution

The $0-1$ distribution of some of the sequences mentioned in the above lemmas can be described as follows.

Theorem 1 *Given a binary sequence $A = (a_1, a_2, \ldots, a_m)$ where m is odd. When the sequence $\underline{a}$ is of the form $(1, A, \overleftarrow{A} + \underline{1})$, there are $m+1$ 1's and m 0's in $\underline{a}$. If the sequence is of the form $(1, A, 0, \overleftarrow{A} + \underline{1})$, there are $m+1$ 0's and 1's in $\underline{a}$.*

Proof: Let $w(\underline{a})$ be the weight of $\underline{a}$. When $\underline{a} = (1, A, \overleftarrow{A} + \underline{1})$, we have

$$\begin{aligned} w(\underline{a}) &= w(1) + w(A) + w(\overleftarrow{A} + \underline{1}) \\ &= 1 + w(A) + m - w(A) = 1 + m. \end{aligned}$$

The proof for the case $\underline{a} = (1, A, 0, \overleftarrow{A} + \underline{1})$ is similar. ■

2.2.2 Linear Span

Our major result is described as follows.

Theorem 2 *For the sequences of the forms*

$$\underline{a} = \begin{cases} (1, A, \overleftarrow{A} + \underline{1}) \\ (1, A, 0, \overleftarrow{A} + \underline{1}) \\ (1, A, \mathrm{Tr}(c_6), \overleftarrow{A} + \underline{1}) \\ (1, A, \mathrm{Tr}(c_6), \overleftarrow{A} + (1010 \cdots 101)) \\ (1, A, \mathrm{Tr}(c_6), \overleftarrow{A} + (010 \cdots 101)) \end{cases},$$

the linear span of $\underline{a}$, $\mathrm{LS}(\underline{a})$, is bounded by

$$\mathrm{LS}(\underline{a}) \geq \lceil \frac{r}{4} \rceil.$$

The following result is well-known.

Proposition 2 *Let $f(x)$ be the minimal polynomial of $S = (s_0, s_1, \ldots, s_{m-1})$ over $\mathrm{GF}(2)$. The minimal polynomial $g(x)$ of $\overleftarrow{S} = (s_{m-1}, s_{m-2}, \ldots, s_0)$ is the reciprocal polynomial of S. That is,*

$$g(x) = x^n f(\frac{1}{x}).$$

Proof: [of theorem 2] We will show the case for the sequence $\underline{a} = (1, A, 0, \overleftarrow{A} + \underline{1})$, where $A = (a_1, a_2, \ldots . a_m)$. We have $L(\underline{a}) = (A, 0, \overleftarrow{A} + \underline{1}, 1)$, where L is the left-shift operator. Note that $\mathrm{LS}(\underline{a}) = \mathrm{LS}(L(\underline{a}))$.

Let $f(x)$ be the minimal polynomial of $\underline{a}$. If $\deg(f) \geq \frac{r}{4}$, then we are done.

Suppose $\deg(f) \leq \frac{r}{4} - 1$. By the Berlekamp-Massey algorithm [1, 8] running on shifts of the sequence, we must have $f(x)$ being the minimal polynomial for $(A, A, \ldots)$ as well as $(\overleftarrow{A} + \underline{1}, \overleftarrow{A} + \underline{1}, \ldots)$.

If $(x+1)$ is a factor of $f(x)$, then there exists $\underline{a}_i$ such that $\underline{a}_i = (11 \cdots 1)$. However, we then have

$$\overleftarrow{A} + \underline{1} = \overleftarrow{\underline{a}}_1 + \overleftarrow{\underline{a}}_2 + \cdots + \overleftarrow{\underline{a}}_{i-1} + \overleftarrow{\underline{a}}_{i+1} + \cdots + \overleftarrow{\underline{a}}_s,$$

which leads to a contradiction, as the minimal polynomial for $\overleftarrow{A} + \underline{1}$ is now of degree one less.

Otherwise, if $(x+1)$ is not a factor of $f(x)$, then

$$\overleftarrow{A} + \underline{1} = \overleftarrow{\underline{a}}_1 + \overleftarrow{\underline{a}}_2 + \cdots + \overleftarrow{\underline{a}}_s + \underline{1},$$

which means that the minimal polynomial for $\overleftarrow{A} + \underline{1}$ is of one degree higher, and hence a contradiction.

The case for $\underline{a} = (1, A, \overleftarrow{A} + \underline{1})$ is similar.

For the case $\underline{a} = (1, A, \mathrm{Tr}(c_6), \overleftarrow{A} + (1010 \cdots 101))$, the minimal polynomial for $(1010 \cdots 101)$ is $(x^m + 1)$, which contains the factor $(x+1)$. Hence the proof applies as above. The other cases are similar. ■

Remark: The result above can be applied to any sequences of the types prescribed.

Remark: Some cases of the sequences discussed in this paper but do not satisfy 2. These include

$$\underline{a} = \begin{cases} (1, A, \overleftarrow{A}) \\ (1, A, \mathrm{Tr}(c_6), \overleftarrow{A}) \end{cases}.$$

One main concern is that we may have $A = \overleftarrow{A}$, which then implies that the linear span of the sequence depends on A. These cases are currently open.

2.2.3 Period

The results of the linear span of the sequences constructed induce the following observations to the period of the sequences.

Corollary 1 *The period of the sequences can only be one of the following:*

1. *For*

$$\underline{a} = \begin{cases} (1, A, \overleftarrow{A} + \underline{1}) \\ (1, A, 0, \overleftarrow{A} + \underline{1}) \end{cases}$$

the period is either $\frac{r}{3}$ *or* r.

2. *For*

$$\underline{a} = \begin{cases} (1, A, \mathrm{Tr}(c_6), \overleftarrow{A} + \underline{1}) \\ (1, A, \mathrm{Tr}(c_6), \overleftarrow{A} + (1010 \cdots 101)) \\ (1, A, \mathrm{Tr}(c_6), \overleftarrow{A} + (010 \cdots 101)) \end{cases}$$

the period can be $\frac{r}{4}$, $\frac{r}{3}$, $\frac{r}{2}$ *or* r.

Proof: The period of a sequence is at least the degree of the linear span of the sequence. We also know that the period of the sequence must divide the length r, by way of generation.

For the case where the sequence can be represented by $(1, A, \overleftarrow{A} + \underline{1})$, note that the possible values of the period of the sequence can only be $\frac{r}{3}$ or r. As the length of the sequence is odd.

The case $(1, A, 0, \overleftarrow{A} + \underline{1})$ is similar, note that $a_0 \neq a_{\frac{r}{2}}$ rules out the possibility that the period would be divisible by $\frac{r}{2}$ and hence the only possible values for the period are $\frac{r}{3}$ and r.

Now for the only case left in lemma 3 where we have $(1, A, \mathrm{Tr}(c_6), \overleftarrow{A} + (1010 \cdots 101))$. Note that if $\mathrm{Tr}(c_6) \neq 1$ we can conclude as before that the only possible values for the period are $\frac{r}{3}$ and r. However, when $\mathrm{Tr}(c_6) = 1$, the same argument cannot apply and we can only conclude that the possible values for the period to be $\frac{r}{4}$, $\frac{r}{3}$, $\frac{r}{2}$ or r. ■

Remark: The results also apply to the construction described in [3], where the curves used are supersingular curves. Our sequence is taken as an interleaved sequence from the original construction by taking only the trace of the y-coordinates instead of the trace of both the $x-$ and $y-$coordinates. Thus we have shown that the linear span and period of the sequences generated in [3] are at least 1/8 of the length of the sequence, which is also the order of the generating point P.

2.3 The behaviour of Subsequence A

In cryptographic applications, seldom do we apply the whole sequence. Hence it is desirable to exploit the behaviour of the subsequences. While there are no theoretical results on the subsequences, we have some experimental results on the subsequence A.

2.4 Application

In most elliptic curve cryptosystems we have a generating point Q of order n which is of public knowledge. For each communication session, we randomly select a secret key k with $P = kQ$ as our seed for the elliptic curve pseudorandom sequence generator. As the order of Q is always known, we can determine the order of P easily. In fact, if $\gcd(k, n) = 1$, then $\mathrm{ord}(P) = n$. If this point P satisfy the our conditions, we can use P to generate the elliptic curve sequence as demonstrated.

3 Conclusion

We have demonstrated a construction of pseudorandom sequences based on elliptic curves. A special case is discussed, in which the linear span and period of the sequence is least $1/4$ of the order of the generating point P. For applications in practice, curves are of order 2^{160} or higher, thus producing sequences of high linear span and period. Relating open problems include the sequences not covered by theorem 2.

References

[1] E. R. Berlekamp, *Algebraic coding theory*, McGraw-Hill, New York, 1968.

[2] S.W. Golomb, *Shift Register Sequences,* Holden-Day, Inc., 1967. Revised edition, Aegean Park Press, 1982.

[3] G. Gong, T. Berson and D. Stinson, Elliptic Curve Pseudorandom Sequence Generator, *Proceedings of the Sixth Annual Workshop on Selected Areas in Cryptography*, LNCS 1758, Springer-Verlag, 34-48, 1999.

[4] G. Gong, C. Lam, Linear Recursive Sequences over Elliptic Curves, *Proceedings of Sequences and their Applications - SETA '01*, DMTCS series, Springer-Verlag, 182-196, 2001.

[5] E.E. Mahassni and I. Shparlinski, On the Uniformity of Distribution of Congruential Generators over Elliptic Curves, *preprint*, November 2000.

[6] N. Koblitz, Elliptic Curve Cryptosystems, *Mathematics of Computation (48)*, 203-209, 1987.

[7] B. Kaliski, *A Pseudo-Random Bit Generator Based on Elliptic Logarithms*, Advances in Cryptology–Crypto '86, Lecture Notes in Computer Science, LNCS 263(1986), Springer-Verlag, 84-103.

[8] J. L. Massey, *Shift-register synthesis and BCH decoding*, IEEE Trans. Information Theory, January 1969.

[9] A. J. Menezes, *Elliptic Curve Public Key Cryptosystems*, Kluwer Academic Publishers, 1993.

[10] V. Miller, Uses of Elliptic Curves in Cryptography, *Advances in Cryptology – CRYPTO '85*, LNCS 218, Springer-Verlag, 417-426, 1986.

[11] I. Samaev, Evaluation of discrete logarithms in a group of p-torsion points of an elliptic curve in characteristic p, *Mathematics of Computation*, 67 (1998), 353-356.

[12] T. Satoh and K. Araki, Fermat quotients and the polynomial time discrete log algorithm for anomalous elliptic curves, *Commentarii Mathematici Universitatis Sancti Paul*, 47 (1998), 81-92.

[13] N. Smart, The discrete logarithm problem on elliptic curves of trace one, *Journal of Cryptology*, 12 (1999), 193-196.

[14] J. Solinas, Improved Algorithms for Arithmetic on Anomalous Binary Curves, *CACR Technical Reports*, CORR 99-46, University of Waterloo, 1999.