Inferring sequences produced by a linear congruential generator on elliptic curves missing high–order bits

Jaime Gutierrez, Álvar Ibeas

Faculty of Sciences, University of Cantabria, Santander E–39071, Spain

Abstract. Let p be a prime and let $E(\mathbb{F}_p)$ be an elliptic curve defined over the finite field \mathbb{F}_p of p elements. For a given point $G \in E(\mathbb{F}_p)$ the linear congruential genarator on elliptic curves (EC-LCG) is a sequence (U_n) of pseudorandom numbers defined by the relation

 $U_n = U_{n-1} \oplus G = nG \oplus U_0, \quad n = 1, 2, \dots,$

where \oplus denote the group operation in $E(\mathbb{F}_p)$ and $U_0 \in E(\mathbb{F}_p)$ is the initial value or seed. We show that if G and sufficiently many of the most significants bits of two consecutive values U_n, U_{n+1} of the EC-LCG are given, one can recover the seed U_0 (even in the case where the elliptic curve is private) provided that the former value U_n does not lie in a certain small subset of exceptional values. We also estimate limits of a heuristic approach for the case where G is also unknown. This suggests that for cryptographic applications EC-LCG should be used with great care. Our results are somewhat similar to those known for the linear and non-linear pseudorandom number congruential generator.

Keywords: Pseudorandom congruential generators, Lattice reduction, Elliptic curves.

1 Introduction

For a prime p, denote by \mathbb{F}_p the field of p elements and always assume that it is represented by the set $\{0, 1, \ldots, p-1\}$. Accordingly, sometimes, where obvious, we treat elements of \mathbb{F}_p as integer numbers in the above range.

Let E be an elliptic curve defined over \mathbb{F}_p given by an *affine Weierstrass* equation, which for gcd(p, 6) = 1 takes form

$$Y^2 = X^3 + aX + b, (1)$$

for some $a, b \in \mathbb{F}_p$ with $4a^3 + 27b^2 \neq 0$.

We recall that the set $E(\mathbb{F}_p)$ of \mathbb{F}_p -rational points forms an abelian group, with the *point at infinity* \mathcal{O} as the neutral element of this group (which does not have affine coordinates).

For a given point $G \in E(\mathbb{F}_p)$ the Linear Congruential Generator on Elliptic Curves, EC-LCG is a sequence U_n of pseudorandom numbers defined by the relation

$$U_n = U_{n-1} \oplus G = nG \oplus U_0, \quad n = 1, 2, \dots,$$
 (2)

where \oplus denote the group operation in $E(\mathbb{F}_p)$ and $U_0 \in E(\mathbb{F}_p)$ is the *initial* value or seed. We refer to G as the *composer* of the EC-LCG.

It is clear that the period of the sequence (2) is equal to the order of G. The EC-LCG provides a very attractive alternative to linear and non-linear congruential generators with many applications to cryptography and it has been extensively studied in the literature, see [4, 17, 22, 23, 25, 26, 42, 43]. A very recent survey of related problems is the paper [44].

In the cryptographic setting, the initial value $U_0 = (x_0, y_0)$ and the constants G, a, and b are assumed to be the secret key, and we want to use the output of the generator as a stream cipher. Of course, if two consecutive values U_n are revealed, it is almost always easy to find U_0 and G. So, we output only the most significant bits of each U_n in the hope that this makes the resulting output sequence difficult to predict. The main result of this paper is that not too many bits can be output at each stage: the Linear Congruential Generator on Elliptic Curves is unfortunately polynomial time predictable if sufficiently many bits of its consecutive elements are revealed. We rigorously demonstrate our approach in the special case when the composer G is public. We show that if G and sufficiently many of the most significant bits of two consecutive values U_n, U_{n+1} of the EC-LCG are given, one can recover the seed U_0 (even in the case where the elliptic curve is private) provided that the first coordinate x_0 of the former value $U_n = (x_n, y_n)$ does not lie in a certain small set. Of course, the assumption that G is public reduces the relevance of the problem to cryptography, but we believe that the strength of the result we obtain makes this situation of interest in its own right. We also believe that this approach can be extended to the case where G is secret and we present a heuristic approach for this case. Concretely, we show that if sufficiently many of the most significants bits of three consecutive values U_n, U_{n+1}, U_{n+2} of the EC-LCG are given, one can recover the seed U_0 and the composer G provided that the first value U_n for which an approximation is used does not lie in a certain small set of exceptional values.

This suggests that for cryptographic applications EC-LCG should be used with great care.

Assume that the sequence (U_n) is not known, but for some n, approximations W_j of two consecutive values U_{n+j} , j = 0, 1 are given. We show that the value U_n can be recovered from this information if the approximations W_j are sufficiently good.

For the *linear congruential generator* similar problems have been introduced by Knuth [32] and then considered in [10, 11, 18, 29, 33]; see also the surveys [12,

34]. The quadratic congruential generator and the inverse congruential generator have been studied in [5–7, 19, 21]. Several problems of pseudorandom number generators appear in [39, 40].

On the other hand, our results are substantially weaker than those known for the linear and nonlinear congruential generators. In some sense, the problem we solve can be considered as a special case of the problem of finding small solutions of multivariate polynomial congruences. For polynomial congruences in one variable such an algorithm has been given by Coppersmith [14], see also [15, 16, 20, 27, 30]. However in the general case only heuristic results are known. Here, due to the special structure of the polynomials involved, we are able to obtain rigorous results.

Throughout the paper the term polynomial time means polynomial in log p. Our results involve another parameter Δ which measures how well the values W_j approximate the terms U_{n+j} . This parameter is assumed to vary independently of p subject to satisfying the inequality $\Delta < p$ (and is not involved in the complexity estimates of our algorithms).

More precisely, we say that $W = (x_W, y_W) \in \mathbb{F}_p^2$ is a Δ -approximation to $U = (x_U, y_U) \in \mathbb{F}_p^2$ if there exists integers e, f satisfying:

$$|e|, |f| \le \Delta, \ x_W + e = x_U, \ y_W + f = y_U.$$

In all of our results, the case where Δ grows like a fixed power p^{δ} where $0 < \delta < 1$ corresponds to the situation where a positive proportion δ of the least significant bits of terms of the output sequence remain hidden.

The remainder of the paper is structured as follows: we start with a short outline of some basic facts about lattices and the abelian group associated to an elliptic curve in Section 2. In Section 3 we formulate our main result and outline the plan of the proof Subsection 3.1, which is given in Subsection 3.2. Section 4 is dedicated to study the case when G is also private. Then, in Section 5 we discus the results of numerical tests of our approaches. Finally, we conclude with Section 6 which makes some final comments and poses open questions.

Acknowledgment.

This work is partially supported by the Spanish Ministry of Science grant MTM2004-07086.

2 Background

2.1 Integer Lattices

Here we collect several well-known facts about lattices which form the background to our algorithms.

We review several results and definitions of concepts related to lattices which can be found in [24]. For more details and more recent references, we also recommend consulting [1, 29, 37].

Let $\{\boldsymbol{b}_1,\ldots,\boldsymbol{b}_s\}$ be a set of linearly independent vectors in \mathbb{R}^r . The set

$$\mathcal{L} = \{c_1 \boldsymbol{b}_1 + \dots + c_s \boldsymbol{b}_s : c_1, \dots, c_s \in \mathbb{Z}\}$$

is called (s-dimensional) *lattice* with *basis* $\{b_1, \ldots, b_s\}$. If s = r, the lattice \mathcal{L} is of *full rank*.

To each lattice \mathcal{L} one can naturally associate its *volume*:

$$\operatorname{vol}(\mathcal{L}) = \left(\det\left(\langle \boldsymbol{b}_i, \boldsymbol{b}_j \rangle\right)_{i,j=1}^s\right)^{1/2}$$

where $\langle \boldsymbol{a}, \boldsymbol{b} \rangle$ denotes the inner product. This definition does not depend on the choice of the basis $\{\boldsymbol{b}_1, \ldots, \boldsymbol{b}_s\}$.

For a vector \boldsymbol{u} , let $\|\boldsymbol{u}\|$ denote its *Euclidean norm*. The first Minkowski theorem, see Theorem 5.3.6 in [24], gives the upper bound

$$\min\{\|\boldsymbol{z}\|: \boldsymbol{z} \in \mathcal{L} \setminus \{\boldsymbol{0}\}\} \le s^{1/2} \operatorname{vol}(\mathcal{L})^{1/s}$$
(3)

on the shortest nonzero vector in any s-dimensional lattice \mathcal{L} in terms of its volume.

The Minkowski bound (3) motivates a natural question, the Shortest Vector Problem (SVP): how to find a shortest nonzero vector in a lattice. Unfortunately, there are several indications that this problem is **NP**-hard when the dimension grows. This study has suggested several definitions of a reduced basis $\{b_1, \ldots, b_s\}$ for a lattice, trying to obtain a shortest vector by the first basis element b_1 . The celebrated *LLL algorithm* of Lenstra, Lenstra and Lovász [36] provides a concept of reduced basis and an approximate solution, enough in many practice applications.

Another related question is the *Closest Vector Problem (CVP)*: given a lattice $\mathcal{L} \subseteq \mathbb{R}^r$ and a shift vector $\mathbf{t} \in \mathbb{R}^r$, the goal consists on finding a vector in the set $\mathbf{t} + \mathcal{L}$ with minimum norm. This problem is usually expressed in an equivalent way: finding a vector in \mathcal{L} closest to the target vector $-\mathbf{t}$. It is well known that CVP is **NP**-hard when the dimension grows.

However, both computational problems SVP and CVP are known to be solvable in deterministic polynomial time provided that the dimension of \mathcal{L} is fixed (see [31], for example). The lattices in this paper are of fixed (and low) dimension.

In fact, lattices in this paper consist of integer solutions $\boldsymbol{x} = (x_0, \ldots, x_{s-1}) \in \mathbb{Z}^s$ of a system of congruences

$$\sum_{i=0}^{s-1} a_{ij} x_i \equiv 0 \mod q_j, \qquad j = 1, \dots, m,$$

modulo some positive integers q_1, \ldots, q_m . Typically (although not always) the volume of such a lattice is the product $Q = q_1 \cdots q_m$. Moreover, all the aforementioned algorithms, when applied to such a lattice, become polynomial in log Q. If $\{b_1, \ldots, b_s\}$ is a basis of the above lattice, by the Hadamard inequality we have:

$$\prod_{i=1}^{s} \|\boldsymbol{b}_i\| \ge \operatorname{vol}(\mathcal{L}).$$
(4)

2.2 The Group Associated to an Elliptic Curve

In this subsection we recall some basic facts about the group law on elliptic curves.

Let E be an elliptic curve defined over \mathbb{F}_p given by the affine Weierstrass equation (1).

The operation \oplus acts over the points $P = (x_P, y_P)$ and $Q = (x_Q, y_Q)$ of $E(\mathbb{F}_p)$ with $P, Q \neq \mathcal{O}$ as follows:

$$P \oplus Q = R = (x_R, y_R)$$

- If $x_P \neq x_Q$, then

$$x_R = m^2 - x_P - x_Q, \qquad y_R = m(x_P - x_R) - y_P, \quad \text{where} \quad m = \frac{y_Q - y_P}{x_Q - x_P}.$$
(5)

- If $x_P = x_Q$ but $y_P \neq y_Q$, then $P \oplus Q = \mathcal{O}$. - If P = Q and $y_P \neq 0$, then

$$x_R = m^2 - 2x_P, \qquad y_R = m(x_P - x_R) - y_P, \quad \text{where} \quad m = \frac{3x_P^2 + a}{2y_P}.$$
 (6)
- If $P = Q$ and $y_P = 0$, then $P \oplus Q = \mathcal{O}.$

Our context is a pseudorandom number generator which outputs affine points in an elliptic curve. One obtains recursively them by operating a fixed composer G to the previous value. So, almost always, the above equations in the first case (5) will determine the process.

The set $E(\mathbb{F}_p)$ of \mathbb{F}_p -rational points forms an abelian group satisfying the Hasse-Weil inequality:

$$|\# (E(\mathbf{F}_p) - p - 1)| \le 2\sqrt{p}.$$
 (7)

It is well known that the group $E(\mathbb{F}_p)$ is of the form

$$E(\mathbf{I} \mathbf{F}_p) \cong \mathbb{Z} / L\mathbb{Z} \times \mathbb{Z} / M\mathbb{Z},$$

where the integers L and M are uniquely determined with M divides L, see [2, 8, 47] for these and other general properties of elliptic curves.

3 Predicting Result for Known Composer

In this section we formulate and prove our main result on predicting the linear pseudorandom number generator on elliptic curve, when the composer G and the modulus p are both public.

3.1 Formulation and Plan of Proof

Assume that a, b are unknown, but the prime p and $G = (x_G, y_G) \in E(\mathbb{F}_p)$ are given to us. We show that when we are given Δ -approximations W_n, W_{n+1} to (respectively) two consecutive affine values U_n, U_{n+1} produced by the EC-LCG; we can recover the exact values, provided that the first component x_n of $U_n = (x_n, y_n)$ does not lie in a certain set, whose size is bounded by $O(\Delta^6)$. Note that once two affine points in a curve as (1) are given, such that their first component is different, the curve (the parameters a and b) are determined. Then, after discovering the values U_n and U_{n+1} , we can reproduce (backwards and forwards) the whole sequence. To simplify the notation, we assume that n = 0 from now on.

We write $W_j = (\alpha_j, \beta_j), U_j = (x_j, y_j)$, for j = 0, 1; and so there exist integers e_j, f_j with:

$$x_j = \alpha_j + e_j, \quad y_j = \beta_j + f_j$$

$$|e_j|, |f_j| \le \Delta, \quad j = 0, 1.$$
(8)

Theorem. With the above notations and definitions, there exists a set $\mathcal{U}(\Delta; a, x_G, y_G) \subseteq$ \mathbb{F}_p of cardinality $\#\mathcal{U}(\Delta; a, x_G, y_G) = O(\Delta^6)$ with the following property: whenever $x_0 \notin \mathcal{U}(\Delta; a, x_G, y_G)$ then, given Δ -approximations W_0 , W_1 to two consecutive affine values U_0, U_1 produced by linear congruential generator on elliptic curves (2), and given the value of $G = (x_G, y_G)$, one can recover the seed U_0 in deterministic polynomial time.

An outline of the algorithm given in the proof of this Theorem goes as follows. The algorithm is divided into two stages.

- Stage 1: We include the value x_G in $\mathcal{U}(\Delta; a, x_G, y_G)$. We construct a certain lattice \mathcal{L} (see (14) below) of dimension 7; this lattice depends on the approximations W_0 , W_1 and the composer G. We also show that a certain vector \boldsymbol{E} directly related to missing information about U_0 and U_1 is a very short vector. A closest vector \boldsymbol{F} is found; see [31] for the appropriate algorithm.
- Stage 2: We show that F provides the required information about E for all initial values $U_0 = (x_0, y_0)$ except when x_0 lies in a certain exceptional set $\mathcal{U}(\Delta; a, x_G, y_G) \subseteq \mathbb{F}_p$ of cardinality $\#\mathcal{U}(\Delta; a, x_G, y_G) = O(\Delta^6)$ (which is defined as set of zeroes of a certain parametric family of polynomials).

3.2 Proof

We assume that $x_0 \in \mathbb{F}_p$ is chosen so as not to lie in a certain subset $\mathcal{U}(\Delta; a, x_G, y_G)$ of \mathbb{F}_p . The cardinality of this set is bounded by $O(\Delta^6)$. It consists of the solutions of a certain polynomial together with two extra values. It is explained through the proof.

We place the value $x_G \in \mathcal{U}(\Delta; a, x_G, y_G)$, so that U_0 is not G or -G. Then, clearing denominators in equations (5), we can translate

$$U_1 = U_0 \oplus G \tag{9}$$

into the following identities in the field ${\rm I\!F}_p$:

$$L_1 = L_1(x_0, y_0, x_1) \equiv 0 \mod p, \qquad L_2 = L_2(x_0, y_0, x_1, y_1) \equiv 0 \mod p,$$

where

$$L_{1} = x_{G}^{3} + x_{1}x_{G}^{2} - x_{0}x_{G}^{2} - 2x_{1}x_{G}x_{0} - x_{G}x_{0}^{2} + x_{0}^{3} + 2y_{G}y_{0} + x_{1}x_{0}^{2} - y_{2}^{G} - y_{0}^{2},$$

$$L_{2} = y_{1}x_{G} - y_{1}x_{0} - y_{G}x_{0} + y_{G}x_{1} - y_{0}x_{1} + y_{0}x_{G}.$$
(10)

Using the equalities $x_j = \alpha_j + e_j$ and $y_j = \beta_j + f_j$ for j = 0, 1, equations (10) become

$$- (-2x_{G}\alpha_{0} - 2x_{G}\alpha_{1} + 3\alpha_{0}^{2} + 2\alpha_{1}\alpha_{0} - x_{G}^{2})e_{0} + (\alpha_{0}^{2} - 2x_{G}\alpha_{0} + x_{G}^{2})e_{1} + (2y_{G} - 2\beta_{0})f_{0} + (3\alpha_{0} + \alpha_{1} - x_{G})e_{0}^{2} + (2\alpha_{0} - 2x_{G})e_{0}e_{1} + [e_{0}^{3} + e_{0}^{2}e_{1} - f_{0}^{2}] = x_{G}^{2}\alpha_{0} - x_{G}^{2}\alpha_{1} + x_{G}\alpha_{0}^{2} - \alpha_{1}\alpha_{0}^{2} + 2x_{G}\alpha_{0}\alpha_{1} - \alpha_{0}^{3} - x_{G}^{3} + \beta_{0}^{2} + y_{G}^{2} - 2y_{G}\beta_{0},$$
$$- (-\beta_{1} - y_{G})e_{0} + (y_{G} - \beta_{0})e_{1} + (x_{G} - \alpha_{1})f_{0} + (x_{G} - \alpha_{0})f_{1} - [e_{0}f_{1} + e_{1}f_{0}] = \beta_{1}\alpha_{0} - x_{G}\beta_{1} + y_{G}\alpha_{0} - y_{G}\alpha_{1} + \beta_{0}\alpha_{1} - x_{G}\beta_{0}.$$

Now, we linearize this polynomial system. Writing

$$\begin{aligned} A_{0} &\equiv x_{G}^{2} \alpha_{0} - x_{G}^{2} \alpha_{1} + x_{G} \alpha_{0}^{2} - \alpha_{1} \alpha_{0}^{2} + 2x_{G} \alpha_{0} \alpha_{1} - \alpha_{0}^{3} - x_{G}^{3} + \beta_{0}^{2} + y_{G}^{2} - 2y_{G} \beta_{0} \mod p \\ A_{1} &\equiv -2 x_{G} \alpha_{1} - 2 x_{G} \alpha_{0} + 3 \alpha_{0}^{2} + 2 \alpha_{1} \alpha_{0} - x_{G}^{2} \mod p, \quad A_{2} &\equiv \alpha_{0}^{2} + x_{G}^{2} - 2 x_{G} \alpha_{0} \mod p, \\ A_{3} &\equiv 2 y_{G} - 2 \beta_{0} \mod p, \quad A_{4} \equiv 0 \mod p, \quad A_{5} \equiv \alpha_{1} + 3 \alpha_{0} - x_{G} \mod p \\ A_{6} &\equiv -2 x_{G} + 2 \alpha_{0} \mod p, \quad A_{7} \equiv 0 \mod p, \quad A_{8} \equiv 1 \mod p \\ B_{0} &\equiv \beta_{1} \alpha_{0} - x_{G} \beta_{1} + y_{G} \alpha_{0} - y_{G} \alpha_{1} + \beta_{0} \alpha_{1} - x_{G} \beta_{0} \mod p, \quad B_{1} \equiv -\beta_{1} - y_{G} \mod p, \\ B_{2} &\equiv y_{G} - \beta_{0} \mod p, \quad B_{3} \equiv x_{G} - \alpha_{1} \mod p, \quad B_{4} \equiv x_{G} - \alpha_{0} \mod p, \\ B_{5} &\equiv 0 \mod p, \quad B_{6} \equiv 0 \mod p, \quad B_{7} \equiv -1 \mod p, \quad B_{8} \equiv 0 \mod p, \end{aligned}$$

$$(11)$$

we obtain that vector

$$E = (\Delta^2 e_0, \Delta^2 e_1, \Delta^2 f_0, \Delta^2 f_1, \Delta e_0^2, \Delta e_0 e_1, \Delta (e_1 f_0 + e_0 f_1), e_0^3 + e_0^2 e_1 - f_0^2) = (\Delta^2 E_1, \Delta^2 E_2, \Delta^2 E_3, \Delta^2 E_4, \Delta E_5, \Delta E_6, \Delta E_7, E_8)$$

is a solution to the following linear system of congruences:

$$\sum_{i=1}^{4} A_i X_i + \sum_{i=5}^{7} \Delta A_i X_i + \Delta^2 A_8 X_8 \equiv \Delta^2 A_0 \mod p,$$

$$\sum_{i=1}^{4} B_i X_i + \sum_{i=5}^{7} \Delta B_i X_i + \Delta^2 B_8 X_8 \equiv \Delta^2 B_0 \mod p,$$

$$X_1 \equiv X_2 \equiv X_3 \equiv X_4 \equiv 0 \mod \Delta^2,$$

$$X_5 \equiv X_6 \equiv X_7 \equiv 0 \mod \Delta.$$
(12)

Moreover, \boldsymbol{E} is a relatively short vector. We have:

 $|E_i| \leq \Delta, i = 1, 2, 3, 4, |E_i| \leq \Delta^2, i = 5, 6, |E_7| \leq 2\Delta^2, |E_8| \leq 3\Delta^3; ||E|| \leq \sqrt{19}\Delta^3.$ (13) Let \mathcal{L} be the lattice consisting of integer solutions $\mathbf{X} = (X_1, X_2, \dots, X_8) \in \mathbb{Z}^8$ of the system of congruences:

$$\sum_{i=1}^{4} A_i X_i + \sum_{i=5}^{7} \Delta A_i X_i + \Delta^2 A_8 X_8 \equiv 0 \mod p,$$

$$\sum_{i=1}^{4} B_i X_i + \sum_{i=5}^{7} \Delta B_i X_i + \Delta^2 B_8 X_8 \equiv 0 \mod p,$$

$$X_1 \equiv X_2 \equiv X_3 \equiv X_4 \equiv 0 \mod \Delta^2,$$

$$X_5 \equiv X_6 \equiv X_7 \equiv 0 \mod \Delta.$$
(14)

We compute a solution T of the system of congruences (12), using linear diophantine equations methods. Applying an algorithm solving the CVP for the shift vector T and the lattice \mathcal{L} , we obtain a vector

$$\boldsymbol{F} = (\Delta^2 F_1, \Delta^2 F_2, \Delta^2 F_3, \Delta^2 F_4, \Delta F_5, \Delta F_6, \Delta F_7, F_8)$$

satisfying equations (12) and

$$|F_i| \le \sqrt{19\Delta}, i = 1, 2, 3, 4, \quad |F_i| \le \sqrt{19\Delta^2}, i = 5, 6, 7, \quad |F_8| \le \sqrt{19\Delta^3}$$
$$\|F\| \le \sqrt{19\Delta^3}.$$
(15)

Note that we can compute F in polynomial time from the information we are given. We might hope that E and F are the same, or at least, that we can recover the approximations errors from F. If not, we will show that x_0 belongs to a subset $\mathcal{U}(\Delta; a, x_G, y_G) \subseteq \mathbb{F}_p$ of cardinality $\#\mathcal{U}(\Delta; a, x_G, y_G) = O(\Delta^6)$. Let us bound the "bad" possibilities for which this process does not succeed. Vector D = E - F lies in \mathcal{L} :

$$D = (\Delta^2 D_1, \Delta^2 D_2, \Delta^2 D_3, \Delta^2 D_4, \Delta D_5, \Delta D_6, \Delta D_7, D_8), \quad D_i = E_i - F_i, i = 1, \dots, 8$$

Bounds (13) and (15) imply $||D|| \le 2\sqrt{19}\Delta^3$ and

$$|D_i| \le 2\sqrt{19}\Delta, i = 1, 2, 3, 4, \quad |D_i| \le 2\sqrt{19}\Delta^2, i = 5, 6, 7, \quad |D_8| \le 2\sqrt{19}\Delta^3.$$
(16)

If $D_1 \equiv 0 \mod p$ and $D_3 \equiv 0 \mod p$, then $U_0 = (x_0, y_0) = (\alpha_0 + F_1, \beta_0 + F_3) \in \mathbb{F}_p^2$ and we can recover the original values U_0 and U_1 .

By the same argument, if $D_2 \equiv 0 \mod p$ and $D_4 \equiv 0 \mod p$, we have $U_1 = (x_1, y_1) = (\alpha_1 + F_2, \beta_1 + F_4)$. In order to recover $U_0 = U_1 \oplus (-G)$, we need $U_1 \neq -G$. So, let us include the first component of -2G, namely $\left(\frac{3x_G^2 + a}{2y_G}\right)^2 - 2x_G$ (see equation (6)), in the set $\mathcal{U}(\Delta; a, x_G, y_G)$.

So, we can assume $(D_1 \neq 0 \text{ or } D_3 \neq 0)$, and $(D_2 \neq 0 \text{ or } D_4 \neq 0)$. Substituting D in Equations (14) defining lattice \mathcal{L} we obtain:

$$\sum_{i=1}^{8} A_i D_i \equiv 0 \mod p, \quad \sum_{i=1}^{8} B_i D_i \equiv 0 \mod p.$$
 (17)

Using the definition of $A_i, B_i, i = 1, ..., 9$ and after the substitutions $\alpha_i = x_i - e_i$ and $\beta_i = y_i - f_i, i = 0, 1$ in the second congruence of (17), we find

$$N(x_0, y_0, x_1, y_1) = N_0 - D_4 x_0 - D_2 y_0 - D_3 x_1 - D_1 y_1 \equiv 0 \mod p,$$
(18)

where

$$N_0 = D_7 + D_4 e_0 + D_2 f_0 + D_2 y_G + D_1 f_1 - D_1 y_G + D_4 x_G + D_3 e_1 + D_3 x_G.$$

We claim that

$$F(x_0) \equiv 0 \bmod p \tag{19}$$

for some nonconstant polynomial of degree at most 18 of the form:

$$F(X) = \sum_{i=0}^{18} C_i X^i,$$

where the coefficients $C_i \in \mathbb{F}_p[N_0, D_1, D_2, D_3, D_4]$, i = 0, ..., 18. Then, for every choice of $D_1, D_2, D_3, D_4, D_7, e_0, e_1, f_0$, and f_1 only a constant number of values x_0 are possible.

In order to proof this last claim, we distinguish two cases: $D_1 \not\equiv 0 \mod p$ and $D_1 \equiv 0 \mod p$.

Case: $D_1 \not\equiv 0 \mod p$. From (18) we obtain that

$$y_1 = -\frac{-N_0 + D_2 y_0 + D_3 x_1 + D_4 x_0}{D_1}.$$

Substituting this expression in the following equation:

$$E_1(x_1, y_1) = y_1^2 - x_1^3 - ax_1 - b$$

and clearing denominators, we obtain a polynomial $E'_1(x_0, y_0, x_1)$ in the variables x_0, y_0 and y_1 :

$$E_1'(x_0, y_0, x_1) = E_1\left(-\frac{-N_0 + D_2 y_0 + D_3 x_1 + D_4 x_0}{D_1}, y_1\right) D_1^2.$$

Solving x_1 from equation (10), substituting it in $E'_1(x_0, y_0, x_1)$ and clearing denominators (we note that $x_0 = x_G$ belongs to the bad set $\mathcal{U}(\Delta; a, x_G, y_G)$, we obtain a polynomial $A(x_0, y_0)$ of degree 6 with respect the variable y_0 :

$$A(x_0, y_0) = E'_1(x_0, y_0, \left(\frac{y_G - y_0}{x_G - x_0}\right)^2 - x_0 - x_G)(x_G - x_0)^6 = -D_1^2 y_0^6 + \cdots$$

Let $F(x_0)$ be the resultant of $A(x_0, y_0)$ and the polynomial

$$E_0(x_0, y_0) = y_0^2 - x_0^3 - ax_0 - b$$

with respect to the variable y_0 :

$$F(x_0) = \operatorname{resultant}_{y_0}(A(x_0, y_0), E_0(x_0, y_0)) = \sum_{i=0}^{18} C_i x_0^i.$$

Using MAPLE we have computed the coefficients C_i explicitly, and we present some of these expressions below:

$$C_{18} = D_2^4,$$

$$C_{17} = -2 D_2^2 \left(6 x_G D_2^2 + D_4^2 \right),$$

$$C_{16} = 2 D_2^2 \left(33 D_2^2 x_G^2 + D_2^2 a - 4 D_3 y_G D_2 + 2 D_4 N_0 + 12 x_G D_4^2 - 2 x_G D_3 D_4 \right) + D_4^4.$$
(20)

Now, we need to prove that $F(x_0)$ is a nonconstant polynomial for every choice of D_1, D_2, D_3, D_4 and N_0 . Clearly, if $D_2 \not\equiv 0 \mod p$, then degree of $F(x_0)$ is 18. Otherwise, we obtain from bounds in (16) and equation (20) that

$$C_{18} = 0, C_{17} = 0, C_{16} = D_4^4.$$

Since $D_2 = 0$, then $D_4 \neq 0$ and the degree of $F(x_0)$ is 16.

Case: $D_1 \equiv 0 \mod p$. ¿From (18) we obtain that

$$x_1 = -\frac{-N_0' + D_2 y_0 + D_4 x_0}{D_3},$$

where $N'_0 = D_7 + D_4 e_0 + D_2 f_0 + D_2 y_G + D_4 x_G + D_3 e_1 + D_3 x_G$. Substituting this expression in Equation (10): $L_1(x_0, y_0, x_1)$, we derive a polynomial $B(x_0, y_0)$ of degree 2 with respect the variable y_0 :

$$B(x_0, y_0) = L_1\left(x_0, y_0, -\frac{-N_0' + D_2 y_0 + D_4 x_0}{D_3}\right) D_3 = -D_3 y_0^2 + \cdots$$

Let $F(x_0)$ be the resultant of $B(x_0, y_0)$ and $E_0(x_0, y_0) = y_0^2 - x_0^3 - ax_0 - b$ with respect the variable y_0 :

$$F(x_0) = \text{resultant}_{y_0}(B(x_0, y_0), E_0(x_0, y_0)) = -D_2^2 x_0^7 + (4 x_G D_2^2 + D_4^2) x_0^6 + \cdots$$

Again, we need to prove that $F(x_0)$ is a non constant polynomial for every choice of D_1, D_2, D_3, D_4 and N_0 . Firstly, we note that the resultant specialize well, because the leadings coefficients of $B(x_0, y_0)$ and $E_0(x_0, y_0)$ with respect y_0 are non zero. Secondly, if $D_2 \neq 0 \mod p$, then degree of $F(x_0)$ is 7. Otherwise, we have that $F(x_0)$ is a polynomial of degree 6 because its leading coefficient is $D_4^2 \neq 0$.

Since F is a non-constant polynomial in x_0 of degree at most 18, the congruence (19) can be satisfied for at most 18 values of x_0 once D_i , $i = 1, \ldots, 4$, and N_0 have been chosen. By (16) the total number of possible choices for D_1, D_2, D_3, D_4 is $O(\Delta^4)$. On the other hand, N_0 can take $O(\Delta^2)$ distinct values, because writing N_0 as:

$$N_0 = D_7 + D_4 e_0 + D_2 f_0 + D_1 f_1 + D_3 e_1 + (D_2 - D_1) y_G + (D_3 + D_4) x_G.$$

From bounds in (8) and (16) we obtain that $D_7 + D_4 e_0 + D_2 f_0 + D_1 f_1 + D_3 e_1 = O(\Delta^2)$. And fixed D_1, D_2, D_3 and D_4 then is fixed $D_2 - D_1$ and $D_3 + D_4$. Hence there are only $O(\Delta^6)$ values of x_0 that satisfy some congruence (19). We place these $O(\Delta^6)$ values of x_0 in $\mathcal{U}(\Delta; a, x_G, y_G)$. So all short vectors satisfying (12) lead to discover the approximation errors whenever $x_0 \notin \mathcal{U}(\Delta; a, x_G, y_G)$. Finally, if that is not the case, we can trivially calculated e_0, e_1 and then U_n for $n = 0, 1, \ldots$, which finishes the proof.

4 Unknown Composer

In the previous section we have provided an upper bound (namely, 1/6) for the fraction of bits one should hide from each value obtained with EC-LCG in order to avoid lattice attacks which could reproduce the sequence. However, it has been assumed that the cryptanalyst has access to the composer G, which places his task in a quite optimistic frame. So, in this section we suppose that the parameter G is also private. In this case we require three approximations, instead of two.

We assume that the sequence (U_n) is not known, but for some n, approximations W_j of 3 consecutive values U_{n+j} , j = 0, 1, 2 are given. We show that the value $U_n = (x_n, y_n)$ can be recovered from this information if the approximations W_j are sufficiently good. We can suppose that n=0.

We write $W_j = (\alpha_j, \beta_j)$ where $e_j = x_j - \alpha_j$, $f_j = y_j - \beta_j$ for j = 0, 1, 2 verifying

$$|e_j|, |f_j| \le \Delta, \quad j = 0, 1, 2 \tag{21}$$

So, our input of this new algorithm consists of $\alpha_0, \alpha_1, \alpha_2, \beta_0, \beta_1, \beta_2 \in \mathbb{F}_p$ and the positive integer Δ .

The first attempt to design a such procedure would be, as in the previous section, to suppose that $U_0, U_1 \notin \{G, -G\}$ and use the addition formulae (5) to

derive a closest vector problem instance whose solution may lead to recover the three values, and the secret parameter G.

In that case, the polynomial equations obtained grow significatly in degree and number of monomials involved; so in order to keep dealing with lowdimensional lattices, we have followed a different approach.

We just consider the information given as approximations to arbitrary points in the same elliptic curve, in such a way that we are not taking advantage from the knowledge of the procedure which has generated them. In other words, we give a method to recover three points lying in an elliptic curve in the form (1), given corresponding approximations. And we use that method in the frame of an EC-LCG and three values partially revealed.

The starting point is:

$$\begin{split} y_0^2 &= x_0^3 + a x_0 + b, \\ y_1^2 &= x_1^3 + a x_1 + b, \\ y_2^2 &= x_2^3 + a x_2 + b. \end{split}$$

Eliminating the curve parameters a, b and assuming that $U_0 \notin \{U_1, -U_1\}$ (that is, $x_0 \neq x_1$), we obtain the following equation:

$$-y_{2}^{2}x_{1}+y_{2}^{2}x_{0}+x_{2}^{3}x_{1}-x_{2}^{3}x_{0}-x_{2}y_{0}^{2}+x_{2}x_{0}^{3}+x_{2}y_{1}^{2}-x_{2}x_{1}^{3}-y_{1}^{2}x_{0}+x_{1}^{3}x_{0}+x_{1}y_{0}^{2}-x_{1}x_{0}^{3}=0.$$
(22)

Following the same process as in Section 3, we substitute $x_i = \alpha_i + e_i$, $y_i = \beta_i + f_i$ for i = 0, 1, 2 in the above equation and obtaining a linear system of congruence equations:

$$\sum_{i=1}^{6} A_i X_i + \sum_{i=7}^{15} A_i \Delta X_i + \sum_{i=16}^{21} A_i \Delta^2 X_i + A_{22} \Delta^3 X_{22} \equiv A_0 \Delta^3 \mod p,$$

$$X_i \equiv 0 \mod \Delta^3, \ i = 1, \dots, 6 \qquad (23)$$

$$X_i \equiv 0 \mod \Delta^2, \ i = 7, \dots, 15$$

$$X_i \equiv 0 \mod \Delta, \ i = 16, \dots, 21$$

with at least a solution bounded by $\sqrt{42}\Delta^4$:

$$\boldsymbol{E} = (\Delta^3 E_1, \dots, \Delta^3 E_6, \Delta^2 E_7, \dots, \Delta^2 E_{15}, \Delta E_{16}, \dots, \Delta E_{21}, E_{22}),$$
(24)

which first six components contain the approximation errors, and the other ones are polynomial expressions on those:

The coefficients A_i , i = 1, ..., 22 describing the system are easily obtained from the known infomation $\alpha_i, \beta_i, i = 0, 1, 2$ and Δ . Now, we can find a particular solution T to the system (23) and then apply the CVP algorithm for the shift vector T and the homogenization lattice obtained from system (23):

$$\sum_{i=1}^{6} A_i X_i + \sum_{i=7}^{15} A_i \Delta X_i + \sum_{i=16}^{21} A_i \Delta^2 X_i + A_{22} \Delta^3 X_{22} \equiv 0 \mod p,$$

$$X_i \equiv 0 \mod \Delta^3, \ i = 1, \dots, 6 \qquad (25)$$

$$X_i \equiv 0 \mod \Delta^2, \ i = 7, \dots, 15$$

$$X_i \equiv 0 \mod \Delta, \ i = 16, \dots, 21$$

Then, we obtain an smaller vector \boldsymbol{F} in polynomial time from the given information. We might hope that \boldsymbol{E} and \boldsymbol{F} are the same. This time, we are not giving a rigorous proof to bound the number of possibilities for which this method could fail.

The so-called "Gaussian heuristic" suggests that and *s*-dimensional lattice \mathcal{L} with volume $vol(\mathcal{L})$ is unlikely to have a nonzero vector which is substantially shorter than $vol(\mathcal{L})^{1/s}$. Moreover, if it is known that such a very short vector does exist, then up to a scalar factor it is likely to be the only vector with this property.

Then, vector \boldsymbol{E} is likely to be the one founded whenever $\Delta^4 < p^{1/22} \Delta^{42^{1/22}}$, this is,

$$\Delta < p^{1/46} = p^{0,0217...}.$$

5 Empirical results

We have proposed two algorithms to recover a sequence of pseudorandom numbers produced by EC-LCG. The input required by both algorithms include approximations to some pseudorandom values. The first one requires additionally precise knowledge of the parameter G. The quality of those approximations is the measure used to characterise when the algorithms output the expected sequence.

In the first case, a "bad" set of values for the component x_0 is described, proving that whenever that value lies outside the set, the algorithm works correctly. Furthermore, the size of the set is asymptotically bounded with Δ^6 . This means that if $\Delta < p^{1/6}$ and p is large enough, assuming a uniform distribution of probabilities for $x_0 \in \mathbb{F}_p$, the method is unlikely to fail.

However, two aspects must be taken into account before considering $p^{1/6}$ as the threshold for the error tolerance upon which the algorithm fails. On the one side, the constants hidden in the asymptotic reasoning (namely, the size of the prime p). On the other one, the threshold could be higher, as the "bad" set does not guarantee that the methods indeed fails.

We have performed some numerical tests with a C++ implementation of the main Theorem, using NTL library [41]. Firstly, we generate an elliptic curve over

a prime finite field of a desired size by chosing ramdomly in \mathbb{F}_p parameters a, b to fix Equation (1). Then, we generate randomly points in the curve by choosing their first coordinate and trying to solve Equation (1). For several pairs of points, an EC-LCG is simulated, and approximations to some consecutive values are given as input to our algorithms.

We summarize its results in the following table. We have selected primes of several sizes, and note the obtained success threshold. As we can see, for the first method 1/6 appears as the correct threshold:

$\log_2(p)$	50	100	500	1000
$\log_p(\Delta)$	0.15	0.156	0.164	0.165

As for the second algorithm proposed, the threshold has been obtained using the so-called Gaussian heuristic. As the dimesion of the employed lattice is significantly bigger, the prime size must be also bigger to obtain results according our predictions.

6 Remarks and Open Questions

Obviously our result is nontrivial only for $\Delta = O(p^{1/6})$. Thus increasing the size of the admissible values of Δ (even at the cost of considering more consecutive approximations) is of prime importance.

It would be interesting to provide a proof of the heuristic method for the case that the composer G is secret. Unfortunately, we do not know how to predict the EC-LCG when the modulus p is secret as well. Certainly both of these questions deserves further study.

Finally, it is no clear how to extend these results to the Power Generator on elliptic curves and to the Naor-Reingold Generator on Elliptic curves, see [35, 38, 45, 46].

References

- M. Ajtai, R. Kumar and D. Sivakumar, "A sieve algorithm for the shortest lattice vector problem", Proc. 33rd ACM Symp. on Theory of Comput. (STOC 2001), Association for Computing Machinery, 2001, 601–610.
- 2. R. Avanzi, H. Cohen, C. Doche, G. Frey, T. Lange and K. Nguyen, "Elliptic and hyperelliptic curve crytography: Theory and practice", CRC Press, 2005.
- L. Babai, "On Lovasz Lattice Reduction and the Nearest Lattice Point Problem", Combinatorica, 6, 1986, 1–6.
- P. Beelen and J. Doumen, 'Pseudorandom sequences from elliptic curves', Finite Fields with Applications to Coding Theory, Cryptography and Related Areas, Springer-Verlag, Berlin, 2002, 37-52.
- S. R. Blackburn, D. Gomez-Perez, J. Gutierrez and I. E. Shparlinski, "Predicting nonlinear pseudorandom number generators", *Math. Computation*, **74** (2005), 1471–1494.

- S. R. Blackburn, D. Gomez-Perez, J. Gutierrez and I. E. Shparlinski, "Predicting the inversive generator", *Proc. Coding and Cryptography*, *IMA-03*, LNCS 2898, Springer-Verlag, Berlin 2003, 264–275.
- S. R. Blackburn, D. Gomez-Perez, J. Gutierrez and I. E. Shparlinski, "Reconstructing noisy polynomial evaluation in residue rings", *Journal of Algorithms*. Vol 61(12) (2006), 45–57.
- I. Blake, G. Seroussi and N. Smart, "Elliptic curves in cryptography", London Math. Soc., Lecture Note Series, 265, Cambridge Univ. Press, 1999.
- J. Bloemer, A. May, "A tool kit for Finding small roots of Bivariate Polynomial over the Integers", Advances in Cryptology-Crypto 2003, LNCS 2729, Springer Verlag, 2003, 27–43.
- 10. J. Boyar, 'Inferring sequences produced by pseudo-random number generators', J. ACM, **36** (1989), 129–141
- J. Boyar, 'Inferring sequences produces by a linear congruential generator missing low-order bits', J. Cryptology 1 (1989) 177–184.
- E. F. Brickell and A. M. Odlyzko, 'Cryptanalysis: A survey of recent results', *Contemp. Cryptology*, IEEE Press, NY, 1992, 501–540.
- J.W.S. Cassels, "An Introduction to the Geometry of Numbers". Springer-Verlag, New York, 1971.
- 14. D. Coppersmith: "Small solutions to polynomial equations and low exponent RSA vulnerabilities". J. Cryptology 10 (4), 1997, 233–260.
- D. Coppersmith: "Finding a Small Root of a Bivariate Integer Equations; Factoring with High Bits Known". U. Maurer (Ed), Proc. EUROCRYPT-96, LNCS 1070, Springer-Verlag, Berlin 1996, 155–156.
- J-S Coron, "Finding small roots of Bivariate Integer Polynomial Equations Revisted", Proc. Advances in Cryptology- Eurocrypt'04, LNCS 3027, Springer Verlag, 2004, 492–505.
- E. El Mahassni and I. E. Shparlinski, 'On the uniformity of distribution of congruential generators over elliptic curves', Proc. Intern. Conf. on Sequences and their Applications, Bergen 2001, Springer-Verlag, London, 2002, 257–264.
- A. M. Frieze, J. Håstad, R. Kannan, J. C. Lagarias and A. Shamir, 'Reconstructing truncated integer variables satisfying linear congruences', SIAM J. Comp., 17 (1988), 262–280.
- D. Gomez-Perez, J. Gutierrez and A. Ibeas, "Cryptanalysis of the Quadratic generator", *Proceedings in Cryptology-INDOCRYPT 2005*, LNCS **3797**, Springer Verlag, Berlin 2005, 118–129.
- D. Gomez-Perez, J. Gutierrez and A. Ibeas, "An Algorithm for Finding Small Roots of Multivariate Polynomials over the Integers", Faculty of Science, University of Cantabria, Preprint, 2006.
- D. Gomez-Perez, J. Gutierrez and A. Ibeas, "Attacking the Pollard Generator", IEEE. Trans. Information Theory, vol. 52, n. 12, 2006.
- G. Gong, T. A. Berson and D. A. Stinson, 'Elliptic curve pseudorandom sequence generators', *Lect. Notes in Comp. Sci.*, Springer-Verlag, Berlin, 1758 (2000), 34–49.
- G. Gong and C. C. Y. Lam, 'Linear recursive sequences over elliptic curves', Proc. Intern. Conf. on Sequences and their Applications, Bergen 2001, Springer-Verlag, London, 2002, 182-196.
- 24. M. Grötschel, L. Lovász and A. Schrijver, *Geometric algorithms and combinatorial optimization*, Springer-Verlag, Berlin, 1993.
- S. Hallgren, 'Linear congruential generators over elliptic curves', Preprint CS-94-143, Dept. of Comp. Sci., Cornegie Mellon Univ., 1994, 1-10.

- F. Hess and I. E. Shparlinski, 'On the linear complexity and multidimensional distribution of congruential generators over elliptic curves', *Designs, Codes and Cryptography*, 35 (2005), 111–117.
- N. A. Howgrave-Graham, 'Finding small roots of univariate modular equations revisited', Proc. 6th IMA Intern. Conf on Cryptography and Coding, Lect. Notes in Comp. Sci., vol. 1355, Springer-Verlag, Berlin, 1997, 131–142.
- J.W.S. Gruber and C.G. Lekkerkerker, "Geometry of Numbers". North-Holland, 1987.
- A. Joux and J. Stern, "Lattice reduction: A toolbox for the cryptanalyst", J. Cryptology, 11 (1998), 161–185.
- 30. E. Jochemz and A. May, "A Strategy for Finding Roots of Multivariate Polynomials with New Applications in Attacking RSA Variants" n Advances in Cryptology (Asiacrypt 2006), Lecture Notes in Computer Science, Springer-Verlag, 2006.
- R. Kannan, "Minkowski's convex body theorem and integer programming", Math. Oper. Res., 12 (1987), 415–440.
- D. E. Knuth, 'Deciphering a linear congruential encryption', *IEEE Trans. Inf. Theory* 31 (1985), 49–52.
- H. Krawczyk, 'How to predict congruential generators', J. Algorithms, 13 (1992), 527–545.
- J. C. Lagarias, 'Pseudorandom number generators in cryptography and number theory', Proc. Symp. in Appl. Math., Amer. Math. Soc., Providence, RI, 42 (1990), 115–143.
- T. Lange and I. E. Shparlinski, 'Certain exponential sums and random walks on elliptic curves', *Canad. J. Math.*, 57 (2005), 338–350.
- A. K. Lenstra, H. W. Lenstra and L. Lovász, "Factoring polynomials with rational coefficients", *Mathematische Annalen*, 261 (1982), 515–534.
- D. Micciancio and S. Goldwasser, "Complexity of lattice problems", Kluwer Acad. Publ., 2002.
- M. Naor and O. Reingold, "Number theoretic constructions of efficient pseudorandom functions", Proc 38th IEEE Symp. on Found. of Comp. Sci., IEEE, 1997, 458–467.
- H. Niederreiter, 'New developments in uniform pseudorandom number and vector generation', in: H. Niederreiter and P.J. Shiue (Eds), Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing, Lect. Notes in Statistics Vol. 106, Springer-Verlag, Berlin, 1995, 87–120.
- 40. H. Niederreiter, 'Design and analysis of nonlinear pseudorandom number generators', in G.I. Schueller and P. D. Spanos (Eds) *Monte Carlo Simulation*, A.A. Balkema Publishers, Rotterdam, 2001, 3–9.
- V. Shoup, "Number theory C++ library (NTL)", version 5.3.1, available at http://www.shoup.net/ntl/.
- I. E. Shparlinski, 'Cryptographic applications of analytic number theory ', Birkhauser, 2003.
- I. E. Shparlinski, "Orders of points on elliptic curves", Affine Algebraic Geometry, Amer. Math. Soc., 2005, 245–252.
- 44. I. E. Shparlinski, "Pseudorandom Points on Elliptic Curves over Finite Fields", *Recent trends in Cryptography*, Contemporary Mathematics, Amer.Math. Soc., to appear.
- 45. I. E. Shparlinski, 'On the Naor-Reingold pseudo-random function from elliptic curves', *Appl. Algebra in Engin.*, Commun. and Computing, 11 (2000), 27–34.

- I. E. Shparlinski and J. H. Silverman, 'On the linear complexity of the Naor-Reingold pseudorandom function from elliptic curves', *Designs, Codes and Cryp*tography, 24 (2001), 279–289.
- 47. J. H. Silverman, "The arithmetic of elliptic curves", Springer-Verlag, Berlin, 1995.