Point Decomposition Problem in Binary Elliptic Curves

Koray Karabina(✉)

Florida Atlantic University, Boca Raton, FL 33431, USA
kkarabina@fau.edu

Abstract. We analyze the point decomposition problem (PDP) in binary elliptic curves. It is known that PDP in an elliptic curve group can be reduced to solving a particular system of multivariate non-linear equations derived from the so called Semaev summation polynomials. We modify the underlying system of equations by introducing some auxiliary variables. We argue that the trade-off between lowering the degree of Semaev polynomials and increasing the number of variables provides a significant speed-up.

Keywords: Semaev polynomials · Elliptic curves · Point decomposition problem · Discrete logarithm problem

1 Introduction

The *point decomposition problem* (PDP) in an additive abelian group $\mathbb{G}$ with respect to a *factor base* $\mathcal{B} \subset \mathbb{G}$ is the following: Given a point[1] $R \in \mathbb{G}$, find $P_i \in \mathcal{B}$ such that

$$R = \sum_{i=1}^{m} P_i$$

for some positive integer m; or conclude that R cannot be decomposed as a sum of points in $\mathcal{B}$. The *discrete logarithm problem* (DLP) in $\mathbb{G}$ with respect to a base $P \in \mathbb{G}$ is the following: Given P and $Q = aP \in \mathbb{G}$ for some secret integer a, compute a. DLP can be solved using the *index calculus algorithm* in two main steps. In the *relation collection* step, fix a factor base $\mathcal{B}$, and find a set of points $R_i = a_iP + b_iQ$ for some randomly chosen integers a_i, b_i, such that R_i can be decomposed with respect to $\mathcal{B}$, i.e.,

$$R_i = \sum_j P_{ij}, \ P_{ij} \in \mathcal{B}.$$

Here, we may assume for convenience that P_{ij} are not necessarily distinct. Note that each decomposition induces a modular linear dependence on the discrete

[1] We prefer to use *point* rather than *element* because elliptic curve group elements are commonly called points.

© Springer International Publishing Switzerland 2016
S. Kwon and A. Yun (Eds.): ICISC 2015, LNCS 9558, pp. 155–168, 2016.
DOI: 10.1007/978-3-319-30840-1_10

logarithms of $Q \in \mathbb{G}$ and $P_{ij} \in \mathcal{B}$ with respect to the base P. After collecting sufficiently many relations[2], *linear algebra* step solves for the discrete logarithm of $Q \in \mathbb{G}$, as well as the discrete logarithms of the factor base elements. Clearly, the success probability and the running time of the index calculus algorithm heavily depend on the decomposition probability of a random element in $\mathbb{G}$, the cost of the decomposition step, and the size of the factor base. In particular, the overall cost of the relation collection and the linear algebra steps must be optimized with a non-trivial success probability.

In 2004, Semaev [11] showed that solving PDP in an elliptic curve group is equivalent to solving a particular system of multivariate non-linear equations derived from the so called *Semaev summation polynomials*. Semaev's work triggered the possibility of the existence of an index calculus type algorithm which is more efficient than the Pollard's rho algorithm to solve the discrete logarithm problem in elliptic curves defined over $\mathbb{F}_{q^n}$, which we denote ECDLP(q, n). Note that Pollard's rho algorithm is a general purpose algorithm that solves DLP in a group $\mathbb{G}$, and runs in time $\mathcal{O}(\sqrt{|\mathbb{G}|})$. Gaudry [7] showed that, for a fixed n, Semaev summation polynomials can be effectively used to solve ECDLP(q, n) in heuristic time $\mathcal{O}(q^{2-\frac{2}{n}})$, where the constant in $\mathcal{O}(\cdot)$ is exponential in n. For example, Gaudry's algorithm and Pollard's rho algorithm solve ECDLP$(q, 3)$ in time $\mathcal{O}(q^{1.33})$ and $\mathcal{O}(q^{1.5})$, respectively. Due to the exponential in n constant in the running time of Gaudry's algorithm, his attack is expected to be more effective than Pollard's rho algorithm if $n \geq 3$ is relatively small and q is large. Diem [2] rigorously showed that ECDLP(q, n) can be solved in an expected subexponential time when $a(\log q)^{\alpha} \leq n \leq b(\log q)^{\beta}$ for some $a, b, \alpha, \beta > 0$. On the other hand, Diem's method has expected exponential running time $\mathcal{O}(e^{n(\log n)^{1/2}})$ for solving ECDLP$(2, n)$. As a result, the index calculus type algorithms presented in [2,7] do not yield ECDLP solvers which are more effective than Pollard's rho method when $q = 2$ and n is prime. The ideas for choosing an appropriate factor base in [2] have been adapted in [5,10], and the complexity of the relation collection step have been analyzed. In both papers [5] and [10], a positive integer m, which we call the *decomposition constant*, is fixed to represent the number of points in the decomposition of a random point in the relation collection step. The factor base consists of elliptic curve points whose x-coordinates belong to an n'-dimensional subspace $V \subset \mathbb{F}_{2^n}$ over $\mathbb{F}_2$, where n' is chosen such that $mn' \approx n$. We refer to PDP in this setting by PDP(n, m, n') throughout the rest of this paper.

Faugère et al. [5] showed, under a certain assumption, that ECDLP$(2, n)$ can be solved in time $\mathcal{O}(2^{\omega n/2})$, where $2.376 \leq \omega \leq 3$ is the linear algebra constant. The running time analysis in [5] considers the linearization technique to solve the system of multivariate nonlinear of equations which are derived from the $(m+1)$'st Semaev polynomial S_{m+1} during the relation collection step to solve PDP(n, m, n'). Faugère et al. further argue that, Groebner basis techniques may improve the running time by a factor m in the exponent, where m is the decomposition constant. This last claim has been confirmed in the experiments in [5] for elliptic curves defined over $\mathbb{F}_{2^n}$ with $n \in \{41, 67, 97, 131\}$ and $m = 2$.

[2] This is roughly when the number of relations exceeds $|\mathcal{B}|$.

Petit and Quisquater's heuristic analysis in [10] claims that ECDLP$(2, n)$ can asymptotically be solved in time $\mathcal{O}(2^{cn^{2/3}\log n})$ for some constant $0 < c < 2$. The subexponential running time in [10] is based on a rather strong assumption on the behavior of the systems of equations that arise from Semaev polynomials. In particular, it is assumed in [10] that the degree of regularity D_{reg} and the first fall degree $D_{\mathsf{FirstFall}}$ of the underlying polynomial systems to solve PDP(n, m, n') are approximately equal. The analysis in [10] also assumes that $n' = n^{\alpha}$ and $m = n^{1-\alpha}$ for some positive constant α. Experiments with a very limited set of parameters (n, m, n'), $n \in \{11, 17\}$, $m \in \{2, 3\}$, $n' = \lceil n/m \rceil$ were conducted in [10] in the favor of their assumption.

A recent paper by Shantz and Teske [13] presented some extended experimental results on solving PDP(n, m, n') for the same setting as in the Petit and Quisquater's paper [10]. In particular, [13] validates the degree of regularity assumption in [10] for the set of parameters (n, m, n') such that $n \in \{11, 13, 15, 17, 19, 23, 29\}$, $m = 2$, $n' = \lceil n/m \rceil$; and for (n, m, n') such that $n \in \{11, 13, 15, 17, 19, 21\}$, $m = 3$, $n' = \lceil n/m \rceil$. Shantz and Teske [13] were able to extend their experimental data for the parameters (n, m, n', Δ), $n \leq 48$, $m = 2$, and where $\Delta = n - mn'$ is chosen appropriately to possibly improve the running time of ECDLP$(2, n)$. In another recent paper [8], Huang et al. exploit the symmetry in Semaev polynomials, and improve on the running time and memory requirements of the PDP(n, m, n') solver in [5]. The efficiency of the method in [8] is tested for parameters (n, m, n') such that $n \leq 53$, $m = 3$, $n' = 3, 4, 5, 6$.

Petit and Quisquater's heuristic analysis [10] claims that index calculus methods for solving ECDLP$(2, n)$ is more effective than the Pollard's rho method for $n > 2000$, $m \geq 4$ and $mn' \approx n$. However, all the experiments reported so far on solving PDP(n, m, n') for the set of parameters (n, m, n', Δ) with $\Delta = n - mn' \leq 1$ and $m = 3$ are limited to $n \leq 19$; see [8,13]. Similarly, all the experiments for the set of parameters (n, m, n', Δ) with $m = 3$ are limited to $n' \leq 6$, which forces $\Delta \geq 2$ for $n \geq 20$. In general, it is desired to have n' increasing as a function of n, rather than having some upper bound on n', so that $n \approx mn'$ as assumed in the running time analysis of ECDLP$(2, n)$ solvers in [5,10]. Therefore, it remains a challenge to run experiments on an extensive set of parameters (n, m, n') with larger prime n values, $m \geq 4$, and $mn' \approx n$. For example, it is stated in [8, Sect. 4.1] that

> On the other hand, the method appears unpractical for $m = 4$ even for very small values of n because of the exponential increase with m of the degrees in Semaev's polynomials.

In a more recent paper [6], Galbraith and Gebregiyorgis introduce a new choice of variables and a new choice of factor base, and they are able to solve PDP with various $n \geq 17$, $m = 4$, $n' = 3, 4$ using Groebner basis algorithms; and also with various $n \geq 17$, $m = 4$, $n' \leq 7$ using SAT solvers.

In this paper, we modify the system of equations, that are derived from Semaev polynomials, by introducing some auxiliary variables. We show that

PDP(n, m, n') can be solved by finding a solution to a system of equations derived from several third Semaev polynomials S_3 each of which has at most three variables. For a comparison, PDP(n, m, n') in $E(\mathbb{F}_{2^n})$ with decomposition constant $m = 5$ would be traditionally attacked via considering the Semaev polynomial S_6 with 5 variables, which is likely to have a root in V^5, where $V \subset \mathbb{F}_{2^n}$ is a subspace of dimension $n' = \lfloor n/5 \rfloor$. On the other hand, when $m = 5$, our polynomial system consists of third Semaev polynomials $S_{3,i}$ $(i = 1, 2, 3, 4)$, and a total of 8 variables which is likely to have a root in $V^5 \times \mathbb{F}_{2^n}^3$, where $V \subset \mathbb{F}_{2^n}$ is a subspace of dimension $\lfloor n/5 \rfloor$. As a result, our technique overcomes the difficulty of dealing with the $(m + 1)$'st Semaev polynomial S_{m+1} when solving PDP(n, m, n') with $m \geq 4$. We should emphasize that choosing $m \geq 4$ is desirable for an index calculus based ECDLP$(2, n)$ solver to be more effective than a generic DLP solver such as Pollard's rho algorithm. Our method introduces an overhead of introducing some auxiliary variables. However, we argue that the trade-off between lowering the degree of Semaev polynomials and increasing the number of variables provides a significant speed-up. In particular, we present some experimental results on solving PDP(n, m, n') for the following parameters:

- $n \leq 19$, $m = 4, 5$, and $n' = \lfloor n/m \rfloor$. We are not aware of any previous experimental data for $n > 15$ and $m = 5$.
- $n \leq 26$, $m = 3$, $n' = \lfloor n/m \rfloor$. We are not aware of any previous experimental data for $n > 21$, $m = 3$, and $\Delta = n - mn' \leq 2$.

We observe in our experiments that regularity degrees of the underlying systems are relatively low. We also observe that running time and memory requirement of algorithms can be improved significantly if the Groebner basis computations are first performed on a subset of polynomials and if the ReductionHeuristic parameter in Magma is set to be a small number; see Sect. 5 for more detail. We would like to emphasize that these techniques are applied for the first time in this paper to solve the point decomposition problem. As a result, we gain significant improvement over the recently published experimental results [12]. For a comparison, we are able to solve PDP$(15, 5, 3)$ instances in about 7 s (with 256 MB memory). Note that, PDP$(15, 5, 3)$ is solved in about 175 s (with 2635 MB memory) in [12]. In general, our experimental findings with $m = 3, 4, 5$ extend and improve on the recently reported results in [8,12,13].

The rest of this paper is organized as follows. In Sect. 2, we recall Semaev polynomials and their application to ECDLP$(2, n)$. In Sect. 3, we describe and analyze a new method to solve PDP(n, m, n') in $E(\mathbb{F}_{2^n})$. In Sect. 4, we present our experimental results. In Sect. 5, we extend our results from Sect. 3.

2 Semaev Polynomials and ECDLP

Let $\mathbb{F}_{2^n} = \mathbb{F}_2[\sigma]/\langle f(\sigma) \rangle$ be a finite field with 2^n elements, where $f(\sigma)$ is a monic irreducible polynomial of degree n over the field $\mathbb{F}_2 = \{0, 1\}$. Let E be a non-singular elliptic curve defined by the short Weierstrass equation

$$E/\mathbb{F}_{2^n} : \; y^2 + xy = x^3 + ax^2 + b, \; a, b \in \mathbb{F}_{2^n}.$$

We denote the identity element of E by ∞. The i'th Semaev polynomial associated with E is defined as follows:

$$S_i(x_1, x_2, \ldots, x_i) = \begin{cases} (x_1x_2 + x_1x_3 + x_2x_3)^2 + x_1x_2x_3 + b & \text{if } i = 3 \\ \mathrm{Res}_X(S_{i-j}(x_1, \ldots, x_{i-j-1}, X), S_{j+2}(x_{i-j}, \ldots, x_i, X)) & \text{if } i \geq 4, \end{cases} \tag{1}$$

where $1 \leq j \leq i-3$.

For $n' \leq n$, let

$$V = \{a_0 + a_1\sigma + \cdots + a_{n'-1}\sigma^{n'-1} : \ a_i \in \mathbb{F}_2\} \subset \mathbb{F}_{2^n}$$

and define the factor base

$$\mathcal{B} = \{P = (x, y) \in E : \ x \in V\}.$$

Recall that in PDP(n, m, n'), we are looking for $P_i = (x_i, y_i) \in \mathcal{B}$ such that

$$P_1 + \cdots + P_m = R, \tag{2}$$

for some given point $R = (x_R, y_R) \in E$. We refer to (2) as an m-decomposition of R in $\mathcal{B}$. We expect that, on average, a random point $R \in E$ has an m-decomposition in $\mathcal{B}$ with probability $2^{mn'}/2^n m!$ simply because $|\mathcal{B}| \approx 2^{n'}$ and permuting P_i does not change the sum $\sum P_i$ (see [7]). As described in Sect. 1, the DLP in E can be solved via an index-calculus based approach by computing about $|\mathcal{B}|$ explicit m-decompositions and solving a sparse linear system of about $|\mathcal{B}|$ equations. Therefore, the cost of solving ECDLP$(2, n)$ may be estimated as

$$2^{n'}\frac{2^n m!}{2^{mn'}}C_{n,m,n'} + 2^{\omega' n'}, \tag{3}$$

where $C_{n,m,n'}$ is the cost of solving PDP(n, m, n'), and $\omega' = 2$ is the sparse linear algebra constant. Semaev [11] showed that a decomposition of the form (2) exists if and only if the x-coordinates of P_i and R are zeros of the $(m+1)$'st Semaev polynomial, that is, $S_{m+1}(x_1, \ldots, x_m, x_R) = 0$. In the rest of this paper, we focus on solving PDP(n, m, n') (and estimating $C_{n,m,n'}$) via modifying the equation induced by S_{m+1}.

3 A New Approach to Solve the Point Decomposition Problem

Let $E/\mathbb{F}_{2^n}$, V, and $\mathcal{B}$ be as defined in Sect. 2. Recall that an m-decomposition of a point

$$R = P_1 + \cdots + P_m,$$

where $R = (x_R, y_R) \in E$, $P_i = (x_i, y_i) \in \mathcal{B}$, can be computed (if exists) by identifying a tuple $(x_1, \ldots, x_m) \in V^m$ that satisfies

$$S_{m+1}(x_1, \ldots, x_m, x_R) = 0 \tag{4}$$

Note that x_i belong to an n'-dimensional subspace of $\mathbb{F}_{2^n}$. Therefore, (4) defines a system Sys_1 of a single equation over $\mathbb{F}_{2^n}$ in m variables. In [5,10], the Weil descent technique is applied, and a second system Sys_2 of n equations over $\mathbb{F}_2$ in mn' boolean variables is derived from Sys_1. The cost $C_{n,m,n'}$ of solving PDP(n, m, n') in [5,10] is estimated through the analysis of solving Sys_2 using linearization and Groebner basis techniques. Next, we describe a new approach to derive another system Sys_3 of boolean equations such that a solution of Sys_3 yields an m-decomposition of a point R.

Notation. Throughout the rest of this paper, we distinguish between two classes Semaev polynomials. The first class of Semaev polynomials is denoted by $S_{m,1}(x_1, \ldots, x_m)$, which represents the m'th Semaev polynomial with m variables. The second class of Semaev polynomials is denoted by $S_{m,2}(x_1, \ldots, x_{m-1}, x_R)$, which represents the m'th Semaev polynomial with $m-1$ variables (i.e., the last variable x_m is evaluated at a number x_R).

3.1 The Case: $m = 3$

Let $R = (x_R, y_R) \in E$. Notice that there exist $P_i \in \mathcal{B}$ such that

$$P_1 + P_2 + P_3 - R = \infty$$

if and only if there exist $P_i \in \mathcal{B}$ and $P_{12} \in E$ such that

$$\begin{cases} P_1 + P_2 - P_{12} = \infty \\ P_3 + P_{12} - R = \infty \end{cases} \tag{5}$$

Therefore, a 3-decomposition of $R = P_1 + P_2 + P_3$ may be found as follows:

1. Define the following system of equations derived from Semaev polynomials

$$\begin{cases} S_{3,1}(x_1, x_2, x_{12}) = 0 \\ S_{3,2}(x_3, x_{12}, x_R) = 0. \end{cases} \tag{6}$$

   Note that this system is defined over $\mathbb{F}_{2^n}$ and has 4 variables x_1, x_2, x_3, x_{12}.
2. Introduce boolean variables $x_{i,j}$ such that

$$x_i = \sum_{j=0}^{n'-1} x_{i,j}\sigma^j,$$

   for $i = 1, 2, 3$, and

$$x_{12} = \sum_{j=0}^{n} x_{12,j}\sigma^j.$$

   Apply the Weil descent technique to (6) and define an equivalent system of $2n$ equations over $\mathbb{F}_2$ with $3n' + n$ boolean variables

$$\{x_{i,j} : \ i = 1, 2, 3, \ j = 0, \ldots n' - 1\} \cup \{x_{12,j} : \ j = 0, \ldots n - 1\}.$$

Solve this new system of boolean equations and recover $x_1, x_2, x_3 \in \mathbb{F}_{2^n}$ from $x_{i,j} \in \mathbb{F}_2$.

Note that the proposed method solves a system of $2n$ equations over $\mathbb{F}_2$ with $3n' + n$ boolean variables rather than solving a system of n equations over $\mathbb{F}_2$ with $3n'$ boolean variables.

3.2 The Case: $m = 4$

Let $R = (x_R, y_R) \in E$. Notice that there exist $P_i \in \mathcal{B}$ such that

$$P_1 + P_2 + P_3 + P_4 - R = \infty$$

if and only if there exist $P_i \in \mathcal{B}$ and $P_{12} \in E$ such that

$$\begin{cases} P_1 + P_2 - P_{12} = \infty \\ P_3 + P_4 + P_{12} - R = \infty \end{cases} \tag{7}$$

Therefore, a 4-decomposition of $R = P_1 + P_2 + P_3 + P_4$ may be found as follows:

1. Define the following system of equations derived from Semaev polynomials

$$\begin{cases} S_{3,1}(x_1, x_2, x_{12}) = 0 \\ S_{4,2}(x_3, x_4, x_{12}, x_R) = 0 \end{cases} \tag{8}$$

   Note that this system is defined over $\mathbb{F}_{2^n}$ and has 5 variables $x_1, x_2, x_3, x_4, x_{12}$.
2. Introduce boolean variables $x_{i,j}$ such that

$$x_i = \sum_{j=0}^{n'-1} x_{i,j}\sigma^j,$$

   for $i = 1, 2, 3, 4$, and

$$x_{12} = \sum_{j=0}^{n} x_{i,j}\sigma^j.$$

   Apply the Weil descent technique to (8) and define an equivalent system of $2n$ equations over $\mathbb{F}_2$ with $4n' + n$ boolean variables

$$\{x_{i,j} : \ i = 1, 2, 3, 4 \ j = 0, \ldots n' - 1\} \cup \{x_{12,j} : \ j = 0, \ldots n - 1\}.$$

   Solve this new system of boolean equations and recover $x_1, x_2, x_3, x_4 \in \mathbb{F}_{2^n}$ from $x_{i,j} \in \mathbb{F}_2$.

Note that the proposed method solves a system of $2n$ equations over $\mathbb{F}_2$ with $4n' + n$ boolean variables rather than solving a system of n equations over $\mathbb{F}_2$ with $4n'$ boolean variables.

3.3 The Case: $m = 5$

Let $R = (x_R, y_R) \in E$. Notice that there exist $P_i \in \mathcal{B}$ such that

$$P_1 + P_2 + P_3 + P_4 + P_5 - R = \infty$$

if and only if there exist $P_i \in \mathcal{B}$ and $P_{123} \in E$ such that

$$\begin{cases} P_1 + P_2 + P_3 - P_{123} = \infty \\ P_4 + P_5 + P_{123} - R = \infty \end{cases} \tag{9}$$

Therefore, a 5-decomposition of $R = P_1 + P_2 + P_3 + P_4 + P_5$ may be found as follows:

1. Define the following system of equations derived from Semaev polynomials

$$\begin{cases} S_{4,1}(x_1, x_2, x_3, x_{123}) = 0 \\ S_{4,2}(x_4, x_5, x_{123}, x_R) = 0 \end{cases} \tag{10}$$

   Note that this system is defined over $\mathbb{F}_{2^n}$ and has 6 variables x_1, x_2, x_3, x_4, x_5, x_{123}.
2. Introduce boolean variables $x_{i,j}$ such that

$$x_i = \sum_{j=0}^{n'-1} x_{i,j}\sigma^j,$$

   for $i = 1, 2, 3, 4, 5$, and

$$x_{123} = \sum_{j=0}^{n} x_{123,j}\sigma^j.$$

   Apply the Weil descent technique to (10) and define an equivalent system of $2n$ equations over $\mathbb{F}_2$ with $5n' + n$ boolean variables

$$\{x_{i,j} : \ i = 1, 2, 3, 4, 5 \ j = 0, \ldots n' - 1\} \cup \{x_{123,j} : \ j = 0, \ldots n - 1\}.$$

   Solve this new system of boolean equations and recover $x_1, x_2, x_3, x_4, x_5 \in \mathbb{F}_{2^n}$ from $x_{i,j} \in \mathbb{F}_2$.

Note that the proposed method solves a system of $2n$ equations over $\mathbb{F}_2$ with $5n' + n$ boolean variables rather than solving a system of n equations over $\mathbb{F}_2$ with $5n'$ boolean variables.

3.4 Analysis of the New Polynomial Systems

One of the methods to solve a system of multivariate non-linear equations is to compute the Groebner basis of the underlying ideal. Groebner basis computations can be performed using Faugère's algorithms [3,4], which reduce the

problem to Gaussian elimination of Macaulay-type matrices M_d of degree d. The Macaulay matrix M_d encodes degree (at most) d polynomials, that are generated during Groebner basis computation. Therefore, the cost of solving a system of equations is determined by the maximal degree D (also known as the degree of regularity of the system) reached during the computation. If N is the number of variables in the system, then the cost is estimated as $O\left(\binom{N+D-1}{D}^{\omega}\right)$, where $\binom{N+D-1}{D}$ is the maximum number of columns in M_D and ω is the linear algebra constant. In general, it is hard to estimate D. In the recent paper [10], it is conjectured that the degree of regularity D_{reg} of systems arising from PDP(n, m, n') satisfies $D_{\mathsf{reg}} = D_{\mathsf{FirstFall}} + o(1)$, where $D_{\mathsf{FirstFall}}$ is the first fall degree of the system and defined as follows.

Definition 1. [10] Let R be a polynomial ring over a field K. Let $F := \{f_1, \ldots, f_\ell\} \subset R$ be a set of polynomials of degrees at most $D_{\mathsf{FirstFall}}$. The first fall degree of F is the smallest degree $D_{\mathsf{FirstFall}}$ such that there exist polynomials $g_i \in R$ with $\max_i (\deg(f_i) + \deg(g_i)) = D_{\mathsf{FirstFall}}$, satisfying $\deg(\sum_{i=1}^{\ell} g_i f_i) < D_{\mathsf{FirstFall}}$ but $\sum_{i=1}^{\ell} g_i f_i \neq 0$.

Experimental studies in recent papers [10,13] give supporting evidence that $D_{\mathsf{reg}} \approx D_{\mathsf{FirstFall}}$. However, experimental data is yet very limited (see Sect. 1) to verify this conjecture. In this section, we compute the first fall degree of the systems proposed in Sects. 3.1, 3.2, and 3.3. Our experimental results in Sect. 4 support that $D_{\mathsf{reg}} \approx D_{\mathsf{FirstFall}}$.

***D*$_{\mathsf{FirstFall}}$ of the system when $m = 3$.** In this case, one needs to solve the system of $2n$ equations over $\mathbb{F}_2$ with $3n' + n$ boolean variables. The system of equations is derived by applying Weil descent to (6) that consists of two Semaev polynomials $S_{3,1}$ and $S_{3,2}$. The monomial set of $S_{3,1}(x_1, x_2, x_{12})$ is

$$\{1, x_1^2 x_2^2, x_1^2 x_{12}^2, x_2^2 x_{12}^2, x_1 x_2 x_{12}\}.$$

Therefore, the Weil descent of $S_{3,1}(x_1, x_2, x_{12})$ yields a $2n' + n$ variable polynomial set $\{f_i\}$ over $\mathbb{F}_2$ such that $\max_i(\deg(f_i)) = 3$. On the other hand, the monomial set of $x_1 \cdot S_{3,1}(x_1, x_2, x_{12})$ is

$$\{x_1, x_1^3 x_2^2, x_1^3 x_{12}^2, x_2^2 x_{12}^2, x_1^2 x_2 x_{12}\}.$$

Therefore, the Weil descent of $x_1 \cdot S_{3,1}(x_1, x_2, x_{12})$ yields a polynomial set $\{F_i\}$ over $\mathbb{F}_2$ such that $\max_i(\deg(F_i)) = 3$. It follows from the definition that $D_{\mathsf{FirstFall}}(S_{3,1}) \leq 4$ because the maximum degree of polynomials obtained from the Weil descent of x_1 is 1. Similarly, the monomial set of $S_{3,2}(x_3, x_{12}, x_R)$ is

$$\{1, x_3^2 x_{12}^2, x_3^2, x_{12}^2, x_3 x_{12}\}.$$

Therefore, the Weil descent of $S_{3,2}(x_3, x_{12}, x_R)$ yields a $n'+n$ variable polynomial set $\{f_i\}$ over $\mathbb{F}_2$ such that $\max_i(\deg(f_i)) = 2$. On the other hand, the monomial set of $x_3^3 \cdot S_{3,2}(x_3, x_{21}, x_R)$ is

$$\{x_3^3, x_3^5 x_{12}^2, x_3^5, x_3^3 x_{12}^2, x_3^4 x_{12}\}.$$

Therefore, the Weil descent of $x_3^3 \cdot S_{3,2}(x_3, x_{12}, x_R)$ yields a polynomial set $\{F_i\}$ over $\mathbb{F}_2$ such that $\max_i(\deg(F_i)) = 3$. It follows from the definition that $D_{\mathsf{FirstFall}}(S_{3,2}) \leq 4$ because the maximum degree of polynomials obtained from the Weil descent of x_3^3 is 2. We conclude that $D_{\mathsf{FirstFall}} \leq 4$.

$D_{\mathsf{FirstFall}}$ of the system when $m = 4$. In this case, one needs to solve the system of $2n$ equations over $\mathbb{F}_2$ with $4n' + n$ boolean variables. The system of equations is derived by applying Weil descent to (8) that consists of two Semaev polynomials $S_{3,1}$ and $S_{4,2}$. From our above discussion, $D_{\mathsf{FirstFall}}(S_{3,1}) \leq 4$. Now, analyzing the monomial set of $S_{4,2}(x_3, x_4, x_{123}, x_R)$, we can see that the Weil descent of $S_{4,2}(x_3, x_4, x_{123}, x_R)$ yields a $2n' + n$ variable polynomial set $\{f_i\}$ over $\mathbb{F}_2$ such that $\max_i(\deg(f_i)) = 6$ (this follows from the Weil descent of the monomial $(x_3x_4x_{123})^3$). On the other hand, analyzing the monomial set of $x_3 \cdot S_{4,2}(x_3, x_4, x_{123}, x_R)$, we see that the Weil descent of $x_3 \cdot S_{4,2}(x_3, x_4, x_{123}, x_R)$ yields a polynomial set $\{F_i\}$ over $\mathbb{F}_2$ such that $\max_i(\deg(F_i)) = 6$. It follows from the definition that $D_{\mathsf{FirstFall}}(S_{4,2}) \leq 7$ because the maximum degree of polynomials obtained from the Weil descent of x_3 is 1. We conclude that $D_{\mathsf{FirstFall}} \leq 7$.

$D_{\mathsf{FirstFall}}$ of the system when $m = 5$. In this case, one needs to solve the system of $2n$ equations over $\mathbb{F}_2$ with $5n' + n$ boolean variables. The system of equations is derived by applying Weil descent to (10) that consists of two Semaev polynomials $S_{4,1}$ and $S_{4,2}$. From our above discussion, $D_{\mathsf{FirstFall}}(S_{4,2}) \leq 7$. Now, analyzing the monomial set of $S_{4,1}(x_1, x_2, x_3, x_{123})$, we can see that the Weil descent of $S_{4,1}(x_1, x_2, x_3, x_{123})$ yields a $3n' + n$ variable polynomial set $\{f_i\}$ over $\mathbb{F}_2$ such that $\max_i(\deg(f_i)) = 8$ (this follows from the Weil descent of the monomial $(x_1x_2x_3x_{123})^3$). On the other hand, analyzing the monomial set of $x_3 \cdot S_{4,1}(x_1, x_2, x_3, x_{123})$, we see that the Weil descent of $x_3 \cdot S_{4,1}(x_1, x_2, x_3, x_{123})$ yields a polynomial set $\{F_i\}$ over $\mathbb{F}_2$ such that $\max_i(\deg(F_i)) = 8$. It follows from the definition that $D_{\mathsf{FirstFall}}(S_{4,1}) \leq 9$ because the maximum degree of polynomials obtained from the Weil descent of x_3 is 1. We conclude that $D_{\mathsf{FirstFall}} \leq 9$.

4 Experimental Results

We implemented the methods proposed in Sect. 3 on a desktop computer (Intel(R) Xeon(R) CPU E31240 3.30GHz) using Groebner basis algorithms in Magma [1]. For each parameter set (n, m, n'), we solved 5 random instances of PDP over a randomly chosen elliptic curve $E/\mathbb{F}_{2^n}$. In Table 1, we report on our experimental results for solving PDP$(n, m, n' = \lfloor n/m \rfloor)$ with $m = 3, 4, 5$. In particular, for each of these 5 computations, we report on the maximum CPU time (seconds) and memory (MB) required for solving PDP. We also report on the maximum of the maximum step degrees D in the Groebner basis computations. Recall that in Sect. 3, we estimated $D_{\mathsf{FirstFall}} \leq 4$ when $m = 3$; $D_{\mathsf{FirstFall}} \leq 7$ when $m = 4$; and $D_{\mathsf{FirstFall}} \leq 9$ when $m = 5$. In our experiments, we observe that $D_{\mathsf{reg}} = 4$ when $m = 3$; $D_{\mathsf{reg}} = 7$ when $m = 4$; and $D_{\mathsf{reg}} \leq 8$ when $m = 5$.

Let $m = 5$ and $n' = \lfloor n/m \rfloor$. Based on our experimental data, it is tempting to assume that the underlying system of polynomial equations has $D_{\mathsf{reg}} \leq 9$.

Table 1. Experimental results on solving PDP$(n, m, n' = \lfloor n/m \rfloor)$. Time in seconds; Memory in MB; D is the maximum step degree.

	$m = 3$			$m = 4$			$m = 5$		
n	Time	Memory	D	Time	Memory	D	Time	Memory	D
11							0.520	25.8	7
12							0.670	33.0	7
13							0.890	42.8	7
14							4.260	126.7	8
15							350.100	1839.5	8
16				414.320	5100.7	7	408.270	2633.9	8
17	1.690	38.8	4	1395.170	5632.8	7	506.340	4050.3	8
18	26.680	264.5	4	497.770	5632.8	7	920.790	6186.9	8
19	15.270	321.8	4	509.330	5634.1	7	1265.090	8282.9	8
20	49.350	397.6	4						
21	163.100	1228.3	4						
22	126.290	1413.2	4						
23	248.820	1668.7	4						
24	1266.610	5142.2	4						
25	1623.180	6363.8	4						
26	1645.78	6596.9	4						

Moreover, the system has $N = 5n' + n \approx 2n$ boolean variables. Therefore, when $m = 5$, we may estimate the cost of solving ECDLP$(2, n)$ (see (3)) as

$$\begin{aligned} & 2^{n'} \frac{2^n m!}{2^{mn'}} \binom{N + D_{\mathsf{reg}} - 1}{D_{\mathsf{reg}}}^w + 2^{w'n'} \\ & \approx 2^{n/5} m! (2n)^{9w} + 2^{w'n/5} \\ & \approx 2^{34} 2^{n/5} n^{27} + 2^{2n/5}, \end{aligned}$$

where we assume $w = 3$ and $w' = 2$. For example, when $n \approx 1200$, the cost of solving ECDLP$(2, n)$ is estimated to be 2^{550} which is significantly smaller than the cost 2^{600} of square-root time algorithms.

5 Extensions and Optimization

In Sect. 3, we introduced a single auxiliary variable to lower the degree of Semaev polynomials. The degree of polynomials can further be lowered by introducing more auxiliary variables. As an example, we consider the case $m = 5$. Let $R = (x_R, y_R) \in E$, as before. Notice that there exist $P_i \in \mathcal{B}$ such that

$$P_1 + P_2 + P_3 + P_4 + P_5 - R = \infty$$

if and only if there exist $P_i \in \mathcal{B}$ and $P_{12}, P_{34}, P_{50} \in E$ such that

$$\begin{cases} P_1 + P_2 - P_{12} = \infty \\ P_3 + P_4 - P_{34} = \infty \\ P_5 - P_{50} - R = \infty \\ P_{12} + P_{34} + P_{50} = \infty \end{cases} \tag{11}$$

Therefore, a 5-decomposition of $R = P_1 + P_2 + P_3 + P_4 + P_5$ may be found as follows:

1. Define the following system of equations derived from Semaev polynomials

$$\begin{cases} S_{3,1}(x_1, x_2, x_{12}) = 0 \\ S_{3,1}(x_3, x_4, x_{34}) = 0 \\ S_{3,2}(x_5, x_{50}, x_R) = 0 \\ S_{3,1}(x_{12}, x_{34}, x_{50}) = 0 \end{cases} \tag{12}$$

   Note that this system is defined over $\mathbb{F}_{2^n}$ and has 8 variables $x_1, x_2, x_3, x_4, x_5, x_{12}, x_{34}, x_{50}$.
2. Introduce boolean variables $x_{i,j}$ such that

$$x_i = \sum_{j=0}^{n'-1} x_{i,j}\sigma^j,$$

   for $i = 1, 2, 3, 4, 5$, and

$$x_{i,j} = \sum_{k=0}^{n} x_{i,j}\sigma^j,$$

   for $i = 12, 34, 50$. Apply the Weil descent technique to (12) and define an equivalent system of $4n$ equations over $\mathbb{F}_2$ with $5n' + 3n$ boolean variables

$$\{x_{i,j} : i = 1, 2, 3, 4, 5 \; j = 0, \ldots n'-1\} \cup \{x_{i,j} : i = 12, 34, 50, \; j = 0, \ldots n-1\}.$$

   Solve this new system of boolean equations and recover $x_1, x_2, x_3, x_4, x_5 \in \mathbb{F}_{2^n}$ from $x_{i,j} \in \mathbb{F}_2$.

Note that the proposed method solves a system of $4n$ equations over $\mathbb{F}_2$ with $5n' + 3n$ boolean variables rather than solving a system of n equations over $\mathbb{F}_2$ with $5n'$ boolean variables. Similar to the analysis in Sect. 3, we can show that $D_{\mathsf{FirstFall}} \leq 4$.

In Table 2, we report on our experimental results for solving PDP$(n, m, n' = \lfloor n/m \rfloor)$ with $m = 5$ deploying only the third Semaev polynomials; see (12). The time and memory results in the second and third column of Table 2 are obtained using the Groebner basis implementation of Magma with the grevlex ordering of monomials. We observe that the maximum step degree is $D_{\mathsf{reg}} = 4$ for $11 \leq n \leq 19$. The time and memory results in the last two columns of Table 2 are

Table 2. Experimental results on solving PDP$(n, m, n' = \lfloor n/m \rfloor)$. Time in seconds; Memory in MB; D is the maximum step degree; $D_{\mathsf{Heuristic}}$ is set to be 4 in Groebner basis computations.

	$D_{\mathsf{Heuristic}} = 4$				
	$m = 5$			$m = 5$	
n	Time	Memory	D	Time	Memory
11	2.380	58	4		
12	4.150	116.7	4		
13	6.390	124.1	4		
14	9.510	245.2	4		
15	393.170	6421.9	4	7.130	256.3
16	242.500	5911.7	4	6.900	320.4
17	365.460	7063.8	4	6.660	320.4
18	836.080	8619.4	4	11.700	394.6
19	531.420	8864.2	4	45.570	2505.3

obtained using the Groebner basis implementation of Magma with the grevlex ordering of monomials in a boolean ring. We also introduced two modifications in the computations: We set the ReductionHeuristic parameter in Magma to 4; and we first computed Groebner bases of partial systems described by single equations in (12), and merged them later. These two techniques yield non-trivial optimization both in time and memory. For a comparison, when $n = 15$ and $m = 3$, (Time, Memory) values decrease from $(393.170, 6421.9)$ to $(7.130, 256.3)$ when this modification is deployed in the computation; see Table 2. For the same parameters ($n = 15$ and $m = 3$), (Time, Memory) values are reported as $(174.47, 2635.4)$ in [12].

Based on our experimental data, it is tempting to assume that the underlying system of polynomial equations has $D_{\mathsf{reg}} \leq 4$ for all n. Moreover, the system has $N = 5n' + 3n \approx 4n$ boolean variables. Therefore, when $m = 5$, we may estimate the cost of solving ECDLP$(2, n)$ (see (3)) as

$$\begin{aligned} & 2^{n'} \frac{2^n m!}{2^{mn'}} \binom{N + D_{\mathsf{reg}} - 1}{D_{\mathsf{reg}}}^w + 2^{w'n'} \\ & \approx 2^{n/5} m! (4n)^{4w} + 2^{w'n/5} \\ & \approx 2^{31} 2^{n/5} n^{12} + 2^{2n/5}, \end{aligned}$$

where we assume $w = 3$ and $w' = 2$. This running time outperforms square-root methods when $n > 457$. For example, when $n \approx 550$, the cost of solving ECDLP$(2, n)$ is estimated to be 2^{250} which is significantly smaller than the cost 2^{275} of square-root time algorithms.

Acknowledgment. I would like to acknowledge two recent papers [9,12]. Semaev [12] claims a new complexity bound $2^{c(\sqrt{n \ln n})}$ for solving ECDLP$(2, n)$ under the assumption that the degree of regularity in Groebner computations of particular polynomial systems is $D_{\mathsf{reg}} \leq 4$. Semaev also shows that ECDLP$(2, n)$ can be solved in time $2^{o(c\sqrt{n \ln n})}$ under a weaker assumption that $D_{\mathsf{reg}} = o(\sqrt{n/\ln n})$ The techniques used in [12] and in this paper are similar. In [9], Kosters and Yeo provide experimental evidence that the degree of regularity of the underlying polynomial systems is likely to increase as a function of n, whence the conjecture $D_{\mathsf{reg}} \approx D_{\mathsf{FirstFall}}$ may be false.

I would like to thank Michiel Kosters and Igor Semaev for their comments on the first version of this paper.

References

1. Bosma, W., Cannon, J., Playoust, C.: The Magma algebra system I: the user language. J. Symbolic Comput. **24**, 235–265 (1997)
2. Diem, C.: On the discrete logarithm problem in elliptic curves II. Algebra Number Theory **7**, 1281–1323 (2013)
3. Faugère, J.-C.: A new efficient algorithm for computing Groebner bases (F4). J. Pure Appl. Algebra **139**, 61–68 (1999)
4. Faugère, J.-C.: A new efficient algorithm for computing Groebner bases without reduction to zero (F5). In: International Symposium on Symbolic and Algebraic Computation, pp. 75–83 (2002)
5. Faugère, J.-C., Perret, L., Petit, C., Renault, G.: Improving the complexity of index calculus algorithms in elliptic curves over binary fields. In: Pointcheval, D., Johansson, T. (eds.) EUROCRYPT 2012. LNCS, vol. 7237, pp. 27–44. Springer, Heidelberg (2012)
6. Galbraith, S.D., Gebregiyorgis, S.W.: Summation polynomial algorithms for elliptic curves in characteristic two. In: Meier, W., Mukhopadhyay, D. (eds.) Progress in Cryptology – INDOCRYPT 2014. LNCS, vol. 8885, pp. 409–427. Springer, Berlin (2014)
7. Gaudry, P.: Index calculus for abelian varieties of small dimension and the elliptic curve discrete logarithm problem. J. Symbolic Comput. **44**, 1690–1702 (2009)
8. Huang, Y.-J., Petit, C., Shinohara, N., Takagi, T.: Improvement of Faugère *et al.*'s Method to Solve ECDLP. In: Terada, M., Sakiyama, K. (eds.) IWSEC 2013. LNCS, vol. 8231, pp. 115–132. Springer, Heidelberg (2013)
9. Kosters, M., Yeo, S.: Notes on summation polynomials (2015). arXiv:1503.08001
10. Petit, C., Quisquater, J.-J.: On polynomial systems arising from a Weil descent. In: Wang, X., Sako, K. (eds.) ASIACRYPT 2012. LNCS, vol. 7658, pp. 451–466. Springer, Heidelberg (2012)
11. Semaev, I.: Summation polynomials and the discrete logarithm problem on elliptic curves, Cryptology ePrint Archive: Report/031 (2004)
12. Semaev, I.: New algorithm for the discrete logarithm problem on elliptic curves, Cryptology ePrint Archive: Report 2015/310 (2015)
13. Shantz, M., Teske, E.: Solving the elliptic curve discrete logarithm problem using Semaev polynomials, Weil descent and Gröbner basis methods – an experimental study. In: Fischlin, M., Katzenbeisser, S. (eds.) Buchmann Festschrift. LNCS, vol. 8260, pp. 94–107. Springer, Heidelberg (2013)