# POINT DECOMPOSITION PROBLEM IN BINARY ELLIPTIC CURVES

KORAY KARABINA

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 system of 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 is worth.

## 1. Introduction

*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 $\mathcal{B}$. *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, and only finitely many of them are non-identity. Note that each decomposition induces a modular linear dependence on the discrete 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 [10] showed that solving PDP in an elliptic curve group is equivalent to solving a particular system of multivariate non-linear system of 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 [6] showed that Semaev summation polynomials can be effectively used to solve ECDLP$(q, n)$ in heuristic time $\mathcal{O}(q^{2-\frac{2}{n}})$, where

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

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

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, index calculus type algorithms presented in [6, 2] 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, 9], and the complexity of the relation collection step have been analyzed. In both papers [5] and [9], 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 $\mathcal{O}(2^{wn/2})$, where $2.376 \leq w \leq 3$ is the linear algebra constant. The running time analysis in [5] considers the linearization technique to solve the multivariate nonlinear system of equations which are derived from the $(m + 1)$'st Seamev 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 $m = 2$. Petit and Quisquater's heuristic analysis in [9] 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 [9] 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 [9] that the degree of regularity $D_{\mathsf{reg}}$ 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 [9] also assumes that $n' = n^{\alpha}$ and $m = n^{1-\alpha}$ for some positive constant $\alpha$. 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 [9] in the favor of their assumption.

A recent paper by Shantz and Teske [12] presented some extended experimental results on solving PDP$(n, m, n')$ for the same setting as in the Petit and Quisquater's paper [9]. In particular, [12] validates the degree of regularity assumption in [9] 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 [12] were able to extend their experimental data for the parameters $(n, m, n', \Delta)$, $n \leq 48$, $m = 2$, and where $\Delta = n - mn'$ is chosen appropriately to possibly improve the running time of ECDLP$(2, n)$. In another recent paper [7], Yun-Ju 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 [7] 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 [9] 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 $(n, m, n', \Delta)$ with $\Delta = n - mn' \leq 1$ and $m = 3$ are limited to $n \leq 19$; see [12, 7]. Similarly, all the experiments for the set of parameters $(n, m, n', \Delta)$ 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 $\mathrm{ECDLP}(2,n)$ solvers in [5, 9]. Therefore, it remains as 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 [7, Section 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 this paper, we modify the system of equations, that are derived from Semaev polynomials, by introducing some auxiliary variables. We show that $\mathrm{PDP}(n,m,n')$ 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, $\mathrm{PDP}(n,m,n')$ 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 $\mathrm{PDP}(n,m,n')$ with $m \geq 4$. We should emphasize that choosing $m \geq 4$ is desirable for an index calculus based $\mathrm{ECDLP}(2,n)$ 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 is worth. In particular, we present some experimental results on solving $\mathrm{PDP}(n,m,n')$ 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 $n > 16$ and $m = 4$; and 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 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 Section 5. We are able to solve $\mathrm{PDP}(15,5,3)$ instances in about 7 seconds (with 256 MB memory). Note that, $\mathrm{PDP}(15,5,3)$ is solved in about 175 seconds (with 2635 MB memory) in [11]. Our experimental findings with $m = 3,4,5$ extend and improve on recently reported results in [12, 7, 11].

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

**Recent papers.** The author of this paper would like to acknowledge two recent papers [11, 8]. Seamev [11] claims a new complexity bound $2^{c(\sqrt{n \ln n})}$ for solving $\mathrm{ECDLP}(2,n)$ 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 $\mathrm{ECDLP}(2,n)$ can be solved in time $2^{o(\sqrt{n \ln n})}$ under a weaker assumption that $D_{\mathsf{reg}} = o(\sqrt{n/\ln n})$ The techniques used in [11] and in this paper are similar. In [8], 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.

## 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 $\infty$. The $i$'th Seamev polynomial associated with $E$ is defined as follows:

$$S_i(x_1, x_2, \ldots, x_i) = \begin{cases} (x_1x_2 + x_1x_3 + x_2x_3)^2 + x_1x_2x_3 + b & \text{if } i = 3 \\ \mathrm{Res}_X(S_{i-j}(x_1, \ldots, x_{i-j-1}, X), S_{j+2}(x_{i-j}, \ldots, x_i, X)) & \text{if } i \geq 4, \end{cases} \tag{2.1}$$

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

Let

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

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

$$P_1 + \cdots P_m = R, \tag{2.2}$$

for some given point $R = (x_R, y_R) \in E$. We refer to (2.2) 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 [6]). As described in Section 1, 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

$$2^{n'} \frac{2^n m!}{2^{mn'}} C_{n,m,n'} + 2^{w'n'}, \tag{2.3}$$

where $C_{n,m,n'}$ is the cost of solving PDP$(n, m, n')$, and $w' = 2$ is the sparse linear algebra constant. Seamev [10] showed that a decomposition of the form (2.2) 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 reduced by $S_{m+1}$.

## 3. A new approach to solve point decomposition problem

Let $E/\mathbb{F}_{2^n}$, $V$, and $\mathcal{B}$ be as defined in Section 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

$$S_{m+1}(x_1, \ldots, x_m, x_R) = 0 \tag{3.1}$$

Note that $x_i$ belong to an $n'$-dimensional subspace of $\mathbb{F}_{2^n}$. Therefore, (3.1) defines a system $\mathsf{Sys}_1$ of a single equation over $\mathbb{F}_{2^n}$ in $m$ variables. In [5, 9], the Weil descent technique is applied, and a second system $\mathsf{Sys}_2$ of $n$ equations over $\mathbb{F}_2$ in $mn'$ boolean variables is derived from $\mathsf{Sys}_1$. The cost $C_{n,m,n'}$ of solving PDP$(n, m, n')$ in [5, 9] is estimated through the analysis of solving $\mathsf{Sys}_2$ using linearization and Groebner basis techniques. Next, we describe a new approach to derive another system $\mathsf{Sys}_3$ of boolean equations such that a solution of $\mathsf{Sys}_3$ 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 $S_{m,1}(x_1, \dots, x_m)$, which represents the $m$'th Semaev polynomial with $m$ variables. The second class of Semaev polynomials is denoted by $S_{m,2}(x_1, \dots, x_{m-1}, x_R)$, 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

$$\begin{cases} P_1 + P_2 - P_{12} = \infty \\ P_3 + P_{12} - R = \infty \end{cases} \tag{3.2}$$

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

$$\begin{cases} S_{3,1}(x_1, x_2, x_{12}) = 0 \\ S_{3,2}(x_3, x_{12}, x_R) = 0. \end{cases} \tag{3.3}$$

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 (3.3) and define an equivalent system of $2n$ equations over $\mathbb{F}_2$ with $3n' + n$ boolean variables

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

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

$$\begin{cases} P_1 + P_2 - P_{12} = \infty \\ P_3 + P_4 + P_{12} - R = \infty \end{cases} \tag{3.4}$$

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

$$\begin{cases} S_{3,1}(x_1, x_2, x_{12}) = 0 \\ S_{4,2}(x_3, x_4, x_{12}, x_R) = 0 \end{cases} \tag{3.5}$$

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 (3.5) and define an equivalent system of $2n$ equations over $\mathbb{F}_2$ with $4n' + n$ boolean variables

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

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

$$\begin{cases} P_1 + P_2 + P_3 - P_{123} = \infty \\ P_4 + P_5 + P_{123} - R = \infty \end{cases} \tag{3.6}$$

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

$$\begin{cases} S_{4,1}(x_1, x_2, x_3, x_{123}) = 0 \\ S_{4,2}(x_4, x_5, x_{123}, x_R) = 0 \end{cases} \tag{3.7}$$

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 (3.7) and define an equivalent system of $2n$ equations over $\mathbb{F}_2$ with $5n' + n$ boolean variables

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

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 new polynomial systems.** One of the methods to solve a multivariate nonlinear system of 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 typically estimated as $O\left(\binom{N+D-1}{D}^w\right)$, where $\binom{N+D-1}{D}$ is the maximum number of columns in $M_D$ and $w$ is the linear algebra constant. In general, it is hard to estimate $D$. In the recent paper [9], it is conjectured that the degree of regularity $D_{\mathsf{reg}}$ of systems arising from $\mathrm{PDP}(n, m, n')$ 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 3.1.** [9] *Let $R$ be a polynomial ring over a field $K$. Let $F := \{f_1, \dots, f_\ell\} \subset R$ 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 [9, 12] give supporting evidence that $D_{\mathsf{reg}} \approx D_{\mathsf{FirstFall}}$. However, experimental data is yet very limited (see Section 1) to verify this conjecture. In this section, we compute the first fall degree of the systems proposed in Section 3.1, Section 3.2, and Section 3.3. Our experimental results in Section 4 indicate 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 (3.3) 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

$$\{1, x_1^2x_2^2, x_1^2x_{12}^2, x_2^2x_{12}^2, x_1x_2x_{12}\}.$$

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

$$\{x_1, x_1^3x_2^2, x_1^3x_{12}^2, x_2^2x_{12}^2, x_1^2x_2x_{12}\}.$$

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

$$\{1, x_3^2x_{12}^2, x_3^2, x_{12}^2, x_3x_{12}\}.$$

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

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

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 (3.5) 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 (3.7) 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 proposed methods in Section 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 $D$ (for which ) in the Groebner basis computations. Recall that in Section 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 $D_{\mathsf{reg}} \approx D_{\mathsf{FirstFall}}$.

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 $D_{\mathsf{reg}} \approx 9$. 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 (2.3)) 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.

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 | $D$ |
| n | | | | | | | 0.520 | 25.8 | 7 |
| n | | | | | | | 0.670 | 33.0 | 7 |
| n | | | | | | | 0.890 | 42.8 | 7 |
| n | | | | | | | 4.260 | 126.7 | 8 |
| n | | | | | | | 350.100 | 1839.5 | 8 |
| n | | | | 414.320 | 5100.7 | 7 | 408.270 | 2633.9 | 8 |
| n | 1.690 | 38.8 | 4 | 1395.170 | 5632.8 | 7 | 506.340 | 4050.3 | 8 |
| n | 26.680 | 264.5 | 4 | 497.770 | 5632.8 | 7 | 920.790 | 6186.9 | 8 |
| n | 15.270 | 321.8 | 4 | 509.330 | 5634.1 | 7 | 1265.090 | 8282.9 | 8 |
| n | 49.350 | 397.6 | 4 | | | | | | |
| n | 163.100 | 1228.3 | 4 | | | | | | |
| n | 126.290 | 1413.2 | 4 | | | | | | |
| n | 248.820 | 1668.7 | 4 | | | | | | |
| n | 1266.610 | 5142.2 | 4 | | | | | | |
| n | 1623.180 | 6363.8 | 4 | | | | | | |
| n | 1645.78 | 6596.9 | 4 | | | | | | |

## 5. Extensions and Optimization

In Section 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

$$\begin{cases} P_1 + P_2 - P_{12} = \infty \\ P_3 + P_4 - P_{34} = \infty \\ P_5 - P_{50} - R = \infty \\ P_{12} + P_{34} + P_{50} = \infty \end{cases} \tag{5.1}$$

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

$$\begin{cases} S_{3,1}(x_1, x_2, x_{12}) = 0 \\ S_{3,1}(x_3, x_4, x_{34}) = 0 \\ S_{3,2}(x_5, x_{50}, x_R) = 0 \\ S_{3,1}(x_{12}, x_{34}, x_{50}) = 0 \end{cases} \tag{5.2}$$

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

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

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 |

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 (5.2) 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 Section 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 (5.2). 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 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 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 (5.2), 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 [11].

Based on our experimental data, we may assume that the underlying system of polynomial equations has $D_{\mathsf{reg}} \approx 4$ 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

(2.3)) 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 $\mathrm{ECDLP}(2, n)$ 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 thank Michiel Kosters and Igor Semaev for their comments on the first version of this paper.

## References

1. W. Bosma, J. Cannon, and C. Playoust, *The magma algebra system. i. the user language*, J. Symbolic Comput. **24** (1997), 235–265.
2. C. Diem, *On the discrete logarithm problem in elliptic curves ii*, Algebra and Number Theory **7** (2013), 1281–1323.
3. J.-C. Faugère, *A new efficient algorithm for computing Groebner bases (f4)*, Journal of Pure and Applied Algebra **139** (1999), 61–68.
4. ———, *A new efficient algorithm for computing Groebner bases without reduction to zero (f5)*, International Symposium on Symbolic and Algebraic Computation (2002), 75–83.
5. J.-C. Faugère, L. Perret, C. Petit, and G. Renault, *Improving the complexity of index calculus algorithms in elliptic curves over binary fields*, Advances in Cryptology – EUROCRYPT 2012, Lecture Notes in Computer Science **7237** (2012), 27–44.
6. P. Gaudry, *Index calculus for abelian varieties of small dimension and the elliptic curve discrete logarithm problem*, Journal of Symbolic Computation **44** (2009), 1690–1702.
7. Y.-J. Huang, C. Petit, N. Shinohara, and T. Takagi, *Improvement of Faugère et al.s method to solve ECDLP*, Advances in Information and Computer Security, Lecture Notes in Computer Science **8231** (2013), 115–132.
8. M. Kosters and S. Yeo, *Notes on summation polynomials*, (2015), arXiv:1503.08001.
9. C. Petit and J.-J. Quisquater, *On polynomial systems arising from a Weil descent*, Advances in Cryptology – ASIACRYPT 2012, Lecture Notes In Computer Science **7658** (2012), 451–466.
10. I. Semaev, *Summation polynomials and the discrete logarithm problem on elliptic curves*, Cryptology ePrint Archive: Report 2004/031, 2004.
11. ———, *New algorithm for the discrete logarithm problem on elliptic curves*, (2015), Cryptology ePrint Archive: Report 2015/310.
12. M. Shantz and E. Teske, *Solving the elliptic curve discrete logarithm problem using Semaev polynomials, Weil descent and Groebner basis methods – an experimental study*, Number Theory and Cryptography. Papers in Honor of Johannes Buchmann on the Occasion of His 60th Birthday, Lecture Notes In Computer Science **8260** (2013), 94–107.

Dept. of Mathematical Sciences, Florida Atlantic University, Boca Raton, Florida, US

*E-mail address*: kkarabina@fau.edu