On index calculus algorithms for subfield curves

Galbraith, Steven D.; Granger, Robert; Merz, Simon-Philipp; Petit, Christophe

DOI:
10.1007/978-3-030-81652-0_5

Document Version
Peer reviewed version

Citation for published version (Harvard):
Galbraith, SD, Granger, R, Merz, S-P & Petit, C 2021, On index calculus algorithms for subfield curves. in O Dunkelman, MJ Jacobson, Jr. & C O'Flynn (eds), *Selected Areas in Cryptography - 27th International Conference, 2020, Revised Selected Papers.* Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 12804, Springer, pp. 115-138, 27th International Conference on Selected Areas in Cryptography, SAC 2020, Virtual, Online, 21/10/20. https://doi.org/10.1007/978-3-030-81652-0_5

On Index Calculus Algorithms for Subfield Curves

Steven D. Galbraith[1], Robert Granger[2], Simon-Philipp Merz[3], and Christophe Petit[4,5]

[1] Mathematics Department, University of Auckland, New Zealand
s.galbraith@auckland.ac.nz
[2] Surrey Centre for Cyber Security,
Department of Computer Science, University of Surrey, UK
r.granger@surrey.ac.uk
[3] Information Security Group, Royal Holloway, University of London, UK
simon-philipp.merz.2018@rhul.ac.uk
[4] Département d'informatique, Université libre de Bruxelles, Belgium
[5] School of Computer Science, University of Birmingham, UK
christophe.f.petit@gmail.com

Abstract. In this paper we further the study of index calculus methods for solving the elliptic curve discrete logarithm problem (ECDLP). We focus on the index calculus for subfield curves, also called Koblitz curves, defined over $\mathbb{F}_q$ with ECDLP in $\mathbb{F}_{q^n}$. Instead of accelerating the solution of polynomial systems during index calculus as was predominantly done in previous work, we define factor bases that are invariant under the q-power Frobenius automorphism of the field $\mathbb{F}_{q^n}$, reducing the number of polynomial systems that need to be solved. A reduction by a factor of $1/n$ is the best one could hope for. We show how to choose factor bases to achieve this, while simultaneously accelerating the linear algebra step of the index calculus method for Koblitz curves by a factor n^2. Furthermore, we show how to use the Frobenius endomorphism to improve symmetry breaking for Koblitz curves. We provide constructions of factor bases with the desired properties, and we study their impact on the polynomial system solving costs experimentally.
This work gives an answer to the problem raised in the literature on how the Frobenius endomorphism can be used to speed-up index calculus on subfield curves.

1 Introduction

Elliptic curve cryptography (ECC) is a classical approach to public-key cryptography based on the algebraic structure of elliptic curves over finite fields. The use of elliptic curves in cryptography was suggested independently by Koblitz and Miller in 1985 [17, 19]. During the last decades, ECC has become increasingly important because shorter keys yield the same security as in cryptographic schemes based on discrete logarithms in plain Galois fields or on factorisation problems. This has allowed to reduce storage and transmission requirements for various cryptographic applications. Today, ECC is ubiquitous and can for example be found in SSL/TLS which secures the majority of connections in the World Wide Web (see e.g. [24]).

Let $\mathbb{F}_{q^n}$ denote the finite field of cardinality q^n, where q is the power of a prime. An elliptic curve is a non-singular plane algebraic curve satisfying an equation of the form

$$E : y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6, \quad a_i \in \mathbb{F}_{q^n}$$

and a point at infinity denoted $\mathcal{O}_E$. The set of points $(x, y) \in \mathbb{F}_{q^n}^2$ on an elliptic curve is an abelian group under the "chord and tangent rule", with $\mathcal{O}_E$ being the identity element. For non-negative integers a we define the multiplication by a map as $[a] : E \to E,\ P \mapsto P + P + \cdots + P$ (a times).

The elliptic curve discrete logarithm (ECDLP) is the following computational problem. Given points $P, Q \in E(\mathbb{F}_{q^n})$, find an integer a, if it exists, such that $Q = [a]P$. Many ECC protocols assume that the ECDLP is computationally infeasible for the curves used.

Like any other discrete logarithm problem, ECDLP can be solved using generic algorithms such as Baby-step-Giant-step, Pollard's ρ and their variants [23]. These algorithms can be parallelised efficiently, but have exponential runtime complexity, i.e. roughly square root of the size r of the cyclic subgroup $\langle P \rangle \subset E(\mathbb{F}_{q^n})$. More precisely, Van Oorschot and Wiener showed how to get a heuristic expected running time of $(\sqrt{\pi/2} + o(1))\sqrt{r}$ group operations using "distinguished points" [28]. Therefore, the difficulty of the ECDLP depends, among other things, on the size of r. Moreover, r should be prime to prevent Pohlig-Hellman attacks [22]. To ensure that $\#E(\mathbb{F}_{q^n})$ has a large subgroup of prime order, n is usually chosen to be either 1 or a prime as $\#E(\mathbb{F}_q)$ divides $\#E(\mathbb{F}_{q^n})$.

Index calculus is another approach to solve a discrete logarithm problem by reducing it to a linear algebra problem. Given an elliptic curve, one defines a subset of the curve named the *factor base* and tries to express points of the form $[a_i]P + [b_i]Q$ as a sum of factor base elements. After collecting sufficiently many linearly independent of these so-called *relations*, one can compute the discrete logarithm by solving a system of linear equations.

A *subfield curve*, or *Koblitz curve*, is an elliptic curve defined over a small finite field $\mathbb{F}_q$ which is considered over a large extension field. Put differently, a Koblitz curve is defined using coefficients from $\mathbb{F}_q$ with ECDLP in $E(\mathbb{F}_{q^n})$. Koblitz curves were used in practice because the q-th power Frobenius endomorphism $\pi : E \to E$, $(x, y) \mapsto (x^q, y^q)$ can be used to devise fast point multiplication algorithms via Frobenius expansion [27]. While half of the elliptic curves that were standardised by NIST in the current NIST SP 800-186 draft are Koblitz curves defined over $\mathbb{F}_2$, it is highlighted there that they are now deprecated. It was shown that ECDLP in a Koblitz curve $E(\mathbb{F}_{q^n})$ can be solved at $1/\sqrt{n}$ of the cost compared to a general elliptic curve [30].

The slowest part of index calculus on elliptic curves is the relation collection. Accelerating this step of the algorithm by exploiting the additional structure provided by Koblitz curves compared to general elliptic curves was mentioned as an open problem in the literature [9]. An interesting approach to this question was considered by Gorla and Massierer [12], but it did not lead to a dramatic speed-up. Their idea was to represent the ECDLP instance as a trace zero variety on which index calculus is performed directly. For curves of genus $g > 1$, Gaudry gives one example of a Frobenius invariant factor base and describes how this yields a speed-up [10]. We answer the open question by emulating this idea to the harder case where $g = 1$.

Our contribution: We describe how the relation collection can be sped up when using factor bases that are carefully chosen with respect to the Frobenius endomorphism acting on Koblitz curves. First, we show how this allows to improve symmetry breaking. Then we focus on factor bases that are closed under the Frobenius endomorphism. We show that we can reduce such a factor base easily to a smaller one generating the same number of relations. As a consequence, after finding one relation in the larger factor base, one can rewrite them in terms of a reduced factor base consisting of representatives for each Frobenius orbit in the larger factor base. Under certain conditions it is then sufficient to collect $1/n$ as many relations leading to a speed-up of n for the relation collection. Moreover, the dimension of the sparse matrix in the linear algebra step decreases by a factor of $1/n^2$ reducing the cost of the linear algebra step during index calculus by $1/n^2$.

We provide concrete constructions for factor bases with the necessary properties for some classes of elliptic curves and examine experimentally how some of the choices influence the complexity of computations during index calculus.

While Pollard's ρ algorithm remains the fastest method to solve ECDLP instances used in cryptography in practice, this work shows that the performance gap between index calculus and Pollard's ρ is smaller for Koblitz curves. More precisely, the ECDLP in a Koblitz curve $E(\mathbb{F}_{q^n})$ can be solved faster than for general elliptic curves by a factor of $\sqrt{n}$ using Pollard's ρ [30] compared to our speed-up by roughly n.

The paper is organised as follows. Section 2 recalls the index calculus framework, provides details on the current state-of-the-art index calculus methods for ECDLP and references related work. In Section 3 we present our improvements for Koblitz curves. This is achieved by choosing factor bases satisfying certain properties with respect to the Frobenius endomorphism. To make these results more concrete, we provide constructions for such factor bases in Section 4. In Section 5, we display our experimental results before concluding the paper in Section 6.

2 Index calculus

Apart from generic algorithms such as Pollard's ρ, baby-step-giant-step or kangaroo algorithms, one approach to solve discrete logarithm problems (in any cyclic group $\langle P \rangle$) is index calculus. This method tries to reduce the discrete logarithm problem to linear algebra. In this section we recall the index calculus framework to solve discrete logarithms and we recollect how it is applied to elliptic curves in practice.

2.1 Framework of index calculus

The basic framework of index calculus is as follows:

1. Define a subset $\mathcal{F}$ of $\langle P \rangle \subset E$, called the factor base.
2. Collect relations:
   (a) Pick random integers a_j and b_j and compute $R = [a_j]P + [b_j]Q$.
   (b) Try to decompose R as a sum of elements of $\mathcal{F}$.
   (c) If the decomposition is successful, call $[a_j]P + [b_j]Q = \sum_{P_i \in \mathcal{F}} [e_{ij}]P_i$ a relation and store both the vector (e_{ij}) as a row of a matrix and the integers (a_j, b_j).
   (d) Repeat the collection of relations until there are $|\mathcal{F}|$ linearly independent ones.
3. Use linear algebra modulo r to compute a non-zero column vector $(\gamma_1, \ldots, \gamma_{|\mathcal{F}|})^T$ in the right kernel of the matrix (e_{ij}) with $1 \leq i, j \leq |\mathcal{F}|$.
4. Compute the discrete logarithm of Q as $-(\sum_{j=1}^{|\mathcal{F}|} a_j\gamma_j)(\sum_{j=1}^{|\mathcal{F}|} b_j\gamma_j)^{-1} \mod r$, if $\sum_{j=1}^{|\mathcal{F}|} b_j\gamma_j$ is invertible modulo r, otherwise return to Step 2.

The efficiency of the index calculus approach depends on the choice of the factor base $\mathcal{F}$. While it should be possible to write a large proportion of group elements as a sum of elements in $\mathcal{F}$ (to prevent step 2(b) from failing to often), the set $\mathcal{F}$ should not be too large, as we need to collect $\#\mathcal{F}$ relations. Moreover, the decomposition of group elements into a sum of elements in $\mathcal{F}$ should be efficient if it exists.

To tackle these problems, index calculus in elliptic curves requires two crucial ingredients: Semaev's summation polynomials and the Weil restriction of scalars.

One can use the formulae of the group law to decompose a point R into a sum of points of the factor base, $R = P_1 + \cdots + P_k$, if $\mathcal{F}$ has a nice algebraic description. To compute this in practice, Semaev's summation polynomials are used. Our improvements in this paper concern subfield curves and most such curves used in practice are defined over a field of characteristic 2. For simplicity, we therefore recall the results due to Semaev [25] only for this case. However,

we want to emphasise that all our improvements presented later in the paper apply to general subfield curves defined over fields of size q.

A subfield curve E with solutions in $\mathbb{F}_{2^n}$ defined over $\mathbb{F}_2$ is specified by an equation

$$E: y^2 + xy = x^3 + ax^2 + 1, \text{ where } a \in \{0, 1\}. \tag{1}$$

Semaev's summation polynomials $\{S_m \in \mathbb{F}_{q^n}[x_1, \ldots, x_m]\}_{m \in \mathbb{N}}$ have the defining property that there is a root at $(X_1, \ldots, X_m) \in \overline{\mathbb{F}}_{2^n}^m$, i.e. $S_m(X_1, \ldots, X_m) = 0$, if and only if there exist $(Y_1, \ldots, Y_m) \in \overline{\mathbb{F}}_{2^n}^m$ such that $(X_i, Y_i) \in E(\overline{\mathbb{F}}_{2^n})$ for all $1 \leq i \leq m$ and $(X_1, Y_1) + (X_2, Y_2) + \cdots + (X_m, Y_m) = 0$ on the curve.

Theorem 2.1. *[25] The summation polynomials of E given by Equation (1) are recursively defined by*

$$S_2(x_1, x_2) := x_1 + x_2,$$
$$S_3(x_1, x_2, x_3) := (x_1x_2 + x_1x_3 + +x_2x_3)^2 + x_1x_2x_3 + 1,$$

and for $m \geq 4$ and any k, $1 \leq k \leq m-3$, the m-th summation polynomial is

$$S_m(x_1, \ldots, x_m) := \mathsf{Res}_X\left(S_{m-k}(x_1, \ldots, x_{m-k-1}, X),\ S_{k+2}(x_{m-k}, \ldots, x_m, X)\right)$$

where Res_X denotes the resultant with respect to X. For $m \geq 2$ the polynomial S_m is symmetric and has degree 2^{m-2} in each variable x_i.

For general elliptic curves defined over fields of even and odd characteristic such formulas exist as well but we omit the details here. For a general formula see Lemma 3.4 of [3].

As a result of trying to decompose points as a sum of factor base elements using Semaev polynomials, we will obtain a system of polynomial equations defined over $\mathbb{F}_{q^n}$. Using Weil restriction of scalars, this can be converted into equations over $\mathbb{F}_q$. The basic idea hereby is to rewrite a polynomial equation over an extension field $\mathbb{F}_{q^n}$ as n polynomial equations over $\mathbb{F}_q$.

Lemma 2.2. *Let q be a prime power, $n \geq 1$ an integer and fix a vector space basis $\{\theta_1, \ldots, \theta_n\}$ for $\mathbb{F}_{q^n}$ over $\mathbb{F}_q$. Let $f \in \mathbb{F}_{q^n}[x_1, \ldots, x_m]$. There exist unique polynomials $f_k \in \mathbb{F}_q[y_{i,j}]$, $1 \leq i \leq m, 1 \leq j \leq n$, for $1 \leq k \leq n$ such that*

$$f(y_{1,1}\theta_1 + \cdots + y_{1,n}\theta_n, \ldots, y_{m,1}\theta_1 + \cdots + y_{m,m}\theta_n) = \sum_{k=1}^{n} \theta_k f_k(y_{i,j}).$$

If $f(x_1, \ldots, x_m) = 0$ for some $x_1, \ldots, x_m \in \mathbb{F}_{q^n}$ then there exist $y_{i,j} \in \mathbb{F}_q$ such that $x_i = \sum_{j=1}^{\ell} y_{i,j}\theta_j$ and $f_k(y_{i,j}) = 0$ for all $1 \leq k \leq n$.

Now, we are ready to describe the index calculus for elliptic curves in more detail.

2.2 Index calculus for elliptic curves

Semaev was the first one to sketch a framework for index calculus on elliptic curves using the summation polynomials and Weil descent [25], which was fully developed by Gaudry [11] and Diem [3] later.

The approach works as follows: choose an $\mathbb{F}_q$-vector subspace $V = \langle v_1, \ldots, v_{n'} \rangle \subset \mathbb{F}_{q^n}$ of dimension $1 \leq n' \leq n$ and define the factor base to be

$$\mathcal{F} = \{P \in E(\mathbb{F}_{q^n}) : x(P) \in V\}, \tag{2}$$

where $x(P)$ refers to the x-coordinate of a point P.

To collect relations, choose random integers a, b modulo the order of $\langle P \rangle$ and compute the point $R = aP + bQ$ in $\langle P \rangle$. To decompose such a point R as a sum over the factor base, i.e. $R = P_1 + \cdots + P_m$ with $P_i \in \mathcal{F}$, one tries to find roots of the $(m+1)$-th summation polynomials $S_{m+1}(x_1, \ldots, x_m, x(R)) \in \mathbb{F}_{q^n}[x_1, \ldots, x_{m+1}]$ with $x_i \in V$.

To make sure that the P_i in the decomposition lie in the factor base $\mathcal{F}$, one rewrites the summation polynomial using the linear constraints $x_i = v_1 y_{i,1} + v_2 y_{i,2} + \cdots + v_{n'} y_{i,n'}$. This yields a new polynomial in $\mathbb{F}_{q^n}[y_{1,1}, \ldots, y_{m,n'}]$, i.e. in $m \cdot n'$ variables, with $y_{i,j} \in \mathbb{F}_q$. More precisely, we can look at the polynomial as an element f in $\mathbb{F}_{q^n}[y_{1,1}, \ldots, y_{m,n'}]/\langle y_{i,j}^q - y_{i,j} \rangle$, where $\langle y_{i,j}^q - y_{i,j} \rangle$ for $1 \leq i \leq m$ and $1 \leq j \leq n'$ denotes the ideal generated by the field equations.

In order to find solutions to this polynomial f, previous work deemed it most efficient to look at it as a system of polynomials over $\mathbb{F}_q$ and use Gröbner basis methods [14]. That is to apply Weil restriction of scalars to get a system of n polynomials in $\mathbb{F}_q[y_{i,j}]$, $1 \leq i \leq m, 1 \leq j \leq n$. Namely, let $\{\theta_1, \ldots, \theta_n\}$ be a basis for $\mathbb{F}_{q^n}$ as $\mathbb{F}_q$ vector space. By Lemma 2.2, we can decompose f as

$$f = f_1\theta_1 + f_2\theta_2 + \cdots + f_n\theta_n$$

for some $f_i \in \mathbb{F}_q[y_{1,1}, \ldots, y_{m,n'}]$. Due to the linear independence of $\{\theta_1, \ldots, \theta_n\}$, finding a solution to f is equivalent to solving the polynomial system of n equations

$$f_1 = f_2 = \cdots = f_n = 0$$

in mn' variables. According to [6, 16], the best way to solve this system is to compute a Gröbner basis with respect to the graded reverse lexicographical monomial order using Faugère's F4 or F5 algorithm [4, 5] and apply the FGLM [7] algorithm to transform the basis into a Gröbner basis with respect to the lexicographical order.

Remark 2.3. Choosing n' and m is an important decision. In general, one chooses $n'm \approx n$ in order to have as many equations as indeterminates in the polynomial system. This is the natural choice in the sense that it is the smallest value for $n'm$ where one expects to get a solution. In the following, we will assume that $n'm \approx n$.

For a discussion of over- and under-determined systems, we refer to Galbraith and Gaudry [9].

We want to give an estimation of the two most expensive steps of index calculus, the relation collection and the linear algebra steps. To analyse the cost of the relation collection, let $H_1(n, n', m)$ denote the cost of solving one polynomial system with n equations and mn' variables of the above form.

Theorem 2.4. *Let $n = n' \cdot m + k$, $k \geq 0$. Under the heuristic assumption that a factor base defined by Equation (2) has size $\#V = q^{n'}$, the cost of the relation collection step during index calculus is approximately*

$$\frac{q^{n'+k}}{2^m} \cdot m! \cdot H_1(n, n', m).$$

Proof. First, we determine the probability that a random point R can be written as a sum of m points from $\mathcal{F}$. Since the group operation on elliptic curves is commutative, if one can write $R = P_1 + \cdots + P_m$ then there are, in general, $m!$ such decompositions. Therefore, we can estimate the number of points that can be represented as sums of m points of our factor base as $\frac{q^{n'm}}{m!}$. Since E has roughly q^n points, we can estimate the probability of a relation naively as $\frac{q^{n'm}}{q^n m!} = (q^k \cdot m!)^{-1}$.

When solving one polynomial system during the relation collection step, we solve for the x-coordinates of points on the curve. Since the factor base $\mathcal{F}$ has some symmetry by negation, there are 2^m choices for the signs for points corresponding to the same m x-coordinates in the polynomial system (2^{m-1} of which are linearly independent). Similarly, we can restrict the computation to half of the targeted points due to negation. Therefore, the probability to find a relation when solving one polynomial system is $\frac{2^m \cdot q^{n'm}}{q^n m!} = \frac{2^m}{q^k \cdot m!}$. Intuitively, solving one polynomial system with respect to the x-coordinates allows to check for 2^m relations of points on the elliptic curve simultaneously.

Hence, we expect to solve $\frac{q^k}{2^m} \cdot m!$ polynomial systems to find a single relation. Assuming roughly the same cost for the solution of all the polynomial systems, $H_1(n, n', m)$, the collection of $q^{n'}$ relations costs $\frac{q^{n'+k}}{2^m} \cdot m! \cdot H_1(n, n', m)$. □

Note that the preceding proof assumes the same cost for solving polynomial systems with and without solutions. This is a simplification and not true in general, e.g. for systems without a solution we might already reach a contradiction earlier in the computation.

To estimate the cost of the linear algebra step, we note that the number of non-zero coefficients for all the points in the factor base per relation is very low, namely at most m. Therefore, it is possible to compute the linear algebra step using sparse linear algebra techniques. Algorithms such as the Wiedemann algorithm [29] allow to find solutions to a matrix of dimension N^2 containing m entries per row in $\mathcal{O}(mN^2)$ multiplications. Let $|\mathcal{F}|$ denote the size of the factor base. As we aim to collect roughly $|\mathcal{F}|$ relations, we get the following theorem.

Theorem 2.5. *The cost of the linear algebra step is* $\mathcal{O}(m \cdot |\mathcal{F}|^2)$.

Under the heuristic assumption that a factor base as defined by Equation (2) has size approximately $\#V = q^{n'}$, the cost of the linear algebra step is $\mathcal{O}(m \cdot q^{2n'})$.

In practice, there are various tricks to lower the degree of the polynomial system we need to solve during the decomposition step in the relation collection. For a summary of different approaches we refer to [9]. Amongst ideas like breaking the symmetry, which we will address in the next subsection, there are approaches to introduce additional variables to lower the degree of the polynomial systems. These are sometimes referred to as "the splitting trick" or "unrolling the resultant" [13, 14, 26].

2.3 Breaking symmetries

As addition on elliptic curves is commutative, the symmetric group acts on solutions $(P_1, \ldots, P_m)$ to the point decomposition problem in the relation search. This leads to an inconvenient $m!$ factor in the complexity statements. To "break symmetry" usually refers to removing this redundancy. One approach is to rewrite the summation polynomial in terms of generators of the ring of invariants under the action of the symmetric group, i.e. the elementary symmetric polynomials, as was done in [6]. With respect to these new variables, the polynomial system arising from the summation polynomial has usually $m!$ fewer solutions and the degrees of the polynomials are potentially lower. Given a solution to this system, one needs to recompute a solution in the original variables. However, this only needs to be done for the successful systems and is relatively fast.

Another idea to mitigate the factor of $m!$ in the success probability during the point decomposition is attributed in [20] to Matsuo. He suggested the use of m disjoint factor bases $\mathcal{F}_i$ of size $q^{n'}$ and forcing P_i to be in different factor bases $\mathcal{F}_i$. While this approach allows to avoid the

factor $m!$ in the probability to decompose a point over our factor base (see Theorem 2.4), the factor base is then a union of m sets of size $q^{n'}$ and thus we need to collect m times as many relations. Consequently, the relation collection step is only accelerated by a factor of $(m-1)!$.

Moreover, the cost of the linear algebra step (see Theorem 2.5) increases by a factor m^2, although the linear system is a block diagonal matrix and some optimisation may be possible.

3 Index calculus for Koblitz curves

In this section we present our main improvements for index calculus methods on Koblitz curves. We start by presenting a new approach for symmetry breaking in Koblitz curves. Then, we show how to exploit Frobenius invariant factor bases to speed up the decomposition of points into factor base elements. We conclude with a brief comparison of performance of the different methods.

3.1 Improved symmetry breaking for Koblitz curves

We present a novel way of symmetry breaking for Koblitz curves that builds on top of Matsuo's idea described in Section 2.3, but allows for a full saving of $m!$ in the relation collection step and does not have an increased linear algebra cost by a factor of m^2.

As we mention in the introduction, elliptic curves that are interesting in cryptography have one large cyclic subgroup $\langle P \rangle$ of prime order r containing the ECDLP.

Lemma 3.1. *Let E be a Koblitz curve defined over $\mathbb{F}_q$ with one large cyclic subgroup $\langle P \rangle$ of prime order r and let $\pi : E(\mathbb{F}_{q^n}) \to E(\mathbb{F}_{q^n})$, $(x, y) \mapsto (x^q, y^q)$ denote the q-th power Frobenius endomorphism. Then there exists some $\lambda \in \mathbb{Z}$, $1 \leq \lambda \leq r-1$ such that $\pi(Q) = [\lambda]Q$ for all $Q \in \langle P \rangle$.*

Proof. The Frobenius endomorphism preserves the order of points on the curve. Therefore, $\pi(P)$ is another point on the Koblitz curve of order r. Since there is only one large cyclic subgroup of order r, it follows that $\pi(P) \in \langle P \rangle$ and there exists some scalar $1 \leq \lambda \leq r-1$ such that $\pi(P) = [\lambda]P$. As scalar multiplication commutes with the Frobenius map, we have $\pi(Q) = [\lambda]Q$ for all $Q \in \langle P \rangle$. □

It is known that for any point $P \in E$, the Frobenius endomorphism π satisfies

$$\pi^2(P) \pm [t]\pi(P) + [q]P = \mathcal{O}_E,$$

where t is the trace of the curve E, i.e. the integer satisfying $|E(\mathbb{F}_{q^n})| = q^n + 1 - t$. Therefore, it can be shown that the value λ of the preceding lemma is one of the roots of the quadratic congruence

$$X^2 \pm tX + q \equiv 0 \pmod{r},$$

which makes it efficiently computable.

The key of our method is to choose the vector space defining the factor bases in a specific way. As before, let V be an $\mathbb{F}_q$ vector subspace of $\mathbb{F}_{q^n}$ of dimension n' and let $\mathcal{F} = \{P \in E(\mathbb{F}_{q^n}) : x(P) \in V\}$.

Let V_i be vector spaces that give rise to m pairwise "disjoint" factor bases with $\mathcal{F}_1 = \mathcal{F}$, $\mathcal{F}_2 = \pi(\mathcal{F}), \ldots, \mathcal{F}_m = \pi^{m-1}(\mathcal{F})$. Given a normal basis of $\mathbb{F}_{q^n}$ over $\mathbb{F}_q$, it is easy to construct such V_i. Namely, let $\dim V = n'$ be the desired dimension of V with $n'm \leq n$ and let β be a normal basis element of $\mathbb{F}_{q^n}$ over $\mathbb{F}_q$. We can take the vector space $V_i := \langle \beta^{q^{mj+i}} | j \in \{0 \ldots, n'-1\} \rangle$ to define the factor base $\mathcal{F}_i$ for $i = 1, \ldots, m$.

Given such factor bases one can break symmetry as follows:

1. Decompose points as sums of the form

$$R = P_1 + \cdots + P_m, \quad \text{where} \quad P_i \in \mathcal{F}_i. \tag{3}$$

2. Rewrite the relation as

$$R = P_1' + \lambda P_2' + \cdots + \lambda^{m-1} P_m' \quad \text{where all } P_i' \in \mathcal{F}_1 = \mathcal{F}.$$

The restriction $P_i \in \mathcal{F}_i$ is equivalent to the constraints $x(P_i) \in V_i$ that become linear in the Weil restriction. The first part is therefore just like Matsuo's idea.

The second step is possible as a consequence of Lemma 3.1 and the relation between the different factor bases $\mathcal{F}_i$: for any $P_i \in \mathcal{F}_i$, $i = 1, \ldots, m$, we can find some $P_i' \in \mathcal{F}_1$ such that $P_i = \pi^{(i-1)}(P_i') = \lambda^{(i-1)}(P_i')$.

The analysis given in the proof of Theorem 2.4 applies still except for saving the full $m!$ term in the probability of finding a relation. Let $H_2(n, n', m)$ denote the cost of solving the polynomial systems during point decomposition for the described factor bases. We arrive at the following theorem.

Theorem 3.2. *Let $n = n' \cdot m + k$ and let E be a Koblitz curve with coefficients in $\mathbb{F}_q$ and ECDLP in $E(\mathbb{F}_{q^n})$. The cost of the relation collection for index calculus on E is*

$$\frac{q^{n'+k}}{2^m} \cdot H_2(n, n', m).$$

Theorem 2.5 still applies and the cost of the linear algebra step remains $\mathcal{O}(m \cdot q^{2n'})$, as $|\mathcal{F}_1| = q^{n'}$.

Note that this construction applies to every Koblitz curve. Under the assumption that $H_1 = H_2$, the choice of a factor base as described above allows to reduce the necessary work in the relation collection by a factor $m!$ without inflating the size of the factor base. Therefore, the approach also does not suffer any additional cost in the linear algebra step.

3.2 Frobenius invariant factor bases

In this section, we give another idea to accelerate index calculus for Koblitz curves $E(\mathbb{F}_{q^n})$ defined over $\mathbb{F}_q$ even further. The underlying idea is similar to the one used by Joux and Lercier [15] to solve discrete logarithms in the multiplicative group of extension fields. There, the idea was to use factor bases that are Galois invariant.

Our goal in this section is to show how this can be done for Koblitz curves. We use factor bases closed under Frobenius endomorphisms, i.e. $\pi(\mathcal{F}) = \mathcal{F}$. Once a point has been decomposed, we can rewrite the relation in terms of a smaller factor base consisting of representatives for each Frobenius orbit in the factor base. This allows to reduce the number of linearly independent relations we need to collect by a factor n, accelerating the relation collection accordingly. Simultaneously, the dimension of the matrix in the linear algebra step is reduced, decreasing the cost of this step by n^2.

One can define a unique representative of each Frobenius orbit in an ad-hoc manner to get a factor base $\mathcal{F}'$ of size approximately $|\mathcal{F}|/n$. An immediate consequence of Lemma 3.1 is that every point $P \in \mathcal{F}$ can be written as $P = \pi^j(P') = \lambda^j P'$ for some $P' \in \mathcal{F}'$ and $j \in \{0, \ldots, n-1\}$.

In the relation collection, one uses the summation polynomials to write $R = P_1 + \cdots + P_m$ where $P_i \in \mathcal{F}$. Then one reduces each relation to one consisting entirely of points in the reduced factor base $\mathcal{F}'$.

However, one Frobenius invariant factor base will in general not be enough to generate n linearly independent relations under the action of the Frobenius endomorphism, as we will show in Lemma 3.3.

First, we recall some background. Let the degree of the extension n be prime to the characteristic of the field $\mathbb{F}_{q^n}$. The Frobenius map π generates the Galois group $\mathsf{Gal}(\mathbb{F}_{q^n}/\mathbb{F}_q)$ and its characteristic polynomial acting on $\mathbb{F}_{q^n}$, considered as a $\mathbb{F}_q$ vector space, is $x^n - 1$. The polynomial $x^n - 1$ factors over $\mathbb{F}_q$ and a unique $\mathbb{F}_q$ subspace $V_f \subset \mathbb{F}_{q^n}$ corresponds to every $\mathbb{F}_q$-irreducible factor f dividing $x^n - 1$ in $\mathbb{F}_q[x]$. By Schur's Lemma, any Galois invariant subspace of $\mathbb{F}_{q^n}$ is a direct sum of such V_f.

Now, we are ready to show that we do not always get n independent relations under the Frobenius action from a single Galois invariant subspace.

Lemma 3.3. *Let $\mathcal{F}$ be a Frobenius invariant factor base of dimension n' defined by the polynomial $\ell(X) = \sum_{j=0}^{n'} c_j X^j$ dividing $X^n - 1$ in $\mathbb{F}_q[X]$. Given a decomposition of a point $R = \sum_{i=1}^{m} P_i$ as sum of elements in $\mathcal{F}$, the Frobenius action provides at most n' linearly independent relations.*

Proof. This is an immediate consequence of

$$\pi^{n'}\left(\sum_{i=1}^{m} P_i\right) = \sum_{i=1}^{m} \pi^{n'}(P_i) = \sum_{i=1}^{m}\sum_{j=0}^{n'-1} c_j \pi^j(P_i) = \sum_{j=0}^{n'-1} c_j \pi^j\left(\sum_{i=1}^{m} P_i\right) \tag{4}$$

Even though the order of the Frobenius endomorphism is n, when applying the endomorphism to a relation with respect to a Frobenius invariant vector space one only gets n' independent relations. Let $H_3(n, n', m)$ denote the cost of solving the polynomial systems during point decomposition for the described factor bases.

Theorem 3.4. *As before, let $n = n' \cdot m + k$ and let E be a Koblitz curve with coefficients in $\mathbb{F}_q$ and ECDLP in $E(\mathbb{F}_{q^n})$.*

Given a single Frobenius invariant factor base $\mathcal{F}$ defined by a Galois invariant vector subspace of $\mathbb{F}_{q^n}$ of size $q^{n'}$ such that applying the Frobenius endomorphism to a relation provides n' linearly independent relations, the cost of relation collection for index calculus on E is

$$\frac{1}{n'} \cdot \frac{q^{n'+k}}{2^m} \cdot m! \cdot H_3(n, n', m)$$

and the cost of the linear algebra step is in $\mathcal{O}(\frac{m}{n^2} \cdot q^{2n'})$.

Proof. As in Theorem 2.4, the probability that a random point R can be written as a sum of m points of $\mathcal{F}$ is $\frac{2^m}{q^k \cdot m!}$. However, because applying Frobenius yields $n' - 1$ additional relations for free, we only need to find $\frac{1}{n'} q^{n'}$ such relations. Thus, the cost of the relation search becomes $\frac{1}{n'} \cdot \frac{q^{n'+k}}{2^m} \cdot m! \cdot H_3(n, n', m)$.

For the linear algebra step, we can choose a unique representative in each Frobenius orbit of factor base elements and reduce each relation to one consisting only of such representatives as described at the beginning of this section. Hence, Theorem 2.5 implies that the cost of the linear algebra step is reduced by a factor n^2 to $\mathcal{O}(\frac{m}{n^2} \cdot q^{2n'})$. □

Note, that the linear dependence in Lemma 3.3 is a consequence of all the points in the sum having been chosen from the same factor base. To get the full saving of $1/n$ in the relation collection, we combine Matsuo's idea of using m different factor bases for symmetry breaking with Frobenius invariance of the factor bases.

Assume we have m Galois invariant factor bases $\mathcal{F}_i$ such that applying the Frobenius endomorphism to a single relation of the form $R = P_1 + \cdots + P_m$ with $P_i \in \mathcal{F}_i$ yields n linearly independent relations. Then not only do we save the symmetry factor $(m-1)!$, as described by Matsuo, but by the Frobenius action we also get n independent relations instead of n'. Intuitively, the reason for this is that in Equation (4) the coefficients c_j will be different for each index i.

In each factor base we can choose a unique representative in the Galois orbits, but there are m distinct factor bases so the cost of linear algebra is only reduced by a factor $\left(\frac{n}{m}\right)^2$. Let $H_4(n, n', m)$ be the cost of solving one polynomial system during point decomposition in this case.

Theorem 3.5. *As before, let $n = n' \cdot m + k$ and let E be a Koblitz curve with coefficients in $\mathbb{F}_q$ and ECDLP in $E(\mathbb{F}_{q^n})$. Given m distinct Frobenius invariant factor bases $\mathcal{F}$ of size $q^{n'}$ such that applying Frobenius to sums of m points from distinct factor bases yields linearly independent relations. Then, the cost of the relation collection for index calculus on E is*

$$\frac{1}{n} \cdot \frac{q^{n'+k}}{2^m} \cdot m \cdot H_4(n, n', m)$$

and the cost of the linear algebra step is in $\mathcal{O}(\frac{m^3}{n^2} \cdot q^{2n'})$.

In Section 4, we will discuss some constructions of factor bases that are closed under the Frobenius endomorphism. They have a nice algebraic description, which allows them to be easily substituted for the variables in the summation polynomial to restrict the polynomial system to factor base elements.

Remark 3.6. Note that the overall speed-up factor to solve an instance of ECDLP is relative to the original cost $R+L$, where R is the cost of the relation collection and L is the cost of the linear algebra prior to the improvements presented in this paper. By reducing the relation collection cost by n and linear algebra cost by n^2, the overall speed-up factor is $n^2(R+L)/(nR+L)$. Depending on the relation between R and L, this expression corresponds to a speed-up between n and n^2.

3.3 Comparison of different variants

We summarize the results on the performance of the different variations of the index calculus methods described in this section.

We follow the convention of the previous sections that $n = n'm + k$ and the $(m+1)$-th summation polynomial is used to decompose points during the relation collection. Table 1 summarises the cost of various approaches. The relation collection column shows the number of polynomial systems that need to be solved to collect the required number of relations. Each of these polynomial systems arises from the Weil descent of the $(m+1)$-th summation polynomial with variables restricted to the x-coordinates defining the factor bases (as in Section 2.2), has n equations and $n'm$ variables. While our methods clearly reduce both the number of systems to solve and the linear algebra costs, we stress that using different factor bases may a priori lead to polynomial systems that are harder (or easier) to solve using Gröbner basis methods. We study these costs experimentally in Section 5. Finally, we note that the lower three methods in Table 1 are only applicable to subfield curves defined over $\mathbb{F}_q$.

Method	Relation collection	Linear algebra	Reference
General index calculus	$m! \cdot \frac{q^{n'+k}}{2^m} \cdot H_1(n, n', m)$	$\mathcal{O}(m \cdot q^{2n'})$	[3, 8, 25]
Sym. breaking with m factor bases	$m \cdot \frac{q^{n'+k}}{2^m} \cdot H_1(n, n', m)$	$\mathcal{O}(m^3 \cdot q^{2n'})$	[20]
Sym. breaking for Koblitz curves	$\frac{q^{n'+k}}{2^m} \cdot H_2(n, n', m)$	$\mathcal{O}(m \cdot q^{2n'})$	this paper
One Frobenius inv. factor base from a Galois inv. vector space	$\frac{m!}{n'} \cdot \frac{q^{n'+k}}{2^m} \cdot H_3(n, n', m)$	$\mathcal{O}(\frac{m}{n^2} \cdot q^{2n'})$	this paper
Frobenius inv. and symmetry breaking combined	$\frac{m}{n} \cdot \frac{q^{n'+k}}{2^m} \cdot H_4(n, n', m)$	$\mathcal{O}(\frac{m^3}{n^2} \cdot q^{2n'})$	this paper

Table 1. Cost of the relation collection and linear algebra steps for different index calculus methods solving ECDLP in $E(\mathbb{F}_{q^n})$

We want to emphasise that there have been various improvements and variants of index calculus since the work listed in the references in Table 1. We refer to Section 9 of Galbraith-Gaudry [9] for further references. However, we note that these works were targeting improvements in the complexity of solving polynomial systems arising in index calculus rather than to lower the number of systems that require solving as displayed in Table 1. Depending on the specific construction used to define the Frobenius invariant factor bases, these improvements carry over to the reduced number of polynomial systems that need to be solved.

4 Frobenius invariant factor bases

In the previous section, we have seen how factor bases closed under the Frobenius endomorphism can be used to accelerate both the relation collection and the linear algebra steps when using index calculus to solve ECDLP instances. Therefore, an important problem is to find such factor bases of suitable size. Moreover, we want the factor bases to have a nice algebraic description in order to get polynomial systems that are similarly fast to solve as the ones described in Section 2.2.

In this section, we present the construction of such Frobenius invariant factor bases for some classes of elliptic curves. The first construction is a new idea based on linearised polynomials, while the others use Galois invariant subsets of finite fields constructed from isogenies between commutative algebraic groups, as described by Couveignes and Lercier [2].

4.1 Linearised polynomials

When using $\mathbb{F}_q$-vector subspaces V of $\mathbb{F}_{q^n}$ to define the factor base with respect to the abscissa of points on the curve, a factor base is Frobenius invariant if and only if the vector space V is Galois invariant. In this subsection we want to address this case.

Let $\mathbb{F}_{q^n}$ be an extension field of $\mathbb{F}_q$. A *linearised polynomial* $f \in \mathbb{F}_{q^n}[x]$ is a polynomial for which the exponents of all the constituent monomials are powers of q, e.g.

$$f(x) = \sum_{i=0}^{k} a_i x^{q^i}, \text{ with } a_i \in \mathbb{F}_{q^n}.$$

This class of polynomials has one particular property that makes it interesting for this work: the set of roots of a linearised polynomial is a $\mathbb{F}_q$ vector space and is closed under the q-th power Frobenius map if and only if $a_i \in \mathbb{F}_q$.

As we wrote prior to Lemma 4, the irreducible factors of $x^n - 1$ in $\mathbb{F}_q[x]$ correspond to unique Galois invariant $\mathbb{F}_q$ subspaces $V_f \subset \mathbb{F}_{q^n}$ and every Galois invariant subspace is a direct sum of such V_f. Thus, finding Galois invariant $\mathbb{F}_q$ vector subspaces of $\mathbb{F}_{q^n}$ depends on the factorisation of $x^n - 1$ over $\mathbb{F}_q[x]$.

We use the remainder of this section to demonstrate how to obtain Frobenius invariant factor bases in the case of Koblitz curves defined over characteristic 2 fields, which are the most common ones in practice. However, the method can be generalised easily to fields of odd characteristics.

Lemma 4.1. *([18], Lemma 7) Let n be an odd prime, let ℓ denote the multiplicative order of 2 modulo n and let $n = s\ell + 1$. The polynomial $x^n - 1$ factors in $\mathbb{F}_2[x]$ as $(x-1)f_1 f_2 \dots f_s$, where the f_i are distinct irreducible polynomials of degree ℓ.*

When defining a factor base using some vector space V corresponding to the polynomial f_i, the size of V depends on the degree ℓ of f_i. Restricting ourselves to Galois invariant vector spaces, means losing the fine control over the size of $\mathcal{F}$, as we will see in the following.

Suppose the order of 2 modulo n is ℓ. Then by Lemma 4.1 $x^n - 1$ has factors $f_j(x) = \sum_k f_{j,k} x^k$ of degree ℓ. Introducing the linearised polynomial

$$F_j(X) = \sum_k f_{j,k} X^{2^k}$$

we have that $F_j(X) \mid X^{2^n} - X$ and so we can define

$$\mathcal{F} = \{P \in E(\mathbb{F}_{2^n}) : F_j(x(P)) = 0\}.$$

As F_j is a linearised polynomial with coefficients in $\mathbb{F}_q$, its set of roots is Galois invariant and thus we have $\pi(\mathcal{F}) = \mathcal{F}$. Moreover, from the degree of f_j we can estimate the size of $|\mathcal{F}|$ to be approximately 2^ℓ. Note, that one can group several polynomial factors together to get larger sets $\mathcal{F}$ of size $2^{k\cdot\ell+\epsilon}$, where $\epsilon \in \{0, 1\}$ is due to the factor $(x-1)$ in Lemma 4.1 and $k \leq s$, $k \in \mathbb{Z}$, is the number of polynomial factors of degree ℓ.

Consequently, using linearised polynomials we cannot get Galois invariant subspaces of arbitrary size. In particular, we do not have any fine control over the size of $\mathcal{F}$ when the order of 2 modulo n is large. However, if the order of 2 modulo n is small, i.e. $\ell \ll n - 1$, which happens for example when $n = 2^l - 1$ is a Mersenne prime, or if ℓ divides the desirable dimension of our factor base, then linearised polynomials allow for a very easy construction of Frobenius invariant factor bases.

If the factor base $\mathcal{F}$ is defined using a linearised polynomial, it is specified by linear constraints in the Weil restriction and thus it is easily deployable in index calculus. In Section 5, we compare the complexity of solving polynomial systems arising from index calculus when using Frobenius invariant vector spaces instead of the vector spaces that are commonly used so far to define the factor base.

4.2 Factor bases from isogenies between algebraic groups

In this subsection we recall results due to Couveignes and Lercier, who gave a framework on how to construct Galois invariant flags of subsets of $\mathbb{F}_{q^n}$ using isogenies between algebraic groups [2].

First, we recall the basic idea behind their results. Then, we show how their results for two specific algebraic groups, algebraic tori of dimension 1 and elliptic curves, give rise to factor bases for index calculus in elliptic curves using Semaev polynomials.

Let G be a commutative algebraic group defined over the finite field $\mathbb{F}_q$. Moreover, let $T \subset G(\mathbb{F}_q)$ be a non-trivial finite group of $\mathbb{F}_q$-rational points in G of cardinality n. Then, the quotient isogeny $I : G \to H$ of G by T is of degree n.

Given a point $a \in H$ such that the preimage $I^{-1}(a)$ is irreducible over $\mathbb{F}_q$, any point $b \in G(\overline{\mathbb{F}_q})$ with $I(b) = a$ defines a cyclic degree n extension of $\mathbb{F}_q$, i.e. $\mathbb{F}_{q^n} = \mathbb{F}_q(b)$. We refer to (5) for an example of what it means for the set $I^{-1}(a)$ to be irreducible in this context.

Under this identification of $\mathbb{F}_{q^n}$ with $\mathbb{F}_q(b)$, any Galois action permutes elements in $G(\mathbb{F}_{q^n})$ that have the same image under I. In other words, we can identify $T = \ker(I)$ with the Galois group of $\mathbb{F}_{q^n}$ over $\mathbb{F}_q$.

In particular, the elements $b \oplus_G t$ are all Galois conjugates of b for all $t \in T$ as

$$I(b \oplus_G t) = I(b) \oplus_H I(t) = a \oplus_H 0 = a.$$

Conversely, all Galois conjugates of b are obtained this way. Here, $\oplus_G$ and $\oplus_H$ denote the addition in G and H respectively.

Consequently, if there exist formulae to describe the translations $P \to P \oplus_G t$ for $t \in T$ in G, we get an explicit description of the Galois action on $\mathbb{F}_{q^n}$. This can be used to derive descriptions of Galois invariant subsets of $\mathbb{F}_{q^n}$.

During the remainder of this section, we will see that different commutative algebraic groups bring their own contribution to this general construction of Galois invariant subsets. We show that this can be used to obtain Frobenius invariant factor bases that can be used for index calculus.

Isogenies between algebraic tori The simplest commutative algebraic groups beyond the ones underlying Kummer and Artin-Schreier theory are algebraic tori of dimension 1. They can be used to obtain Galois invariant sets in $\mathbb{F}_{q^n}$ through the previously described construction by Couveignes and Lercier. First, we will recall the details in this case. Then, we compute the size of the resulting sets and show how this leads to practical Frobenius invariant factor bases for Koblitz curves with ECDLP defined over field extensions $\mathbb{F}_{q^n}/\mathbb{F}_q$ whenever the extension degree n divides $q + 1$.

Let $\mathbb{F}_q$ be a finite field of characteristic different from 2, let $D \in \mathbb{F}_q^*$ and let $\mathbf{G}$ denote the open subset of the projective line $\mathbb{P}^1(\mathbb{F}_q)$ defined by

$$U^2 - DV^2 \neq 0,$$

where $[U, V]$ are the projective coordinates. We associate affine coordinates to points on the projective line using the map $u : \mathbb{P}^1 \to \mathbb{F}_q \cup \{\infty\},\ [U, V] \mapsto \frac{U}{V}$.

On $\mathbf{G}$ we have a group structure: for $P_1, P_2 \neq 0_{\mathbf{G}}$, addition is given by

$$u(P_1 \oplus_{\mathbf{G}} P_2) = \frac{u(P_1)u(P_2) + D}{u(P_1) + u(P_2)} \quad \text{and} \quad u(\ominus_{\mathbf{G}}) = -u(P_1).$$

The neutral element $0_{\mathbf{G}}$ is the point with projective coordinates $[1, 0]$ and affine coordinate ∞. Assuming D is not a square in $\mathbb{F}_q$, the group $\mathbf{G}(\mathbb{F}_q)$ is cyclic of order $q + 1$.

Let a be a generator of $\mathbf{G}(\mathbb{F}_q)$ and let $[n] : \mathbf{G} \to \mathbf{G}$ denote the multiplication by n isogeny. Furthermore, let $A(X)$ be the polynomial annihilating the associated affine coordinates of all points in the preimage of a by $[n]$, i.e.

$$A(X) = \prod_{b \in [n]^{-1}(a)} (X - u(b)). \tag{5}$$

Couveignes and Lercier show that this degree n polynomial is irreducible in $\mathbb{F}_q[X]$. Hence, we have $\mathbb{F}_q[X]/A(X) = \mathbb{F}_{q^n}$.

Using the general framework, they show that every $\mathbb{F}_q$-automorphism of $\mathbb{F}_{q^n}$ transforms $\omega := X \bmod A(X)$ into a linear rational fraction of ω. This proves that for every integer k such that $0 \leq k < n$, the subset

$$V_k := \left\{ \frac{u_0 + u_1\omega + u_2\omega^2 + \cdots + u_k\omega^k}{v_0 + v_1\omega + v_2\omega^2 + \cdots + v_k\omega^k} \middle| (u_0, u_1, \ldots, u_k, v_0, v_1, \ldots, v_k) \in \mathbb{F}_q^{2k+2} \right\} \subset \mathbb{F}_{q^n} \tag{6}$$

is Galois invariant.

The results of Couveignes and Lercier give rise to Frobenius invariant factor bases for index calculus on Koblitz curves.

Assume we want to solve the discrete logarithm problem in an elliptic curve $E(\mathbb{F}_{q^n})$ where n divides $q + 1$. We can choose some $D \in \mathbb{F}_q^*$ that is not a square in $\mathbb{F}_q$ to define an algebraic torus $\mathbf{G}$ of dimension 1. Then, we find a generator a of $\mathbf{G}(\mathbb{F}_q)$. Using the exponentiation formulae in $\mathbf{G}$, we can compute the polynomial $A(X)$ given in Equation (5) explicitly. As before, we set $\omega = X \bmod A(X)$ and define V_k as in (6).

Lemma 4.2. *Let $k < n/2$. Then, we have $|V_k| = q^{2k+1}$.*

Proof. Identifying elements of the set V_k given in (6) with vectors in $\mathbb{F}_q^{2k+2}$, we see that elements are counted multiple times if and only if the numerator and denominator are not coprime. Hence we want to count the number of coprime pairs of polynomials of degree $\leq k$ with one polynomial corresponding to the non-zero denominator. Theorem 3 in [1] proves that $\frac{q-1}{q}$ of pairs of polynomials not both of degree 0 are coprime in $\mathbb{F}_q[x]$. Note that this contains those pairs where the polynomial corresponding to the denominator is 0, which we don't want to count. Moreover, being coprime in $\mathbb{F}_q[x]$ is up to multiplication by units of the field $\mathbb{F}_q$, e.g. $\gcd(2x, 2) = 1$ if q is no power of 2. Thus, we need to reduce the count again under orbits of multiplication by the $(q-1)$ units in F_q. Hence, the $q^{2k+2} - q^2$ elements in V_k given by fractions of polynomials not both of degree 0 represent

$$\frac{1}{q-1} \cdot \frac{q-1}{q} \cdot (q^{2k+2} - q^2) = q^{2k+1} - q$$

distinct elements in $\mathbb{F}_{q^n}$. Additionally, we get the q field elements in our set by considering the fractions of degree 0 elements, which finishes the proof. □

Let $V = V_k$ for some k such that V_k is of the size we desire our factor base to be, which can be computed using the previous lemma. We can define a factor base for index calculus on some curve E by requiring the x-coordinates of factor base elements to be in the set $V \subset \mathbb{F}_{q^n}$, as it was done before in the case of vector spaces, i.e.

$$\mathcal{F} := \left\{ P \in E(\mathbb{F}_{q^n}) : x(P) \in V \right\}.$$

The Galois invariance of V in $\mathbb{F}_{q^n}/\mathbb{F}_q$ implies the invariance under the q-th power Frobenius of $\mathcal{F}$ in E.

As described in Section 2, we use the $(m+1)$-th Semaev summation polynomial and Weil restriction of scalars to solve the decomposition problem during the relation search in the index calculus. However, instead of restricting the variables of the Semaev polynomial to some vector space, we restrict them to V. After subsitution of expressions of the form (6) into the Semaev polynomial, we clear the denominator.

The resulting polynomial systems can be solved using the same methods as applied in the case of factor bases arising from vector spaces. While the systems have some additional structure, i.e. they are homogeneous with respect to certain blocks of variables before the Weil descent, they are generally of larger degree. This means it is harder to make a direct comparison of the complexity of solving those systems with the case where factor bases arising from vector spaces are used. We discuss this matter more together with our experimental evidence in Section 5.

Isogenies between elliptic curves Different commutative algebraic groups allow to construct Galois invariant subsets for different classes of finite field extensions. In addition to the construction using algebraic tori, Couveignes and Lercier give an explicit description using ordinary elliptic curves [2]. It allows to write down Galois invariant subsets of field extensions $\mathbb{F}_{q^n}$ of characteristic p, whenever n has a squarefree multiple N such that $N \neq 1 \pmod{p}$ and

$$q + 1 - 2\sqrt{q} < N < q + 1 + 2\sqrt{q}.$$

Their results also give rise to Frobenius invariant factor bases for index calculus for Koblitz curves with ECDLP problem defined in $\mathbb{F}_{q^n}$. Our description of their result follows [2] and we reproduce it here for completeness.

First, we use another elliptic curve H to construct a Galois invariant subset of $\mathbb{F}_{q^n}$. We choose H to have N rational points over $\mathbb{F}_q$ and trace $t = q + 1 - N$. Such a curve can be found using exhaustive search or complex multiplication theory. Note that this precomputation needs to be done only once for any field. The ideal $(\pi - 1) \subset \mathrm{End}(H)$ has a degree n factor $\mathfrak{i}$. Therefore, $\mathrm{End}(H)/\mathfrak{i}$ is cyclic of order n and H contains a cyclic subgroup $T := \ker(\mathfrak{i})$ of order n.

Let $I : H \to F$ be the quotient isogeny with kernel T. As the quotient $F(\mathbb{F}_q)/I(E(\mathbb{F}_q))$ is isomorphic to T and thus cyclic of order n, we can take a generator a of this quotient. As in the case of algebraic tori, the preimage of a under I is an irreducible divisor. Thus, there are n geometric points in the preimage of a that are defined in $\mathbb{F}_{q^n}$ and permuted by Galois action.

Let $B := I^{-1}(a)$ denote the corresponding prime divisor. Then $\mathbb{F}_{q^n}$ is the residue extension of E at B, i.e. the elements of $\mathbb{F}_{q^n}$ can be represented as residues of functions on E at B that do not have a pole at B.

For a function f in $H(\mathbb{F}_q)$, let the degree of f be the number of poles of f counted with multiplicities. We denote the set of functions in $H(\mathbb{F}_q)$ of degree $\leq k$ having no pole at B by $\mathcal{F}_k$ for every $k \geq 0$. Define V_k to be the corresponding set of residues at B in $\mathbb{F}_{q^n}$:

$$V_k := \{f \pmod{B} \mid f \in \mathcal{F}_k\}.$$

As shown in [2], we have $\mathbb{F}_q = V_0 = V_1 \subset V_2 \subset \cdots \subset V_d = \mathbb{F}_{q^n}$ and translations by an element in $T = \ker(\mathfrak{i})$ do not change the number of poles of functions in $H(\mathbb{F}_q)$. Consequently, V_k is invariant under the action of $\mathrm{Gal}(\mathbb{F}_{q^n}/\mathbb{F}_q)$.

As in the case of algebraic tori, these results give rise to a construction of Frobenius invariant factor bases for Koblitz curves. Note that functions in the sets $\mathcal{F}_k$ have at most k poles and thus can be written as a quotient of two homogeneous polynomials of degree $\left\lceil \frac{k+1}{3} \right\rceil$ polynomials.

Evaluating $\mathcal{F}_k$ at B gives the values of V_k. As before, we define our factor base for index calculus to be the set of points on the curve containing the ECDLP problem with x-coordinates in V_k.

As conjectured by Couveignes and Lercier [2], other commutative algebraic groups might bring their own contribution to the construction of Galois invariant subsets of $\mathbb{F}_{q^n}$ which give rise to factor bases for index calculus in various classes of Koblitz curves.

5 Experimental Results

We implemented the construction of Frobenius invariant factor bases emerging both from linearised polynomials and from isogenies between algebraic tori. We used this to construct and solve the polynomial systems arising from Weil descent after constraining the variables in Semaev's summation polynomials to these factor bases. All the experiments were executed using the 64-bit version of MAGMA. Gröbner bases were computed using Faugère's F4 algorithm, as implemented in MAGMA.

5.1 Frobenius invariant vector spaces

In the case of factor bases from vector spaces defined by linearised polynomials, we tested our implementation on a Koblitz curve over $\mathbb{F}_{q^n} = \mathbb{F}_{2^{31}}$, using $m = 3$. Since the order of 2 in $\mathbb{Z}_{31}$ is 5, the vector space arising from our construction using linearised polynomials is of dimension 5. More precisely, the vector space is generated by $\{x, x^2, x^4, x^8, x^{16}\}$, where x is a symbolic root of an irreducible polynomial of degree n over the base field $\mathbb{F}_q$. We implemented the index calculus methods of Faugère et al. [8] to compute the polynomial system arising from Weil descent of the $(m+1)$-th Semaev polynomial, S_4. We timed the set-up of the polynomial system via Weil descent and the solving of the resulting system for different factor bases. We compared our Frobenius invariant factor base to factor bases from randomly chosen vector spaces and "standard" vector spaces, i.e. spanned by $\{1, x, x^2, \ldots, x^k\}$, of the same dimension. Our measurements over 200 runs of the experiments are displayed in Table 2.

factor base from vector space	$t_{\text{set-up}}$	t_{Groeb}
standard	0.19	2.95
random	0.19	3.15
Frobenius invariant	0.19	3.15

Table 2. Time in seconds for setting up the polynomial system and the Gröbner basis computation during point decomposition in index calculus for factor bases arising from different vector spaces. Parameters $m = 3$, $p = 2$, $n = 31$ and vector spaces are of dimension 5.

As was observed previously, using factor bases from standard vector spaces is faster than from randomly chosen vector spaces [14]. However, the results also show that our Frobenius invariant vector spaces do not behave worse than a randomly chosen one on average. Considering that we would only need to solve $1/n = 1/31$ as many such systems to compute an ECDLP instance when using the Frobenius invariant vector space, the experiments suggest that the full speed-up by a factor n in the relation collection could be (nearly) reached for these parameters.

5.2 Factor bases from isogenies between algebraic tori

In the case of factor bases arising from isogenies between algebraic tori, our experimental results are less promising and less clear.

The polynomials S_{m+1} are symmetric and have degree 2^{m-1} in each of their variables. Consider sums of the form $\sum_{i=0}^{k} x_i\omega^i$ for some fixed $\omega \in \mathbb{F}_{q^n}$ as one *block of variables*. When substituting vector space constraints into the Semaev polynomial S_{m+1} with one fixed variable, we get an equation containing m blocks of variables.

In the case of fractions of polynomials, we have $2m$ (shorter) blocks of variables. After clearing the denominator, we are left with a homogeneous polynomial in terms of these blocks. The degree of every monomial equals $m \cdot 2^{m-1}$. Therefore, we have more monomials of a large degree compared to the case of vector spaces. However, the blocks of variables are shorter compared to a factor base of the same size from a vector space.

The following example compares the degree of the polynomial systems after Weil descent for vector spaces and the polynomial fractions of (6) in the case $m = 3$.

Example 5.1. By Lemma 4.2, the size of a vector space of dimension 3 equals the number of elements of the set (6) for $k = 1$. For $m = 3$ we substitute three blocks of variables into the Semaev polynomial. In the case of vector spaces, blocks containing 3 variables are taken to the $2^{m-1} = 4$-th power. According to the multinomial theorem, the generalisation of the binomial theorem, in characteristic $p = 2$ this only leads to mixed terms containing monomials with at most two variables over the base field from the same block. After reducing modulo the field equations, we are left with polynomials of degree at most $m \cdot 2 = 6$.

In the case of polynomial fractions, each block of variables to the 3-rd and 4-th power also contains monomials with at most two variables. After substituting the variables into S_4 and clearing denominators, the summation polynomial is homogeneous of degree 12 with respect to the blocks of variables. Multiplying out and reducing modulo the field equations, this leads to an upper bound of 9 on the degree of the polynomials after Weil descent. Both values of the degree match the degrees observed in our experiments.

However, the degree of the system alone does not determine the difficulty of the Gröbner basis computation.

We did not find instances fulfilling the requirements for factor bases as defined in (6), i.e. where n divides $q + 1$, that were tractable for experiments. However, we computed and solved the polynomial systems for $p = q = 2$, $m = 3$ and various primes n when using vector spaces of dimension 5 and 7 and polynomial fractions as in (6) for $k = 2$ and $k = 3$. Note that we paired the factor bases of the same size by Lemma 4.2. The results of our experiments are shown in Table 3. Note that for these parameters the factor bases are *not* Galois invariant and the experiments are merely to observe the computational impact of using factor bases defined by the given fractions instead of vector spaces.

For the chosen parameters it is apparent that both the transformation from Semaev polynomial to binary polynomial system and the computation of Gröbner bases are significantly slower when using the polynomial fractions to express the factor bases. The increased cost does not seem to be justified by needing $1/n$ fewer relations. Moreover, we were not able to run experiments on less underdetermined systems in the case of $\log_q(|\mathcal{F}|) = 7$. This was because the solution of the polynomial systems in the fraction case took either too long or met our memory limit of 1.4 TB. While the factor bases from isogenies between algebraic tori do not seem to give very hopeful results, we want to point out that the algorithm used to solve the polynomial systems did not exploit the homogeneous structure of the polynomial system and therefore speed-ups might be possible.

m=3, $\log_q(\lvert\mathcal{F}\rvert) = 5$			
n		$t_{\text{set-up}}$	t_{Groeb}
17	vector spaces	0.16	5.32
	fractions	1.43	1390.34
23	vector spaces	0.20	3.94
	fractions	2.09	1290.14
29	vector spaces	0.23	3.66
	fractions	2.31	3764.22
31	vector spaces	0.24	3.57
	fractions	2.48	950.64
37	vector spaces	0.30	6.41
	fractions	3.18	853.79
41	vector spaces	0.32	6.26
	fractions	3.34	1130.67
43	vector spaces	0.33	3.03
	fractions	3.30	972.56

m=3, $\log_q(\lvert\mathcal{F}\rvert) = 7$			
n		$t_{\text{set-up}}$	t_{Groeb}
101	vector spaces	3.53	115.05
	fractions	67.64	701.33
103	vector spaces	4.02	35.55
	fractions	76.63	827.00
107	vector spaces	4.42	43.50
	fractions	83.00	338.75
109	vector spaces	4.57	15.85
	fractions	82.70	252.14
113	vector spaces	5.00	15.26
	fractions	84.88	243.68
127	vector spaces	5.27	0.91
	fractions	92.48	235.91
131	vector spaces	5.46	0.84
	fractions	96.53	237.70

Table 3. Average time of Weil descent and Gröbner basis over 5 instances for parameters $p = q = 2$ and $\log_q(|\mathcal{F}|) = 5$ and $\log_q(|\mathcal{F}|) = 7$ in s using factor bases defined by vector spaces and fractions of the form given in (6).

We leave the research of the combined effects of larger degrees, smaller block sizes and the homogeneous structure on the complexity of solving the polynomial systems for future work. It will be interesting to study the impact in practice for different characteristics and asymptotically.

6 Conclusion

This work presents multiple ideas on how to accelerate index calculus on Koblitz curves using careful choices of factor bases with respect to the q-power Frobenius endomorphism. This allows for better symmetry breaking, and Frobenius invariant factor bases enable us to reduce both the computational effort in the relation collection and the linear algebra step of index calculus.

While a lot of work in the literature has been directed at improving the complexity for solving polynomial systems arising during index calculus, our speed-ups are achieved by reducing the number of such polynomial systems that need to be solved in the first place.

For suitable parameters, we define Frobenius invariant factor bases from Galois invariant vector spaces that can be constructed using linearised polynomials. We can rewrite every relation found in terms of a reduced factor base consisting only of representatives for each Frobenius orbit. As a consequence, we need to find n times fewer relations, which leads to a speed-up by a factor of roughly n in the relation collection and by n^2 during the linear algebra step. Our experimental evidence supports that the polynomial systems arising in this way appear to be roughly as hard to solve as the ones arising from a more standard choice used in index calculus.

For further parameter sets, we construct Frobenius invariant factor bases using the work of Couveignes and Lercier. Yet, the polynomial systems arising in the relation collection in this

case were more expensive to solve in our experiments. The experiments suggest that this second approach is less promising.

Given the uncertainties in computing the exact cost of Gröbner basis algorithms, a precise complexity estimate of index calculus methods for Koblitz curves is beyond the scope of this work. Nevertheless, based on previous work such as [21] our improvements do not lead to index calculus algorithms faster than Pollard ρ on instances used in practice. However, this work shows that index calculus on Koblitz curves can be accelerate beyond the $\sqrt{n}$ previously achieved for Pollard ρ compared to the same algorithm on general curves and thus can be used to narrow the gap in performance for Koblitz curves. Moreover, this paper answers an open problem raised in [9] about how the Frobenius endomorphism on Koblitz curves can be exploited to accelerate index calculus.

We hope the new ideas described in this work may be used as another building block for more efficient index calculus based methods to solve ECDLP in theory and practice.

Acknowledgements. We thank Jean-Marc Couveignes and Reynald Lercier for their work on Galois invariant smoothness bases [2] and helpful conversations about the topic. Furthermore, we would like to thank the anonymous reviewers for their helpful comments on the submitted manuscript of this paper. Christophe Petit's work was supported by EPSRC grant EP/S01361X/1.

Bibliography

[1] Arthur T Benjamin and Curtis D Bennett. The probability of relatively prime polynomials. *Mathematics Magazine*, 80(3):196–202, 2007.

[2] Jean-Marc Couveignes and Reynald Lercier. Galois invariant smoothness basis. In *Algebraic Geometry And Its Applications: Dedicated to Gilles Lachaud on His 60th Birthday*, pages 142–167. World Scientific, 2008.

[3] Claus Diem. On the discrete logarithm problem in elliptic curves. *Compositio Mathematica*, 147(1):75–104, 2011.

[4] Jean-Charles Faugère. A new efficient algorithm for computing Gröbner bases (F4). *Journal of pure and applied algebra*, 139(1-3):61–88, 1999.

[5] Jean-Charles Faugère. A new efficient algorithm for computing Gröbner bases without reduction to zero (F5). In *Proceedings of the 2002 international symposium on Symbolic and algebraic computation*, pages 75–83, 2002.

[6] Jean-Charles Faugère, Pierrick Gaudry, Louise Huot, and Guénaël Renault. Using symmetries in the index calculus for elliptic curves discrete logarithm. *Journal of Cryptology*, 27(4):595–635, 2014.

[7] Jean-Charles Faugère, Patrizia Gianni, Daniel Lazard, and Teo Mora. Efficient computation of zero-dimensional Gröbner bases by change of ordering. *Journal of Symbolic Computation*, 16(4):329–344, 1993.

[8] Jean-Charles Faugère, Ludovic Perret, Christophe Petit, and Guénaël Renault. Improving the complexity of index calculus algorithms in elliptic curves over binary fields. In *Advances in Cryptology - EUROCRYPT 2012 - 31st Annual International Conference on the Theory and Applications of Cryptographic Techniques, Cambridge, UK, April 15-19, 2012. Proceedings*, pages 27–44, 2012.

[9] Steven D Galbraith and Pierrick Gaudry. Recent progress on the elliptic curve discrete logarithm problem. *Designs, Codes and Cryptography*, 78(1):51–72, 2016.

[10] Pierrick Gaudry. An algorithm for solving the discrete log problem on hyperelliptic curves. In *Advances in Cryptology - EUROCRYPT 2000, International Conference on the Theory and Application of Cryptographic Techniques, Bruges, Belgium, May 14-18, 2000, Proceeding*, pages 19–34, 2000.

[11] Pierrick Gaudry. Index calculus for abelian varieties of small dimension and the elliptic curve discrete logarithm problem. *Journal of Symbolic Computation*, 44(12):1690–1702, 2009.
[12] Elisa Gorla and Maike Massierer. Index calculus in the trace zero variety. *Adv. Math. Commun.*, 9(4):515–539, 2015.
[13] Ming-Deh A. Huang, Michiel Kosters, and Sze Ling Yeo. Last fall degree, hfe, and weil descent attacks on ECDLP. In *Advances in Cryptology - CRYPTO 2015 - 35th Annual Cryptology Conference, Santa Barbara, CA, USA, August 16-20, 2015, Proceedings, Part I*, pages 581–600, 2015.
[14] Yun-Ju Huang, Christophe Petit, Naoyuki Shinohara, and Tsuyoshi Takagi. Improvement of Faugère et al.'s Method to Solve ECDLP. In *International Workshop on Security*, pages 115–132. Springer, 2013.
[15] Antoine Joux and Reynald Lercier. The function field sieve in the medium prime case. In *Advances in Cryptology - EUROCRYPT 2006, 25th Annual International Conference on the Theory and Applications of Cryptographic Techniques, St. Petersburg, Russia, May 28 - June 1, 2006, Proceedings*, pages 254–270, 2006.
[16] Antoine Joux and Vanessa Vitse. Elliptic curve discrete logarithm problem over small degree extension fields. *Journal of Cryptology*, 26(1):119–143, 2013.
[17] Neal Koblitz. Elliptic curve cryptosystems. *Mathematics of Computation*, 48(177):203–209, 1987.
[18] Alfred Menezes and Minghua Qu. Analysis of the Weil descent attack of Gaudry, Hess and Smart. In *Cryptographers' Track at the RSA Conference*, pages 308–318. Springer, 2001.
[19] Victor S. Miller. Use of elliptic curves in cryptography. In *Advances in Cryptology - CRYPTO '85, Santa Barbara, California, USA, August 18-22, 1985, Proceedings*, pages 417–426, 1985.
[20] Koh-ichi Nagao. Decomposition formula of the Jacobian group of plane curve. 2013.
[21] Christophe Petit and Jean-Jacques Quisquater. On polynomial systems arising from a weil descent. In *Advances in Cryptology - ASIACRYPT 2012 - 18th International Conference on the Theory and Application of Cryptology and Information Security, Beijing, China, December 2-6, 2012. Proceedings*, pages 451–466, 2012.
[22] Stephen Pohlig and Martin Hellman. An improved algorithm for computing logarithms over GF(p) and its cryptographic significance (corresp.). *IEEE Transactions on Information Theory*, 24(1):106–110, 1978.
[23] John M Pollard. Kangaroos, monopoly and discrete logarithms. *Journal of Cryptology*, 13(4):437–447, 2000.
[24] Eric Rescorla and Tim Dierks. The transport layer security (TLS) protocol version 1.3. 2018.
[25] Igor Semaev. Summation polynomials and the discrete logarithm problem on elliptic curves. *IACR Cryptology ePrint Archive*, 2004:31, 2004.
[26] Igor Semaev. New algorithm for the discrete logarithm problem on elliptic curves. Cryptology ePrint Archive, Report 2015/310, 2015. https://eprint.iacr.org/2015/310.
[27] Nigel P Smart. Elliptic curve cryptosystems over small fields of odd characteristic. *Journal of Cryptology*, 12(2):141–151, 1999.
[28] Paul C Van Oorschot and Michael J Wiener. Parallel collision search with cryptanalytic applications. *Journal of Cryptology*, 12(1):1–28, 1999.
[29] Douglas Wiedemann. Solving sparse linear equations over finite fields. *IEEE Transactions on Information Theory*, 32(1):54–62, 1986.
[30] Michael J Wiener and Robert J Zuccherato. Faster attacks on elliptic curve cryptosystems. In *International workshop on Selected Areas in Cryptography*, pages 190–200. Springer, 1998.