An Index Calculus Algorithm for Plane Curves of Small Degree

Claus Diem

University of Leipzig, Germany

Abstract. We present an index calculus algorithm which is particularly well suited to solve the discrete logarithm problem (DLP) in degree 0 class groups of curves over finite fields which are represented by plane models of small degree. A heuristic analysis of our algorithm indicates that asymptotically for varying q, "almost all" instances of the DLP in degree 0 class groups of curves represented by plane models of a fixed degree $d \geq 4$ over $\mathbb{F}_q$ can be solved in an expected time of $\tilde{O}(q^{2-2/(d-2)})$. Additionally we provide a method to represent "sufficiently general" (non-hyperelliptic) curves of genus $g \geq 3$ by plane models of degree $g+1$. We conclude that on heuristic grounds, "almost all" instances of the DLP in degree 0 class groups of (non-hyperelliptic) curves of a fixed genus $g \geq 3$ (represented initially by plane models of bounded degree) can be solved in an expected time of $\tilde{O}(q^{2-2/(g-1)})$.

1 Introduction

In recent works by Gaudry, Thomé, Thériault and the author ([13]) as well as Nagao ([22]), a double large prime variation for index calculus in degree 0 class groups of curves of small genus over finite fields has been introduced.

In this work, we present a different double large prime variation algorithm which is particularly well suited for the computation of the discrete logarithm problem (DLP) in degree 0 class groups of curves which are represented by plane models of *small degree*.

A heuristic analysis of our algorithm indicates (see Section 4):

Heuristic Result 1 *Let $d \geq 4$ be fixed. Let us consider the DLP in degree 0 class groups of curves of a fixed genus $g \leq (d-1)(d-2)/2$ represented by plane models of degree d over finite fields $\mathbb{F}_q$. Then "almost all" instances of the DLP in such groups can be solved in an expected time of $\tilde{O}(q^{2-\frac{2}{d-2}})$.*

Here, the $\tilde{O}$-notation means that we suppress logarithmic factors.

Additionally to the index calculus algorithm, we present a method to find plane models of degree $g+1$ of "sufficiently general" (non-hyperelliptic) curves of genus $g \geq 3$ (see Section 6).

By applying our algorithm to such a plane model, we obtain that on heuristic grounds "almost all" instances of the DLP in degree 0 class groups of (non-hyperelliptic) curves of a fixed genus $g \geq 3$ (initially represented by plane models of

bounded degree) can be solved in an expected time of

$$\tilde{O}(q^{2-\frac{2}{g-1}}).$$

This result should be compared with the following provable result which can be obtained with a variant of one of the algorithms in [13] (see [7]).

Let $g \geq 2$ be fixed. Then the DLP in cyclic degree 0 class groups of curves of genus g represented by plane models of bounded degree can with a randomized algorithm be solved in an expected time of $\tilde{O}(q^{2-2/g})$.

An important special case for our algorithm is constituted by the DLP in degree 0 class groups of *non-hyperelliptic curves of genus 3* over finite fields $\mathbb{F}_q$: Every such curve can (via the canonical embedding) be represented as a plane quartic. By applying our algorithm to such a model, we obtain a heuristic running time of $\tilde{O}(q)$.

This result is of particular importance because the DLP in degree 0 class groups of non-hyperelliptic genus 3 curves has recently received considerable attention as a potential cryptographic primitive; it is studied in detail in the related article [10] in which also some experimental data is presented.

Even though the DLP in degree 0 class groups of non-hyperelliptic curves of genus larger than 3 has not received much attention as a potential cryptographic primitive, our algorithm has yet another important application in cryptanalysis:

The method of "covering attacks" (a.k.a. Weil descent attacks) (cf. [8, Appendix], [9], [17], [12, Section 4.4]) allows to transfer the DLP in groups of rational points of certain elliptic curves (or in degree 0 class groups of certain curves of small genus) over extension fields into the DLP in degree 0 class groups of curves of rather small genus over smaller fields. The results in the present work suggest that it is advantageous for the attack if the resulting curves are not hyperelliptic.

2 Setting and First Remarks

Preliminaries

In this work, if not stated otherwise, a *curve* is always non-singular, projective and geometrically irreducible.

In the presentation above we implicitly used the following conventions concerning the representation of curves, divisors and divisor classes:

Let q be a prime power. We let $\mathbb{P}^2_{\mathbb{F}_q} := \mathrm{Proj}(\mathbb{F}_q[X,Y,Z])$; we thus have the canonical "homogeneous coordinate system" $X, Y, Z \in \Gamma(\mathbb{P}^2_{\mathbb{F}_q}, \mathcal{O}(1))$.

We think of every curve in question as being the normalization of a possibly singular curve in $\mathbb{P}^2_{\mathbb{F}_q}$. We distinguish the two by calling the latter one a *plane model* of the curve, denoted by $\mathcal{C}_{pm}$. We use a defining homogeneous polynomial to represent the plane model (and thus the curve itself).

By a *divisor* on a curve $\mathcal{C}$ over $\mathbb{F}_q$ we mean a divisor over $\mathbb{F}_q$. We think of divisors as being represented as a formal sum of closed points in $\mathcal{C}$. (This is called the *free representation* in [16].)

For some divisor D on $\mathcal{C}$, we denote the corresponding divisor class by $[D]$. We denote the degree 0 class group of $\mathcal{C}$ over $\mathbb{F}_q$ by $\mathrm{Cl}^0(\mathcal{C})$.

For fixed genus g and $q \gg 0$, $\mathcal{C}(\mathbb{F}_q)$ is non-empty; we assume that this is the case and fix some $P_0 \in \mathcal{C}(\mathbb{F}_q)$. An effective divisor D on $\mathcal{C}$ is called *maximally reduced along* P_0 if the linear system $|D - P_0|$ is empty. By the Riemann-Roch theorem, maximally reduced divisors have degree $\leq g$, and $D \mapsto [D] - \deg(D) \cdot [P_0]$ defines a bijection between the effective maximally reduced divisors and the elements of the degree 0 class group $\mathrm{Cl}^0(\mathcal{C})$ (see [16, Prop. 8.2.]).

It is by now a classical result that with this representation of the elements of the degree 0 class group, the arithmetic in $\mathrm{Cl}^0(\mathcal{C})$ can – for curves represented by plane models of bounded degree – be carried out in randomized polynomial time (cf. e.g. [26], [18], [16], [20], [19]).

Further notation and conventions

We use the same notation for functions on $\mathbb{P}^2_{\mathbb{F}_q}$, their restriction to $\mathcal{C}_{pm}$, their pull-back to $\mathcal{C}$ as well as the induced element in the function field $\mathbb{F}_q(\mathcal{C})$. Moreover, if $\varphi : \mathcal{C} \longrightarrow \mathbb{P}^2_{\mathbb{F}_q}$ is the (fixed) morphism from $\mathcal{C}$ to $\mathbb{P}^2_{\mathbb{F}_q}$, we use the same notation for elements of $\Gamma(\mathbb{P}^2_{\mathbb{F}_q}, \mathcal{O}(1))$ and their pull-backs to $\Gamma(\mathcal{C}, \varphi^*(\mathcal{O}(1)))$.

We identify zero-dimensional closed subschemes on $\mathcal{C}$ with effective divisors. To distinguish the divisor of zeros of an element of $W \in \Gamma(\mathbb{P}^2_{\mathbb{F}_q}, \mathcal{O}(1))$ from the divisor of zeros of the induced element in $\Gamma(\mathcal{C}, \varphi^*(\mathcal{O}(1)))$, we write $\mathrm{div}_{\mathcal{C}}(W)$ for the latter. (See [15, II, §7] for information about the divisor of zeros.)

Calculating the group order

We assume that the order of the degree 0 class group is known. From a theoretical point of view this is however not an obstacle because it can be shown that the L-polynomials of curves over $\mathbb{F}_q$ represented by plane models of bounded degree can be calculated in (deterministic) polynomial time in $\log(q)$. (This result follows from [24, Theorem H] which in turn relies on Pila's extension of the point counting algorithm by Schoof ([25]) to abelian varieties ([23]).) Moreover, in cryptographic situations, the order of the cyclic subgroup in question is always known, and this suffices for practical applications of our algorithm.

Overview over the new algorithm

Our algorithm can be viewed as a variant of the recent double large prime variation algorithms by Gaudry, Thomé, Thériault and the author ([13]) as well as Nagao ([22]) (see also [3]).

The main difference is that we use principal divisors to construct the graph of large prime relations, whereas in [13] and [22] random linear combinations of the two input elements in the degree 0 class group have been used.

More concretely, we find relations by intersecting the plane model with lines running through two elements of the factor base. We advice the reader to have the

following *intuitive idea* about the algorithm and its heuristic analysis in mind: Every line which runs through the non-singular part of the plane model defines a divisor of degree d on the curve. If we now intersected the plane model with arbitrary lines, heuristically we would obtain a running time which is analogous to the running time of the previous double-large prime-variation algorithms with g substituted by d. As we however only consider lines which already run through two points of the factor base, we obtain a running time which is analogous to the running time of the previous algorithms with g substituted by $d-2$.

We recall that there are two algorithms in [13]: the "full algorithm" and the "simplified algorithm". Our algorithm is closer to the "full algorithm" but there is an essential difference: In the full algorithm in [13], recombined relations over the factor base are already obtained during the construction of the graph. In contrast, we first try to construct a sufficiently dense graph, and after that we construct what is known as a *shortest path tree*. Then we use random linear combinations of the two input elements to generate recombined relations over the factor base with the help of the tree.

The heuristic nature of our results

The analysis of the algorithm presented in this work is heuristic. It is conceivable that there is a sequence of instances which violates the stated running times. This is why we talk about "almost all" instances.

A rigorous interpretation of our claims can be given as follows:

Let us fix the degree d and the genus $g \leq (d-1)(d-2)/2$. Now for a prime power q, let $S(q)$ be the set of all instances of the DLP in curves of genus $g \leq (d-1)(d-2)/2$ over $\mathbb{F}_q$ represented by plane models of degree d. (With the representations described above.)

The (conjectural) claim is now that there exist subsets $S_1(q)$ of $S(q)$ with $\#S_1(q)/\#S(q) \longrightarrow 1$ ($q \longrightarrow \infty$) such that the instances in $S_1(q)$ can be solved in the stated time.

Above, we also used the term "sufficiently general". This term will be defined in Section 6.

Historical remarks and comparison

The idea to use principal divisors to generate relations in class groups is not new. For example, the same approach was taken in the work by Adleman, DeMarrais, Huang ([1]), in which the first algorithm with a heuristic subexponential running time for the computation of the DLP in degree 0 class groups of hyperelliptic curves of large genus was given.

We note that to our knowledge, all known index calculus algorithms which rely on the consideration of principal divisors are analyzed only heuristically. With our two-step procedure to generate relations we have however eliminated a crucial hypothesis which previously occurred in the analyses of such algorithms: the hypothesis that "sufficiently many" of the relations generated are linearly independent.

3 The Algorithm

We consider curves over $\mathbb{F}_q$ represented by plane models of a fixed degree $d \geq 4$. Let $\mathcal{C}$ be such a curve with a fixed plane model $\mathcal{C}_{pm}$ in $\mathbb{P}^2_{\mathbb{F}_q}$, given by

$$F(X, Y, Z) = 0\,.$$

Let $a, b \in \mathrm{Cl}^0(\mathcal{C})$ such that $b \in \langle a \rangle$. The goal is to compute an $x \in \mathbb{N}$ with $x \cdot a = b$.

Let $D_\infty := \mathrm{div}_{\mathcal{C}}(Z)$. Note that this is a divisor of degree d on $\mathcal{C}$. (This divisor will appear in the description of the algorithm, it is however not necessary to compute it.)

Let $\mathcal{C}_{ns}$ be the non-singular part of $\mathcal{C}_{pm}$.

We now describe how the partial relations used to construct the graph of large prime relations are obtained.

The following classical statement from the theory of linear systems is crucial:

Lemma 1. *Let $W \in \Gamma(\mathbb{P}^2_{\mathbb{F}_q}, \mathcal{O}(1))$ ($W \neq 0$), and let $D := \mathrm{div}_{\mathcal{C}}(W)$. Then D is linearly equivalent to D_∞.*

Sketch of the proof. $D - D_\infty$ is the principal divisor of $\frac{W}{Z} \in \mathbb{F}_q(\mathcal{C})$. □

As a reformulation of this we obtain: Let $c_X, c_Y, c_Z \in \mathbb{F}_q$, not all 0, and let L be the line defined by $c_X X + c_Y Y + c_Z Z = 0$. Let $D := L \cap \mathcal{C}_{pm}$ be the (scheme-theoretic) intersection. If then D is contained in $\mathcal{C}_{ns}$, we can regard D as a divisor on $\mathcal{C}$, and we have

$$[D] - [D_\infty] = 0\ . \tag{1}$$

Lemma 2. *Given $c_X, c_Y, c_Z \in \mathbb{F}_q$, not all 0, one can decide in randomized polynomial time in $\log(q)$ if the support of the intersection of $\mathcal{C}_{pm}$ with L consists of $\mathbb{F}_q$-rational points of $\mathcal{C}_{ns}$ and – if this is the case – compute the (completely split) intersection divisor D.*

Proof. Let us (w.l.o.g.) assume that $c_Y = 1$. Then the point $(0 : 1 : 0)$ does not lie on L. The homogeneous polynomial $F(X, -c_X X - c_Z Z) \in \mathbb{F}_q[X, Z]$ now defines the image of the intersection under the projection to the (X, Z)-coordinates (with multiplicities). The support of the intersection of $\mathcal{C}_{pm}$ with L consists of $\mathbb{F}_q$-rational points of $\mathcal{C}_{pm}$ if and only if this polynomial factors completely. This factorization can be computed in randomized polynomial time in $\log(q)$. The Y-coordinates of the intersection points can then easily be obtained by using the equation for the line L. Finally, one can check whether the intersection points lie in $\mathcal{C}_{ns}$ by evaluating the partial derivatives of F. □

Let us now fix a *factor base* $\mathcal{F} = \{F_1, F_2, \ldots\} \subset \mathcal{C}_{ns}(\mathbb{F}_q)$. Let $\mathcal{L} := \mathcal{C}_{ns}(\mathbb{F}_q) - \mathcal{F}$ be the set of the so-called *large primes*. Analogously to [13] we define:

Definition 1. *A relation (1) (with $D \geq 0$) is called a* Full relation *if D is a sum of elements of the factor base. It is called an* FP relation *if D is a sum of elements of the factor base and the non-trivial multiple of one large prime. It is called a* PP relation *if D is the sum of elements of the factor base and non-trivial multiples of two large primes.*

In the first phase of the algorithm, we construct a *graph of large prime relations* on $\mathcal{L} \dot{\cup} \{*\}$ using FP and PP relations.

We find such relations by intersecting the curve with lines $L : c_X X + c_Y Y + c_Z Z = 0\,(c_X, c_Y, c_Z \in \mathbb{F}_q)$ running through two points of the factor base.

For the construction of the graph of large prime relations, we proceed as follows:

If we have a Full relation, we do nothing. If we have an FP relation with a large prime P, we consider the edge between $*$ and P, if we have a PP relation with two large primes P and Q, we consider the edge between P and Q. If the edge does not yet occur in the graph, we insert it, labeled with the data for the relation.

Remark 1. The graph we construct here can have many cycles. In contrast, the graph constructed in the "full algorithm" in [13] is acyclic.

After having constructed a graph with a sufficiently large connected component containing $*$, we construct what is known as a *shortest path tree* with root $*$.

Definition 2. *Let G be an undirected (unweighted) graph, and let $*$ be a vertex in G. Then a* shortest path tree *with root $*$ is a tree on a subset of the set of vertices of G with the following properties:*

- *The vertices in T are the vertices in the connected component of $*$ in G.*
- *For any vertex V in T, the distance between $*$ and V in G is equal to the distance between $*$ and V in T.*

Notation 1 *The set of vertices of a tree T is also denoted by T.*

It is easy to construct a shortest-path tree algorithmically with the so-called breadth-first search (see [6, Section 22.2]).

As written in Section 2, for every element $c \in \mathrm{Cl}^0(\mathcal{C})$ there is a unique along F_1 maximally reduced effective divisor D such that $[D] - \deg(D) \cdot [F_1] = c$ (here as above, F_1 is the first element of the factor base).

We use this representation of the elements of the degree 0 class group and proceed as in Phase 2 of the "simplified algorithm" in [13]. Provided that the degree 0 class group is cyclic and generated by a this means that we consider random linear combinations of the inputs a and b which we try to express as sums of elements of $\mathcal{F} \cup T$. We then use the tree to substitute the vertices of T involved by sums of (possibly negative) multiples of elements in the factor

base and D_∞. Finally, we solve the DLP with an algorithm from sparse linear algebra.

We are now ready to give the complete algorithm. For simplicity we thereby assume that the group order ℓ is prime. (If the group is cyclic but not of prime order or the group is arbitrary but its structure is known, Steps 5 and 6 should be modified according to the descriptions in [13] and [11].)

The algorithm

Input: A curve $\mathcal{C}/\mathbb{F}_q$, given by a plane model of degree d,
the group order $\ell := \#\mathrm{Cl}^0(\mathcal{C})$ and two elements $a, b \in \mathrm{Cl}^0(\mathcal{C})$ with $\langle a \rangle = \mathrm{Cl}^0(\mathcal{C})$.

1. Enumerate $\mathcal{C}_{ns}(\mathbb{F}_q)$ and choose a factor base $\mathcal{F} = \{F_1, F_2, \ldots\}$ uniformly at random from the set of all subsets of $\mathcal{C}_{ns}(\mathbb{F}_q)$ with $\lceil (4\cdot(d-2)!)^{1/(d-2)} \cdot q^{1-1/(d-2)} \rceil$ elements.
   (If $\mathcal{C}_{ns}(\mathbb{F}_q)$ has fewer elements, terminate.)
2. Construct a graph G on $\mathcal{L} \dot\cup \{*\}$ (where $\mathcal{L} := \mathcal{C}_{ns}(\mathbb{F}_q) - \mathcal{F}$) as follows:
   For all $i < j$ do
   Compute the line L through F_i and F_j.
   If $D := L \cap \mathcal{C}_{pm}$ is contained in $\mathcal{C}_{ns}$
   and splits completely into points of $\mathcal{C}_{ns}(\mathbb{F}_q)$, then
   if it defines an FP or a PP relation, then
   if the corresponding edge does not yet occur in the graph, then
   insert the edge in the graph.
3. Construct a shortest path tree T with root $*$ in G.
4. If T has less than $\frac{1}{\log(q)} \cdot q$ vertices or the depth of T is $> \log^2(q)$, go back to 1.
5. Construct a sparse matrix R over $\mathbb{Z}/\ell\mathbb{Z}$ as follows:
   For $i = 1, \ldots, \#\mathcal{F} + 1$ do
   Repeat
   Choose uniformly and independently randomly α_i and β_i and compute the unique along F_1 maximally reduced effective divisor D with $[D] - \deg(D) \cdot [F_1] = \alpha_i a + \beta_i b$.
   Until D splits into elements of $\mathcal{F} \cup T$.
   Use the tree T to substitute these elements
   by sums of multiples of elements of $\mathcal{F} \cup \{D_\infty\}$.
   If this substitution leads to the relation $\sum_j r_{i,j}[F_j] + r_i[D_\infty] = \alpha_i a + \beta_i b$,
   store $(r_{i,j})_j$ as the i-th row of R.
6. Compute a non-zero vector γ over $\mathbb{Z}/\ell\mathbb{Z}$ with $\gamma R = 0$ with an algorithm from sparse linear algebra.
7. If $\sum_i \gamma_i\beta_i \in (\mathbb{Z}/\ell\mathbb{Z})^*$, let
$$x \longleftarrow -\frac{\sum_i \gamma_i\alpha_i}{\sum_i \gamma_i\beta_i} ,$$
   otherwise go back to 5.
   Output x.

Proposition 1. *If the algorithm outputs x, we have $x \cdot a = b$.*

Proof. With the notation in Steps 5 and 6, we have

$$\sum_i \gamma_i \alpha_i a + \sum_i \gamma_i \beta_i b = \sum_{i,j} \gamma_i r_{i,j}[F_j] + \sum_i \gamma_i r_i [D_\infty] = \sum_i \gamma_i r_i [D_\infty].$$

As $\sum_i \gamma_i \alpha_i a + \sum_i \gamma_i \beta_i b$ has degree 0, we have $\sum_i \gamma_i r_i = 0$, i.e. $\sum_i \gamma_i \alpha_i a + \sum_i \gamma_i \beta_i b = 0$. This implies $x \cdot a = b$. □

4 Heuristic Analysis

The following heuristic analysis is for fixed degree d and fixed genus $g \leq (d-1)(d-2)/2$ and $q \longrightarrow \infty$. We note that even though the genus is bounded if we fix the degree (which suffices for our heuristic analysis), we fix the genus additionally to the degree because we want to derive statements on almost all instances for every fixed degree and genus.

A "randomized" factor base as in Step 1 can be found in an expected time of $\tilde{O}(q)$ as follows:

First all points of $\mathcal{C}_{ns}(\mathbb{F}_q)$ are enumerated. By iterating over the (X, Z)-coordinates and considering the possible Y-coordinates, this can be done in a time of $\tilde{O}(q)$. After this, a factor base as in Step 1 of the algorithm can be constructed by uniformly randomly choosing points of $\mathcal{C}(\mathbb{F}_q)$. The expected running time is then again in $\tilde{O}(q)$.

We now come to the task to analyze the size of the tree T as well as its depth. This task seems to be very difficult, and our analysis relies on several heuristic assumptions. A key technique of our approach is to use the randomization of the factor base and to rely on a heuristic comparison of the graph which is constructed in Step 2 with an appropriate "random graph".

We will use these notations:

Definition 3. *Let $(a_n)_{n\in\mathbb{N}}$ and $(b_n)_{n\in\mathbb{N}}$ be two sequences of real numbers. Then we write*

$$a_n \gtrsim b_n$$

if $\liminf \frac{a_n}{b_n} \geq 1$.

Definition 4. *For $P, Q \in \mathcal{C}_{ns}(\mathbb{F}_q)$ with $P \neq Q$, let p_{PQ} be the probability that $P, Q \in \mathcal{L}$ and the unordered pair $\{P, Q\}$ occurs as an edge in the graph (if we choose the factor base uniformly at random from the set of all factor bases with $\lceil (4 \cdot (d-2)!)^{1/(d-2)} \cdot q^{1-1/(d-2)} \rceil$ elements). Let*

$$p_{av} := \frac{1}{\#\mathcal{C}_{ns}(\mathbb{F}_q) \cdot (\#\mathcal{C}_{ns}(\mathbb{F}_q) - 1)} \cdot \sum_{P,Q \in \mathcal{C}_{ns}(\mathbb{F}_q) \text{ with } P \neq Q} p_{PQ} \ .$$

Note that p_{av} can be seen as the *average probability* that an (unordered) pair of distinct points in $\mathcal{C}_{ns}(\mathbb{F}_q)$ occurs as an edge in the graph.

Lemma 3. *For $P, Q \in \mathcal{C}_{ns}(\mathbb{F}_q)$ with $P \neq Q$ such that the line through P and Q intersects $\mathcal{C}_{pm}$ only in $\mathcal{C}_{ns}$ and the intersection divisor splits completely into a sum of distinct points of $\mathcal{C}_{ns}(\mathbb{F}_q)$, we have*

$$p_{PQ} \sim 4 \cdot (d-2)! \cdot \frac{1}{q} \ .$$

Proof. Let $D = P + Q + R$ be the intersection divisor. Then the probability p_{PQ} is equal to the probability that the factor base contains all $d-2$ points from R and does not contain P and Q.

The probability p_{PQ} is thus

$$\frac{\binom{\#\mathcal{C}_{ns}(\mathbb{F}_q)-d}{\lceil (4\cdot(d-2)!)^{1/(d-2)} \cdot q^{1-1/(d-2)} \rceil - (d-2)}}{\binom{\#\mathcal{C}_{ns}(\mathbb{F}_q)}{\lceil (4\cdot(d-2)!)^{1/(d-2)} \cdot q^{1-1/(d-2)} \rceil}} \ .$$

For $q \longrightarrow \infty$ this is asymptotically equivalent to

$$\left(\frac{(4 \cdot (d-2)!)^{1/(d-2)} \cdot q^{1-1/(d-2)}}{q} \right)^{d-2} = 4 \cdot (d-2)! \cdot \frac{1}{q} \ .$$

□

By the Hasse-Weil bounds, there are $\sim q^d$ divisors of degree d on $\mathcal{C}$ of whose $\sim \frac{1}{d!} q^d$ split completely. The probability that a uniformly randomly chosen divisor on $\mathcal{C}$ of degree d is completely split is thus asymptotically equal to $\frac{1}{d!}$. This motivates:

Heuristic Assumption 1 *For almost all instances, the probability that a uniformly randomly chosen divisor in the linear system $|D_\infty|$ is completely split is $\geq \frac{1}{2} \cdot \frac{1}{d!}$.*

Remark 2. In the case non-hyperelliptic curves of genus 3 (given as plane quartics), it is possible to prove via an effective Chebotarev theorem that the probability that a uniformly randomly chosen divisor in $|D_\infty|$ is completely split is asymptotically equal to $\frac{1}{4!}$. Thus Heuristic Assumption 1 is satisfied in this case (see [10]).

Proposition 2. *Under Heuristic Assumption 1, for almost all instances, $p_{av} \cdot q \gtrsim 2$, and the expected number of edges in the graph of large prime relations is $\gtrsim q$.*

Proof. We restrict ourselves to instances for which Heuristic Assumption 1 is satisfied.

We first note that the number of divisors in $|D_\infty|$ which split completely into sums of distinct points of $\mathcal{C}_{ns}(\mathbb{F}_q)$ is $\gtrsim \frac{1}{2} \frac{1}{d!} \cdot q^2$.

Indeed, the number of completely split divisors is by assumption $\geq \frac{1}{2} \frac{1}{d!} \cdot \frac{q^3-1}{q-1}$. By the formulae for the arithmetic and the geometric genus, the number of singular points in $(\mathcal{C}_{pm})_{\overline{\mathbb{F}}_q}$ is $\leq (d-1)(d-2)/2$, thus the number of lines in $\mathbb{P}^2_{\overline{\mathbb{F}}_q}$

through singular points is in $O(q)$. Moreover, every divisor in $|D_\infty|$ which has the form $\sum_P n_P P$ with $n_P \geq 2$ for some $P \in \mathcal{C}_{ns}(\mathbb{F}_q)$ is defined by a line in $\mathbb{P}^2_{\mathbb{F}_q}$ which is tangential to $\mathcal{C}_{pm}$. This means that the total number of such divisors is also in $O(q)$.

For any divisor D in $|D_\infty|$ which splits completely into a sum of distinct points of $\mathcal{C}_{ns}(\mathbb{F}_q)$, there are $d\cdot(d-1)$ ordered pairs of distinct points in $\mathcal{C}_{ns}(\mathbb{F}_q)$ in the support of the divisor. Each of these pairs of points fulfills the assumptions of Lemma 3 (and conversely, any pair of points fulfilling the assumptions of Lemma 3 determines uniquely such a divisor D). Thus there are $\gtrsim \frac{1}{2\cdot d!}\cdot d(d-1)\cdot q^2 = \frac{1}{2\cdot(d-2)!}\cdot q^2$ ordered pairs of distinct points of $\mathcal{C}_{ns}(\mathbb{F}_q)$ which fulfill the assumption of Lemma 3.

The average probability p_{av} is thus

$$\gtrsim \frac{1}{q^2}\cdot(\frac{1}{2\cdot(d-2)!}\cdot q^2)\cdot(4\cdot(d-2)!\cdot\frac{1}{q}) = \frac{2}{q}\ .$$

If one multiplies the average probability p_{av} by the number of unordered pairs of points of $\mathcal{C}_{ns}(\mathbb{F}_q)$, one obtains the claimed asymptotic lower bound on the expected number of edges. □

It does not seem to be easy to study the number of vertices in the connected component of $*$ of G (which is equal to the number of vertices in the tree T) as well as the depth of the tree.

We note however the following result from the theory of random graphs: Let $G(n,p)$ denote a random graph on n vertices in which each unordered pair of vertices appears (independently of the other pairs of vertices) as an edge with probability p (this is called a *Bernoulli random graph* in [27]). Then we have (see [4, Theorem 6.11] together with [4, Theorem 2.2 a)] as well as [5]):

Proposition 3. *Let $c > 1$ be a constant. Then for $p\cdot n \geq c$, with probability converging to 1 for $n \longrightarrow \infty$, $G(n,p)$ has a "giant connected component" of size $\Theta(n)$, and the diameter of the graph is in $O(\log(n))$.*

We now have the following situation: As in the conclusion of Proposition 2, let $p_{av}\cdot q \gtrsim 2$. Then with probability converging to 1, a random graph $G(\#\mathcal{L}\cup\{*\}, p_{av})$ has a "giant connected component" of size $\Theta(q)$ and diameter $O(\log(q))$.

Clearly, there are three essential differences between Bernoulli random graphs and the situation we have here:

1. In contrast to Bernoulli random graphs, many of the pairs of vertices are never drawn.
2. In contrast to Bernoulli random graphs, the probabilities of two pairs of vertices appearing as edges in the graph are not independent.
3. In contrast to Bernoulli random graphs, we have the "special vertex" $*$ which heuristically occurs in much more edges than the vertices in $\mathcal{L}$.

The analysis now relies on the heuristic assumption that analogous to a random graph $G(\#\mathcal{L} \dot\cup \{*\}, p_{av})$, for almost all instances, "sufficiently often" our graph has a "giant connected component" of "sufficient size" and "sufficiently small" diameter containing $*$. As an approach to cope with possible distortions, we require only that with a probability of $\tilde\Omega(1)$, we have $\geq \frac{1}{\log(q)} \cdot q$ vertices and the maximal distance to $*$ is $\leq \log^2(q)$ (cf. the conditions in Step 4). (The $\tilde\Omega$-notation should be understood analogously to the $\tilde O$-notation.)

The above considerations motivate:

Heuristic Result 2 *For almost all instances, Step 5 of the algorithm is reached after at most $\tilde O(1)$ iterations of 1 – 4.*

As there are $\Theta(q^{2-2/(d-2)})$ iterations within Step 2, this step has a running time of $\tilde O(q^{2-2/(d-2)})$.

With the breadth-first algorithm, given a graph on n vertices with m edges represented by numbers whose bit-length is polynomial in $\log(n)$, a shortest-path tree can be computed in a time of $\tilde O(n+m)$. As the graph clearly contains $O(q)$ vertices and $O(q^{2-2/(d-2)})$ edges, the running time of Step 3 is in $\tilde O(q^{2-2/(d-2)})$.

This means that on the basis of Heuristic Result 2, for almost all instances, Step 5 of the algorithm can be reached in a time of $\tilde O(q^{2-2/(d-2)})$.

Under the assumption that the degree 0 class group is cyclic or the group structure is known, the rest of the algorithm can be analyzed rigorously. For simplicity, as in the description of the algorithm, we stick to the case that the degree 0 class group has prime order ℓ. For modifications for the general case, we refer to [11] and [13].

We have the following general lemma.

Lemma 4. *Let us consider curves $\mathcal{C}$ over $\mathbb{F}_q$ of a fixed genus g together with a point $P_0 \in \mathcal{C}(\mathbb{F}_q)$ and a set of rational points $S \subset \mathcal{C}(\mathbb{F}_q)$ such that $\#S = \tilde\Omega(q)$. Then there are $\tilde\Omega(q^g)$ effective divisors D which split completely into sums of elements of S and are maximally reduced along P_0.*

Proof. If D is a non-special effective divisor of degree g, then the unique effective divisor D' which does not have P_0 in its support and satisfies $D' + (g - \deg(D')) \cdot P_0 = D$ is maximally reduced along P_0. Clearly there are $\tilde\Omega(q^g)$ effective divisors of degree g which split completely into sums of elements of S, and by the Hasse-Weil bounds, there are only $O(q^{g-1/2})$ special divisors of degree g. This is asymptotically negligible against $\tilde\Omega(q^g)$. □

This lemma implies that with a probability of $\tilde\Omega(1)$ one choice of α_i and β_i in Step 5 leads to a divisor D which splits over $\mathcal{F} \cup T$. Step 5 then has an expected running time of $\tilde O(q^{1-1/(d-2)})$.

Because of the condition that the depth of T is $\leq \log^2(q)$, the expected average number of elements in each row of the relation matrix is in $O(\log(q)^2)$. This implies that Step 6 has a running time of $\tilde O(q^{2-2/(d-2)})$. Finally, as argued in [13], $\sum_i \gamma_i \beta_i$ is uniformly randomly distributed over the group $\mathbb{Z}/\ell\mathbb{Z}$.

All in all, we have the following heuristic result:

Heuristic Result 3 *For almost all instances, the DLP in* $\mathrm{Cl}^0(\mathcal{C})$ *can be computed in a time of* $\tilde{O}(q^{2-2/(d-2)})$.

This is essentially the heuristic result stated in the introduction.

We note however that in the introduction we did not assume that the group is cyclic or the group structure is known. We have to make an additional heuristic assumption if the relation generation takes place in a proper subgroup of the degree 0 class group.

5 Practical Aspects

In this section, we briefly discuss some practical aspects of our algorithm and possible variants for concrete computations.

1. For practical purposes it might be advisable not to first construct the graph, then the shortest path tree and then to use this tree to derive relations via random linear combinations of the input values. Instead, one can proceed as follows:
   - First, one computes representatives of multiples αa and βb of the input values a and b which split completely into sums of points of $\mathcal{C}_{ns}(\mathbb{F}_q)$.
   - One chooses the factor base, thereby inserting the points in $\mathcal{C}_{ns}(\mathbb{F}_q)$ for the representatives for αa and βb.
   - One generates relations by considering lines through points of the factor base as described in Section 3 but otherwise one proceeds as in the "full algorithm" of [13]. This means that every time one would obtain a cycle, one does not insert the corresponding edge in the graph but instead tries to use this cycle to obtain a relation over the factor base.
   - One stops if one has found enough "sufficiently light" cycles. Then one solves the DLP via linear algebra.

   With this approach only for the initial computation of multiples of a and b one needs an algorithm for arithmetic in the degree 0 class group. If this initial computation is not time-critical, this might simplify the implementation.

   The approach presented above is particularly advantageous if g is much larger than d (for example if the plane model itself is non-singular and therefore $g = (d-1)(d-2)/2$). Note that the initial computation of multiples of a and b might even dominate the running time.
2. The number of points in the factor base ($\lceil (4 \cdot (d-2)!)^{1/(d-2)} \cdot q^{1-1/(d-2)} \rceil$) was chosen such that we expect the graph of large prime relations to be large enough for fixed degree d and $q \gg 0$. It might be necessary to choose the factor base slightly larger for concrete computations. This applies in particular if one follows the variant presented above.
3. If every pair of points in the factor base is considered to generate the graph, a line through the factor base defining a PP relation is usually considered $\binom{d-2}{2}$ times. To decrease the occurrence of such "repeated selections", it might be advisable to choose the factor base larger than necessary.

4. For the graph of large prime relations to be large enough, one needs at least about q divisors in $|D_\infty|$ which split completely. This implies that *one should have $q > d!$ if one applies the algorithm.* The case $d! \approx q$ can be considered as a boundary case. In this case, one could try to apply the variant presented in Point 1 by choosing the factor base equal to $\mathcal{C}_{ns}(\mathbb{F}_q)$ and ignoring the large prime variation.
5. To reduce the storage requirements, it might be advisable to combine our relation generation with the "simplified algorithm" of [13], i.e. when constructing the graph of large prime relations, one discards all edges which are not connected to $*$. The factor base then has to be enlarged by a logarithmic factor.

6 Finding Plane Models of Degree g+1

In this section we start off with some curve $\mathcal{C}$ of genus ≥ 3 over an "effective field" k. The goal is to find a plane model of degree $g+1$ (provided such a model exists). In order to bound the time for computation of this plane model we assume that the curve $\mathcal{C}$ is initially given by a plane model of bounded degree.

The idea is to define a morphism $\mathcal{C} \longrightarrow \mathbb{P}^2_k$ via a special linear system of degree $g+1$. The case of non-hyperelliptic genus 3 curves is particularly easy: the canonical system $|K|$ itself defines an embedding into $\mathbb{P}^2_k$. For the general case we have the following proposition (see Point (b) in the introduction of [14]):

Proposition 4. *A general linear system of degree d and (projective) dimension ≥ 2 on a general curve of genus g has dimension 2, no base-points and defines a morphism to $\mathbb{P}^2$ which is birational onto its image.*

Here as usual, by a *general curve* we mean a curve which is obtained by base-change from the curve corresponding to the generic point of the (coarse) moduli space $\mathcal{M}_g$. (This space exists by [21, Corollary 7.14.].) A *general effective divisor* of degree d is the divisor on $\mathcal{C}_{k(\mathcal{C}_d)}$ corresponding to the generic point of $\mathcal{C}_d$. Here, following [2] and [14], $\mathcal{C}_d$ denotes the d-fold symmetric power of $\mathcal{C}$.

Let us say that a property holds for *sufficiently general* curves (of a prescribed genus) and / or for *sufficiently general* linear systems of divisors (of a prescribed degree and dimension) if it holds for curves and divisors in an open part of the corresponding moduli space.

We can then conclude that the linear system of any sufficiently general linear system of degree d and dimension ≥ 2 on any sufficiently general curve defines a morphism to $\mathbb{P}^2$ which is birational onto its image. (As usual, the morphism is unique up to an automorphism of $\mathbb{P}^2$.)

Following [2] and [14], let us denote the locus of complete linear systems of degree d and (projective) dimension $\geq n$ (in a twist of the Jacobian) by $W^n_d(\mathcal{C})$. We have the following proposition.

Proposition 5. *Let $\mathcal{C}$ be any curve of genus $g \geq 3$. Then we have birational morphisms*

$$\begin{array}{ccccc} \mathcal{C}_{g-3} & \longrightarrow & W^0_{g-3}(\mathcal{C}) & \longrightarrow & W^2_{g+1}(\mathcal{C}) \\ D & \mapsto & |D| & \mapsto & |K-D|. \end{array}$$

In particular, for any (sufficiently) general effective divisor D of degree $g-3$ on a (sufficiently) general curve $\mathcal{C}$, $|K-D|$ has no base points, (projective) dimension 2 and defines a morphism to $\mathbb{P}^2$ which is birational onto its image.

Proof. The morphism $\mathcal{C}_{g-3} \longrightarrow W^0_{g-3}(\mathcal{C})$, $D \mapsto |D|$ is birational because for any curve, the linear system of any general effective divisor of degree $< g$ has dimension 0.

By the Riemann-Roch theorem ([15, IV, Theorem 1.3]) and the fact that $\deg(K) = 2g-2$, we have an isomorphism $W^0_{g-3}(\mathcal{C}) \longrightarrow W^2_{g+1}(\mathcal{C})$, $|D| \mapsto |K-D|$. □

Remark 3. Not every curve of genus g has a plane model of degree $g+1$. For example, no hyperelliptic curve has such a model.

We have the following method to compute plane models of degree $g+1$:

Computation of a plane model of degree $g+1$

Input: Any curve $\mathcal{C}/k$.

1. Compute a canonical divisor K on $\mathcal{C}$.
2. Select any effective divisor D on $\mathcal{C}$ of degree $g-3$.
3. Compute a basis $b_1, \ldots, b_n$ of the Riemann-Roch space $L(K-D)$.
4. If the basis has more than 3 elements, terminate.
5. Compute a homogeneous polynomial $F(X,Y,Z) \in k[X,Y,Z]$ of minimal degree with $F(b_1, b_2, b_3) = 0$.
6. If $\deg(F) < g+1$, terminate.
7. Output $(F; b_1, b_2, b_3)$.

The necessary computations of divisors and Riemann-Roch spaces can be carried out with the algorithms in [16]. Step 5 can be performed by computing (successively for $i = 1, \ldots, g+1$) the functions $b_1^{i_1} \cdot b_2^{i_2} \cdot b_3^{i_3}$ with $i = i_1 + i_2 + i_3$ and trying to find a linear relation between them. The latter task is a linear algebra problem. Over finite fields we have:

Proposition 6. *There exists a specification of the above method such that for curves over finite fields $\mathbb{F}_q$ initially represented by plane models of bounded degree the expected running time is polynomial in* $\log(q)$.

Example 1. At the end of [8], an elliptic curve $\mathcal{E}$ over $\mathbb{F}_{p^7}$ with $p = 10000019$ is given such that the DLP in $\mathcal{E}(\mathbb{F}_{p^7})$ can be transferred into a DLP in the degree 0 class group of a certain curve $\mathcal{C}$ of genus 7 over $\mathbb{F}_p$. An explicit equation for $\mathcal{C}$ is also given. Using the method outlined above, we computed various degree 8 models of this curve.

Acknowledgments

With great pleasure, I thank G. Frey, P. Gaudry, F. Hess, K. Khuri-Makdisi, J. Pila, J. Scholten, N. Thériault, E. Thomé and E. Viehweg for discussions and comments. I also thank the anonymous referee for suggestions.

References

[1] L. Adleman, J. DeMarrais, and M.-D. Huang. A Subexponential Algorithm for Discrete Logarithms over the Rational Subgroup of the Jacobians of Large Genus Hyperelliptic Curves over Finte Fields. In L. Adleman and M.-D. Huang, editors, *Algorithmic Number Theory – ANTS I*, LNCS, pages 28–40, Berlin, 1994. Springer-Verlag.

[2] E. Arbarello, M. Cornalba, P. Griffiths, and J. Harris. *Geometry of Algebraic Curves*. Springer-Verlag, 1985.

[3] R. Avanzi and N. Thériault. Index Calculus for Hyperelliptic Curves. In H. Cohen and G. Frey, editors, *Handbook of Elliptic and Hyperelliptic Curve Cryptogrpahy*, chapter 21. Chapman & Hall/CRC, Boca Raton, 2006.

[4] B. Bollobas. *Random Graphs*. Cambridge University Press, Cambridge, 2001.

[5] F. Chung and L. Lu. The diameter of sparse random graphs. *Adv. in Appl. Math.*, 26:257–279, 2001.

[6] T. Cormen, C. Leiserson, R. Rivest, and C. Stein. *Introduction to Algorithms*. McGraw-Hill and The MIT Press, 2001. Second Edition.

[7] C. Diem. Index calculus with double large prime variation for arbitrary curves of small genus. Forthcoming.

[8] C. Diem. The GHS Attack in odd Characteristic. *J. Ramanujan Math. Soc.*, 18:1–32, 2003.

[9] C. Diem and J. Scholten. Cover attacks. A report for the AREHCC project, available under http://www.arehcc.com/documents.htm, 2003.

[10] C. Diem and E. Thomé. Index calculus in class groups of non-hyperelliptic curves of genus 3. Forthcoming.

[11] A. Enge and P. Gaudry. A general framework for subexponential discrete logarithm algorithms. *Acta. Arith.*, 102:83–103, 2002.

[12] S. Galbraith and A. Menezes. Algebraic curves and cryptography. *Finite fields and applications*, 11:544–577, 2005.

[13] P. Gaudry, E. Thomé, N. Thériault, and C. Diem. A double large prime variation for small genus hyperelliptic index calculus. accepted for publication in Math. Comp., 2005.

[14] P. Griffiths and J. Harris. On the variety of special linear systems on a general algebraic curve. *Duke Math. J.*, 47(1):233–272, 1980.

[15] R. Hartshorne. *Algebraic Geometry*. Springer-Verlag, New York, 1977.

[16] F. Heß. Computing Riemann-Roch spaces in algebraic function fields and related topics. *J. Symbolic Computation*, 11, 2001.

[17] F. Heß. Weil descent attacks. In G. Seroussi, I. Blake, and N. Smart, editors, *Advances in Elliptic Curve Cryptography*. Cambridge University Press, 2004.

[18] M.-D. Huang and D. Ierardi. Efficient Algorithms for the Riemann-Roch Problem and for Addition in the Jacobian of a Curve. *J. Symbolic Computation*, 18:519–539, 1994.

[19] K. Khuri-Makdisi. Asymptotically fast group operations on Jacobians of general curves. available on the arXiv under math.NT/0409209, 2004.
[20] K. Khuri-Makdisi. Linear algebra algorithms for divisors on an algebraic curve. *Math. Comp.*, 73:333–357, 2004.
[21] D. Mumford. *Geometric Invariant Theory*. Springer-Verlag, Berlin, 1965.
[22] K. Nagao. Improvement of Thériault Algorithm of Index Calculus of Jacobian of Hyperelliptic Curves of Small Genus. Cryptology ePrint Archive, Report 2004/161, http://eprint.iacr.org/2004/161, 2004.
[23] J. Pila. Frobenius maps of abelian varieties and fining roots of unity in finite fields. *Math. Comp.*, 55:745–763, 1990.
[24] J. Pila. Counting points on curves over families in polynomial time. available on the arXiv under math.NT/0504570, 1991.
[25] R. Schoof. Elliptic curves over finite fields and the compuation of square roots mod p. *Math. Comp.*, 44:483–494, 1985.
[26] E. Volcheck. Computing in the Jacobian of a Plane Algebraic Curve. In L. Adleman and M.-D. Huang, editors, *Algorithmic Number Theory – ANTS I*, LNCS, pages 221–233, Berlin, 1994. Springer-Verlag.
[27] N. Wormald. Random Graphs. In I. Gross and J. Yellen, editors, *Handbook of Graph Theory*, chapter 8.2. CRC Press, Boca Raton, 2004.