Index calculus in class groups of non-hyperelliptic curves of genus three

Claus Diem and Emmanuel Thomé

November 13, 2007

Abstract

We study an index calculus algorithm to solve the discrete logarithm problem (DLP) in degree 0 class groups of non-hyperelliptic curves of genus 3 over finite fields. We present a heuristic analysis of the algorithm which indicates that the DLP in degree 0 class groups of non-hyperelliptic curves of genus 3 can be solved in an expected time of $\tilde{O}(q)$. This heuristic result relies on one heuristic assumption which is studied experimentally.

We also present experimental data which show that a variant of the algorithm is faster than the Rho method even for small group sizes, and we address practical limitations of the algorithm.

Key words: Index calculus, non-hyperelliptic curves, class groups, Jacobians

MSC2000: Primary: 11Y16; Secondary 14G50, 94A60

1 Introduction

In addition to the discrete logarithm problem (DLP) in elliptic curves and degree 0 class groups (also named Picard groups or Jacobian groups) of hyperelliptic curves, it has recently been proposed by various authors to use the DLP in degree 0 class groups of *non-hyperelliptic curves of genus 3* over finite fields as a primitive for public-key cryptographic protocols. Particular families of such curves proposed for cryptography include Picard curves [5, 11, 18, 26] and more generally $C_{3,4}$-curves [3, 4]. For these families of curves, effort has been put into providing efficient construction means and explicit formulae for computations in the degree 0 class group. In spite of these efforts, for a fixed group size, the computational requirements for the setup of cryptosystems based on such curves (either directly via point counting or through other construction methods) remains higher than for curves of genus 1 or 2, and the arithmetic in the degree 0 class group remains slower. Moreover, we are not aware of any cryptographic protocol exploiting special properties of non-hyperelliptic curves of genus 3, giving for some purposes a compelling reason to use such curves rather than curves of genus 1 or 2.

We argue in this work that, still in comparison with curves of genus 1 or 2 with a degree 0 class group of comparable group size, both from an asymptotic as well as from a practical standpoint, the discrete logarithm problem is considerably easier or, equivalently, the group size has to be considerably increased in order to keep the same level of security. This implies that given the current knowledge about the discrete logarithm problem in degree 0 class groups of curves, the DLP in degree 0 class groups of non-hyperelliptic curves of genus 3 is not a recommended primitive for public-key cryptographic systems.

In [9], an index calculus algorithm with double large prime variation which is well-suited for the solution of the DLP in class groups of curves over finite fields represented by plane models of small degree has been introduced. Using the fact that any non-hyperelliptic genus 3 curve is isomorphic to a plane quartic, a heuristic analysis of the algorithm in [9] gives rise to:

Asymptotically for $q \longrightarrow \infty$, the DLP in degree 0 class groups of non-hyperelliptic genus 3 curves over $\mathbb{F}_q$ can be solved in an expected time of $\tilde{O}(q)$.

Here, the $\tilde{O}$-notation captures logarithmic factors.

The heuristic expected running time of $\tilde{O}(q)$ should be compared with the expected running time of "generic methods" like the Rho method: As asymptotically for $q \longrightarrow \infty$, degree 0 class groups of genus 3 curves over $\mathbb{F}_q$ have $\sim q^3$ elements, generic methods to solve the DLP have a (heuristic) expected running time of $\Theta(q^{3/2})$ group operations, provided that the group order is nearly prime.

In this work, we study the application of variants of the algorithm in [9] to non-hyperelliptic genus 3 curves in detail. We

- prove a crucial heuristic assumption from the analysis in [9] (Assumption 1 in [9]),
- study the remaining heuristic assumption experimentally,
- present experimental data which show that a practical variant of the algorithm is indeed faster than the Rho method even for relatively small group orders and address practical limitations of this variant on current off-the-shelf hardware.

The algorithm is given in Section 3. The heuristic expected running time is derived in Section 4, with a crucial ingredient of the heuristic analysis being proved in Section 5. In Section 6 the remaining heuristic assumption is studied experimentally. Finally, in Section 7 a variant of the algorithm which is well suited for practical computations is given and studied experimentally.

2 Setting and terminology

This work can be read from the point of view of the exposition in Chapters I and II of [21] as well as from the point of view of scheme theory as in [14]. For Section 5, familiarity with linear systems, function field theory and Galois theory is required.

We use the following terminology and notations, following the usual terminology in scheme theory: The *projective n-space* over a field k is denoted by $\mathbb{P}^n_k$. Given a curve $\mathcal{C}$ (which is always assumed to be projective, smooth and geometrically irreducible) over a perfect field k, a *closed point* P of $\mathcal{C}$ (denoted $P \in \mathcal{C}$) is a Galois orbit of points in $\mathcal{C}(\overline{k})$. If $\lambda|k$ is a field extension, we denote the curve obtained by base-change to λ by $\mathcal{C}_\lambda$ (that is, $\mathcal{C}_\lambda$ is denoted $\mathcal{C}/\lambda$ in [21]). Moreover, when speaking of a *divisor* D on $\mathcal{C}$, we implicitly assume that D is k-rational.

Via the canonical embedding, any non-hyperelliptic genus 3 curve $\mathcal{C}$ over $\mathbb{F}_q$ is isomorphic to a non-singular quartic in $\mathbb{P}^2_{\mathbb{F}_q}$. Conversely any non-singular plane quartic is a non-hyperelliptic genus 3 curve, and the linear system cut out by lines is the canonical linear system [14, Example IV.5.2.1.]. We fix a homogeneous coordinate system X, Y, Z on $\mathbb{P}^2_{\mathbb{F}_q}$ and assume that the curve is given by an equation $F(X, Y, Z) = 0$, where $F(X, Y, Z)$ is a homogeneous polynomial of degree 4. We assume that the order of the degree 0 class group is known. In cryptographic applications this is always the case, and by Pila's variant of Schoof's algorithm [20], it can be computed in polynomial time in $\log(q)$.

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

Let us fix some $P_0 \in \mathcal{C}(\mathbb{F}_q)$; by the Hasse-Weil bound such a point exists if $q \geq 36$. Following [15], we call an effective divisor D on $\mathcal{C}$ *maximally reduced along* P_0 if $D - P_0$ is not linearly equivalent to an effective divisor. By the Riemann-Roch Theorem, maximally reduced effective divisors have degree ≤ 3 (see [15, Proposition 8.2.]). We have a bijection $D \mapsto [D] - [\deg(D)] \cdot [P_0]$ between the divisors maximally reduced along P_0 and the elements of the degree 0 class group.

There are various natural ways to represent divisors and divisor classes on non-hyperelliptic genus 3 curves. To make things precise, we assume that the input elements for the algorithm are given by effective divisors maximally reduced along a fixed point P_0, and the divisors themselves are given in *free representation*, that is, as formal sums of closed points.

It is well known that with this representation the arithmetic in the degree 0 class group can be carried out in randomized polynomial time in $\log q$ (this follows for example from the algorithms in [15] for the computation of Riemann-Roch spaces in a function field / ideal theoretic setting). In [1, 3, 4, 11, 12] efficient special purpose algorithms for various classes of non-hyperelliptic genus 3 curves have been developed. These algorithms are however of no relevance for our work: On the one hand, for a theoretical analysis the algorithms in [15] suffice. On the other hand, for a practical variant of our algorithm we only use the addition in the degree 0 class group to find one multiple of each of the two input elements which is represented by a completely split divisor. For this we only need about 12 additions in total.

3 The algorithm

In this section, we first give an overview about basic strategies for index calculus in the context of non-hyperelliptic genus 3 curves. Then we present one possible algorithm which lends itself well to a heuristic analysis and gives rise to the heuristic complexity result stated in the introduction.

Let $\mathcal{C}$ be a non-singular plane quartic, given by $F(X, Y, Z) = 0$. Let $a, b \in \mathrm{Cl}^0(\mathcal{C})$ with $b \in \langle a \rangle$. The goal is to compute an $x \in \mathbb{N}$ with $x \cdot a = b$. The general approach of *index calculus* for curves of small genus is as follows: One fixes a *factor base* $\mathcal{F} = \{F_1, F_2, \ldots\} \subseteq \mathcal{C}(\mathbb{F}_q)$. Then one generates relations between the elements of $\mathcal{F}$ and a, b (and possibly one other fixed divisor). If one has obtained enough relations, one solves the discrete logarithm problem of a with respect to b with an algorithm from sparse linear algebra.

Let us for simplicity assume that $\#\mathrm{Cl}^0(\mathcal{C})$ is square-free and $\mathrm{Cl}^0(\mathcal{C})$ is generated by a. There are two natural ways to generate relations:

1. Let $P_0 \in \mathcal{C}(\mathbb{F}_q)$ be a fixed point.

   One considers a random linear combination $\alpha a + \beta b$ of the input elements, which one represents by a divisor of the form $D - \deg(D) \cdot P_0$ with D effective and maximally reduced along P_0. This gives rise to the relation

$$[D_{\alpha,\beta}] - \deg(D_{\alpha,\beta}) \cdot [P_0] = \alpha a + \beta b \, . \tag{1}$$

2. Let D_∞ be the intersection of $\mathcal{C}$ with the line $Z = 0$ (with multiplicities); D_∞ is a divisor on $\mathcal{C}$ of degree 4.

   One selects a pair of distinct points $F_i, F_j \in \mathcal{F}$ and considers the line L through F_i and F_j. Let D be the intersection of L with $\mathcal{C}$ (with multiplicities). Then $D = F_i + F_j + D_{i,j}$ with some effective divisor $D_{i,j}$ of degree 2. By construction, D is linearly equivalent to D_∞, and one has the relation

$$[F_i] + [F_j] + [D_{i,j}] - [D_\infty] = 0 \,. \tag{2}$$

In a basic index calculus one would now require that $D_{\alpha,\beta}$ or $D_{i,j}$ split completely into points of the factor base.

We expand the algorithm with a *double large prime variation.* This means that we also use relations which involve up to two elements of $\mathcal{L} := \mathcal{C}(\mathbb{F}_q) - \mathcal{F}$, the set of the so-called *large primes*. Analogously to the usual terminology recalled in [13] we define:

Definition 1. A relation of the form (1) or (2) is called a *Full*, *FP* or *PP relation* if $D_{\alpha,\beta}$ or $D_{i,j}$ splits completely into elements of $\mathcal{C}(\mathbb{F}_q)$ and it contains zero, one or two large primes, respectively.

In a double large prime variation, one considers FP and PP relations as edges in a *graph of large prime relations* on the vertex set $\mathcal{L} \,\dot\cup\, \{*\}$, where $*$ is a special vertex. An FP relation involving one large prime P corresponds to an edge $*$–P, while a PP relation involving two large primes P and Q corresponds to an edge P–Q. As detailed in [13], this graph is used to obtain *recombined relations* involving only elements of the factor base and a, b (and again possibly one other fixed divisor). Again one solves the DLP by sparse linear algebra.

Generating the graph and the recombined relations can be done by using relations of the form (1) or relations of the form (2) (or both). As argued in [9], the use of relations of the form (2), rather than merely the use of relations of the form (1), leads to a considerably faster algorithm. The intuitive reason for this is that $D_{i,j}$ has degree 2 whereas with a probability converging to 1 for $q \longrightarrow \infty$, $D_{\alpha,\beta}$ has degree 3. (This follows from $\mathrm{Cl}^0(\mathcal{C}) \sim q^3$ and the fact that there are only $\sim q^2$ effective divisors of degree 1 and 2.) The use of relations of the form (2) leads however to several stumbling blocks towards a rigorous analysis. Even though with the algorithm presented in this section we try to overcome this difficulty, the forthcoming analysis in Section 4 relies on a heuristic assumption, which is studied experimentally in Section 6.

Recall the following definition.

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

- The vertices in T are the vertices in G connected to $*$.
- Let V be a vertex connected to $*$ in G. The distance from V to $*$ in G is equal to the distance between V and $*$ in T.

Notation 3. The set of vertices of a tree T is also denoted by T.

A shortest-path tree can be constructed by a breadth-first search, described for instance in [8]. A trivial extension allows to build a tree of limited depth: only vertices in G within a fixed distance of the root are considered. This is used in the algorithm below.

The following algorithm is the algorithm presented in [9] applied to plane quartics with the differences that the size of the factor base is reduced by a factor of $\sqrt{2}$, only a tree of depth $\leq \log^2(q)$ is constructed, and the condition in Step 4 is relaxed. As in [9] we assume for simplicity that the degree 0 class group has prime order. If it is not of prime order but cyclic or the group structure is known, one should modify Steps 5 and 6 according to the descriptions in [10] and [13].

The algorithm

Input: A non-hyperelliptic curve $\mathcal{C}$ of genus 3 over $\mathbb{F}_q$, given by a homogeneous equation $F(X,Y,Z) = 0$ of degree 4, the group order $\ell := \#\mathrm{Cl}^0(\mathcal{C})$ (a prime number) and two elements $a, b \in \mathrm{Cl}^0(\mathcal{C})$ $(a \neq 0)$.

1. Enumerate $\mathcal{C}(\mathbb{F}_q)$ and choose a factor base $\mathcal{F} = \{F_1, F_2, \ldots\}$ uniformly at random from the set of all subsets of $\mathcal{C}(\mathbb{F}_q)$ with $\lceil 2\sqrt{q} \rceil$ elements (if $\mathcal{C}(\mathbb{F}_q)$ has fewer elements, terminate). Let $\mathcal{L} := \mathcal{C}(\mathbb{F}_q) - \mathcal{F}$.
2. Construct a graph G on $\mathcal{L} \,\dot\cup\, \{*\}$ as follows:
   For all $i < j$ do
   - Compute the line L through F_i and F_j.
   - Let $D = F_i + F_j + D_{i,j}$ be the intersection divisor of $\mathcal{C}$ with L (with multiplicities).
   - If $D_{i,j}$ splits into points P, Q of $\mathcal{C}(\mathbb{F}_q)$, if at least one of these points lies in $\mathcal{L}$, and if the corresponding edge (P–Q, P–$*$, or $*$–Q) does not yet occur in the graph, insert the edge in the graph.
3. With a breadth-first search, construct a tree T in G with root $*$, limiting the depth to $\log^2(q)$.
4. If T has fewer than $q^{5/6}$ vertices, 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 effective divisor D maximally reduced along F_1 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 replace these elements by sums of elements of $\mathcal{F} \cup \{D_\infty\}$.
   - 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 $x := -\dfrac{\sum_i \gamma_i \alpha_i}{\sum_i \gamma_i \beta_i}$, otherwise go back to 5.
   Output x.

Proposition 4. *If the algorithm outputs x, we have $x \cdot a = b$.*

Proof. Easy; see also [9]. □

4 Heuristic analysis

The enumeration in Step 1 of the algorithm can be performed in an expected time of $\tilde{O}(q)$ by iterating over the (X, Z)-coordinates and considering the possible Y-coordinates. After this, a factor base as in Step 1 can also be found in an expected time of $\tilde{O}(q)$.

Given i, j, the computation of $D_{i,j}$ is an easy algorithmic task. One first computes the line $L : aX + bY + cZ = 0$ through F_i and F_j. Using the equation for L, it is sufficient to compute either the (X, Z) or (Y, Z) coordinates of $D_{i,j}$. Without loss of generality, assume for example that $b = 1$, such that the (X, Z) coordinates of D are the roots of $F(X, -aX - cZ, Z)$, which has degree 4. Since two known roots are the (X, Z) coordinates of F_i and F_j, the remaining (X, Z) coordinates are obtained by solving a quadratic equation. This implies that given i, j, the divisor $D_{i,j}$ can be computed in randomized polynomial time in $\log q$, hence Step 2 can be performed in an expected time of $\tilde{O}(q)$.

The limited breadth-first search in Step 3 has a complexity bounded by the complexity of the complete breadth-first search, which is $\tilde{O}(q)$ (the graph has $O(q)$ vertices, and $O(q)$ edges). Hence Step 3, as previous steps, can be performed in an expected time of $\tilde{O}(q)$.

Let us postpone for later investigation the probability of passing the test in Step 4.

For estimating the complexity of Steps 5–7, we first prove two lemmata yielding the probability that D splits over $\mathcal{F} \cup T$.

Lemma 5. *Let $\mathcal{C}$ be a non-hyperelliptic curve of genus 3 over $\mathbb{F}_q$, let $P_0 \in \mathcal{C}(\mathbb{F}_q)$, and let $S \subset \mathcal{C}(\mathbb{F}_q)$ such that $\#S \in \Omega(q^{5/6})$. Then there are $\Omega(q^{5/2})$ effective divisors D which split completely into sums of elements of S and are maximally reduced along P_0.*

Proof. The assumption on S implies that there are $\binom{\#S+2}{3} \in \Omega(q^{5/2})$ effective divisors of degree 3 on $\mathcal{C}$ which are completely split into sums of elements of S. We wish to estimate the number of such divisors which are maximally reduced along P_0.

For any effective divisor D on $\mathcal{C}$, let D_{red} be the unique divisor maximally reduced along P_0 such that $D_{\text{red}} + (\deg(D) - \deg(D_{\text{red}})) \cdot P_0$ is linearly equivalent to D. The map $D \mapsto D_{\text{red}}$ is injective for non-special divisors D of degree 3. The following lemma shows that the number of special divisors of degree 3 on $\mathcal{C}$ is in $o(q^{5/2})$, which proves the claim. □

Lemma 6. *The number of special linear systems of degree 3 on a curve of genus 3 over $\mathbb{F}_q$ is $\sim q$, and the number of special divisors of degree 3 is $\sim q^2$.*

Proof. Let K be a canonical divisor on $\mathcal{C}$ (e.g. $K = D_\infty$ if the curve is non-hyperelliptic and given by a plane quartic).

For $P \in \mathcal{C}(\mathbb{F}_q)$, the linear system $|K - P|$ is obviously a special linear system of degree 3. On the other hand, if $|D|$ is a special linear system of degree 3, then $|K - D|$ is non-empty. Clearly the point $P \in \mathcal{C}(\mathbb{F}_q)$ with $|P| = |K - D|$ is unique.

We therefore have a bijection between $\mathcal{C}(\mathbb{F}_q)$ and the special linear systems of degree 3 on $\mathcal{C}$. This proves the first assertion.

By the Riemann-Roch Theorem, for every point $P \in \mathcal{C}(\mathbb{F}_q)$, $|K - P|$ has (projective) dimension $\dim(|P|) + 1 = 1$. Together with the first assertion, this implies the second assertion. □

Proposition 7. *Let $\mathcal{C}$ be a non-hyperelliptic genus 3 curve over $\mathbb{F}_q$, $\mathcal{F}$ a factor base of size $O(q^{1/2})$, and T a tree with root $*$ on a subset of $\mathcal{L} \,\dot\cup\, \{*\}$ (with $\mathcal{L} := \mathcal{C}(\mathbb{F}_q) - \mathcal{F}$) such that T*

has $\Omega(q^{5/6})$ vertices and a depth of $O(\log(q)^{O(1)})$. Let us assume that the order of $\mathrm{Cl}^0(\mathcal{C})$ is prime. Let $a, b \in \mathrm{Cl}^0(\mathcal{C})$ with $a \neq 0$. Then following Steps 5–7 of the algorithm, one can solve the DLP with respect to a and b in an expected time of $\tilde{O}(q)$.

Proof. In Step 5 $\alpha, \beta \in \mathbb{Z}/\ell\mathbb{Z}$ is drawn uniformly and independently at random, and therefore so is $\alpha a + \beta b \in \mathrm{Cl}^0(\mathcal{C})$.

By Lemma 5 and the fact that $\#\mathrm{Cl}^0(\mathcal{C}) \sim q^3$, such a combination has then a probability of $\Omega(q^{-1/2})$ to split completely into elements of $\mathcal{F} \cup T$. Furthermore, since the depth of T is in $O(\log(q)^{O(1)})$, we conclude that Step 5 has a complexity of $\tilde{O}(q^{1/2} \cdot \#\mathcal{F}) = \tilde{O}(q)$.

The bound on the depth of T also implies that rows of the relation matrix have no more than $O(\log(q)^{O(1)})$ elements. As the matrix has size $\Theta(\sqrt{q}) \times \Theta(\sqrt{q})$, Step 6 can thus also be performed in an expected time of $\tilde{O}(q)$ (see [10, Theorem 3]).

Finally, as argued in [10], if γ is the vector obtained in Step 6, $\sum_i \gamma_i \beta_i$ is uniformly randomly distributed over the group $\mathbb{Z}/\ell\mathbb{Z}$. Step 7 therefore succeeds with probability $1 - \frac{1}{\ell}$. $\square$

Remark. This proposition also holds if $\mathrm{Cl}^0(\mathcal{C})$ is cyclic or its structure is known provided that the algorithm is modified according to the descriptions in [10] and [13].

Estimating the size of the tree T

It remains to prove that Step 4 of the algorithm is passed with sufficiently high probability. In order to derive the desired result that the expected running time of the algorithm is in $\tilde{O}(q)$, we would need to prove that with a probability of $\Omega(\frac{1}{\log(q)^{O(1)}})$, the set of vertices of the graph of large prime relations which have distance $\leq (\log(q))^2$ to $*$ contains $\geq q^{5/6}$ elements.

We do not know how to prove this result, and therefore our analysis relies on a heuristic comparison with appropriate random graphs in standard models. Recall that random graphs are mostly studied for two models in the literature. The first one is the *Bernoulli* (or *binomial*) random graph $\mathbb{G}(n, p)$. A set of n vertices is fixed, and each unordered pair of distinct vertices appears (independently of the other pairs) with a probability p as an edge of the graph. The second one is the *uniform* random graph $\mathbb{G}(n, m)$. Again, a set of n vertices is fixed, and the set of edges is drawn uniformly from the set of subsets of unordered pairs of distinct vertices with m elements. Here and in the following, we use the notations of [16].

In the following paragraphs, we examine the properties which the graph of large prime relations would enjoy if it were a random graph either in the Bernoulli or uniform model. Let E be the expected number of edges in the graph of large prime relations at the end of Step 2. As a first approach, we compare our graph with a Bernoulli random graph $\mathbb{G}(\#(\mathcal{L} \dot{\cup} \{*\}), p)$, where $p := \frac{E}{\binom{\#(\mathcal{L} \dot{\cup} \{*\})}{2}}$. Note that just as our graph, this random graph has an expected number of E edges.

As usual we call a set S with a fixed point $* \in S$ a *pointed set*. We call a graph on a pointed vertex set $(V, *)$ a *pointed graph*; if G is a graph on V, we denote the corresponding pointed graph by $(G, *)$. These definitions extend naturally to random graphs (where we still view $*$ as being fixed). The following proposition follows from results in [16] and [7]:

Proposition 8. *For two positive constants c_1 and c_2, consider the following properties of graphs G and pointed graphs $(G, *)$ on a set of n vertices:*

- $(\mathcal{Q}_{c_1,c_2})$ *There exists a connected subgraph of G of size $\geq c_1 n$ and diameter $\leq c_2 \log(n)$.*

- ($\mathcal{Q}^*_{c_1,c_2}$) *There exists a connected subgraph of G of size $\geq c_1 n$, diameter $\leq c_2 \log(n)$ containing $*$.*

Let $c > 1$ be a constant. Then there exist positive constants c_1, c_2 such that for $p \geq \frac{c}{n}$ the Bernoulli random graph $\mathbb{G}(n,p)$ satisfies property $\mathcal{Q}_{c_1,c_2}$ with a probability converging to 1 for $n \longrightarrow \infty$ and the pointed Bernoulli random graph $(\mathbb{G}(n,p),*)$ satisfies property $\mathcal{Q}^*_{c_1,c_2}$ with a probability of $\Omega(1)$ for $n \longrightarrow \infty$.

Proof. By [16, Theorem 5.4], there exists a positive constant c_1 such that with a probability converging to 1 for $n \longrightarrow \infty$, the graph has a "giant connected component" of size $\geq c_1 n$. By the results of [7], there exists a positive constant c_2 such that with a probability converging to 1 for $n \longrightarrow \infty$, the graph has diameter $\leq c_2 \log(n)$. This proves the statement on Bernoulli random graphs.

For the second statement, let us first assume that $*$ is chosen uniformly and independently of the other choices (rather than fixed beforehand). Then the statement follows because the probability that property $\mathcal{Q}^*_{c_1,c_2}$ is satisfied is $\geq c_1$ times the probability that property $\mathcal{Q}_{c_1,c_2}$ is satisfied. The statement on pointed Bernoulli random graphs follows because the property $\mathcal{Q}_{c_1,c_2}$ is invariant under graph isomorphism. □

In Section 5, we will prove the following proposition.

Proposition 9. *The expected number E of edges in the graph of large prime relations is $\sim q$. The expected number of edges around vertex $*$ is $\sim \frac{4}{3} q^{1/2}$.*

This proposition implies that the probability p is $\sim \frac{q}{q^2/2} = \frac{2}{q}$. As $\#(\mathcal{L} \,\dot{\cup}\, \{*\}) \sim q$, it is reasonable to assume that the conclusion of Proposition 8 is also satisfied for the graph of large prime relations at the end of Step 2.

We note however that there are of course essential differences between our graph and a Bernoulli random graph $\mathbb{G}(\#(\mathcal{L} \,\dot{\cup}\, \{*\}), p)$. In particular:

- The set of vertices $\mathcal{L} \,\dot{\cup}\, \{*\} = (\mathcal{C}(\mathbb{F}_q) \,\dot{\cup}\, \{*\}) - \mathcal{F}$ of the random graph is not fixed (but its cardinality is), and the set of edges is uniquely determined by the set of vertices.
- Regarded as a graph on $\mathcal{C}(\mathbb{F}_q) \,\dot{\cup}\, \{*\}$, many pairs of vertices are never drawn, and the probability that a particular edge is drawn is not independent of other edges being drawn.
- The expected value of the number of edges around the special vertex $*$ is much larger than for the corresponding Bernoulli random graph.

Note that the third point suggests that with a very large probability $*$ is contained in the largest connected component of the graph. Together with Proposition 8 one might conjecture that there are c_1 and c_2 such that our graph has property $\mathcal{Q}^*_{c_1,c_2}$ with a probability converging to 1 for $q \longrightarrow \infty$. We do however not need this condition for our heuristic analysis.

So far we have considered the expected number of edges and compared the graph of large prime relations with a Bernoulli random graph. To give further heuristic evidence that Step 4 is passed with sufficiently high probability, we now would like to compare our graph with a uniform random graph. Analogously to Bernoulli random graphs we have the following result.

Proposition 10. *Let $c > 1$ be a constant. Then the conclusions of Proposition 8 also hold for the uniform random graph $\mathbb{G}(n,m)$ (and the corresponding pointed uniform random graph) with $m \geq \frac{cn}{2}$.*

Proof. Both properties are monotone. (If $G_1 \subset G_2$ are graphs on the set of vertices $\{1, \ldots, n\}$ and G_1 satisfies $\mathcal{Q}_{c_1,c_2}$ (resp. $\mathcal{Q}^*_{c_1,c_2}$) then G_2 also satisfies $\mathcal{Q}_{c_1,c_2}$ (resp. $\mathcal{Q}^*_{c_1,c_2}$).)

Because of this we only have to prove the statement for $m(n) = \lceil \frac{cn}{2} \rceil$.

Let $p := \frac{m(n)}{\binom{n}{2}}$. We wish to prove the statement for $m(n)$ by applying the general "comparison result" [16, Proposition 1.15] to the Bernoulli random graph $\mathbb{G}(n,p)$ on the one hand and to the uniform random graph $\mathbb{G}(n,m)$ on the other hand.

Note first that as the property $\mathcal{Q}_{c_1,c_2}$ is monotone it is in particular convex (that is, if $G_1 \subseteq G_2 \subseteq G_3$ are three graphs on the set of vertices $\{1, \ldots, n\}$ and both G_1 and G_3 satisfy $\mathcal{Q}_{c_1,c_2}$ then G_2 also satisfies $\mathcal{Q}_{c_1,c_2}$).

Let c_1 and c_2 be such that the first property of Proposition 8 is satisfied. Note that as $p \geq \frac{cn}{n(n-1)} = \frac{c}{n-1} \geq \frac{c}{n}$ this means in particular that the Bernoulli random graph $\mathbb{G}(n,p)$ satisfies property $\mathcal{Q}_{c_1,c_2}$ with a probability converging to 1 for $n \longrightarrow \infty$. With [16, Proposition 1.15] we conclude that $\mathcal{Q}_{c_1,c_2}$ also holds for the uniform random graph $\mathbb{G}(m,n)$ with a probability converging to 1 for $n \longrightarrow \infty$.

Again the statement on $\mathcal{Q}^*_{c_1,c_2}$ follows easily. $\square$

In contrast to Proposition 10, for any $c < 1$, there exists a positive constant c_1 such that with a probability converging to 1 for $n \longrightarrow \infty$, all components of the uniform random graph $\mathbb{G}(n,m)$ with $m \leq \frac{cn}{2}$ contain fewer than $c_1 \log(n)$ vertices. This follows again from [16, Theorem 5.4] together with [16, Proposition 1.15].

This dichotomy of uniform random graphs is called "phase transformation". The following proposition guarantees that with a probability of $\Omega(1)$, the random graphs constructed in the algorithm have a number of edges which is "above the phase transformation", and thus the conclusions of Propositions 8 and 10 apply to the uniform random graph with the same number of vertices and edges. This gives further heuristic evidence that the conclusions of these propositions also apply to the graph of large prime relations at the end of Step 2.

Proposition 11. *If the factor base $\mathcal{F}$ is chosen uniformly at random from the set of all subsets of $\mathcal{C}(\mathbb{F}_q)$ with $\lceil 2q^{1/2} \rceil$ elements, with a probability of $\Omega(1)$ we have more than $\frac{2}{3}q$ edges in the graph of large prime relations.*

Proof. Let c be the number of lines drawn through two factor base elements which give rise to FP or PP relations. Then $c \leq 2(\sqrt{q}+1)^2$. By letting the factor base vary, the quantity $\frac{c}{2(\sqrt{q}+1)^2}$ becomes a random variable with values in $[0,1]$. Let us call this random variable X. Then the probability in question is greater than or equal to $P := \mathbb{P}(X > \frac{1}{3})$. We have

$$\mathbb{E}(X) \leq \mathbb{P}(X \leq \frac{1}{3}) \cdot \frac{1}{3} + \mathbb{P}(X > \frac{1}{3}) \cdot 1 = (1-P)\frac{1}{3} + P,$$

$$\text{hence:} \qquad P \geq \frac{\mathbb{E}(X) - 1/3}{2/3}.$$

By Proposition 9 we have $\mathbb{E}(X) \sim \frac{1}{2}$ for $q \longrightarrow \infty$, thus $\liminf \mathbb{P}(X > \frac{1}{3}) \geq \frac{1/6}{2/3} = \frac{1}{4}$, where the limes inferior is taken over all curves. $\square$

The above comparisons of the graph of large prime relations with Bernoulli and uniform random graphs motivate that the conclusions of Propositions 8 and 10 are valid for the graph of large prime relations at the end of Step 2. The derivation of the complexity result relies on the following weaker assumption.

Heuristic Assumption 12. *With a probability of* $\Omega(\frac{1}{\log(q)^{O(1)}})$*, the set of vertices of the graph of large prime relations which have distance* $\leq (\log(q))^2$ *to* $*$ *contains* $\geq q^{5/6}$ *elements.*

This assumption implies that the expected number of iterations of Steps 1 – 4 until Step 4 is passed is in $O(\log(q)^{O(1)})$. Putting this together with Proposition 7 and the initial arguments of this section, we finally have:

Heuristic Result 13. *One can solve the DLP in degree 0 class groups of non-hyperelliptic genus 3 curves in an expected time of* $\tilde{O}(q)$*, provided that the class group is cyclic or the group structure is known.*

The result holds rigorously for any class of non-hyperelliptic curves of genus 3 for which Heuristic Assumption 12 is satisfied and the class group is cyclic or the group structure is known.

Again on a heuristic basis one should expect this result to hold even if one calculates discrete logarithms in proper subgroups of the degree 0 class groups. One then obtains the heuristic result stated in the introduction: *One can calculate the DLP in degree 0 class groups of non-hyperelliptic genus 3 curves in an expected time of* $\tilde{O}(q)$.

5 On the number of edges in the graph of large prime relations

As above, let $\mathcal{C}$ be a non-hyperelliptic genus 3 curve over $\mathbb{F}_q$, given as a plane quartic.

The purpose of this section is to prove Proposition 9 in the previous section. For this, we first derive the following result.

Proposition 14. *The number of lines in* $\mathbb{P}^2_{\mathbb{F}_q}$ *intersecting the curve in 4 distinct* $\mathbb{F}_q$*-rational points is in*

$$\frac{1}{24}q^2 + O(q^{3/2}) \, .$$

Remark. A reformulation of this proposition is:

If we choose a tuple of distinct points of $\mathcal{C}(\mathbb{F}_q)$ uniformly at random, the probability that the line running through P and Q intersects $\mathcal{C}$ in 4 distinct $\mathbb{F}_q$-rational points is in

$$\frac{1}{2} + O(q^{-1/2}) \, .$$

Indeed, by the Hasse-Weil bound, there are $\sim q^2$ ordered tuples of points of $\mathcal{C}(\mathbb{F}_q)$, and every line which intersects $\mathcal{C}$ in 4 distinct $\mathbb{F}_q$-rational points is defined by $4 \cdot 3 = 12$ different such tuples.

The *proof* is based on an effective Chebotarev density theorem in the "geometric" or "function field theoretic" setting.

As above, let D_∞ be the intersection of $\mathcal{C}$ with $Z = 0$. Let us now fix a point $P \in \mathcal{C}(\mathbb{F}_q)$. We wish to estimate the number of lines over $\mathbb{F}_q$ intersecting the curve $\mathcal{C}$ in 4 distinct $\mathbb{F}_q$-rational points one of which is P.

By associating to each line in $\mathbb{P}^2_{\mathbb{F}_q}$ over $\mathbb{F}_q$ its intersection divisor with $\mathcal{C}$ one obtains a bijection between the set of lines in $\mathbb{P}^2_{\mathbb{F}_q}$ over $\mathbb{F}_q$ and the linear system $|D_\infty|$, the canonical linear system. Now, under this bijection, the lines passing through P are in bijection with the divisors in $|D_\infty|$ containing P. By subtracting P, these divisors are in turn in bijection with the divisors in the complete linear system $|D_\infty - P|$ (over $\mathbb{F}_q$). We thus wish to estimate the number of completely split divisors in this linear system.

The complete linear system $|D_\infty - P|$ is base-point free, has degree 3 and (projective) dimension 1. It thus gives rise to a covering $\mathcal{C} \longrightarrow \mathbb{P}^1_{\mathbb{F}_q}$ of degree 3 (unique up to an automorphism of $\mathbb{P}^1_{\mathbb{F}_q}$). We recall that the divisors in $|D_\infty - P|$ are (by definition of the covering $\mathcal{C} \longrightarrow \mathbb{P}^1_{\mathbb{F}_q}$) exactly the preimages of the points in $\mathbb{P}^1(\mathbb{F}_q)$. We have the following proposition.

Proposition 15.

- *If the covering $\mathcal{C} \longrightarrow \mathbb{P}^1_{\mathbb{F}_q}$ associated to $|D_\infty - P|$ has an automorphism of order 3, the number of completely split divisors in $|D_\infty - P|$ is in*

$$\frac{1}{3}q + O(q^{1/2}) \; .$$

- *If the covering $\mathcal{C} \longrightarrow \mathbb{P}^1_{\mathbb{F}_q}$ does not have an automorphism of order 3 but the covering $\mathcal{C}_{\overline{\mathbb{F}}_q} \longrightarrow \mathbb{P}^1_{\overline{\mathbb{F}}_q}$ over $\overline{\mathbb{F}}_q$ has such an automorphism, there are no completely split divisors in $|D_\infty - P|$.*
- *If the covering $\mathcal{C}_{\overline{\mathbb{F}}_q} \longrightarrow \mathbb{P}^1_{\overline{\mathbb{F}}_q}$ over $\overline{\mathbb{F}}_q$ does not have an automorphism of order 3, then the number of completely split divisors in $|D_\infty - P|$ is in*

$$\frac{1}{6}q + O(q^{1/2}) \; .$$

Proof. Note first that the extension $\mathbb{F}_q(\mathcal{C})|\mathbb{F}_q(\mathbb{P}^1)$ is separable. Indeed, as the extension degree is prime, if it was not separable it would be purely inseparable. This would however imply that $g(\mathcal{C}) = g(\mathbb{P}^1_{\mathbb{F}_q}) = 0$ (see [23, III.9.2 (c) (2) and (3)] or [14, IV, Proposition 2.5]).

The completely split divisors in $|D_\infty - P|$ are in bijection with the elements in $\mathbb{P}^1(\mathbb{F}_q)$ which are completely split in $\mathcal{C}$.

We consider the three cases in the statement separately.

Let us first assume that the covering $\mathcal{C} \longrightarrow \mathbb{P}^1_{\mathbb{F}_q}$ has an automorphism of order 3. Then it is a Galois covering in the sense that the associated extension of function fields is Galois. By the effective Chebotarev theorem in [19] the number of elements in $\mathbb{P}^1(\mathbb{F}_q)$ which are completely split is in $\frac{1}{3}q + O(q^{1/2})$.

Now let us assume that we are in either the second or the third case. Let M be the Galois closure of the extension of function fields $\mathbb{F}_q(\mathcal{C})|\mathbb{F}_q(\mathbb{P}^1)$. Then $M|\mathbb{F}_q(\mathbb{P}^1)$ is a Galois extension with Galois group isomorphic to S_3. Moreover, let L be the unique quadratic extension of $\mathbb{F}_q(\mathbb{P}^1)$ in M. In the second case, we have $L = \mathbb{F}_{q^2}(\mathbb{P}^1)$ and $M = \mathbb{F}_{q^2}(\mathcal{C})$, and in the third case, $\mathbb{F}_q$ is the exact constant field of M.

In the following, we use the same notation for closed points of the curves and the corresponding places of the function fields.

For a place Q of $\mathbb{F}_q(\mathbb{P}^1)$ of degree 1 which is unramified in $\mathbb{F}_q(\mathcal{C})$ (and thus in M), we have the following possibilities:

- Q splits completely in M and thus also in $\mathbb{F}_q(\mathcal{C})$ and L.

- Q splits as Q_1+Q_2 in M, where the Q_i are distinct places of degree 3. Then the splitting fields of both Q_1 and Q_2 are L (the unique subfield of M of index 3). In particular, Q is inert in $\mathbb{F}_q(\mathcal{C})$.

- Q splits as $Q_1+Q_2+Q_3$ in M, where the Q_i are distinct places of degree 2. Then the decomposition groups of Q_1, Q_2, Q_3 are of order 2, and (as always in a Galois extension) they form a complete conjugacy class. This means that these groups are equal to the three subgroups of order 2. Thus exactly one of the decomposition fields is equal to $\mathbb{F}_q(\mathcal{C})$. In particular, Q splits as P_1+P_2 in $\mathbb{F}_q(\mathcal{C})$ where P_1 and P_2 are places of degree 1 and 2 respectively.

Note here that Q cannot be inert in M because $M|\mathbb{F}_q(\mathbb{P}^1)$ is not cyclic. We see in particular that Q splits completely in $\mathbb{F}_q(\mathcal{C})$ if and only if it splits completely in M.

Now, if $L=\mathbb{F}_{q^2}(\mathbb{P}^1)$ (second case), every place of $\mathbb{F}_q(\mathbb{P}^1)$ of degree 1 is inert in L, thus it cannot split completely in M. If $\mathbb{F}_q$ is the exact constant field of M (third case), again by the effective Chebotarev theorem in [19], the number of places of $\mathbb{F}_q(\mathbb{P}^1)$ of degree 1 which are completely split in M is in $\frac{1}{6}q+O(q^{1/2})$ (the other two numbers corresponding to the items above are in $\frac{1}{3}q+O(q^{1/2})$ and $\frac{1}{2}q+O(q^{1/2})$ respectively). □

By [22], the cardinality of the automorphism group of $\mathcal{C}_{\overline{\mathbb{F}}_q}$ is bounded by $16\cdot g(\mathcal{C})^4=1296$. There are thus at most 647 subgroups of the automorphism group of order 3. This implies that there are, up to automorphisms of $\mathbb{P}^1_{\overline{\mathbb{F}}_q}$, at most 647 distinct coverings $\mathcal{C}_{\overline{\mathbb{F}}_q}\longrightarrow\mathbb{P}^1_{\overline{\mathbb{F}}_q}$ of order 3 with an automorphism of order 3. In terms of linear systems this means that there are at most 647 base point free linear systems of degree 3 and dimension 1 on $\mathcal{C}_{\overline{\mathbb{F}}_q}$ which define a covering to $\mathbb{P}^1_{\overline{\mathbb{F}}_q}$ which has a non-trivial automorphism. This means that there are at most $647\in O(1)$ points $P\in\mathcal{C}(\mathbb{F}_q)$ for which in Proposition 15 we are in the first or second case.

We are interested in the number of divisors in $|D_\infty|$ which split completely into 4 distinct $\mathbb{F}_q$-rational points.

Every divisor in $|D_\infty|$ which contains a double point is defined by a line which is tangential to the curve. This implies that there are at most $\#\mathcal{C}(\mathbb{F}_q)\sim q$ such divisors. Keeping in mind that every divisor in the canonical system $|D_\infty|$ which splits into 4 distinct $\mathbb{F}_q$-rational points occurs in exactly 4 systems of the form $|D_\infty-P|$, we obtain:

Proposition 16. *The number of divisors in $|D_\infty|$ which split into 4 distinct $\mathbb{F}_q$-rational points is in*

$$\frac{1}{24}q^2+O(q^{3/2})\ .$$

Proposition 14 is a reformulation of this proposition. We now show how one can use Proposition 14 to derive Proposition 9.

Definition 17. For some line L in $\mathbb{P}^2_{\mathbb{F}_q}$ for which the intersection with the curve $\mathcal{C}$ consists of $\mathbb{F}_q$-rational points, let

$$a_L, b_L, c_L$$

be the probabilities (over the possible factor base choices) that L gives rise to a Full, FP or PP relation respectively.

The following proposition contains slightly more information than we need for the proof of Proposition 9.

Proposition 18. *For all lines L for which the intersection with $\mathcal{C}$ consists of 4 distinct $\mathbb{F}_q$-rational points, we have:*

$$a_L \sim \frac{16}{q^2}\ ,\ \ b_L \sim \frac{32}{q^{3/2}}\ ,\ \ c_L \sim \frac{24}{q}\ .$$

Moreover, for all lines L for which the intersection with $\mathcal{C}$ consists of 3 distinct $\mathbb{F}_q$-rational points (one of which is a double point), we have

$$a_L \sim \frac{8}{q^{3/2}}\ ,\ b_L \sim \frac{12}{q}\ ,\ c_L = 0\ .$$

Proof. Let L be a line intersecting $\mathcal{C}$ in s distinct $\mathbb{F}_q$-rational points. The probability that a factor base $\mathcal{F}$ is chosen such that $\#(L \cap \mathcal{F}) = r$ is

$$\frac{1}{\binom{\#\mathcal{C}(\mathbb{F}_q)}{\lceil 2q^{1/2}\rceil}}\binom{s}{r}\binom{\#\mathcal{C}(\mathbb{F}_q)-s}{\lceil 2q^{1/2}\rceil - r} \in \binom{s}{r}\left(\frac{2q^{1/2}+O(1)}{q+O(\sqrt{q})}\right)^r$$
$$\subseteq \binom{s}{r}\left(\frac{2q^{1/2}}{q}\left(1+O(q^{-1/2})\right)\right)^r \subseteq \binom{s}{r}\left(2q^{-1/2}\right)^r\cdot\left(1+O(q^{-1/2})\right)\ .$$

For the line L to be considered by the relation collection step, we must have $r \geq 2$. Therefore, when $s-r$ is 0, 1 or 2 respectively, the estimate above gives the probabilities a_L, b_L, c_L, except that c_L is 0 if $s=3$. (Note for the last assertion that in the algorithm only lines which pass through two different points of the factor base are considered.) The resulting equivalents follow. □

Proposition 9 is now a consequence of the following proposition.

Proposition 19. *If we choose the factor base $\mathcal{F}$ uniformly at random from the set of all subsets of $\mathcal{C}(\mathbb{F}_q)$ with $\lceil 2q^{1/2}\rceil$ elements, the expected values B, C of FP, PP relations satisfy asymptotically for $q \longrightarrow \infty$:*

$$B \in \frac{4}{3}q^{1/2}(1+O(q^{-1/2}))\ ,\ \ C \in q(1+O(q^{-1/2}))\ .$$

Proof. We have

$$B = \sum_L b_L\ ,\ C = \sum_L c_L\ ,$$

where the sums run over all lines in $\mathbb{P}^2_{\mathbb{F}_q}$ which intersect the curve in $\mathbb{F}_q$-rational points. Combining Proposition 14, the proof of Proposition 18, and the fact that the number of lines whose intersection with the curve contain a double point is in $O(q)$, the claims follow. □

6 Experimental study of the heuristic assumption

The analysis in Section 4 relies on Heuristic Assumption 12. In order to test this assumption, for each of the base fields $\mathbb{F}_{2^{19}}$ to $\mathbb{F}_{2^{24}}$, we built 160 graphs of large prime relations, built from 10 (pseudo-)randomly chosen factor bases over 16 random curves of genus 3 given by arbitrary plane quartics. We thereby discarded FP relations, that is, we only considered PP relations. The reason for this is that the number of FP relations is asymptotically negligible, but FP relations might lead to distortions which hide phenomena occurring for $q \longrightarrow \infty$.

We made a comparison with the same number of instances of Bernoulli random graphs $\mathbb{G}(q,p)$, where $p = \frac{2}{q}$. As discussed after Proposition 9, such graphs have an expected number of edges $\sim q$. The graph instances are constructed by first choosing the number of edges according to the corresponding binomial distribution (approximated by a normal distribution), and then picking an instance of a uniform random graph.

Note that a comparison following more closely the statement of Heuristic Assumption 12 would be to construct, for each curve, instances of the Bernoulli random graph $\mathbb{G}(\mathcal{L} \dot\cup \{*\}, p)$, where p is such that the expected number of edges is exactly the same as the number C defined in Proposition 19. However, for computing exactly C we would have to determine the exact number of lines in $\mathbb{P}^2_{\mathbb{F}_q}$ which lead to completely split divisors; we are not aware of any sufficiently fast method for this task.

Our comparison criteria are both derived from the properties stated in Proposition 8, as well as the usage of the graph of large prime relations in the algorithm.

Given a graph G and a vertex x, we define:

$$\begin{aligned} N_k(x) &= \{y \in G,\ d_G(x,y) \leq k\}\,, \\ \mathsf{tree\_depth}(x,S) &= \min\{k \mid \#N_k(x) \geq S\}\,, \\ \mathsf{cc\_depth}(x) &= \max\{k \mid N_k(x) \supsetneq N_{k-1}(x)\}\,. \end{aligned}$$

It is easily seen that when x belongs to a connected component Γ, we have

$$\mathsf{cc\_depth}(x) \leq \mathrm{diameter}(\Gamma) \leq 2\mathsf{cc\_depth}(x).$$

Based on this, we use cc_depth as a rough (indirect) measure of the diameter of the giant connected component.

Furthermore, we also measure $\mathsf{tree\_depth}(x, q^{5/6})$, as the accordance of this quantity between the graph of large prime relations and the random graph case ensures the success of Step 4 of the algorithm of Section 3.

Table 1 gathers these measurements. For each set of graphs, as well as for the corresponding instances of Bernoulli random graphs, we give the extremal values as well as the observed mean for $\mathsf{tree\_depth}(x, q^{5/6})$ and $\mathsf{cc\_depth}(x)$ for vertices x picked at random within the giant connected component. The size of the giant connected component is also given, in millions of vertices. Finally, we give the average number of edges present in the graphs, as a ratio compared to the expected value q. Table 1 shows no noticeable deviation between the graph of large prime relations and the corresponding random graph.

7 Practical aspects and computations

For practical implementation of the algorithm, the following modifications were made.

- The factor base is not chosen at random. Instead, we pick all the $\mathbb{F}_q$-rational points whose abscissa has an integer representation within the interval $[0, B]$, where B is $\lceil 2\sqrt{q}\rceil$ or a nearby bound (experiments were made with $B = \lceil \frac{4}{3}\sqrt{2q}\rceil$, which was sufficient). Then multiples of the input elements are computed which are represented by completely split divisors. (For this, about 6 multiples have to be computed for each input element.) The corresponding points of $\mathcal{C}(\mathbb{F}_q)$ are inserted into the factor base.

q		tree_depth$(x, q^{5/6})$	cc_depth(x)	giant c.c. (in millions)	#edges$/q$
2^{19}	real	14 . . . **16.8** . . . 26	29 . . . **34.7** . . . 47	0.41 . . . **0.42** . . . 0.42	0.99 . . . **0.99** . . . 1.00
	random	14 . . . **16.8** . . . 26	30 . . . **34.5** . . . 42	0.41 . . . **0.42** . . . 0.42	0.99 . . . **0.99** . . . 1.00
2^{20}	real	15 . . . **17.6** . . . 28	32 . . . **36.4** . . . 49	0.83 . . . **0.83** . . . 0.84	0.99 . . . **1.00** . . . 1.00
	random	15 . . . **17.6** . . . 28	32 . . . **36.5** . . . 48	0.83 . . . **0.83** . . . 0.83	0.99 . . . **1.00** . . . 1.00
2^{21}	real	15 . . . **18.4** . . . 30	33 . . . **38.2** . . . 48	1.67 . . . **1.67** . . . 1.67	1.00 . . . **1.00** . . . 1.00
	random	16 . . . **18.4** . . . 29	33 . . . **38.1** . . . 48	1.66 . . . **1.67** . . . 1.67	1.00 . . . **1.00** . . . 1.00
2^{22}	real	16 . . . **19.2** . . . 28	35 . . . **39.8** . . . 51	3.33 . . . **3.34** . . . 3.34	1.00 . . . **1.00** . . . 1.00
	random	16 . . . **19.1** . . . 28	35 . . . **39.7** . . . 50	3.33 . . . **3.34** . . . 3.34	1.00 . . . **1.00** . . . 1.00
2^{23}	real	17 . . . **20.0** . . . 29	37 . . . **41.7** . . . 53	6.67 . . . **6.68** . . . 6.68	1.00 . . . **1.00** . . . 1.00
	random	17 . . . **20.0** . . . 34	36 . . . **41.6** . . . 55	6.67 . . . **6.67** . . . 6.68	1.00 . . . **1.00** . . . 1.00
2^{24}	real	18 . . . **20.8** . . . 31	39 . . . **43.3** . . . 53	13.35 . . . **13.36** . . . 13.37	1.00 . . . **1.00** . . . 1.00
	random	18 . . . **20.8** . . . 29	38 . . . **43.3** . . . 52	13.35 . . . **13.36** . . . 13.36	1.00 . . . **1.00** . . . 1.00

Table 1: Comparison of the graph of large prime relations with a Bernoulli random graph

- The relations used for the matrix construction in Step 5 of the algorithm are the same as relations used for building the graph (following the non-simplified algorithm in [13]).
- As the problem size grows, it becomes cumbersome to deal with the whole graph for memory reasons. Since the previous modification implies that we are interested in *cycles* occurring in this graph (as described in [13]), we first perform a "filtering" pass: All PP relations are gathered, and used to identify a smaller set of relations containing a smaller subgraph with sufficiently many cycles. This "filtering" step is done in the spirit of e.g. [6].

We have been able to carry out discrete logarithm computations in the degree 0 class group of the $C_{3,4}$ curve defined by $Y^4 + Y^3 + Y^2 + X^2Y + X^3 + X + 1 = 0$ over the field $\mathbb{F}_{2^{31}}$. Choosing a curve with a model defined over $\mathbb{F}_2$ avoids the problem of computing the group order, which is readily obtained. The group order has 93 bits, and it has a 90-bit prime factor.

The implementation has been carried out in `C/C++` and run on 2.4GHz Opteron processors. A pair of distinct points $F_i, F_j \in \mathcal{F}$ is processed in 3.4 microseconds, yielding the satisfying pace of 6.7 microseconds per PP relation. This is the only step of the algorithm which is sensible to the choice of the curve, and if the $C_{3,4}$ curve is replaced by a random non-singular plane quartic, a PP relation is produced in 8.1 microseconds on average. In comparison, a step of the algorithm in [13] applied to hyperelliptic curves of genus 3 is performed in 5.0 microseconds on the same hardware, but succeeds in producing a PP relation only with a probability of roughly $\frac{2\#\mathcal{F}}{q}$.

Most processes dealing with the relations produced and the graph of large prime relations are dominated by the input/output costs, as indicated by Table 2. Indeed, the graph considered has roughly $2 \cdot 10^9$ edges, and about as many vertices. This motivates the prime need for reduction of the graph to a smaller subgraph containing sufficiently many cycles. We isolated roughly 380 million relations, yielding about 200 000 recombined relations. This was more than enough, and made it possible to select only the lightest relations.

We eventually produced a $87\,803 \times 87\,803$ matrix with an average of 352 non-zero coefficients per row. The linear system has been solved using the block Wiedemann algorithm in just below a day, using 4 dual-CPU machines. The solutions were checked using MAGMA.

In comparison with this index calculus experiment, we extrapolate on the feasibility of such an attack using Pollard's Rho method, or the parallel collision search algorithm from [25]. For

Stage	Wall-clock time	Time spent on I/O
Relation collection (1 CPU)	8h	$\approx 50\%$
Relation filtering (1 CPU)	1day	$\geq 95\%$
Linear algebra (4×2 CPUs)	1day	$\leq 5\%$

Table 2: Running times for different stages of the computation (one $C_{3,4}$ curve over $\mathbb{F}_{2^{31}}$ with a model over $\mathbb{F}_2$)

such an attack, fast arithmetic in the degree 0 class group is required. Let us count only field multiplications: algorithms from [12, 3] require between 130 and 170 multiplications per operation in the degree 0 class group.[1] Approximately $\sqrt{\pi\#G/2}$ operations in the degree 0 class group G would be required to compute one discrete logarithm, hence at least $1.6\cdot 10^{16}$ field multiplications. In comparison, our implementation requires on average 86 field multiplications to obtain one PP relation (and no exceptional effort has been put into trimming down this number), therefore the total cost of the relation collision step is $1.7\cdot 10^{11}$ field multiplications. This implies that the parallel collision search method can be expected to require about 10^5 times as much CPU time as the relation collection step in our implementation, hence an estimated cost of about 370 000 hours on one CPU.

It should be noted that Pollard's Rho method is also surpassed by the presented algorithm even for tiny experiments. Over the field $\mathbb{F}_{2^{17}}$, all the steps of the discrete logarithm computation by the index calculus approach can be performed in approximately 5 seconds, while Pollard's Rho method would require approximately 10 minutes.

We wish to extrapolate from our index calculus computation to the feasibility of computations in larger groups. The limiting factor is thereby that we allow resources (in hardware and time) comparable to the latest factorization record: the factorization of RSA-200 with the General Number Field Sieve [2]. Note that for this record, both for the relation collection and the linear algebra, only off-the-shelf hardware was used. (According to [2], the relation collection could have been performed on a single 2.2 GHz AMD Opteron CPU in 55 years. The linear algebra took place on a cluster of 40 dual-CPU 2.2 GHz AMD Opteron computers connected with gigabit ethernet and took three months.)

The relation collection step scales with no difficulty. For a field size of $q = 2^{37}$ (hence a group size near 2^{111}), it could be completed in just above three weeks on one machine, including input/output overhead. However the overhead induced by relation filtering and the expectable overhead of linear algebra are not so easily overcome. The amount of PP relations to be considered ($1.3\cdot 10^{11}$) and the size of the linear system to be solved (740 000 unknowns) are comparable in magnitude to recent works. The above mentioned factorization records handled $3\cdot 10^9$ partial relations, and for the records for finite field discrete logarithms [17, 24], linear systems of this size have already been solved.

These records indicate that taking into account the overhead for managing the data size, the presented algorithm can probably be employed until approximately a group size of 2^{111}, using hardware and time comparable to the resources used in the factorization of RSA-200.

[1] These figures are valid for odd characteristic. We assume that the cost for characteristic 2 would be similar.

Acknowledgments

It is a great pleasure to thank G. Frey, P. Gaudry, F. Heß, R. van der Hofstad, W. König, K. Magaard, C. Ritzenthaler, N. Thériault and E. Viehweg for discussions and helpful comments.

References

[1] F. Abu Salem and K. Khuri-Makdisi. Fast Jacobian group operations for $C_{3,4}$ curves over a large finite field. *LMS J. Comput. Math.*, pages 307 – 328, 2006.

[2] F. Bahr, M. Böhm, J. Franke, and T. Kleinjung. Factorization of RSA-200 by GNFS, May 2005. Unpublished electronic mail.

[3] A. Basiri, A. Enge, J.-C. Faugère, and N. Gürel. Implementing the arithmetic of $C_{3,4}$-curves. In *Algorithmic Number Theory — ANTS VI*, Lecture Notes in Comput. Sci., pages 87–101, Berlin, 2004. Springer–Verlag.

[4] A. Basiri, A. Enge, J.-C. Faugère, and N. Gürel. The arithmetic of Jacobian groups of superelliptic cubics. *Math. Comp.*, 74(249):389–410, 2005.

[5] M. Bauer, E. Teske, and A. Weng. Point counting on Picard curves in large characteristic. *Math. Comp.*, 74(252):1983–2005, 2005.

[6] S. Cavallar. Strategies in filtering in the number field sieve. In W. Bosma, editor, *Algorithmic Number Theory — ANTS-IV*, volume 1838 of *Lecture Notes in Comput. Sci.*, pages 209–231. Springer–Verlag, 2000.

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

[8] T. Cormen, C. Leiserson, R. Rivest, and C. Stein. *Introduction to algorithms.* MIT Press – McGraw-Hill, 2001. Second edition.

[9] C. Diem. An Index Calculus Algorithm for Plane Curves of Small Degree. In F. Hess, S. Pauli, and M. Pohst, editors, *Algorithmic Number Theory — ANTS VII*, LNCS 4076, pages 543 – 557, Berlin, 2006. Springer–Verlag.

[10] A. Enge and P. Gaudry. A general framework for subexponential discrete logarithm algorithms. *Acta Arith.*, 102(1):83–103, 2002.

[11] S. Flon and R. Oyono. Fast arithmetic on Jacobians of Picard curves. In F. Bao et al., editors, *Advances in Cryptology — PKC 2004*, volume 2947 of *Lecture Notes in Comput. Sci.*, pages 55–68, Berlin, 2004. Springer–Verlag.

[12] S. Flon, R. Oyono, and C. Ritzenthaler. Fast addition on non-hyperelliptic genus 3 curves. IACR Eprint report 2004/118, available at `http://eprint.iacr.org/2004/118`, 2004.

[13] P. Gaudry, E. Thomé, N. Thériault, and C. Diem. A double large prime variation for small genus hyperelliptic index calculus. *Math. Comp.*, 76(257):475–492, 2007.

[14] R. Hartshorne. *Algebraic Geometry*, volume 52 of *Grad. Texts in Math.* Springer–Verlag, 1977.

[15] F. Heß. Computing Riemann-Roch spaces in algebraic function fields and related topics. *J. Symbolic Comput.*, 33(4):425–445, April 2002.

[16] S. Janson, T. Luczak, and A. Rucinski. *Random Graphs.* Wiley, New York, 2000.

[17] A. Joux and R. Lercier. Discrete logarithms in GF(p) — 130 digits. Electronic mail to the `NMBRTHRY` mailing list. Available at `http://listserv.nodak.edu/archives/nmbrthry.html`, June 2005.

[18] K. Koyke and A. Weng. Construction of CM-Picard curves. *Math. Comp.*, 74(249):499–518, 2005.

[19] V. K. Murty and J. Scherk. Effective versions of the Chebotarev density theorem for function fields. *C. R. Acad. Sci. Paris Sér. I Math.*, 319:523–528, 1994.

[20] J. Pila. Frobenius maps of abelian varieties and finding roots of unity in finite fields. *Math. Comp.*, 55(192):745–763, October 1990.

[21] J. H. Silverman. *The arithmetic of elliptic curves*, volume 106 of *Grad. Texts in Math.* Springer–Verlag, 1986.

[22] H. Stichtenoth. Über die Automorphismengruppe eines algebraischen Funktionenkörpers von Primzahlcharakteristik. I. Eine Abschätzung der Ordnung der Automorphismengruppe. *Arch. Math.*, 24:527–544, 1973.

[23] H. Stichtenoth. *Algebraic function fields and codes.* Universitext. Springer–Verlag, 1993.

[24] E. Thomé. Computation of discrete logarithms in $\mathbb{F}_{2^{607}}$. In C. Boyd and E. Dawson, editors, *Advances in Cryptology — ASIACRYPT 2001*, volume 2248 of *Lecture Notes in Comput. Sci.*, pages 107–124. Springer–Verlag, 2001.

[25] P. C. van Oorschot and M. J. Wiener. Parallel collision search with cryptanalytic applications. *J. Cryptology*, 12:1–28, 1999.

[26] A. Weng. A Low-Memory Algorithm for Point Counting on Picard Curves. *Des. Codes Cryptography*, 38:383–393, 2005.