Discrete logarithm in elliptic curves over extension fields of small degree

Pierrick Gaudry

gaudry@lix.polytechnique.fr

Contents

- Two motivations
- Results and ingredients
- Introductory example
- General algorithm
- Conclusion, extensions

Two motivations

Difficulty of DL on curves over $\mathbb{F}_{p^n}$

Frequently asked question:

> Can I trust a cryptosystem based on an elliptic curve defined over a small/medium extension field $\mathbb{F}_{p^n}$?

This is very important in practice:

- Arithmetic in the finite field is fast (OEF, . . .)
- Well suited to constrained environment.
- (But not present in all the standards).

Difficulty of DL on curves over $\mathbb{F}_{p^n}$

General Weil descent attack (Frey):

- Consider the Weil restriction A of E on $\mathbb{F}_p$;
- Find a curve $\mathcal{C}$ on A of genus as low as possible;
  ($\mathcal{C}$ is a cover of E, that can be defined over $\mathbb{F}_p$)
- Map the DLP on E to a DLP on $\mathrm{Jac}(\mathcal{C})$;
- Use index-calculus to solve the DLP on $\mathrm{Jac}(\mathcal{C})$.

Problems:

- Finding a low genus curve $\mathcal{C}$ is difficult.
- The Jacobian group law for $\mathcal{C}$ is non trivial in general.

Previous work in odd characteristic

The case $p = 2$ has been much studied, but results do not extend easily.

- Arita, Galbraith: Some cases in characteristic 3.
- Diem: *The GHS attack in odd characteristic*
  - Give bounds for genus of GHS-like curves $\mathcal{C}$.
  - Find families of elliptic curves for which there exists a $\mathcal{C}$ with low genus.
- Thériault: *Weil descent attack for Kummer extensions*
  - Find families of elliptic and hyperelliptic curves for which there exists a $\mathcal{C}$ with low genus.
  - The curve $\mathcal{C}$ is explicitly constructed and is hyper- or super- elliptic.

Use Abelian varieties that are not Jacobians?

Main objection: the arithmetic is exponential in g, whereas it is polynomial in g for Jacobians.

	Abelian var. of dim g	Jac. of curve of genus g
Key size	$g \log q$	$g \log q$
Rep of elts	$2^g \log q$	$2g \log q$
Group law	$\Omega(2^g \log q)$	$\mathrm{poly}(g \log q)$
Index Calc.	$O((2^g)!q^{2-\frac{2}{2^g}})$	$O(g!q^{2-\frac{2}{g}})$

For $g = 1, 2, 3$, abelian varieties are more or less always Jacobians of curves.

For $g \geq 4$, this is no longer true.

Abelian varieties of dimension 4

A possible design:

- By CM theory, build the period matrix of a dimension 4 abelian variety A;
- Find a suitable finite field, so that the group order is appropriate and reduce the invariants (ratios of squares of Theta constants);
- Implement the group law using addition law of Theta functions;
- Hope (prove?) that there are no curve of low genus on A.

Abelian varieties of dimension 4

It is possible to have an exponentiation algorithm that takes $(7 \times 2^g - 3) \log n$ multiplications in $\mathbb{F}_q$ for an exponentiation by n.

For dimension 4, this gives: $109 \log n$ multiplications. This is very competitive with Jacobians of hyperelliptic curves *(see talk of R. Avanzi, tomorrow)*.

The security is better than with a genus 4 Jacobian, since the index calculus does not work as well.

Question: is this latter statment true?

Results and ingredients

Main result

Let E be an elliptic curve over $\mathbb{F}_{q^n}$. There exists an algorithm that computes a discrete logarithm in E in time

$$O\left(q^{2-\frac{2}{n}}\right),$$

where the constant hidden in $O()$ depends (exponentially) in n.

$\Longrightarrow$ Better than Pollard-Rho for n fixed ≥ 3.

Interpretation

The very bad dependance in n forces us to keep n fixed (small).

Examples:

- $n = 3$. If E is defined over $\mathbb{F}_{q^3}$, then the discrete log problem can be solved in time $O(q^{4/3}) \approx O(q^{1.33})$.
- $n = 4$. If E is defined over $\mathbb{F}_{q^4}$, then the discrete log problem can be solved in time $O(q^{1.5})$.

Rem. These complexities were already obtained by Diem and Thériault for special families of curves. What is new is that it is true for all elliptic curves.

By-product: 2nd result

Let A be an abelian variety of dimension n over $\mathbb{F}_q$. There exists an algorithm that computes a discrete logarithm in A in time

$$O\left(q^{2-\frac{2}{n}}\right),$$

where the constant hidden in $O()$ depends (exponentially) in n.

This result implies the previous one by considering the Weil restriction of the elliptic curve.

Ingredients

- Index calculus: decomposition of elements of the group into «small» elements, compatible with the group law.
- Semaev's «algorithm»: idea that there is no need for a unique decomposition (before, decomposition relies on factorization). Summations polynomials.
- Weil Descent: E over $\mathbb{F}_{q^n}$ is an abelian variety of dimension n over $\mathbb{F}_q$.
- Thériault's algorithm and its extensions by Nagao and G.–Thomé: use of *large primes* in this very particular context.

Introductory example

An ECDL problem over $\mathbb{F}_{1019^2}$

The polynomial $f(t) = t^2 + 1$ is irreducible over $\mathbb{F}_p$ with $p = 1019$, so we choose

$$\mathbb{F}_{p^2} := \mathbb{F}_p[t]/(t^2 + 1).$$

We consider E of equation $y^2 = x^3 + ax + b$ with $a = 214 + 364t$ and $b = 123 + 983t$.

E is cyclic of order $N = 1\,039\,037$.

Let P and Q be two random points of E:

$$P = (401 + 517t, 885 + 15t) \text{ and } Q = (935 + 210t, 740 + 617t).$$

We want to compute the logarithm of Q in base P in E.

A factor basis

Rem. The word «factor» is not well suited: decomposition basis would be better. *(see the group law on E as a multiplication.)*

Def. The factor basis is:

$$\mathcal{F} = \{P = (x, y) \in E;\ x \in \mathbb{F}_p,\ y \in \mathbb{F}_{p^2}\}.$$

For each $x \in \mathbb{F}_p$, the element $x^3 + ax + b$ is «random» in $\mathbb{F}_{p^2}$, and therefore is a square with probability $1/2$.

We have

$$\#\mathcal{F} = 1011 \approx p.$$

Index calculus algorithm

- Form a linear combination of P and Q:

$$R = \alpha P + \beta Q, \quad \text{where} \quad \alpha, \beta \in_R [0, N-1].$$

- Look for P_1 and P_2 in $\mathcal{F}$ such that

$$R = P_1 + P_2.$$

- If we find such a decomposition of R over $\mathcal{F}$, we call it a relation, that we store in the row of a matrix:

$$R = \alpha P + \beta Q = \sum_{P_i \in \mathcal{F}} m_i P_i.$$

- After $\#\mathcal{F} + 1 = 1012$ relations have been found, use linear algebra to build $\overline{\alpha}$ and $\overline{\beta}$ such that

$$\overline{\alpha} P + \overline{\beta} Q = 0.$$

Decomposition of R over $\mathcal{F}$ – (1)

Let $P_1 = (x_1, y_1)$ and $P_2 = (x_2, y_2)$ be the points of $\mathcal{F}$ in the decomposition that we look for.

Semaev gives a polynomial $f_3(x_1, x_2, x_3)$ that vanishes if and only if the x_i are the abscissae of points P_i whose sum vanishes in E.

Then we have to solve $f_3(x_R, x_1, x_2) = 0$, where $x_R \in \mathbb{F}_{p^2}$ is known and x_1 and x_2 are unknowns in $\mathbb{F}_p$.

After symmetrisation $e_1 = x_1 + x_2$ and $e_2 = x_1 x_2$, we get:

$$(e_1^2 - 4e_2)x_R^2 - 2(e_1 e_2 + ae_1 + 2b)x_R + a^2 + e_2^2 - 2ae_2 - 4be_1 = 0.$$

Decomposition of R over $\mathcal{F}$ – (2)

One equation, two unknowns.

But: the unknowns are in a subfield.

Prop. Using the explicit representation of the extension field, we get two equations in two unknowns.

$\Longrightarrow$ Solvable using resultants or Gröbner basis, followed by factorization in $\mathbb{F}_p[X]$.

Ex. Let R be the following linear combination:

$$R = 459\,328P + 313\,814Q = (415 + 211t, 183 + 288t).$$

Then the equation

$$(e_1^2 - 4e_2)x_R^2 - 2(e_1e_2 + ae_1 + 2b)x_R + a^2 + e_2^2 - 2ae_2 - 4be_1 = 0$$

can be rewritten (modulo $f(t) = t^2 + 1$) in

$$(881e_1^2 + 597e_1e_2 + 31e_1 + 843e_2 + 669)\,t +$$

$$(329e_1^2 + 189e_1e_2 + 971e_1 + e_2^2 + 294e_2 + 740) = 0.$$

Decomposition of R over $\mathcal{F}$ – (4)

From that, we deduce

$$(e_1, e_2) = (845, 1003),$$

and then

$$x_1 = 92 \text{ and } x_2 = 753.$$

Finally, we test the possible y_i, and we get

$$R = 459328P + 313814Q = P_1 + P_2,$$

with

$$P_1 = (92, 779 + 754t) \text{ and } P_2 = (753, 628 + 692t).$$

Complexity

- Forming a linear combination of P and Q and testing if it is decomposable takes a time polynomial in $\log p$.
- The expected number of relations given by each random linear combination is $1/2$ (independant of p).
- Forming the matrix of relations costs $O(p)$ polynomial time operations.
- The linear algebra step costs $O(p^3)$ with Gauß, $O(p^2)$ with Wiedemann, and $O(p)$ using the fact that there are only two non-zero entries per row.

Thm. There exists an index calculus algorithm that can solve an ECDLP over $\mathbb{F}_{p^2}$ in time $O(p)$ up to log factors.

Rem. Same complexity as Pollard Rho.

General algorithm

General algorithm – (1)

We consider now a degree n extension (n is small).

$$\mathbb{F}_{p^n} = \mathbb{F}_p[t]/(f(t)),$$

with $f(t)$ irreducible over $\mathbb{F}_p$, of degree n.

Let E be an elliptic curve over $\mathbb{F}_{p^n}$, and P and Q two points on E.

If $(x, y) \in E$, we note $x = x_0 + x_1 t + \cdots + x_{n-1} t^{n-1}$ and $y = y_0 + y_1 t + \cdots + y_{n-1} t^{n-1}$, with x_i and y_i in $\mathbb{F}_p$.

Def. We define a factor base as follows:

$$\mathcal{F} = \{P = (x, y) \in E, \;\; x_0 = x_1 = \cdots = x_{n-2} = 0\}.$$

Rem. The choice of coordinates that we annihilate is arbitrary.

General algorithm – (2)

Prop. Under a few genericity assumptions, $\mathcal{F}$ is an irreducible variety of dimension 1 (we cut an irreducible variety of dimension n by $n-1$ «random» hyperplanes).

Cor. By Hasse-Weil theorem,

$$\#\mathcal{F} \approx p.$$

We form random linear combinations $R = \alpha P + \beta Q$, that we try to decompose in the sum of n points of $\mathcal{F}$.

Questions:

- expected number of relations obtained for each linear combination
- time required for one decomposition over $\mathcal{F}$.

Decomposition

Like in the example, we reduce the problem to finding solutions of a system of algebraic equations.

Resolution using a Gröbner basis computation over $\mathbb{F}_p$, followed by a factorization of a polynomial in $\mathbb{F}_p[X]$.

$\Longrightarrow$ Cost is polynomial in $\log p$ and exponential in n.

Rem. We'll see later bounds on the degrees.

Probability of finding a relation

We consider the function

$$f \; : \quad \begin{array}{ccc} \mathcal{F}^n/\mathfrak{S}_n & \longrightarrow & E \\ (P_1,\ldots,P_n) & \mapsto & P_1+\cdots+P_n \end{array}$$

The decomposition algorithm computes $f^{-1}(R)$ for a given R.

Hence, the expected number of relations that we find for one R is

$$\sum_{R\in E}\frac{\#f^{-1}(R)}{\#E} = \frac{1}{\#E}\#(\mathcal{F}^n/\mathfrak{S}_n) \approx \frac{1}{p^n}\frac{p^n}{n!} \approx \frac{1}{n!}.$$

Rem. We have neglected the terms with smaller order of magnitude (corresponding for instance to $P_1 = P_2$).

Complexity for fixed n

- Forming a linear combination of P and Q and testing if it is decomposable over $\mathcal{F}$ takes a time polynomial in $\log p$.
- The expected number of relations found for each linear combination is $1/n!$ (independant of p).
- Forming the matrix of relations costs $O(p)$ operations that are polynomial-time in $\log p$.
- The (sparse) linear algebra costs $O(p^2)$ with Wiedemann.

$\Longrightarrow$ The total cost of the algorithm for fixed n is in $O(p^2)$.

Use of Thériault's algorithm – (1)

Let $0 < r < 1$ be a parameter. We choose a subset $\mathcal{F}'$ of $\mathcal{F}$ of cardinality p^r:

$$\mathcal{F}' \subset \mathcal{F} \text{ and } \#\mathcal{F}' \approx p^r.$$

For the construction of the matrix, we keep only the relations that involve $n-1$ points of $\mathcal{F}'$ and 1 point of $\mathcal{F} \setminus \mathcal{F}'$ (*large prime*).

Birthday paradox: If we have k such relations, we can deduce (on average) $\frac{k^2}{p}$ relations involving only elements of $\mathcal{F}'$.

We want to have $\#\mathcal{F}' = p^r$ relations, we need then

$$k = p^{\frac{1+r}{2}}$$

relations with a *large prime*.

Use of Thériault's algorithm – (2)

- Cost for finding a relation with *large prime* (n fixed):

$$\left(\frac{p}{p^r}\right)^{n-1} = p^{(1-r)(n-1)}.$$

- Cost for the construction of the matrix of relations over $\mathcal{F}'$:

$$kp^{(1-r)(n-1)} = p^{(1-r)(n-1)+\frac{1+r}{2}}.$$

- Cost of the linear algebra:

$$(p^r)^2 = p^{2r}.$$

Balancing the costs, we find $r = \frac{2n-1}{2n+1}$, and an overall cost of

$$O(p^{2-\frac{4}{2n+1}}).$$

Improvement to Thériault's algorithm

Idea: Taking two large primes instead of one can only help!

The tricky part is to analyze this double-large-prime variant in a rigourous way.

Done independently, and in different ways by Nagao and G.–Thomé.

Result: The complexity drops to $O(p^{2-\frac{2}{n}})$.

Back to the decomposition – (1)

Def. Semaev's summation polynomials: for a given elliptic curve E, there exist polynomials f_n that are symmetric in n variables and that vanish exactly when evaluated at abscissae of points whose sum is 0 in E.

$$P_1 + P_2 + \cdots + P_n = 0 \Longleftrightarrow f_n(x_{P_1}, x_{P_2}, \ldots, x_{P_n}) = 0.$$

Thm. (Semaev) For $n \geq 3$, f_n is of degree 2^{n-2} in each variable.

Cor. If we write f_n in terms of the elementary symmetric polynomials, f_n is of total degree 2^{n-2}.

Back to the decomposition – (2)

Hence the system we have to consider in order to solve the decomposition problem has n equations in n unknowns. Each equation has degree 2^{n-1}.

$\Longrightarrow$ we expect a degree $O(2^{n(n-1)})$ for the univariate polynomial in a Lex Gröbner basis.

Rem. If $n = 5$, we have an ideal whose degree is about 1 million.

Rem. For $n = 3$, Magma makes a decomposition in 0.1 second.

Comparison with classical Weil Descent – (1)

$$
\begin{tikzcd}
 & A \arrow[drr, "\text{Map DL}"] & & \\
\mathcal{C}^n/\mathfrak{S}_n \arrow[ur, "(*)"] & \mathcal{C} \arrow[l] \arrow[u] \arrow[r] & \mathcal{C}^g/\mathfrak{S}_g \arrow[r, "(*)"'] & \mathrm{Jac}(\mathcal{C})
\end{tikzcd}
$$

The arrows marked by $(*)$ are those where the index calculus takes place.

Comparison with classical Weil Descent – (2)

Advantages of our method:

- Does not require any knowledge of the geometry of $\mathcal{C}$. No arithmetic in its Jacobian needed.
- The factorial component in the complexity is always $n!$, as compared to $g!$, where $g \geq n$ can be exponential in n.

Drawbacks of our method:

- Gröbner basis are not easy to deal with.
- Dependance in n is too bad.

Conclusion, extensions

Conclusion

- Algorithm in $O(p^{2-2/n})$ for ECDL over $\mathbb{F}_{p^n}$, with a constant that depends exponentially in n.
- For $n = 3$, we get an $O(p^{1.33})$ algorithm with a reasonnable constant.
- For $n = 4$, we get an $O(p^{1.5})$ algorithm with a bad constant.
- For larger n, just a theoretical result... But see Diem's variant!

$\Longrightarrow$ The elliptic curves defined over extension of small degree ≥ 3 are asymptotically less secure than was what previously admitted.

Generalisations

- This algorithm extends easily to DLP in Jacobians of hyperelliptic curves of genus g over $\mathbb{F}_{p^n}$. The complexity is $O(p^{2-\frac{2}{ng}})$.
- More generally, we can give a general discrete log algorithm for abelian varieties of dimension n over $\mathbb{F}_p$ in time $O(p^{2-4/n})$.
- For instance, in the Jacobian of a genus 2 curve over $\mathbb{F}_{p^2}$ we have a discrete log algorithm in $O(p^{1.5})$.

Other recent works

- Recent preprint by Arita–Matsuo–Nagao–Shimura.

  They propose an attack against a very large class of elliptic curves over $\mathbb{F}_{q^4}$.

  Use Scholten's form to find a curve of genus 9 in the Weil restriction.

- Diem's subexponential algorithm.

  Use a variant of our algorithm to obtain a $L_{3/4}(n \log q)$ algorithm for ECDL in $\mathbb{F}_{q^n}$, with $n \sim \log q$.