Isogeny-based cryptography

Table of content

1 Introduction

Post-quantum cryptography (PQC) is a young field anticipating the availability of large-scale quantum computers in the context of digital security. Seen by many as the next logical step to supersede classical computers, the quantum paradigm offers algorithms that could endanger classical public key cryptography. Possibly the most famous one, Shor's algorithm [] brings integer factorization down to a polynomial problem. This has obvious dire consequences for classical RSA and Diffie-Hellman (DH) protocols. The performances of current quantum computers are still far from reaching usability in attack scenarios. But cautiousness calls for quantum-resistant algorithms that could secure classical computers. In that regard, the NIST (National Institute of Standards of Technology) has organized competitions to foster developpment in PQC.

Different usages require different cryptosystems and as such there are multiple classes of post-quantum protocols investigated. Of interest in this paper is public-key cryptography (PKC) also known as asymmetric cryptography. The benefit of PKC stems from the fact that no pre-sharing of information is needed to establish a secure channel, hence making it very convinient. To achieve this, a Diffie-Hellman key-exchange protocol is usually used. In our case we use elliptic curve isogenies and associated hard homogeneous spaces (HHS) to build such a protocol.

More precisely, we investigate the Couveignes-Rostovtsev-Stolbunov protocol (CRS). This protocol works with isogenies bewteen ordinary curves with fixed complex multiplication over a finite field of prime order $\sim 2^{256}$. Key-exchanges using this scheme are comparatively slower than their supersingular counterpart such as SIDH protocols. However, CRS heuristically gives less room for potential attacks than SIDH. This is due to the fact that supersingular endomorphism rings embed as orders in quaternion algebras whereas ordinary ones embed in imaginary quadratic fields. As such, more structure is available for exploitation in an attack of SIDH protocols. This kind of structural attack has been demonstrated against the NTRU protocol and led to Prime-NTRU. Another advantage of CRS is that it is truely symmetric compared to SIDH where parties have to agree on specific primes to use.

The main focus of this project was to implement as efficiently as possible the CRS protocol following the 2018 article of Luca De Feo, Jean Kieffer and Benjamin Smith []. Recent publications on the subject of isogeny computation allow for very efficient walking algorithm in the isogeny graphs. On the one hand, *radical isogenies* [] are used to take k steps in the l-isogeny graph in time complexity $\mathcal{O}(k)$ for primes 3, 5 and 7. On the other hand, the $\sqrt{}$*-Velu* algorithm [] is used to take walk in the remaining graphs with time complexity $\mathcal{O}(k\sqrt{l})$. The reason this subject was choosen specifically is after a previous unoptimized JULIA implementation of these techniques. It gave promising results (around 35% faster than the 2018 implementation) but was neither optimized nor using the expected quadratic-twist symmetry. The second issue was completely resolved by our C implementation while drastic optimizations were found.

This paper is divided in three sections. The first one mainly accounts for the minimal mathematical background surrounding isogeny-graphs and complex multiplication. An exposition on the CRS protocol is also given. The second part tackles the architecture of the program. The complete project is around 4000 lines long. Its source code is available at []

along with a documentation. Finally, the third part reports prime-by-prime timing results, optimized bound for the secret keys as well as some inner data observed at runtime.

2 Theoretical background

In this section we introduce the necessary mathematical material surrounding the CRS protocol. It is assumed the reader has some familiarities with elliptic curves and their arithmetic. Still, we recall the fundamentals in order to underline some important aspect of the algorithms.

2.1 Elliptic curves and models

Recall the standard definition of an elliptic curve over k:

Definition. An elliptic curve E over a field k is a smooth curve given by a (dehomogenized) polynomial $F \in k[x, y]$ of the form

$$F = y^2 + a_1xy + a_3y - (x^3 + a_2x^2 + a_4x + a_6).$$

This presentation of the curve is usually called the *Weierstrass* form. It is said to be in *short Weierstrass* form if the coefficients a_1, a_2 and a_3 are null. The following proposition allow one to only consider short Weierstrass equations most of the time.

Proposition. An elliptic curve E over a field k of caracteristic different from 2 and 3 is isomorphic to an elliptic curve in short Weierstrass form.

Proof. See Silverman III.1. [] □

The (short-)Weierstrass models admit explicit formulas that reflect the group law of the elliptic curve. It turns out that other models have better, faster properties when it comes to group operations. The following *Montgomery* model has one of the fastest doubling operation of all. It also has a fast scalar multiplication which we talk about in the next section.

Definition. A *Montgomery curve* or an elliptic curve in *Montgomery* form over a field k is an elliptic curve $E_{A,B}$ given by a polynomial $F \in k[x, y]$ of the form

$$F = By^2 - x(x^2 + Ax + 1)$$

with $A, B \in k$ satisfying $b \neq 0$ and $A^2 \neq 4$.

All Mongomery curves admit a short Weierstrass model via the variable change $x := x/B$ and $y := y/B$. The following proposition shows which short Weierstrass curves admit a Montgomery model.

Proposition. The elliptic curve $y^2 = x^3 + ax + b$ over a field k admits a Montgomery model if

- The polynomial $X^3 + aX + b$ has a root w in k

- $3w^2 + a$ is a quadratic residue in k.

Proof. Under the above conditions, set $u = (3w^2 + a)^{-\frac{1}{2}}$. The (admissible) variable change $x := u(x - w)$, $y := uy$ show that the curve has Montgomery model $E_{A,B}$ with $A = 3wu$ and $B = u$. □

Associated to an elliptic curve E is the fundamental quantity called *j-invariant*. For a Mongomery curve $E_{A,B}$ over a field k of caracteristic different from 2 and 3 one has

$$j = \frac{256(A^3 - 3)^3}{A^2 - 4}.$$

Elliptic curves can be shown to be isomorphic if and only if they have the same j-invariant. The crucial thing to remember is that this classification only happens over the algebraic closure $\bar{k}$ of the base field k. Curves can be isomorphic over $\bar{k}$ while being non-isomorphic over k or a finite extension of k. In fact, the B parameter in the definition of a Mongomery curve account for two isomorphism classes: take $A, B, B' \in k$ and consider the curve $E_{A,B}$ and $E_{A,B'}$. One easily show that they are isomorphic at best over $k(\sqrt{B/B'})$. Hence they are isomorphic over k if and only if B/B' is a quadratic residue. We call $E_{A,B'}$ a *quadratic twist* of $E_{A,B}$.

The next and last model of elliptic curve we use is called the *Tate normal* form. It will be used when dealing with radical isogenies.

Definition. An Elliptic curve in *Tate normal* form over a field k is an elliptic curve E given by a polynomial $F \in k[x, y]$ of the form

$$F = y^2 + (1 - c)xy - by - x^2(x - b)$$

on which the point $(0, 0)$ has order at least 4. We also say E is in Tate normal form when given by

$$F = y^2 + (1 - c)xy - by - x^3$$

where $(0, 0)$ is of order 3.

The following lemma and its proof show how to transform a Mongomery curve along with a point P of order $N \geq 4$ into Tate normal form. We demonstrate the lemma with a general Weierstrass curve but only the Montgomery case with $B = 1$ interests us in the implementation. It is only a matter of setting some a-coefficients to 0 in the formulas.

Lemma 2.1.1. Let E be an elliptic curve over a field k and let $P = (x, y) \in E$ be a point of order $N \geq 3$. There exist a Tate normal model for E such that P is sent to $(0, 0)$..

Proof. We consider E in general Weierstrass form

$$Y^2 + a_1XY + a_3Y = X^3 + a_2X^2 + a_4X + a_6.$$

A translation from P to $(0, 0)$ allows us to remove a_6 and write the curve as

$$Y^2 + a_1XY + b_3Y = X^3 + b_2X^2 + b_4X.$$

where $b_2 = 3x + a_1$, $b_3 = 2y + a_1x + a_3$ and $b_4 = 3x^2 + 2a_1x + a_4$. The coefficient b_3 is non-zero as P does not have order 2. This can be read on the duplication formula. Let us remove the b_4 coefficient by the (admissible) change of variable $Y := Y + b_4/b_3X$. We get the representation

$$Y^2 + c_1XY + b_3Y = X^3 + c_2X^2.$$

where $c_1 = 2b_4/b_3 + a_1$ and $c_2 = b_2 - a_1b_4/b_3$.

Now introduce two free scaling variables $\alpha, \beta \in k$. The scaling $X = \alpha X$, $Y = \beta Y$ gives a model that is Tate normal if

$$\alpha^3 = \beta^2$$
$$\beta b_3 = c_2\alpha^2.$$

Following the standard negation formula and $N = 3$ being equivalent to $2P = -P$ we see that $c_2 = 0$ is also equivalent to $N = 3$. Assuming that $N \geq 4$, one find $\alpha = (b_3/c_2)^2$ and $\alpha = c_2/b_3$. Thus setting $b = -c_2(b_3/c_2)^2$ and $c = 1 - c_1(c_2/b_3)^2$ gives the Tate normal form. If $N = 3$, the curve is already in Tate normal form. □

Notice however that in the case $N = 3$, the coefficient b_3 should be computed to get b and c. This in turns involves knowing the value of y. As we will see with x-only arithmetic, over Mongomery curves it is better to use our free system in α and β. Setting $\alpha = b_3^3$ and $\beta = b_3^2$ we find a presentation which only involves y^2. In the end we find $b = -1/b_3^2$ and $c = 1 - 2b_4/b_3^2$. This will avoid us extracting a square root of $x^3 + ax^2 + 1$ to get a y value. This used to be a time-consuming task in the previous implementation that slowed down the computations.

2.2 Isogenies

We fix once and for all a prime number p. Unless indicated otherwise we assume throughout the section that curves are defined over $\mathbb{F}_q$, a field of caracteristic p with q elements. Recall the definition of an isogeny, which is essentialy an homomorphism between varieties preserving the group structure.

Definition. Let E and E' be two elliptic curves defined over $\mathbb{F}_p$. An isogeny $\phi : E \to E'$ is a non-trivial algebraic map sending O_E to $O_{E'}$.

The isogenies from E to E, together with the trivial map form a ring called the endomorphism ring of E and noted $\mathrm{End}(E)$. For a positive integer m, the *multiplication-by-m* map over E is an endomorphism sending a point P to the sum of m copies of P. This map will be denoted by $[m]$.

Example. Of interest in the CRS protocol is the *Frobenius* endomorphism:

$$\pi : (x, y) \in E \mapsto (x^q, y^q) \in E.$$

This endomorphism satisfies the relation $\pi^2 - t\pi + q = 0$, where t is the trace of E. This trace is linked to the cardinality of $E(\mathbb{F}_q)$ by $\#E(\mathbb{F}_q) = q + 1 - t$. It can be computed effectively using the PARI/GP library for instance. When the trace is divisible by p, we say that the curve is *supersingular*. Otherwise it is called *ordinary*. The CRS protocol only uses ordinary curves.

The *degree* of an isogeny is its degree as an algebraic map. For instance, the multiplication-by-m map is of degree m^2 when $m \wedge p = 1$. The degree of the frobenius endomorphism is q. We say an isogeny of degree l is an *l-isogeny.*

For the purpose of CRS, we mainly consider *separable* isogenies between elliptic curves. These include isogenies of degree coprime to p and are in one-to-one correspondance with their kernels up to isomorphism. This means that for any finite subgroup G of E of order l, there is a unique elliptic curve (up to isomorphism) denoted E/G and an l-isogeny

$$E \to E/G$$

whose kernel is G. Recall finally that for each l-isogeny $\phi : E \to E'$, there is a unique l-isogeny $\hat{\phi} : E' \to E$ called the *dual isogeny* such that

$$\phi \circ \hat{\phi} = [l]_{E'} \text{ and } \hat{\phi} \circ \phi = [l]_E.$$

The important thing to remember from this is that being l-isogenous is a *symmetric relation* and being isogenous is an *equivalence relation.* Isogenous curves share some common properties. For instance, they both have the same number of points over $\mathbb{F}_q$ and thus also have the same trace. This brings us to the core component of the CRS protocol.

2.3 Isogeny graphs

For an integer l, the *l-isogeny graph* over a field k is the graph whose vertices are elliptic curve isomorphism classes (over k) and whose edges represent l-isogenies. Recall that for l coprime to p, the l-torsion group of E denoted $E[l]$ is isomorphic to $\mathbb{Z}/l\mathbb{Z} \times \mathbb{Z}/l\mathbb{Z}$. This way, if l is prime and different from p, $E[l]$ contains $l+1$ distinct cyclic subgroups of order l. This in turns mean that there are exactly $l+1$ (separable) l-isogenies whose domain is E. Thus, a connected component in the l-isogeny graph over $\overline{\mathbb{F}_q}$ is an $l+1$-regular graph.

As mentionned before, isomorphism classes are determined by j-invariant over $\overline{\mathbb{F}_q}$. However in CRS we will use isogeny graphs over $\mathbb{F}_q$ which gives more complex structures. Let us analyze the isogeny graph in this case, starting by computing the incidence degree of the graph's vertices. We consider a prime l different from p.

It can be shown that an isogeny is defined over $\mathbb{F}_q$ if and only if the frobenius π stabilizes its kernel. We are only interested in cyclic isogenies, those with cyclic kernels. As such can see the action of π on the kernel as a scalar multiplication due to the fact that π commutes with multiplication-by-m maps. Notice further that $E[l] = \mathbb{Z}/l\mathbb{Z} \times \mathbb{Z}/l\mathbb{Z}$ is a 2-dimentional $\mathbb{F}_l$-vector space. The problem is hence reduced to understanding the stable subspaces of π as an endomorphism of $E[l]$. This calls for a study of the eigenvalues of π, whose characteristic polynomial is $X^2 - tX + q \bmod l$. Counting the stable subgroups, we get:

- If π has no eigenvalues, then E has no l-isogenies.
- If π has one eigenvalue of geometric multiplicity one, there is exactly one l-isogeny from E.
- If π has one eigenvalue of geometric multiplicity two, there are exactly $l+1$ l-isogenies from E.

- If π has two distinct eigenvalues, there are exactly two l-isogenies from E.

Only that last case is of interest for use. We call primes l that satisfy this condition *Elkies* primes. Note that since isogenous curves have same trace, their frobenius has same characteristic polynomial. As such, isogenous curves have the same Elkies primes. This implies that for l an Elkies prime connected components of the l-isogeny graph over $\mathbb{F}_q$ are 2-cycles.

2.4 The structure of End(E)

In this part we show that endomorphism rings of ordinary curves are orders in imaginary quadratic fields. Let E be an ordinary elliptic curve over $\mathbb{F}_q$ with Frobenius element π. Remember that the ring $\mathbb{Z}$ embeds in $\mathrm{End}(E)$ via multiplication-by-m elements. As such we will identify elements $[m]$ and m. Let us start with the definition of an order in an algebra;

Definition. Let A be a finite dimentional algebra over $\mathbb{Q}$. A subring $\mathcal{O}$ of A is an *order* if:

- $\mathcal{O} \otimes \mathbb{Q} = A$.
- $\mathcal{O}$ is a $\mathbb{Z}$-lattice in A.

This essentialy means that $\mathcal{O}$ is a free abelian group containing a $\mathbb{Q}$-basis of A. The following theorem gives the remarkably simple classification of endomorphism rings of elliptic curves. The heavy lifting of the proof is done using Tate-modules theory. See [theorem Silverman III.9.3].

Theorem 1. *Let E be an elliptic curve over a field k. Then $\mathrm{End}(E)$ is either $\mathbb{Z}$, an order in an imaginary quadratic field or an order in a quaternion algebra.*

We now show that our ordinary elliptic curve E falls in the second category. For that we need some generalities about $\mathrm{End}(E)$.

Proposition. The ring $\mathrm{End}(E)$ is torsion free of characteristic 0 and has no non-trivial zero divisors.

Proof. Recall that for $m \neq 0$, the multiplication-by-m map $[m]$ is non-constant. Take $\phi \in \mathrm{End}(E)$ and m such that

$$[m] \circ \phi = [0].$$

The multiplicativity of the degree map gives

$$\deg([m])\deg(\phi) = 0.$$

Thus either $m = [0]$ or $\phi = [0]$ and $\mathrm{End}(E)$ is torsion free of caracteristic 0. Now suppose for $\phi, \psi \in \mathrm{End}(E)$ that $\phi \circ \psi = [0]$. Then taking the degree,

$$\deg(\phi)\deg(\psi) = 0.$$

Thus either $\phi = [0]$ or $\psi = [0]$ and $\mathrm{End}(E)$ has no non-trivial zero divisors. □

The following lemma already shows that $\mathbb{Z} \subsetneq \mathrm{End}(E)$.

Lemma 2.4.1. The Frobenius endomorphism $\pi \in \mathrm{End}(E)$ is not in $\mathbb{Z}$.

Proof. Assume the converse so that $\pi = [n]$ for some integer n. Then taking the degree gives

$$q = \deg(\pi) = \deg([n]) = n^2.$$

This implies that $d := [\mathbb{F}_q : \mathbb{F}_p]$ is even and that $n = \pm p^{d/2}$. Hence using the roots-coefficients relations on the characteristic polynomial of π we get $t = 0 \bmod p$, which is absurd as E is ordinary. □

Let $\mathrm{End}^0(E) := \mathrm{End}(E) \otimes \mathbb{Q}$. The next lemma will be used to show that every isogeny of $\mathrm{End}^0(E)$ commutes with π.

Lemma 2.4.2. Let $n \geq 1$. There exist $a, b \in \mathbb{Z}$ satisfying $a \neq 0 \bmod p$, $b = 0 \bmod p$ and such that

$$\pi^n = a\pi + b.$$

Proof. We use induction on n. The case for $n = 1$ is verified for $a = 1$ and $b = 0$. Assume the lemma holds for every $1 \leq m \leq n$. Then using the fact that $\pi^2 - t\pi + q = 0$, we have for some $a, b \in \mathbb{Z}$ satisfying the hypothesis

$$\pi^{n+1} = \pi(a\pi + b) = b\pi + a(t\pi - q).$$

Rearranging the equation one gets

$$\pi^{n+1} = c\pi + d$$

where $c = at + b$ and $d = -aq$. These verify the conditions since $c = at \neq 0 \bmod p$, E being ordinary, and $d = 0 \bmod p$. □

This lemma shows that $\pi^n \notin \mathbb{Q}$ for any integer $n \geq 1$. Now any element of $\mathrm{End}^0(E)$ can be written as $m\phi$ with $m \in \mathbb{Q}$ and $\phi \in \mathrm{End}(E)$. Being a rational map, the isogeny ϕ is defined over a finite extension of $\mathbb{F}_q$, say $\mathbb{F}_{q^d}$. Writting ϕ in its reduced form $(r(x), ys(x))$ we have

$$\phi\pi^d = (r(x^{q^d}), y^{q^d}s(x^{q^d})) = (r(x)^{q^d}, (ys(x))^{q^d}) = \pi^d\phi.$$

Hence every element of $\mathrm{End}^0(E)$ commutes with π^d for some $d \geq 1$. And this last element is in fact in $\mathbb{Q}(\pi)$ according to the previous lemma. The next lemma shows that this is enough to prove that $\mathrm{End}^0(E) \subset \mathbb{Q}(\pi)$, hence the equality.

Lemma 2.4.3. Let $\alpha, \beta \in \mathrm{End}^0(E)$. If these elements commute and $\alpha \notin \mathbb{Q}$ then $\beta \in \mathbb{Q}(\alpha)$.

Proof. Introduce the $\mathbb{Q}$-linear trace map $T : \alpha \in \mathrm{End}^0(E) \mapsto \alpha + \hat{\alpha} \in \mathbb{Q}$. Its codomain being justified by equalities

$$T\alpha = 1 - \alpha\hat{\alpha} - (\alpha - 1)\widehat{(\alpha - 1)} = 1 - [\deg(\alpha)] - [\deg(\alpha - 1)] \in \mathbb{Q}.$$

We may replace α by $\alpha - \frac{1}{2}T\alpha$ to have $T\alpha = 0$. Since $T\alpha = 0$ and $\mathrm{End}^0(E)$ is without non-trivial zero divisors, we get $\alpha^2 \in \mathbb{Q}^*$. Replacing β by $\beta - \frac{1}{2}T\beta - \frac{1}{2\alpha^2}T(\alpha\beta)\alpha$, we can assume further that

$$T\beta = T\alpha\beta = 0.$$

The combined equations $T\alpha = T\beta = T(\alpha\beta) = 0$ now give $\alpha\beta = -\beta\alpha$. All the substitutions were done while keeping the commutativity hypothesis thus we have $2\alpha\beta = 0$. As $\mathrm{End}^0(E)$ is without non-trivial zero divisors, we find $\beta = 0 \in \mathbb{Q}(\alpha)$ since $\alpha \notin \mathbb{Q}$. □

In the end we have proved that $\mathrm{End}(E) \otimes \mathbb{Q} = \mathbb{Q}(\pi)$. The field $\mathbb{Q}(\pi)$ is quadratic imaginary as the Frobenius' characteristic polynomial is of degree 2. This concludes the proof that ordinary curves have endomorphism rings isomorphic to orders in imaginary quadratic fields.