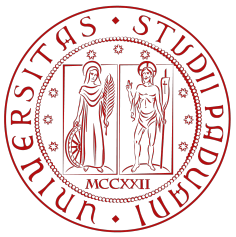


UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Università degli Studi di Padova

DIPARTIMENTO DI MATHEMATICA " TULLIO LEVI-CIVITA"

Corso di Laurea Magistrale in Matematica

The mathematical aspects of the Castryck-Decru key recovery attack on SIDH

Relatore:
Prof. Remke Nanne Kloosterman

Laureanda:
Riya Parankimamvila Mamachan (2070558)

Anno Accademico 2022/2023

19.07.2023

Acknowledgements

First and foremost, I express my deepest gratitude to my advisor Prof. Remke Nanne Kloosterman, for proposing this fascinating thesis topic and for his invaluable supervision throughout the project. His immense knowledge and patience helped me all the time. Thank you so much.

I would take this opportunity to thank all the faculty members in the Department of Mathematics at the Università degli Studi di Padova and the Universität Duisburg-Essen for their help and support through out this journey.

Finally, a very special thanks to my family and closest friends who unceasingly supported me despite anything. Thank you so much for the constant encouragement all along the way.

Contents

Introduction

A well-studied hard problem in number theory is to find an unknown high degree isogeny between two supersingular elliptic curves over a finite field. This lead to the development of isogeny based cryptosystems. Supersingular Isogeny Diffie-Hellman (SIDH) is a key exchange protocol proposed in 2011 that makes use of isogenies between supersingular elliptic curves. Isogeny based cryptosystems are believed to be hard for even quantum computers. One of the most influential primitive in the field of post-quantum cryptographic standards is Supersingular Isogeny Key Encapsulation (SIKE) which is the incarnation of SIDH that recently advanced to the fourth round of NIST's ongoing standardization process. This project deals with a brief study of the mathematical aspects of an efficient key recovery attack on SIDH proposed by Wouter Castryck and Thomas Decru in 2022.

In comparison with other cryptosystems based on the pure isogeny problem, the hardness of SIDH is weaker due to the availability of torsion point information under the secret isogeny that is made public during the protocol. Then the problem is called supersingular isogeny with torsion (SSI-T). The main mathematical tool used behind this key recovery attack is the "glue-and-split" theorem due to Ernst Kani. This attack also exploits the existence of a small non-scalar endomorphism on the starting curve.

Including the preliminaries, the first chapter is mainly devoted for studying abelian varities whose classical examples are elliptic curves. This includes the Picard scheme, the dual and polarizations of an abelian variety. Later the chapter also introduces elliptic curve cryptography and SIDH in detail. The second chapter is based on Kani's paper, where we introduce reducible subgroup and its correspondence with an isogeny factorization configuration. The main aim of this chapter is to see how to employ the reducibility criterion proposed by Ernst Kani to set the decision criteria through out the attack.

In the last two chapters, we will finally turn to see the setup, strategy and the key recovery algorithm presented by Wouter Castryck and Thomas Decru. In particular, we will focus on discussing the mathematical aspects of the steps that are involved in the iteration. The complexity of the algorithm and some generalizations of the attack are also added in the final chapter. It is then concluded by giving a brief study of an algorithm emerged from the same source of inspiration, proposed to attack SIDH without any information the endomorphism ring of the starting curve.

The main reference used for this project is the paper *An efficient key recovery attack on SIDH* by Wouter Castryck and Thomas Decru [1] published in 2022.

Chapter 1

Pre-requisites

1.1 Abelian variety

An abelian variety is a complete algebraic variety whose points form a group, in such a way that the maps defining the group structure are given by morphisms. The book *Abelian Varieties* by Bas Edixhoven, Gerard van der Geer, Ben Moonen [3] is referred for this section.

Definition 1.1. *A group (G, m, i, e) consists of a set G together with the map*

$$m : G \times G \to G \text{ the group law}$$
$$i : G \to G \text{ the inverse}$$
$$e \in G \text{ the identity element}$$

such that following holds

1. *Associativity: $m \circ (m \times id_G) = m \circ (id_G \times m) : G \times G \times G \to G$*
2. *The defining property of the identity element:*

$$m \circ (id_G \times e) = j_1 : G \times \{e\} \to G \text{ and}$$
$$m \circ (e \times id_G) = j_2 : \{e\} \times G \to G$$

   where j_1, j_2 denotes the corresponding projection maps and e denotes the inclusion map $\{e\} \hookrightarrow G$.

3. *Let $\pi : G \to G$ be the constant map $g \mapsto e$ and $\Delta_G : G \to G \times G$ denotes the diagonal map, then i gives the two-sided inverse if it satisfies*

$$\pi = m \circ (id_G \times i) \circ \Delta_G = m \circ (i \times id_G) \circ \Delta_G$$

Translating the same definition to the category of varieties, we obtain the definition of a group variety

Definition 1.2. *A group variety over a field k is a k-variety X together with the k-morphisms*

$$m : X \times X \to X \text{ the group law}$$
$$i : X \to X \text{ the inverse}$$
$$e \in X(k) \text{ a } k\text{-rational point to be the identity element}$$

such that following equality of morphisms holds

1. *Associativity:* $m \circ (m\times\ id_X) = m \circ (id_X \times m) : X \times X \times X \to X$
2. *Let* $j_1 : Spec(k) \times X \to X$, *and* $j_2 : X\times\ Spec(k) \to X$ *be the projection maps, then*

$$m \circ (e\times\ id_X) = j_1 \text{ and } m \circ (id_X \times e) = j_2.$$

3. *Let* $\pi : X \to Spec(k)$ *be the structure morphism, then*

$$e \circ \pi = m \circ (id_X \times i) \circ \Delta_{X/k} = m \circ (i \times id_X) \circ \Delta_{X/k} : X \to X$$

If X is a group variety, then its k-rational points naturally form a group. In general, X yields a contravariant functor from the category of k-schemes to the category of groups by mapping any k-scheme T to the set $X(T)$, of T valued points of X.

Non-singularity ensures that the group variety behaves well with respect to intersection theory, which allows for the definition of divisors, line bundles, and simplifies the study of rational points and the computation of cohomology groups.

Definition 1.3. *An abelian variety is a group variety which, as a variety is complete.*

An abelian variety is smooth, connected, commutative and can also be proved as a projective algebraic group. The classical examples of abelian varieties are *elliptic curves*. An elliptic curve is a complete, non-singular curve of genus 1 over a field k.

Remark 1.4. Non-abelian group varieties are varieties equipped with a non-commutative group structure. For example: General linear group $GL(n)$, Special linear group $SL(n)$, Symmetric groups S_n, Orthogonal group $O(n)$.

The definition of group schemes is a variation on that of a group variety where arbitrary schemes are considered rather than varieties. In a less explicit way, one can define group scheme as follows:

Definition 1.5. *Let S be a scheme. An S-group scheme is a scheme G over S together with an S-morphism $m : G \times G \to G$ such that the induced law of composition $G(T) \times G(T) \to G(T)$ makes $G(T)$ a group for every S-scheme T.*

An equivalent definition is that an S-group scheme is a contravariant functor from the category of schemes over S to the category of groups such that the underlying functor to the category of sets is representable.

Definition 1.6. *Let (X_1, m_1, i_1, e_1) and (X_2, m_2, i_2, e_2) be two group schemes over a common base k. A homomorphism of group schemes $f : X_1 \to X_2$ is a morphism of k-schemes such that*

$$f \circ e_1 = e_2, \;\; m_2 \circ (f \times f) = f \circ m_1, \;\; i_2 \circ f = f \circ i_1$$

If $f : X_1 \to X_2$ is a group scheme homomorphism, then ker(f) is defined as the fibre product

$$\begin{array}{ccc} ker(f) & \longrightarrow & X_1 \\ \downarrow & & \downarrow f \\ Spec\ k & \xrightarrow{e_2} & X_2 \end{array}$$

and ker f is then a subgroup scheme of X_1.

When A, B are abelian varieties we shall say that f is a homomorphism of abelian varieties, or simply a homomorphism.

Definition 1.7. *Let $f : X \to Y$ be a finite surjective morphism between algebraic varieties over a field k. The degree of f is the degree of the finite field extension of the function field $k(X)$ over $f^*k(Y)$.*

Definition 1.8. *Let X, Y be abelian varieties over an algebraically closed field k. A k-isogeny between X and Y is a homomorphism $f : X \to Y$ defined over k and such that ker (f) is finite. The degree of an isogeny f is its degree as a homomorphism.*

An isogeny between two abelian varieties can be characterized in many equivalent ways.

Proposition 1.9. *For a homomorphism $f : A \to B$ of abelian varieties, the following statements are equivalent:*

1. *f is an isogeny.*
2. *$\dim A = \dim B$ and f is surjective.*
3. *$\dim A = \dim B$ and ker (f) is a finite group (scheme).*
4. *f is finite, flat, and surjective.*

1.2 The Picard scheme of an abelian variety

Given a scheme X, the Picard group of a scheme,

$$\mathrm{Pic}(X) = H^1(X, \mathcal{O}_X^*) = \{\text{isomorphism classes of line bundles on } X\}$$

Let X be an abelian variety and $t_x : X \to X$ is the translation defined by $t_x(y) = m(x, y)$. The theorem of square states that if L is a line bundle on X, then for all $x, y \in X(k)$, $t_{x+y}^* L \otimes L \cong t_x^* L \otimes t_y^* L$, it can be shown that the map $\phi_L : X(k) \to Pic(X/k)$ given by $x \mapsto [t_x^* L \otimes L^{-1}]$ is a homomorphism.

Remark 1.10. Since abelian varieties are non-singular, we have the natural isomorphism $\mathrm{Cl}(X) \xrightarrow{\sim} \mathrm{Pic}(X)$ for any abelian variety X and its Jacobian denoted as Jac(X) can be considered as the abelian variety which parametrize the degree zero divisor classes on X by the degree map,

$$(0) \to \mathrm{Jac}(X) \to \mathrm{Pic}(X) \to \mathbb{Z} \to (0)$$

Let X be a scheme over some basis S. Consider the contravariant functor $P_{X/S} : (\mathrm{Sch}_S) \to Ab$ given by

$$P_{X/S} : T \mapsto Pic(X_T) = H^1(X \times_S T, \mathbb{G}_m)$$

This functor is not representable since $P_{X/S}$ is not a sheaf for the Zariski topology on Sch_S.

Definition 1.11. *The relative Picard functor $Pic_{X/S} : (Sch_S) \to Ab$ is defined to be the sheaf associated to the presheaf $P_{X/S}$. So if $Pic_{X/S}$ is representable then an S-scheme representing $Pic_{X/S}$ is called the relative Picard scheme of X over S.*

Concretely, if T is an S-scheme, then an element of $\mathrm{Pic}_{X/S}(T)$ can be described by giving $T' \to T$ and a line bundle L on $X_T \otimes_T T'$ such that the two pull backs of L to $X_T \otimes_T (T' \otimes_T T')$ are isomorphic. $\mathrm{Pic}_{X/S}$ cannot be expected to be representable in general conditions unless we impose certain conditions to the structure morphism $X \to S$. The most important general results about representability all work under the assumption that $X \to S$ is proper, flat and of finite presentation. If so, the representing scheme is unique up to S-isomorphism and comes with the structure of an S-group scheme, locally of finite over k.

Remark 1.12. Let C be a complete curve over a field k. Then $\mathrm{Pic}_{C/k}$ is group scheme locally of finite type over k. It can be shown that $\mathrm{Pic}_{C/k}$ is smooth over k. Assume $C(k) \neq \emptyset$, since $\mathrm{Pic}_{C/k}$ is locally of finite type over k, it suffices to show that any point of $\mathrm{Pic}_{C/k}$ with values in $R_0 := k[t]/(t^n)$ can be lifted to a point with values in

$R := k[t]/(t^{n+1})$. So if we have a line bundle L_0 on $C \otimes R_0$, then the obstruction for extending L_0 to a line bundle on $C \otimes R$ lies in $H^2(C, \mathcal{O}_C)$ which is zero since C is a curve.

Definition 1.13. *If C is a smooth curve over k with $C(k) \neq \emptyset$, then $Pic_{C/k}$ is representable by a smooth group scheme over k whose connected components are complete. In particular, the identity component $Pic^0_{C/k}$ is an abelian variety over k. It is called the **Jacobian of** C denoted as $Jac(C)$.*

Remark 1.14. For an elliptic curve E defined over a finite field $\mathbb{F}_q$ have $E(\mathbb{F}_q) \neq \emptyset$. More precisely, a theorem of Hasse states that if $E/\mathbb{F}_q$ is defined over a finite field, then

$$|\#E(\mathbb{F}_q) - q - 1| \leq 2\sqrt{q}$$

Poincaré line bundle

Let X be a scheme over S with the following situation,

1. $f : X \to S$, the structure morphism is quasi-compact and quasi-separated.
2. $f_*(\mathcal{O}_{X \times_S T}) = \mathcal{O}_T$ and f has a section $\epsilon : S \to X$.

This situation holds when X is a complete k-variety with $X(k) \neq \emptyset$. Rather than sheafifying $P_{X/S}$ there is also a way to rigidify the objects as follows: If L is a line bundle on X_T for some S-scheme T, then *rigidification* of L along ϵ_T for $\epsilon_T : T \to X_T$ induced by ϵ refers to the isomorphism $\alpha : \mathcal{O}_T \to \epsilon_T^* L$. Now define the functor $P_{X/S,\epsilon} : Sch_S \to Ab$ by

$$P_{X/S,\epsilon} : T \mapsto \{ \text{ isomorphism classes of rigidified line bundles } (L, \alpha) \text{ on } X \times_S T\}$$

with the group structure $(L, \alpha).(M, \beta) = (L \otimes M, \gamma)$ where $\gamma : \mathcal{O}_T \to \epsilon^*(L \otimes M)$.

Definition 1.15. *Suppose $P_{X/S,\epsilon}$ is representable by a scheme. Then on $X \times_S P_{X/S,\epsilon}$, there is a universal rigidified line bundle $(\mathcal{P}, \nu)$ called Poincaré line bundle with the property that, if (L, α) is a line bundle on $X \times_S T$ with rigidification along the section ϵ, then there exist a unique morphism $g : T \to P_{X/S,\epsilon}$ such that $(L, \alpha) \cong (id_X \times g)^*(\mathcal{P}, \nu)$ as rigidified line bundles on X_T.*

Remark 1.16. Note that $Pic_{X/k}$ also represents the functor $P_{X/S,0}$ of line bundles with rigidification along the zero section. The identification between the two functors is given by sending the class of a line bundle L on $X \times_k T$ to the class of $L \otimes pr_T^* \epsilon^* L^{-1}$ with its canonical rigidification along $\{0\} \times T$. So in particular, this shows the existence of a Poincaré line bundle $\mathcal{P}$ on $X \times_k Pic_{X/k}$ with a rigidification along $\alpha : \mathcal{O}_{Pic_{X/k}} \to \mathcal{P}_{\{0\} \times Pic_{X/k}}$.

Dual of an abelian variety

Definition 1.17. *Let L be a line bundle on an abelian variety X. The Mumford line bundle $\Lambda(L)$ on $X \times X$ is defined as*

$$\Lambda(L) = m^*L \otimes p_1^*L^{-1} \otimes p_2^*L^{-1}$$

The restriction of $\Lambda(L)$ to both vertical fibre $\{x\} \times X$ and horizontal fibre $X \times \{x\}$ is $t_x^*L \otimes L^{-1}$. So $\Lambda(L)$ is trivial on $\{0\} \times X$ and on $X \times \{0\}$. With the same notation, define $K(L) \subset X$ as the maximal closed subscheme such that $\Lambda(L)|_{X \times K(L)}$ is trivial over $K(L)$ i.e $\Lambda(L)|_{X \times K(L)} = pr_2^*M$ for some line bundle M on $K(L)$. So a point belongs to $K(L)$ if L is invariant under translation by this point.

Proposition 1.18. *The subscheme $K(L)$ is a subgroup scheme of X. In particular, if L is ample then $K(L)$ is a finite group scheme.*

Definition 1.19. *A line bundle L on an abelian variety is said to be non-degenerate if $K(L)$ is finite.*

Remark 1.20. The map $\phi_L : X(k) \to Pic_{X/k}$ given by $x \mapsto [t_x^*L \otimes L^{-1}]$ is the unique morphism with the property $(id_X \times \phi_L)^*(\mathcal{P}) = \Lambda(L)$ on $X \times_S X$. And as X is connected, ϕ_L factors through thee identity component $Pic^0_{X/S}$ with $\phi_L(0) = 0$.

Theorem 1.21. *Let X be an abelian variety over a field k. Then $Pic^0_{X/k}$ is reduced, hence it is an abelian variety. For every ample line bundle L, the homomrphism $\phi_L : X \to Pic^0_{X/k}$ is an isogeny with kernel $K(L)$. We have* $\dim(Pic^0_{X/k}) = \dim(X) = \dim H^1(X, \mathcal{O}_X)$.

Proof. Since ϕ_L has kernel $K(L)$ which is a finite group scheme for any ample line bundle L, it follows that ϕ_L is a finite map with $\dim(Pic^0_{X/k}) \geq \dim(X)$. For any proper variety, the tangent space of $Pic^0_{X/k}$ is isomorphic to $H^1(X, \mathcal{O}_X)$, in addition $Pic^0_{X/k}$ is smooth over k if and only if $\dim Pic^0_{X/k} = \dim H^1(X, \mathcal{O}_X)$. Altogether $\dim Pic^0_{X/k} = \dim H^1(X, \mathcal{O}_X) \leq \dim(X)$ due to the natural structure of graded algebra on the group variety X. Hence $\dim(Pic^0_{X/k}) = \dim(X)$. □

Therefore, for an abelian variety X, $\mathrm{Pic}^0_{X/k}$ is an abelian variety of same dimension as X, which parametrizes precisely the translation-invariant line bundles on X.

Definition 1.22. *The abelian variety $\hat{X} := Pic^0_{X/k}$ is called the dual of X. The restriction of the Poincaré line bundle to $X \times \hat{X}$ is denoted as $\mathcal{P}_X$. If $f : X \to Y$ is*

a homomorphism of abelian varities over k, then $\hat{f} : \hat{Y} \to \hat{X}$ called the dual of f, is the unique homomorphism such that

$$(id \times \hat{f})^* \mathcal{P}_X \cong (f \times id)^* \mathcal{P}_Y$$

as line bundles on $X \times \hat{Y}$ with rigidification along $\{0\} \times \hat{Y}$.

Proposition 1.23. *Let $f : X \to Y$ be a homomorphism. Let M be a line bundle on Y and write $L = f^*M$. Then $\phi_L : X \to \hat{X}$ is the composition of*

$$X \xrightarrow{f} Y \xrightarrow{\phi_M} \hat{Y} \xrightarrow{\hat{f}} \hat{X}$$

If f is an isogeny and M is non-degenerate then L is non-degenerate too and $\mathrm{rank}(K(L)) = \deg(f)^2.\mathrm{rank}(K(M))$.

The Poincaré line bundle on $X \times \hat{X}$ comes with a rigidification along $\{0\} \times \hat{X}$. As $\mathcal{P}|_{X\times\{0\}} \cong \mathcal{O}$, we can choose a rigidification along $X \times \{0\}$ which is unique up to an element of $k^* = \Gamma(X, \mathcal{O}^*)$ and the two rigidifications agree at $(0, 0)$.

Remark 1.24. If we view $\mathcal{P}$ as a family of line bundles on $\hat{X}$ parametrized by X, then there is a morphism $\kappa_X : X \to \hat{\hat{X}}$ with $\kappa_X(0) = 0$ and $\phi_L = \hat{\phi_L} \circ \kappa_X$. It can be also shown that if X is an abelian variety, then κ_X is an isomorphism.

1.2.1 Polarizations

Polarization is the class of an ample line bundle modulo an algebraic equivalence which carries the same information as the associated homomorphism $\lambda = \phi_L : X \to \hat{X}$.

Proposition 1.25. *Let X be an abelian variety with $\lambda : X \to \hat{X}$ be a homomorphism. Consider the line bundle $M := (id, \lambda)^* \mathcal{P}_X$ on X. Then $\phi_M = \lambda + \hat{\lambda}$. In particular, if λ is symmetric then $\phi_M = 2\lambda$.*

Remark 1.26. It can be shown that the homomorphism $\lambda : X \to \hat{X}$ being symmetric is equivalent to the existence of a finite separable field extension $k \subset K$ and L be a line bundle on X_K such that $\lambda_K = \phi_L$.

Proposition 1.27. *Let X be an abelian variety over a field k. Let $\lambda : X \to \hat{X}$ be a homomorphism such that λ is symmetric and let $M := (id, \lambda)^* \mathcal{P}_X$. Let $k \subset K$ be a field extension and L be a line bundle on X_K such that $\lambda_K = \phi_L$. Then we have the following*

1. *λ is an isogeny $\iff$ L is non-degenerate $\iff$ M is non-degenerate.*
2. *If λ is an isogeny, then L is effective if and only if M is effective.*

3. *L is ample $\iff$ M is ample.*

Corollary 1.28. *Let X/k be an abelian variety with $\lambda : X \to \hat{X}$ be a homomorphism, then the following are equivalent*

1. *λ is a symmetric isogeny and the line bundle $(id, \lambda)^*\mathcal{P}_X$ on X is ample.*
2. *λ is a symmetric isogeny and the line bundle $(id, \lambda)^*\mathcal{P}_X$ on X is effective.*
3. *there exists a field extension $k \subset K$ and an ample line bundle L on X_K such that $\lambda_K = \phi_L$.*
4. *there exists a finite separable field extension $k \subset K$ and an ample line bundle L on X_K such that $\lambda_K = \phi_L$.*

Definition 1.29. *A polarization of an abelian variety X is an isogeny $\lambda : X \to \hat{X}$ that satisfies the above equivalent conditions. If λ is an isomorphism, then it is called a principal polarization.*

Proposition 1.30. *Let $f : X \to Y$ be an isogeny. If $\mu : Y \to \hat{Y}$ is a polarization of Y, then $f^*\mu := \hat{f} \circ \mu \circ f$ is a polarization of X of degree* $\deg(f^*\mu) = \deg(f)^2.\deg(\mu)$.

Pairings

Any isogeny f gives a pairing e_f between $\mathrm{Ker}(f)$ and $\mathrm{Ker}(\hat{f})$. In particular for $f = [n]_X$ and for a polarization $\lambda : X[n] \to \hat{X}[n]$, we can obtain a bilinear form e_n^λ on $X[n]$ called the *Weil Pairing*. Let f be an isogeny of abelian varieties, denote β to be the canonical isomorphism of group schemes $\mathrm{Ker}(\hat{f})$ and the Cartier dual $\mathrm{Ker}(f)^D$.

Remark 1.31. For a commutative group scheme G over a basis S, its Cartier dual G^D represents the contra-variant functor $\mathrm{Hom}(G, \mathbb{G}_{m,S}) : Sch_{/S} \to Gr$ given by

$$T \to \mathrm{Hom}_{Sch/T}(G_T, \mathbb{G}_{m,T})$$

Definition 1.32. *Let $f : X \to Y$ be an isogeny of abelian varieties over a field k. Define*

$$e_f : Ker(f) \times Ker(\hat{f}) \to \mathbb{G}_{m,k}$$

to be the bilinear pairing on the points given by $e_f(x, y) = \beta(y)(x)$. If $Ker(f)$ is killed by $n \in \mathbb{Z}_{\geq 1}$, then e_f takes values in $\mu_n \subset \mathbb{G}$. In particular if $f = [n]_X$, then the pairing is called the Weil pairing.

$$e_n : X[n] \times \hat{X}[n] \to \mu_n$$

Let $\lambda : X \to \hat{X}$ is a homomorphism, then

$$e_n^\lambda : X[n] \times X[n] \to \mu_n$$

is a bilinear pairing given by $e_n^\lambda(x_1, x_2) = e_n(x_1, \lambda(x_2))$. If $\lambda = \phi_L$, then e_n^λ is also denoted as e_n^L.

Remark 1.33. If L is a non-degenerate line bundle over an abelian variety X, as the associated isogeny ϕ_L is a symmetric homomorphism, we have $K(L) = \mathrm{Ker}(\phi_L) \cong \mathrm{Ker}(\hat{\phi}_L)$ and

$$e_{\phi_L} : K(L) \times K(L) \to \mathbb{G}_m$$

Proposition 1.34. *1. Let $f : X \to Y$ be a homomorphism of abelian varieties over k. Then the following diagram is commutative for any integer $n \geq 1$*

$$\begin{array}{ccc} X[n] \times \hat{Y}[n] & \xrightarrow{1\times\hat{f}} & X[n] \times \hat{X}[n] \\ {\scriptstyle f\times 1}\downarrow & & \downarrow{\scriptstyle e_n} \\ Y[n] \times \hat{Y}[n] & \xrightarrow{e_n} & \mu_n \end{array}$$

2. Let $f : X \to Y$ and $g : Y \to Z$ be isogenies and write $h = g \circ f : X \to Z$ and $'i'$ be the natural inclusion homomorphism of $Ker(f) \subset Ker(h)$, then if T is a k-scheme with $x \in Ker(f)(T)$ and $\eta \in Ker(\hat{h})(T)$, then $e_f(x, \hat{g}(\eta)) = e_h(i(x), \eta)$.

Corollary 1.35. *Let X be an abelian variety wwith the polarization $\lambda : X \to \hat{X}$, then for any integer $n \geq 1$, the pairing $e_n^\lambda : X[n] \times X[n] \to \mu_n$ is alternating, i.e for any $x \in X[n](T)$, we have $e_n^\lambda(x, x) = 1$.*

1.3 Elliptic curves

An elliptic curve is a pair (E, O) where E is a nonsingular curve of genus one and $O \in E$. The elliptic curve is said to defined over k if E as a curve is defined over k and $O \in E$. The book *The Arithmetic of Elliptic Curves* by Joseph H Silverman [4] is referred for this section.
Let E be an elliptic curve given by the Weierstrass equation. Then $E \subset \mathbb{P}^2$ consists of points $P = (x, y)$ satisfying the Weierstrass equation and a point $O = [0, 1, 0]$ at infinity. Then one have a composition law denoted as $\bigoplus$ on E as follows

Composition rule Let $P, Q \in E$, let L be a line through P and Q and let R be the third point of intersection of L with E. Let L' be the line through R and O. Then

L' intersects E at R ,O and a third point. The third point is denoted by $P \bigoplus Q$. This composition law makes E into an abelian group with identity element O.

Definition 1.36. *Let E_1 and E_2 be two elliptic curves. An isogeny from E_1 to E_2 is a morphism $\phi : E_1 \to E_2$ satisfying $\phi(O) = O$.*

Two elliptic curves are said to be isogenous if there is an non-zero isogeny from E_1 to E_2. Except from the zero isogeny, every other isogeny is a finite map of curves, which gives a usual injection of corresponding function fields $\phi^* : \bar{k}(E_2) \to \bar{k}(E_1)$. Hence the degree of an isogeny ϕ denoted by *deg* ϕ is the degree of the finite extension $\bar{k}(E_1)/\phi^*\bar{k}(E_2)$ and the degree of zero isogeny is zero.
Example: For each $m \in \mathbb{Z}$, *the multiplication by* m isogeny is defined as:

$$[m] : E \to E$$
$$\text{if } m > 0,\ P \mapsto P + P + P + ... + P \text{ (m-times)}$$

for $m < 0$, $[m](P) = [-m](-P)$ and $[0](P) = O$.

Definition 1.37. *Let E be an elliptic curve and let $m \in \mathbb{Z}$ with $m \geq 1$. The m-torsion subgroup of E denoted by $E[m]$ is the set of points of E of order m.*

$$E[m] = \{P \in E : [m]P = O\}$$

Scheme-theoretically, $E[m]$ is the kernel of the isogeny $[m] : E \to E$, hence it is a subgroup scheme of $E(\bar{k})$. The torsion subgroup of E, denoted by E_{tors} is the set of points of finite order

$$E_{tors} = \bigcup_m E[m]$$

Proposition 1.38. *Let E be an elliptic curve and let $m \in \mathbb{Z}$ with $m \neq 0$.*

1. *deg* $[m] = m^2$
2. *If $m \neq 0$ in k (if either char(k) = 0 or p = char(k)> 0 and $p \not| m$), then*

$$E[m] = \frac{\mathbb{Z}}{m\mathbb{Z}} \times \frac{\mathbb{Z}}{m\mathbb{Z}}$$

3. *If char(k)=$p > 0$, then one of the following is true:*

$$\textit{(i) } E[p^e] = \{O\} \textit{ for all } e = 1, 2, 3...$$
$$\textit{(ii) } E[p^e] = \frac{\mathbb{Z}}{p^e\mathbb{Z}} \textit{ for all } e = 1, 2, 3, ...$$

More generally, we have

Theorem 1.39. *Let k be a field and X be a g-dimensional abelian variety defined over k. Let m be an integer which is prime to the characteristic of k, then $X[n]$ is isomorphic to $(\mathbb{Z}/m\mathbb{Z})^{2g}$ as an abstract group.*

Therefore, for an elliptic curve E over k and $m \in \mathbb{Z}$ which is non-zero in k, $E[m]$ is a subgroup scheme of length m^2. In positive characteristic, if $E/\mathbb{F}_{p^2}$ is an elliptic curve, then $E[p^e]$ can never have order p^2; the subgroup scheme $E[p^e]$ can either be trivial or isomorphic to $\mathbb{Z}/p^e\mathbb{Z}$.

1.3.1 Supersingular Elliptic curves

Supersingular elliptic curves are a special class of elliptic curves defined over finite fields. These curves have unique properties that distinguish them from ordinary elliptic curves.

Theorem 1.40. *Let E be an elliptic curve defined over a field k of characterisitc p. For each integer $r \geq 1$, let*

$$\phi_r : E \to E^{(p^r)} \text{ and } \hat{\phi}_r : E^{(p^r)} \to E$$

be the p^r-power Frobenius map and its dual. Then the following are equivalent

1. *$E[p^r] = 0$ for all $r \geq 1$.*
2. *The map $\hat{\phi}_r$ is purely inseparable for all $r \geq 1$.*
3. *The isogeny $[p] : E \to E$ is purely inseparable and $j(E) \in \mathbb{F}_{p^2}$.*
4. *End(E) is an order in a quaternion algebra.*

If the equivalent conditions do not hold, then $E[p^r] = \dfrac{\mathbb{Z}}{p^r\mathbb{Z}}$ for all $r \geq 1$. Further if $j(E) \in \bar{\mathbb{F}}_p$, then End$(E)$ is an order of a quadratic imaginary field.

Definition 1.41. *If E is an elliptic curve with the properties given as the equivalent conditions in the above theorem, then it is called a supersingular elliptic curve.*

Remark 1.42. Every supersingular elliptic curve defined over a field of characteristic p admits an isomorphic representation defined over $\mathbb{F}_{p^2}$.

Commutative Isogeny Diagrams

Classical notation of *pushforward* maps can be used to define commutative diagrams of isogenies. Let E_0, E_1, E_2 be three curves and two separable isogenies $\varphi_1 : E_0 \to E_1$ and $\varphi_2 : E_0 \to E_2$ of coprime degrees N_1 and N_2. Then there is a fourth curve E_3 and two *pushforward isogenies* $[\varphi_1]_*\varphi_2$ and $[\varphi_2]_*\varphi_1$ going from E_1 and E_2 toward E_3 as follows

$$\begin{array}{ccc} E_0 & \xrightarrow{\varphi_1} & E_1 \\ \downarrow{\scriptstyle \varphi_2} & \searrow & \downarrow{\scriptstyle [\varphi_1]_*\varphi_2} \\ E_2 & \xrightarrow{[\varphi_2]_*\varphi_1} & E_3 \end{array}$$

The isogenies $[\varphi_2]_*\varphi_1$ and $[\varphi_1]_*\varphi_2$ are defined separable isogenies of respective kernels $\varphi_2(ker(\varphi_1))$ and $\varphi_1(ker(\varphi_2))$ with degrees to be $\deg([\varphi_1]_*\varphi_2) = N_2$ and $\deg([\varphi_2]_*\varphi_1) = N_1$. The two sides of the commutative diagram can be seen as the decomposition of the same isogeny $\psi : E_0 \to E_3$ where $\psi = [\varphi_2]_*\varphi_1 \circ \varphi_2 = [\varphi_1]_*\varphi_2 \circ \varphi_1$.

Definition 1.43. *A finite subgroup scheme $H \subseteq E$ of an elliptic curve E/K is called primitive, if $E[m] \subseteq H$ implies $m = \pm 1$. In general, H is m-primitive if $H[m] = Ker([m]|_H)$ is primitive. equivalently if*

$$E[q] \not\subseteq H \text{ for all primes } q|m.$$

An isogeny $f : E \to E'$ is called *m-primitive* if $Ker(f)$ is m-primitive or f does not factor over $[q]$ for any prime $q|m$.

Proposition 1.44.

1. *If $H \subseteq E$ is a subgroup of order n and $d|n$, then $d|\#H[m]$ and equality holds if $(d, n/d) = 1$.*

2. *If $H \subseteq E$ is an m-primitive subgroup scheme of order n and $d|(n, m)$ then $\#H[d] = d, \#H[k] = k$ and $[d]H = H[k]$ where $k = n/d$.*

3. *Let H_1 and H_2 be two subgroup schemes of E with $H_1 \cap H_2 = (0)$. Then for $n_i = \#H_i$ and $d = (n_1, n_2)$, H_i is d-primitive. If E is supersingular, then $char(K) \nmid d$.*

4. *Let $\tilde{H}_i = [n]^{-1}(H_i)$ and $H'_i = [k_i](\tilde{H}_i)$ where $k_i = n_i/d$ and $n = k_1 + k_2$. Then $\#\tilde{H}_i[k_i] = k_i, \#H'_i = Nn$ where $N = n_1 + n_2$ and we have*

$$H'_1 + H'_2 = E[N] \text{ and } H'_1 \cap H'_2 = E[n].$$

1.4 Elliptic Curve Cryptography

Public key cryptosystems rely on what are known as *one-way trapdoor functions*. These are easy to compute injective functions $f : A \to B$ with the property that f^{-1} is hard to compute in general, but the same f^{-1} becomes quite easy to compute if someone possesses an extra piece of information k.

Typically in cryptography, we have Alice and Bob want to communicate, while Eve, the eavesdropper intercepts and tries to read their messages. Thus, if Alice and Bob use public key cryptosystems with Alice knowing the value of k, then Bob can send her a message a by sending her the quantity $b = f(a)$. Alice easily recovers $a = f^{-1}(b)$, while Eve, who does not know k, is unable to compute $f^{-1}(b)$.

Public key cryptography was invented by Diffie and Hellman in 1976. The first practical public key cryptosystem was devised the following year by Rivest, Shamir, and Adleman. The RSA cryptosystem bases its security on the difficulty of factoring large numbers. However, Diffie and Hellman did describe a key exchange algorithm whose security relies on the discrete logarithm problem in the multiplicative group of $\mathbb{F}_q$, and subsequently ElGamal created a public key cryptosystem based on the same underlying problem. Koblitz and Miller proposed replacing the finite field $\mathbb{F}_q$ with an elliptic curve E, with the hope that the discrete logarithm problem in the elliptic curve group $E(\mathbb{F}_q)$ might be harder to solve than the discrete logarithm problem in $\mathbb{F}_q^*$. Their intuition led to the creation of elliptic curve cryptography.

Definition 1.45. *Let G be a group and let $x, y \in G$ such that y is in the subgroup generated by x. Then discrete logarithm problem (DLP) is the problem of finding an integer $m \geq 1$ such that*

$$x^m = y$$

The elliptic curve discrete logarithm problem (ECDLP) is the problem of determining $'m'$ for the equation $[m]P = Q$ for the given points $P, Q \in E(\mathbb{F}_q)$.

1.4.1 Diffie-Hellman Key Exchange

It is a method to jointly establish a shared secret key over a public channel based on DLP due to Diffie and Hellman. Let Alice and Bob agree on a p be a large prime, a generator of the cyclic group $(\mathbb{Z}/p\mathbb{Z})^* = \langle g \rangle$. It is then followed by each choosing a secret integer called private keys and sending the result of g raised to the power

equal to the secret integer.

$$Alice \xleftarrow{\quad g \quad} Bob$$

$$a \qquad\qquad b$$

Private keys

$$g^a \longrightarrow g^a$$

$$g^b \longleftarrow g^b$$

Now Alice and Bob compute the shared secret g^{ab}. So Eve has to extract a or b from knowing p, g, g^a, g^b. There is no known efficient algorithm to do this on a classical computer, but quantum computer can do this efficiently.

Similarly in Elliptic Curve Diffie-Hellann key exchange, to securely exchange the value of a point on an elliptic curve,

1. Alice and Bob agree on a finite field $\mathbb{F}_q$, an elliptic curve E over $\mathbb{F}_q$, and a point $P \in E(\mathbb{F}_q)$.
2. Alice selects a secret integer a and computes the point $A = [a]P \in E(\mathbb{F}_q)$.
3. Bob selects a secret integer b and computes the point $B = [b]P \in E(\mathbb{F}_q)$.
4. Alice and Bob exchange the values of A and B over a possibly insecure communication line.
5. Alice computes $[a]B$ and Bob computes $[b]A$. They have now shared the value of the point $[ab]P$.

So for Eve to extract the message, she needs to solve the ECDLP, given three points $P, [a]P$ and $[b]P$ in $E(\mathbb{F}_q)$, compute the point $[ab]P$ without knowing a and b.

Post-quantum cryptography

Quantum computers have the potential to break many of the commonly used public-key cryptographic algorithms, such as RSA and elliptic curve cryptography (ECC), due to their ability to efficiently solve certain mathematical problems that underlie these algorithms. Therefore, there is a need to develop new cryptographic algorithms that are resistant to attacks by both classical and quantum computers.

There are several classes of mathematical problems that are believed to be hard for both classical and quantum computers. These problems serve as the foundation for post-quantum cryptographic algorithms. Some of the most promising approaches include: Lattice-based cryptography, Multivariate cryptography, Hash-based cryptography, Supersingular elliptic curve isogeny cryptography etc. It is important to note that post-quantum cryptography is still an active area of research, and no definitive standard has been established yet. The National Institute of Standards and Technology (NIST) in the United States has been leading the standardization process by soliciting and evaluating candidate algorithms. They are working towards selecting one or more post-quantum cryptographic algorithms as standards to replace existing algorithms vulnerable to quantum attacks. For further information on the above, we refer to the book "Introduction to post-quantum cryptography" by Daniel J Bernstein [7].

1.4.2 Supersingular Isogeny Diffie-Hellman (SIDH)

Supersingular Isogeny Diffie-Hellman (SIDH) is a post-quantum key exchange protocol that relies on the hardness of certain mathematical problems related to isogenies between supersingular elliptic curves. It was proposed as an alternative to traditional Diffie-Hellman key exchange, which is vulnerable to attacks by quantum computers. SIDH was created in 2011 by De Feo, David Jao, and J. Plut [6]. It uses conventional elliptic curve operations. Supersingular Isogeny Key Encapsulation SIKE or SIDH is one of the leading candidates for post-quantum cryptography and is currently being standardized by NIST as part of their ongoing effort to develop quantum-resistant cryptographic algorithms.

The setup for SIDH is as follows
Choose a prime $\mathfrak{p}$ be of the form $\mathfrak{p} = 2^a 3^b f - 1$ for a small cofactor f, together with a supersingular elliptic curve E defined over $\mathbb{F}_{p^2}$. This curve has two large torsion subgroups $E[2^a]$ and $E[3^b]$ which are assigned to Alice and Bob respectively. The protocol starts by choosing a secret subgroup of their respective torsion group and computing the corresponding secret isogeny. It is then followed by sharing the equation of the target curve of one's secret isogeny and the images of the generators of the other party's torsion group. This allows them to privately compute a new isogeny from the starting curve E whose kernel is jointly generated by the two secret subgroups and the j-invariant of the codomain curve is the shared secret between them.

Let P_A, Q_A, P_B, Q_B be the auxiliary points such that $\langle P_A, Q_A \rangle = E[2^a]$ and

$\langle P_B, Q_B \rangle = E[3^b]$. Let sk denotes the secret integer chosen.

$$Alice \xleftarrow{P_A, Q_A, P_B, Q_B} Bob$$

Chooses sk_A Chooses sk_B

$$A = \langle P_A + sk_A Q_A \rangle \subseteq E[2^a] \qquad B = \langle P_B + sk_B Q_B \rangle \subseteq E[3^b]$$

Secret Subgroups

$$\varphi_A : E \to E/A \qquad \varphi_B : E \to E/B$$

$$\xhookrightarrow{E/A,\ \varphi_A(P_B),\ \varphi_A(Q_B)}$$

$$\xhookleftarrow{E/B,\ \varphi_B(P_A),\ \varphi_B(Q_A)}$$

Now Alice can compute $\varphi_B(A)$ and Bob can compute $\varphi_A(B)$. The j-invariant of

$$(E/B)/\varphi_B(A) \cong (E/A)/\varphi_A(B) \text{ is the common secret.}$$

Security

The security of SIKE informally lies on the *(Supersingular)* isogeny walk problem: Given two elliptic curves E, E' in the same isogeny class, find a path made of isogenies of small degree between E and E'. The best generic algorithm currently known is due to Galbraith: it is a "meet in the middle" strategy that on average requires a number of elementary steps proportional to the square root of the isogeny class of E and E' with a constant amount of memory. In SIDH, since the secret subgroup $A = \langle P_A + sk_A Q_A \rangle \subset E[2^a]$, Alice's isogeny $\varphi_A : E \to E_A$ can be viewed as a composition of 2-isogenies of smaller degrees i.e if we plot a graph Γ with

$$V(\Gamma) = \{ \text{ Supersingular elliptic curves over } \mathbb{F}_{p^2} \}/\cong$$
$$E(\Gamma) = \{ \ \{E, E'\} | \ \exists \ \phi : E \to E' \text{ 2-isogeny} \}$$

then quotienting out A from E is a secret walk in this graph starting from E.
Security argument for SIDH: Γ *is a Ramanujan graph with rapid mixing properties.*

Key Recovery

In comparison to cryptosystems that rely purely on the isogeny problem, the hardness assumption underlying SIKE is weaker as the image of the auxiliary points under the secret isogeny is also revealed. In particular, on targeting Bob's secret isogeny, the key recovery amounts to finding an instance of $\varphi_B : E \to E_B$ (equivalently the kernel B) with the given information on the images of the auxiliary points $\varphi_B(P_A)$ and $\varphi_B(Q_A)$ that makes it into an atypical isogenic problem called *supersingular isogeny with torsion* (SSI-T).

Definition 1.46. ***SSI-T****: Given two coprime integers l_A and l_B, two supersingular elliptic curves E and E_B over $\mathbb{F}_{p^2}$ connected by an unknown degree l_B-isogeny $\varphi_B : E \to E_B$ and given the restriction of φ_B to the l_A torsion points of E, recover an isogeny φ_B matching these constraints.*

On studying superspecial principally polarized abelian surfaces and (2,2)-isogenies between them, as well as their endomorphisms, Wouter Castryck and Thomas Decru presented as efficient key recovery attack on SIDH. It is a polynomial-time attack and provide an implementation for all the proposed NIST-parameter sets for SIKE. This attack exploits the existence of a small non-scalar endomorphism on the starting curve and is based on the "glue and split" theorem due to Ernst Kani.

Remark 1.47. During the same time period, Luciano Maino and Chole Martindale have also presented an attack on SIDH which does not require any endomorphism information on the starting curve and the algorithm has subexponential complexity [2]. The inspiration of this attack come from an unrealted collaboration of Lucinao Maino with Wouter Castryck and Thomas Decru on studying superspecial principally polarized abelian surfaces.

Chapter 2

The Reducibility Criterion

As we discussed earlier, recovering the secret keys in SIDH amounts to solve SSI-T. One way of solving it is by constructing an explicit isogeny from an abelian surface that contains the starting curve E as a factor such that one of the components of the isogeny reveals the secret isogeny $\varphi_B : E \to E_B$. So the core idea behind this attack is to construct an auxiliary curve C, an isogeny $\gamma : E \to C$ and a polarized isogeny ψ' originating from the abelian surface $C \times E_B$ (where E_B is the codomain of the secret isogeny) such that the codomain of the isogeny is again a product of elliptic curves. Then the kernel of the isogeny ψ' is called "reducible". This phenomenon is characterized by Ernst Kani called Reducibility Criterion. So in this chapter, we are going to see how to employ Kani's reducibility criterion for the attack. For this chapter, Kani's paper [5] is referred.

This study is done by starting with reducible/irreducible anti-isometries associated to maximally isotropic subgroups with respect to the Weil pairings and then its correspondence with an isogeny factorization configuration. The definition of reducible/irreducible anti-isometry depends on the following general identification.

Proposition 2.1. *Let A be an abelian variety with dimension d with a principal polarization $\lambda : A \to \hat{A}$ defined by an ample divisor $\theta \in Div(A)$ and let $p : A \to A'$ be an isogeny. Then the following are equivalent:*

1. *The subgroup $Ker(p) \subset A[N] = K(N\theta)$ is maximally isotropic with respect to the symplectic pairing $e^{N\theta}$.*
2. *$deg(p) = N^d$ and there exist $\theta' \in Div\ (A')$ such that $p^*\theta' \sim N\theta$.*
3. *There is a principal polarization $\lambda' : A' \to \hat{A}'$ such that $\hat{p} \circ \lambda' \circ p = [N] \circ \lambda$.*

In addition, $p \to Ker(p)$ establishes a bijection as follows

1. *the set of equivalence classes of pairs* (p, λ')

2. *the set of maximally isotropic subgroups of* $A[N]$.

Definition 2.2. *Let* (A, λ) *be a principally polarized abelian surface and* $H \subseteq A[N]$ *be a maximally isotropic subgroup. Then* H *is called reducible if the unique principal polarization* λ_H *on* $A_H = A/H$ *is a product polarization i.e*

$$(A_H, \lambda_H) \cong (E_1 \times E_2, \lambda_{E_1,E_2})$$

where the ample divisor $\theta_{E_1,E_2} = pr_1^*(\theta_{E_1}) + pr_2^*(\theta_{E_2})$.

Definition 2.3. *An anti-isometry* $\psi : E_1[N] \to E_2[N]$ *is called reducible if its graph subgroup, Graph* $(\psi) \subseteq A[N]$ *has this property for* $A = E_1 \times E_2$

Classification of reducible maximally isotropic subgroups $H \subset (E_1 \times E_2)[N]$ which are non-diagonal are closely related to the factorization of isogenies $f : E_1 \to E_2$

Definition 2.4. *An isogeny diamond configuration of order* N *from* E_1 *to* E_2 *is a triplet* (f, H_1, H_2) *consisting of* $f : E_1 \to E_2$ *is an isogeny and two subgroup schemes* $H_1, H_2 \subseteq Ker(f)$ *such that*

$$\#H_1 + \#H_2 = N,\ H_1 \bigcap H_2 = 0,\ \#H_1 * \#H_2 = deg(f)$$

Two isogeny configurations (f, H_1, H_2) and (f', H_1', H_2') are said to be equivalent if and only if either one the following happens

$$f = -f, H_1' = H_2 \text{ and } H_1 = H_2' \text{ or } f' = f, H_1' = H_1 \text{ and } H_2' = H_2$$

If $f : E_1 \to E_2$ is any isogeny with two factorizations $f = f_1' \circ f_1 = f_2' \circ f_2$ such that $\deg(f_1) = \deg(f_2')$ where $f_i : E_1 \to E_i'$ and $f_i' : E_i' \to E_2$, then $(f, \text{Ker}(f_1), \text{Ker}(f_2))$ is an isogeny configuration of order $N = \deg(f_1) + \deg(f_2)$. Conversely each isogeny configuration (f, H_1, H_2) with $H_1, H_2 \subseteq \text{Ker}(f)$ yield a factorization $f_i : E_1 \to E_i' = E_1/H_i$. Moreover f_i is uniquely defined by H_i upto isomorphism. The tuple $(f, f_1, f_1', f_2, f_2')$ is called the isogeny factor set representing (f, H_1, H_2).

Theorem 2.5. *Let* E_1 *and* E_2 *be two elliptic curves over* K *and let* $N \geq 2$. *Then there is a natural bijection between the following sets of :*
(a) equivalence classes of isogeny diamond configurations (f, H_1, H_2) *of order* N *from* E_1 *to* E_2, *and*
(b) non-diagonal, reducible, maximally isotropic subgroups $H \subseteq (E_1 \times E_2)[N]$.

Proof. The theorem can be proved by constructing a map from the set of isogeny diamond configurations of order N to the set of non-diagonal, reducible maximally isotropic subgroups of $(E_1 \times E_2)[N]$ and then show it is bijective. Let $\mathbf{f} = (f, f_1, f_1', f_2, f_2')$ be an isogeny factor set representing (f, H_1, H_2), define the isogeny
$p = p_f : E_1 \times E_2 \to E_1' \times E_2'$ as

$$p(x_1, x_2) = (f_1(x_1) - \tilde{f}_1'(x_2), f_2(x_1) + \tilde{f}_2'(x_2)$$

where $\tilde{f}_i' := \lambda_{E_i}^{-1} \circ \hat{f}_i \circ \lambda_{E_2} : E_2 \to E_i'$.
Claim: $H_f = \mathrm{Ker}(p_f)$ is a reducible, maximally isotropic subgroup of $A[N]$ with respect to the product polarization $\lambda_{E_1.E_2}$ on $A = E_1 \times E_2$.
By the proposition, it is enough to show that

$$\hat{p} \circ \lambda_{E_1', E_2'} \circ p = \lambda_{E_1, E_2} \circ [N]_A \tag{2.6}$$

holds. To verify this, note that it is equivalent to the matrix equation

$$\tilde{M}(p) M(p) = \mathrm{diag}\,([N]_{E_1}, [N]_{E_2}) \tag{2.7}$$

where for the isogeny $p \in Hom(E_1 \times E_2, E_1' \times E_2')$, $M(p)$ is the matrix

$$\begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix} \text{ where } p_{ij} \in \mathrm{Hom}\ (E_j \to E_i')$$

and $\tilde{M}(p)$ is the adjoint matrix defined by $\begin{bmatrix} \tilde{p_{11}} & \tilde{p_{12}} \\ \tilde{p_{21}} & \tilde{p_{22}} \end{bmatrix}$. Now by definition for $p = p_f$

$$M(p_f) = \begin{bmatrix} f_1 & -\tilde{f}_1' \\ f_2 & \tilde{f}_2' \end{bmatrix} \text{ and } \tilde{M}(p_f) = \begin{bmatrix} \tilde{f}_1 & \tilde{f}_2' \\ -f_1' & f_2' \end{bmatrix}$$

Hence the equation 2.7 is equivalent to the equations

$$\tilde{f}_1 \circ f_1 + \tilde{f}_2 \circ f_2 = [deg(f_1)]_{E_1} + [deg(f_2)]_{E_1} = [N]_{E_1}$$
$$\tilde{f}_1 \circ (-\tilde{f}_1') + \tilde{f}_2 \circ \tilde{f}_2' = 0$$
$$(-f_1') \circ f_1 + f_2' \circ f_2 = 0$$
$$(-f_1') \circ (-\tilde{f}_1') + f_2' \circ \tilde{f}_2' = [deg(f_1')]_{E_2} + [deg(f_2')]_{E_2} = [N]_{E_2}$$

which holds true by the representing isogeny factor set. Thus $H_f = Ker(p_f)$ is a reducible, maximally isotropic subgroup of $A[N]$. It is not diagonal, since if $H_f = H_1' \times H_2'$ with $H_i' \subset E_i$, then $H_1' \subset Ker(f_1)$ and $H_2' \subset Ker(\tilde{f}_2')$. Since $\#H_f = N^2$, either one of $\#H_i' \geq N$ which is a contradiction because $\deg(f_1) = N - \deg(f_2) < N$ and $\deg \tilde{f}_2' = \deg(f_1) < N$. It can be also shown that this map is compatible with the

choice of the isogeny factor set representing f.
It remains to the prove the map is bijective. Let f and g be two isogeny configurations such that $Ker(p_f) = Ker(p_g)$. Thus p_f and p_g both satisfy 2.6, then there exist an isomorphism ϕ such that $p_g = \phi \circ p_f$ and $\lambda_{E_1',E_2'} = \hat{\phi} \circ \lambda_{E_1'',E_2''} \circ \phi$. Hence they define equivalent isogeny configuration, so it is injective. Given a non-diagonal, reducible, maximally isotropic subgroup H of $A[N]$, define an isogeny $p : A = E_1 \times E_2 \to A' = E_1' \times E_2'$ with $Ker(p) = H$. Let $M(p) = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix}$, then put $f_1 = p_{11}$, $f_2 = p_{21}$, $f_1' = -\tilde{p_{12}}$ and $f_2' = \tilde{p_{22}}$. Define $f = f_1' \circ f_1$. Since p satisfies 2.6, it can be shown that $\mathbf{f} = (f, f_1, f_1', f_2, f_2')$ is an isogeny factor set of order N. □

Corollary 2.8. *If $\boldsymbol{f} = (f, f_1, f_1', f_2, f_2')$ is any isogeny factor set of order N with $Ker(f_1) \cap Ker(f_2) = (0)$, then there is an unique reducible anti-isometry $\psi = \psi_f : E_1[N] \to E_2[N]$ such that*

$$\tilde{f_1'} \circ \psi = f_1|_{E_1[N]} \quad \textit{and} \quad \tilde{f_2'} \circ \psi = -f_2|_{E_2[N]} \tag{2.9}$$

and every reducible anti-isometry arises in this way. Thus the theorem 2.5 restricts to a bijection between the following sets of

1. *equivalence classes of isogeny diamond configuration of order N from E_1 to E_2 and*

2. *reducible anti-isometries $\psi : E_1[N] \to E_2[N]$.*

Proof. Let p_f be the associated isogeny as in 2.5 and let $H = Ker(p_f)$, $H_i = Ker(f_i)$. Since $\mathrm{Ker}(p_f) \cap (E_1 \times (0)) = (H_1 \cap H_2) \times (0)$. From the hypothesis $H_1 \cap H_2 = (0)$, it follows that $(pr_2)|_H : H \to E_2[N]$ is injective and hence bijective since $\#H = N^2 = \#E_2[N]$.
Thus if we put $\psi' := pr_1 \circ (pr_2)^{-1}|_H : E_2[N] \to E_1[N]$, then $H = \{(\psi'(y), y) : y \in E_2[N]\}$. Since H is an isotropic subgroup of $A[N]$, then ψ' is an anti-isometry and hence so its inverse $(\psi = \psi')^{-1}$ and by construction $\mathrm{Ker}(p_f) = \mathrm{Graph}(\psi)$. The map ψ is unique since if ψ_1 is another anti-isometry with $\mathrm{Graph}(\psi_1) \subseteq \mathrm{Ker}(p_f) = \mathrm{Graph}(\psi)$, then $\psi_1 = \psi$.
Conversely, if $\psi : E_1[N] \to E_2[N]$ is a reducible anti-isometry, then by theorem 2.5 there exists an isogeny factor set $\mathbf{f} = (f, f_1, f_1', f_2, f_2')$ of order N such that $\mathrm{Ker}(p_f) = \mathrm{Graph}(\psi)$. Since ψ is an isomorphism, $\mathrm{Graph}(\psi) \cap (E_1 \times (0)) = (0)$ and so $\mathrm{Ker}(f_1) \cap \mathrm{Ker}(f_2) = (0)$. □

Observe that if we apply f_1' to both sides of $\tilde{f_1'} \circ \psi = f_1|_{E_1[N]}$ from 2.9, we obtain

$$[n_2] \circ \psi = f|_{E_1[N]} \text{ where } n_2 = \deg(f_2) = \deg(f_1')$$

which characterizes ψ if and only if $(n_2, N) = 1$. In particular, if N is prime, then an anti-isometry $\psi : E_1[N] \to E_2[N]$ is reducible if and only if there is an isogeny $f : E_1 \to E_2$ of degree $k(N-k)$ such that the above equation holds for $n_2 = N-k$, for any such f give rise to the isogeny diamond configuration $(f, Ker(f)[k], Ker(f)[N-k])$. The following is an explicit description about the complete characterization of the reducible anti-isometries $\psi : E_1[N] \to E_2[N]$ by using the results from the proposition 1.44.

Theorem 2.10. *Let (f, H_1, H_2) be an isogeny diamond configuration of order N from E_1 to E_2 and put $n = N/d$ and $k_i = n_i/d$ where $d = (n_1, n_2)$ and $n_i = \#H_i$. Then $f = \bar{f} \circ [d]$, f factors uniquely over $[d]$ and there is a unique reducible anti-isometry $\psi = \psi_f : E_1[N] \to E_2[N]$ such that*

$$\psi(k_1 x_1 + k_2 x_2) = \bar{f}(x_2 - x_1) \tag{2.11}$$

for all $x_i \in [n]^{-1}(H_i)$ and every reducible anti-isometry is of this form. In addition if $\boldsymbol{f} = (f', H_1', H_2')$ is another isogeny diamond configuration, then we have $\psi_f = \psi_{f'}$ if and only if $\boldsymbol{f} \sim \boldsymbol{f'}$.

Proof. Let $H_0 = \mathrm{Ker}(f)$ and $H_i[d] = H_i \cap E_1[d]$ for $i = 0, 1, 2$. Observe $H_0[d] = H_1[d] \times H_2[d] \leq E_1[d]$. Since $d|n_i$, we have $d|\#H_i[d]$, so $d^2|H_0[d]$. But $\#E_1[d] = d^2$ and so $H_0[d] = E_1[d]$. Thus $E_1[d] \leq H_0$ and hence f factors over $[d]$.
Let $(f, f_1, f_1', f_2, f_2')$ be an isogeny factor set associated to (f, H_1, H_2). Then $\tilde{f_1'} \circ \bar{f} \circ [d] = \tilde{f_1'} \circ f_1' \circ f_1 = n_2 f_1 = k_2 f_1 \circ [d]$ and $\tilde{f_2'} \circ \bar{f} \circ [d] = k_1 f_2 \circ [d]$. Thus

$$\tilde{f_1'} \circ \bar{f} = k_2 f_1 \quad \tilde{f_2'} \circ \bar{f} = k_1 f_2. \tag{2.12}$$

Now by corollary, there is a unique anti-isometry $\psi : E_1[N] \to E_2[N]$ corresponding to f.
Claim: ψ satisfies 2.11.
Let $x_i \in [n]^{-1}(H_i)$, then $k_i x_i \in E_1[N]$ for $N k_i x_i = n_i n x_i = 0$. So left hand side is defined. Observe that $\mathrm{Ker}(\tilde{f_1'}) \cap \mathrm{Ker}(\tilde{f_2'}) = (0)$ for otherwise f_1', f_2' factor over a common isogeny which contradicts $(f, f_1, f_1', f_2, f_2')$ forms a diamond. Thus to verify 2.11 it is enough to show it is true after applying $\tilde{f_i'}$ for $i = 1, 2$. So we have

$$\tilde{f_1'}(\psi(k_1 x_1 + k_2 x_2)) = f_1(k_1 x_1 + k_2 x_2) = f_1(n x_1 + k_2(x_2 - x_1)) = \tilde{f_1'}(\bar{f}(x_2 - x_1))$$
$$\tilde{f_2'}(\psi(k_1 x_1 + k_2 x_2)) = -f_2(k_1 x_1 + k_2 x_2) = f_2(-n x_2 + k_1(x_2 - x_1)) = \tilde{f_2'}(\bar{f}(x_2 - x_1)).$$

Since every reducible anti-isometry ψ satisfies 2.9, we see by the above that every ψ satisfies 2.11. Uniqueness of the ψ follows from the fact that $[k_1][n]^{-1}(H_1) + [k_2][n]^{-1}(H_2) = E_1[N]$. □

Chapter 3

The set-up and strategy

This algorithm is presented using a general base elliptic curve E_0, however the commonly chosen base curves in SIDH is $E_{start} : y^2 = x^3 + x$ or $E_{start} : y^2 = x^3 + 6x^2 + x$ with respective j-invariants 1728 and 287496. Hence on attacking SIDH it becomes $E_0 = E_{start}$. This attack target Bob's private key. The key recovery amounts to finding an instance of 3^b-isogeny $\varphi : E_0 \to E$ (equivalently the secret subgroup B) with the given information on the images of the auxiliary points $\varphi(P_A), \varphi(Q_A)$ that makes it into an atypical isogenic problem. This method can also be used to recover Alice's key and more generally for any choices of l_{Alice} and l_{Bob} instead of just $l_{Alice} = 2$ and $l_{Bob} = 3$.
The initial input for the algorithm is

1. a prime $\mathfrak{p}$ of the form $2^a 3^b f - 1$ for integers $a \geq 2, b, f \geq 1$ with $2^a \simeq 3^b$.
2. an elliptic curve $E_0/\mathbb{F}_{p^2}$ with $\#E_0(\mathbb{F}_{p^2}) = (p+1)^2$.
3. the generators P_0, Q_0 of $E_0[2^a]$.
4. a 3^β-isogeny $\tau : E_0 \to E_{start}$ for some $\beta \geq 0$.
5. the codomain $E/\mathbb{F}_{p^2}$ of a secret cyclic 3^b-isogeny $\varphi : E_0 \to E$.
6. the generators $P = \varphi(P_0)$ and $Q = \varphi(Q_0)$ of $E[2^a]$.

This input returns an isogeny φ as the output which we assume the secret isogeny φ is uniquely determined by the input and it is true with overwhelming probability. This attack runs in heuristic polynomial time on a classical computer.

3.1 Decision strategy via Gluing and Splitting

Given the inputs, the goal is to decide whether or not there exists

$$\text{a } 3^b\text{-isogeny } \varphi : E_0 \to E \text{ such that } P = \varphi(P_0) \text{ and } Q = \varphi(Q_0)$$

Two temporary assumptions are made here, which will be discussed later in the generalization and construction of the algorithm. The assumptions are

1. Suppose that $2^a > 3^b$.
2. Let $c = 2^a - 3^b$. Assume there exists a cyclic c-isogeny $\gamma : E_0 \to C$ for some codomain curve C and can compute the images $P_c = \gamma(P_0)$ and $Q_c = \gamma(Q_0)$.

Let $x \in \mathbb{Z}$ be an integer that satisfies

$$x3^b \equiv 1 \bmod 2^a \text{ and } -xc \equiv 1 \bmod 2^a$$

Suppose there exist a 3^b-isogeny $\varphi : E_0 \to E$ such that $P = \varphi(P_0)$ and $Q = \varphi(Q_0)$, then consider the isogeny

$$\psi = [-1] \circ \varphi \circ \hat{\gamma} : C \to E$$

with $\psi(P_c) = -cP$ and $\psi(Q_c) = -cQ$.
Observe that for all $R, S \in C[2^a]$, we have

$$e_{2^a}(x\psi(R), x\psi(S)) = e_{2^a}(R, S)^{x^2 c 3^b} = e_{2^a}(R, S)^{-1}$$

Hence the group homomorphism $[x] \circ \psi|_{C[2^a]} : C[2^a] \to E[2^a]$ is an anti-isometry with respect to the 2^a-Weil pairing. This implies that the graph subgroup

$$\langle (P_c, x\psi(P_c)), (Q_c, x\psi(Q_c)) \rangle = \langle (P_c, -xcP), (Q_c, -xcQ) \rangle = \langle (P_c, P), (Q_c, Q) \rangle$$

is maximally isotropic with respect to the 2^a-Weil pairing on $C \times E$.
And this subgroup form the kernel of a $(2^a, 2^a)$-isogeny which is a chain of $(2, 2)$-isogenies of length a. Quotienting out this subgroup is a walk in the $(2, 2)$-isogeny graph of principally polarized abelian surfaces over $\bar{\mathbb{F}}_p$ all of whose vertices are defined over $\mathbb{F}_{p^2}$. These vertices comes in two types;

1. about $\mathfrak{p}^2/288$ products of supersingular elliptic curves.
2. about $\mathfrak{p}^3/2880$ Jacobians of superspecial genus-2 curves.

So one can expect most isogenies to be among Jacobians of genus-2 curves which can be computed by Richelot formula, see Section 3.3.2 for the detailed discussion. With a proportion of $10/\mathfrak{p}$, the codomain of the $(2^a, 2^a)$-isogeny can be a product of elliptic curves, in which the situation is called *split*. The role of isogeny γ is to force it into this expectional situation through ψ.

In this case, the kernel of ψ is a cyclic group of order $c3^b$ and it admit two cyclic subgroups $H_1 = \mathrm{Ker}(\hat{\gamma})$ and $H_2 = \gamma(B)$ of respective orders c and 3^b with the following properties

$$\#H_1 + \#H_2 = 2^a \ , \ \#H_1.\#H_2 = \deg\psi \ , \ H_1 \cap H_2 = \{0\}$$

So the triplet (ψ, H_1, H_2) is an isogeny diamond configuration of order 2^a. Then by Kani's theorem 2.10, the anti-isometry $x\psi|_{C[2^a]}$ is reducible such that it satisfies the condition 2.11. Calculate $d = \gcd(c, 3^b) = 1$. To verify 2.11 for all points R_1, R_2 such that $2^a R_1 \in H_1$ and $2^a R_2 \in H_2$, first observe that $\psi(R_1)$ and $\psi(R_2)$ are 2^a-torsion points, hence

$$x\psi(cR_1 + 3^b R_2) = 3^{-b}(2^a - 3^b)\psi(R_1) + 3^{-b}3^b\psi(R_2) = \psi(R_2) - \psi(R_1)$$

Then by Kani's theorem the anti-isometry $x\psi|_{C[2^a]}$ is reducible.

So the decision strategy amounts to find whether or not quotienting $C \times E$ by $\langle (P_c, x\psi(P_c)), (Q_c, x\psi(Q_c))\rangle = \langle (P_c, P), (Q_c, Q)\rangle$ result in a product of elliptic curves. As if there exists a 3^b-isogeny $\varphi : E_0 \to E$ such that $P = \varphi(P_0)$ and $Q = \varphi(Q_0)$, then this results holds true. Particularly, since quotienting out this subgroup is a process of walking in the $(2,2)$-isogeny graph of principally polarized surfaces, with overwhelming probability, the first $(a-1)$ steps will be one gluing step of $C \times E$ to a Jacobian of genus-2 curve followed by $(a-2)$ Richelot isogenies between Jacobians of genus-2 curves and a final test for checking whether the last step splits or not.

3.2 Construction of the auxiliary isogeny γ

Constructing a c-cyclic isogeny γ and computing its images is non-trivial. The properties of chosen curves E_{start} plays here an important role in the construction of the isogeny γ. Both chosen curves E_{start} come with an endomorphism $2\mathbf{i}$ satisfying $(2\mathbf{i})^2 = -4$ as follows

1. For $E_{start} : y^2 = x^3 + x$, let $\mathbf{i} : (x, y) \to (-x, \sqrt{-1}y)$, then $2\mathbf{i} = [2] \circ \mathbf{i}$.
2. For $E_{start} : y^2 = x^3 + 6x^2 + x$, let f be the 2-isogeny from $y^2 = x^3 + 6x^2 + x$ to the curve $y^2 = x^3 + x$, then $2\mathbf{i} = \hat{f} \circ \mathbf{i} \circ f$.

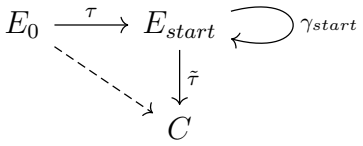
There is a reasonable chance of $1/\sqrt{a}$ for the prime factors of c to be congruent 1 mod 4, so c can be expressed as $c = u^2 + 4v^2 = (u + 2\mathbf{i}v)(u - 2\mathbf{i}v)$.

Remark 3.1. The method of finding u and v in the case of square-free integer c is by using the primes factors l of c to compute $\prod_{l|c} gcd(z_l + \mathbf{i}, l)$ whose outcome is among $\pm(u + 2\mathbf{i}v), \pm\mathbf{i}(u + 2\mathbf{i}v)$ where z_l is any integer such that $z_l^2 \equiv -1 \bmod l$.

Define a degree c endomorphism of E_{start} as

$$\gamma_{start} = [u] + [v] \circ 2\mathbf{i}$$

The input (4) a 3^β-isogeny $\tau : E_0 \to E_{start}$ for some $\beta \geq 0$, is used to construct γ. Let $\tilde{\tau} : E_{start} \to C$ be the isogeny with kernel $\gamma_{start}(\tau(E_0[3^\beta])) = \gamma_{start}(ker(\hat{\tau}))$



Then $\tilde{\tau} \circ \gamma_{start} \circ \tau : E_0 \to C$ is a $3^{2\beta}c$-isogeny vanishing on $E_0[3^\beta]$, so it factors over $[3^\beta]$, then define

$$\gamma = \frac{\tilde{\tau} \circ \gamma_{start} \circ \tau}{3^\beta}$$

Evaluating γ

Now it remains to evaluate γ on the 2^a torsion points P_0 and Q_0.

Case 1: $\beta \leq b$: This is relevant in attacking SIDH when $E_0 = E_{start}$, as in the case of SIKE, while β will grow during search-to-decision reduction, but never beyond b. Since $\ker \hat{\tau} \subset E_0[3^b] \subset E(\mathbb{F}_{p^2})$, let $T \in E_0(\mathbb{F}_{p^2})$ be a generator of $\ker \hat{\tau}$, compute $\tilde{\tau}$ with kernel $\langle \gamma_{start}(T) \rangle$. So evaluating P_0 and Q_0 is by computing the images under $\tilde{\tau} \circ \gamma_{start} \circ \tau$ and then scalar multiplying by the multiplicative inverse of 3^β mod 2^a.

Case 2: $\beta > b$: Observe that the isogeny $\tilde{\tau}$ is the pushforward isogeny $[\gamma_{start}]_* \hat{\tau}$ with degree 3^β. This gives an alternative method to find $\tilde{\tau}$ by using *Deuring Correspondence*. The Deuring correspondence defines bijection between isogenies of supersingular elliptic curves and ideals of maximal orders in a quaternion algebra. For the specific choice of E_{start} in this case, there is an explicit isomorphism

$$i : \mathrm{End}(E_{start}) \to \mathcal{O}_{start}$$

where $\mathcal{O}_{start}$ is a maximal order in the quaternion algebra $B_{\mathfrak{p},\infty} = \langle 1, \mathbf{i}, \mathbf{j}, \mathbf{ij} \rangle$ over $\mathbb{Q}$ with $\mathbf{i}^2 = -1, \mathbf{j}^2 = -\mathfrak{p}$.

Now by using the techniques from the literature [10], one can covert the isogeny $\hat{\tau}$ into a left ideal $I_{\hat{\tau}} \subset \mathcal{O}_{start}$ of norm 3^β. Then compute the left ideal $I_{\tilde{\tau}} = [i(\gamma_{start})]_* I_{\hat{\tau}}$. Finally convert back the left ideal $I_{\tilde{\tau}}$ into a length β chain of 3-isogenies starting from E_{start}, which gives the isogeny $\tilde{\tau}$ of norm 3^β. So here too, evaluating P_0 and Q_0 is by computing the images under $\tilde{\tau} \circ \gamma_{start} \circ \tau$ and then scalar multiplying by the multiplicative inverse of 3^β mod 2^a.

3.3 Computing chains of (2,2)-isogenies

The decision strategy includes the computation of a chain of $(2, 2)$-isogenies starting off by gluing $C \times E$ into a Jacobian of a genus-2 curve, followed by Richelot isogenies. This is done in such a way that it will never run into a product of elliptic curves except possibly at the last step.

Representation of points in J_H

From a geometrical point of view, one can obtain a smooth projective curve C by adding suitable *points at infinity* to the base curve $y^2 = f(x)$. If deg f is odd, there is just one such point, and it is always a rational point. If deg f is even, there are two such points, which correspond to the two square roots of the leading coefficient of f.

Classical Mumford Representation: Let $C : y^2 = f(x)$ be a hyperelliptic curve of odd degree with genus g over k and D be a non-zero effective divisor of degree d such that if $i : C \to C$ is a map which takes $(x, y) \mapsto (x, -y)$ satisfies $i(P_j) \neq P_k$ for all $j \neq k$ in D. Then there exist unique polynomials $u(x), v(x) \in k[x]$ such that

1. $u(x)$ is a monic polynomial of degree d.
2. $\deg(v(x)) < d$.
3. if $P = (x', y') \in C$ then $P \in \text{Supp(D)} \iff u(x') = 0, v(x') = y'$ and in this case, $v_P(D)$ is the multiplicity of x' as a root of $u(x)$.

This representation of D by a pair of polynomials is called the *Mumford representation of D.*

Lemma 3.2. *Let $C : y^2 = f(x)$ be a hyper elliptic curve of odd degree and of genus g over k. Denote its Jacobian as usual by J. Then for every point $P \in J(k)$ there is a unique divisor $D \in Div\ C(k)$ of degree $d = deg(D) \leq g$ such that $P = [D - d.\infty]$.*

Moreover, if the degree of the polynomial f is even, then for every point $P \in J(k)$, there is a unique divisor D such that $P = [D - \frac{d}{2}\infty_1 - \frac{d}{2}\infty_2]$. Throughout, we are interested in genus-2 curve $H : y^2 = h(x) = c_6x^6 + c_5x^5 + ... + c_0$, with $c_6 \neq 0$, so that it has two places ∞_1, ∞_2 at infinity, and all points on its Jacobian J_H that are assumed to be representable as $(\alpha_1, \beta_1) + (\alpha_2, \beta_2) - \infty_1 - \infty_2$ with $\alpha_1 \neq \alpha_2$. Then by the above lemma, points on the Jacobian J_H is assumed to have a Mumford representation of the form $D = [x^2 + u_1x + u_0, v_1x + v_0]$.

3.3.1 Gluing $E \times C$ into a Jacobian

In the first step, $E \times C$ need to be glued into the Jacobian of a genus-2 curve H via the $(2,2)$-subgroup $\langle (2^{a-1}P_c, 2^{a-1}P), (2^{a-1}Q_c, 2^{a-1}Q)\rangle$. Then we need to find the images of the points $(P_c, P), (Q_c, Q)$ of this corresponding isogeny.

Proposition 3.3. *Let $C/K : y^2 = (x-\alpha_1)(x-\alpha_2)(x-\alpha_3)$ and $E : y^2 = (x-\beta_1)(x-\beta_2)(x-\beta_3)$ be elliptic curves over a field K of characterisitc different from two. Write Δ_α for the discriminant of $(x-\alpha_1)(x-\alpha_2)(x-\alpha_3)$ and Δ_β for the discriminant of $(x-\beta_1)(x-\beta_2)(x-\beta_3)$. Define*

1. $a_1 = (\alpha_3-\alpha_2)^2/(\beta_3-\beta_2) + (\alpha_2-\alpha_1)^2/(\beta_2-\beta_1) + (\alpha_1-\alpha_3)^2/(\beta_1-\beta_3)$
2. $b_1 = (\beta_3-\beta_2)^2/(\alpha_3-\alpha_2) + (\beta_2-\beta_1)^2/(\alpha_2-\alpha_1) + (\beta_1-\beta_3)^2/(\alpha_1-\alpha_3)$
3. $a_2 = \alpha_1(\beta_3-\beta_2) + \alpha_3(\beta_2-\beta_1) + \alpha_2(\beta_1-\beta_3)$
4. $b_2 = \beta_1(\alpha_3-\alpha_2) + \beta_3(\alpha_2-\alpha_1) + \beta_2(\alpha_1-\alpha_3)$
5. $A = \Delta_\beta a_1/a_2,\ B = \Delta_\alpha b_1/b_2$
6. $h(x) = -(A(\alpha_2-\alpha_1)(\alpha_1-\alpha_3)x^2 + B(\beta_2-\beta_1)(\beta_1-\beta_3))$
   $.(A(\alpha_3-\alpha_2)(\alpha_2-\alpha_1)x^2 + B(\beta_3-\beta_2)(\beta_2-\beta_1))$
   $.\ (A(\alpha_1-\alpha_3)(\alpha_3-\alpha_2)x^2 + B(\beta_1-\beta_3)(\beta_3-\beta_2))$

Then the (2,2)-isogeny with domain $C\times E$ and kernel $\langle((\alpha_1,0),(\beta_1,0)),((\alpha_2,0)(\beta_2,0))\rangle$ has a codomain as the Jacobian of a genus-2 curve H defined by $y^2 = h(x)$. The degree-2 morphisms of the dual isogeny are given by

$$\begin{gathered}\phi_1 : H \to C \\ (x,y) \mapsto (s_1/x^2 + s_2, (\Delta_\beta/A^3)(y/x^3)) \\ \phi_2 : H \to E \\ (x,y) \mapsto (t_1x^2 + t_2, (\Delta_\alpha/B^3)y)\end{gathered}$$

where $s_1 = -(B/A)(a_2/a_1),\ \ t_1 = -(A/B)(b_2/b_1)$,

$$s_2 = \frac{1}{a_1}(\alpha_1(\alpha_3-\alpha_2)^2/(\beta_3-\beta_2) + \alpha_3(\alpha_2-\alpha_1)^2/(\beta_2-\beta_1) + \alpha_2(\alpha_1-\alpha_3)^2/(\beta_1-\beta_3)),$$

$$t_2 = \frac{1}{b_1}(\beta_1(\beta_3-\beta_2)^2/(\alpha_3-\alpha_2) + \beta_3(\beta_2-\beta_1)^2/(\alpha_2-\alpha_1) + \beta_2(\beta_1-\beta_3)^2/(\alpha_1-\alpha_3))$$

Proof. The proposition is proved in the reference [[8], Proposition 4]. □

The above mentioned morphisms ϕ_i for $i = 1, 2$ can be extended and combined to form a $(2,2)$-isogeny $\phi : J_H \to C \times E$ by mapping

$$[\sum_j P_j] \to \sum_j \phi(P_j)$$

Observe that $\hat{\phi}$ is the isogeny of interest and to compute the image of a point $(P_c, P) \in C \times E$, it suffices to compute for some $[D] \in \phi^{-1}\{(P_c P)\}$

$$2[D] = \hat{\phi}\phi([D]) = \hat{\phi}(P_c, P).$$

Let $D = P_H + Q_H - \infty_1 - \infty_2$ represent a point in J_H. Assume its Mumford representation is of the form $[x^2 + u_1 x + u_0, v_1 x + v_0]$. Note that the divisor $\infty_1 + \infty_2$ maps to ∞ under ϕ_1 and ϕ_2. So it's left to compute $\phi_i(P_H + Q_H)$.

For $i = 2$, the line connecting $\phi_2(P_H)$ and $\phi_2(Q_H)$ has a slope of $\lambda_2 = \dfrac{-(\Delta_\alpha/B^3)v_1}{t_1 u_1}$ and denote $\omega_2 = \lambda_2^2 + \sum_{i=1}^3 \beta_i - t_1(u_1^2 - 2u_0) - 2t_2$, then

$$\phi_2(P_H + Q_H) = (\omega_2, -\lambda_2(\omega_2 - t_2 + (u_0 v_1 - u_1 v_0)t_1/v_1))$$

To calculate ϕ_1, first consider the transformation $\tilde{i} : (x, y) \mapsto (1/x, y/x^3)$. Then let $\tilde{u_0}, \tilde{u_1}, \tilde{v_0}, \tilde{v_1}$ be the Mumford coordinates of $\tilde{P_H} + \tilde{Q_H}$ with

$$\tilde{u_0} = \frac{1}{u_0},\ \tilde{u_1} = \frac{u_1}{u_0},\ \tilde{v_0} = \frac{u_1 v_0 - u_0 v_1}{u_0^2},\ \tilde{v_1} = \frac{u_1^2 v_0 - u_0 v_0 - u_0 u_1 v_1}{u_0^2}$$

Denote λ_1 to be the slope of the line connecting $\phi_1(P_H)$ and $\phi_1(Q_H)$ and $\omega_1 = \lambda_1^2 + \sum_{i=1}^3 \alpha_i - s_1(\tilde{u}_1^2 - 2\tilde{u}_0) - 2s_2$, then

$$\phi_1(P_H + Q_H) = (\omega_1, -\lambda_1(\omega_1 - s_2 + (\tilde{u}_0\tilde{v}_1 - \tilde{u}_1\tilde{v}_0)s_1/\tilde{v}_1))$$

This gives four equations in the unknowns u_0, u_1, v_0, v_1:

$$\begin{aligned} x(\phi_1(P_H + Q_H)) &= x(P_c) \\ y(\phi_1(P_H + Q_H)) &= y(P_c) \\ x(\phi_2(P_H + Q_H)) &= x(P) \\ y(\phi_2(P_H + Q_H)) &= y(P) \end{aligned}$$

Together with expressing $[D] \in J_H$ in the equation of H gives

$$\begin{aligned} 2v_0^2 - 2v_0 v_1 u_1 + v_1^2(u_1^2 - 2u_0) = 2c_0 + (-u_1)c_1 + (u_1^2 - 2u_0)c_2 + (-u_1^3 + 3u_0 u_1)c_3 + \\ (u_1^4 - 4u_1^2 u_0 + 2u_0^2)c_4 + (-u_1^5 + 5u_1^3 u_0 - 5u_1 u_0^2)c_5 + (u_1^6 - 6u_1^4 u_0 + 9u_1^2 u_0^2 - 2u_0^3)c_6 \end{aligned}$$

This system will have four solutions defined over $\mathbb{F}_{p^2}$. Take any of these solutions and double the corresponding point on J_H produces the desired image of (P_c, P).

3.3.2 Richelot isogenies

The next $(a-2)$ steps are assumed to be Richelot isogenies between Jacobians of genus-2 curves. In the case of genus-2 curves, a Richelot isogeny $\phi : J_C \to J_{C'}$ induces an isomorphism between the degree-2 isogeny class of J_C and the degree-2 isogeny class of $J_{C'}$. In general, we have the following definition

Definition 3.4. *Let A be a principally polarized abelian surface. A Richelot isogeny $\phi : A \to A/G$ is an isogeny where $G \cong (\mathbb{Z}/2\mathbb{Z})^2$ is a maximal 2-isotropic subgroup of $A[2]$.*

Let C denotes a hyperelliptic curve of genus two, over a field k of characteristic not two and J_C be the Jacobian of C. It follows from the nondegeneracy of the Weil pairing that the maximal 2-Weil isotropic subgroups of $J_C[2]$ are isomorphic to $(\mathbb{Z}/2\mathbb{Z})^2$.

Lemma 3.5. *Let R be a proper, nontrivial subgroup of $J_C[2]$. If R is the kernel of an isogeny of principally polarised abelian surfaces, then R is a maximal 2-Weil-isotropic subgroup of $J_C[2]$.*

Therefore the lemma implies that if A is a principally polarised abelian surface, and if $\phi : J_C \to A$ is an isogeny respecting the polarisations such that the kernel of ϕ is a proper, nontrivial subgroup of $J_C[2]$, then ϕ is a $(2,2)$-isogeny.

Explicit formula of Richelot isogenies between Jacobians of genus-2 curves are available in the reference [9] chapter 8. Let $H : y^2 = h(x)$ be a hyperelliptic curve and a $(2,2)$-subgroup

$$\langle [g_1(x), 0], [g_2(x), 0] \rangle, \quad g_1(x) = x^2 + g_{11}x + g_{10} \quad g_2(x) = x^2 + g_{21}x + g_{20}$$

of its Jacobian. Let $g_3(x) = h(x)/g_1(x)g_2(x) = g_{32}x^2 + g_{31}x + g_{30}$, then compute

$$\delta = det \begin{bmatrix} g_{10} & g_{11} & 1 \\ g_{20} & g_{21} & 1 \\ g_{30} & g_{31} & g_{32} \end{bmatrix} \text{ and } \quad h'(x) = g_1'(x)g_2'(x)g_3'(x)$$

where $g_i'(x) = \delta^{-1}\left(\frac{dg_j}{dx}g_k - g_j\frac{dg_k}{dx}\right)$ for $(i,j,k) = (1,2,3),(2,3,1),(3,1,2)$. Then the codomain of the corresponding isogeny is the Jacobian of the curve $H' : \mathbf{y}^2 = h'(\mathbf{x})$.

Let $X \subset H \times H'$ be a curve defined by

$$X : g_1(x)g_1'(\mathbf{x}) + g_2(x)g_2'(\mathbf{x}) = y\mathbf{y} - g_1(x)g_1'(\mathbf{x})(x - \mathbf{x}) = 0$$

X is naturally equipped with the projection maps $\pi : X \to H, \pi' : X \to H'$. Then the isogeny is given by

$$J_H \to J_{H'} : [D] \mapsto [\pi'_* \pi^* D]$$

To compute the image of the point $[D] = [x^2 + u_1 x + u_0, v_1 x + v_0]$, eliminate the variables x, y from the system

$$\begin{gathered} x^2 + u_1 x + u_0 = 0 \\ y^2 = v_1 x + v_0 \\ y^2 = h(x) \\ g_1(x) g_1'(\mathbf{x}) + g_2(x) g_2'(\mathbf{x}) = 0 \\ y\mathbf{y} - g_1(x) g_1'(\mathbf{x})(x - \mathbf{x}) \end{gathered}$$

By running a Gröbner basis computation, one can find that

$$[\mathbf{x}^4 + u_3' \mathbf{x}^3 + u_2' \mathbf{x}^2 + u_1' \mathbf{x} + u_0', v_3' \mathbf{x}^3 + v_2' \mathbf{x}^2 + v_1' \mathbf{x} + v_0']$$

are the non-reduced Mumford coordinates for the image on $J_{H'}$.

3.3.3 Last step of the chain: Split or not

The deciding step of the strategy is to check whether quotienting out $C \times E$ will result in a product of elliptic curves i.e the codomain of the last step of $(2, 2)$-isogeny is split or not. It can be shown that the determinant δ of the a-th Richelot isogeny vanishes if and only if the codomain is a product of elliptic curves instead of the Jacobian of a genus-2 curve. Hence the deciding step is reduced to verifying whether or not $\delta = 0$.

Chapter 4

Key Recovery Algorithm and Generalizations

The main aim is to come up with an algorithm that challenge SIDH/SIKE (even in a general case) to recover Bob's secret key.

4.1 Algorithm: Basic Version

Having the input for the algorithm from previous chapter, assume $\beta = 0$, so that $E_0 = E_{start}$. This is the case of SIKE. In the general case, replace the maps $\hat{\kappa} : E_1 \to E_0$, $\widehat{\kappa_2\kappa_1} : E_2 \to E_0$, etc below with their compositions with τ.

4.1.1 Iteration

For the first iteration, choose $\beta_1 \geq 1$ minimal such that there exists an $\alpha_1 \geq 0$ for which

$$c_1 = 2^{a-\alpha_1} - 3^{b-\beta_1}$$

is positive and only has prime factors congruent to 1 mod 4. Let $\kappa_1 : E_0 \to E_1$ be a 3^{β_1}-isogeny followed by $\varphi_1 : E_1 \to E$.

To an attacker, there are 3^{β_1} options for κ_1 and for each of these options, run the decision algorithm on

1. the curve $E_1 = \kappa_1(E_0)$.
2. the generators $P_1 = \kappa_1(2^{\alpha_1}P_0)$ and $Q_1 = \kappa_1(2^{\alpha_1}Q_0)$ of $E_1[2^{a-\alpha_1}]$.
3. the 3^{β_1}-isogeny $\hat{\kappa_1} : E_1 \to E_0$.

4. the codomain E; if the guess is correct then E is connected to E_1 via an unknown isogeny φ_1 of degree $3^{b-\beta_1}$.

5. the generators $2^{\alpha_1}P, 2^{\alpha_1}Q$ of $E[2^{a-\alpha_1}]$.

So according to the decision strategy, only the correct option for κ_1 will pass the test, which will then left a secret isogeny of degree $3^{b-\beta_1}$: the starting step of next iteration.

To begin with, one needs to construct the auxiliary isogeny $\gamma_1 : E_1 \to C_1$ and compute the images P_{c_1}, Q_{c_1} of the points P_1, Q_1.

$$\gamma_{start} \circlearrowright E_0 \xrightarrow{\kappa_1} E_1 \xrightarrow{\varphi_1} E, \quad E_1 \overset{\widehat{\kappa_1}}{\dashrightarrow} C_1 \text{ (with } \gamma_1 \text{ dotted)}$$

Let us denote $\hat{\kappa_1}$ as η_1. So γ_1 auxiliary isogeny of degree c_1

$$\gamma_1 = \frac{\widetilde{\eta_1} \circ \gamma_{start} \circ \eta_1}{3^{\beta_1}}$$

where $\widetilde{\eta_1} : E_{start} \to C_1$ is the isogeny with the kernel $\gamma_{start}(\ker\kappa_1)$. This simplifies the computation to

$$P_{c_1} = 2^{\alpha_1}\tilde{\eta_1}\gamma_{start}(P_0) \text{ and } Q_{c_1} = 2^{\alpha_1}\tilde{\eta_1}\gamma_{start}(Q_0)$$

After computing these points, the final step is to check whether the quotient of $C_1 \times E$ by the $(2^{a-\alpha_1}, 2^{a-\alpha_1})$-subgroup

$$\langle (P_{c_1}, 2^{\alpha_1}P), (Q_{c_1}, 2^{\alpha_1}Q) \rangle$$

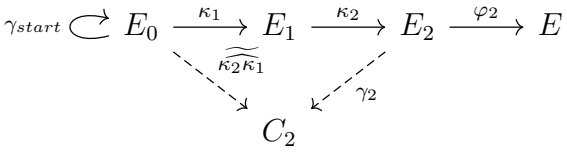
is again a product of elliptic curves or not. This is done by computing the corresponding $(a-\alpha_1)$ chain of $(2,2)$-isogenies. As discussed from the previous chapter, it start with the gluing of $C_1 \times E$ into Jacobian of a genus-2 curve followed by $a-\alpha_1-2$ Richelot isogenies between Jacobian of genus-2 curves and a final easy "$\delta = 0$ test" which confirm whether the last step split or not. If the "$\delta = 0$ test" fails, then try again with a different guess for κ_1.

Remark 4.1. Even in the case of wrong guess for κ_1, the subgroup $\langle (P_{c_1}, 2^{\alpha_1}P), (Q_{c_1}, 2^{\alpha_1}Q) \rangle$ remains maximally isotropic with respect to the Weil pairing. So detecting wrong guesses using Weil pairing is not the way to use. Hence the computation of chain of $(2,2)$-isogeny which includes gluing and its successive Richelot walk is compulsary through out.

If the test passes, i.e when the right option for κ_1 is found, the next iteration starts from E_1. Let $\beta_2 > \beta_1$ be minimal such that there is an α_2 for which

$$c_2 = 2^{a-\alpha_2} - 3^{b-\beta_2}$$

is positive and all its prime factors are congruent to 1 mod 4. Let $\kappa_2 : E_1 \to E_2$ be a $3^{\beta_2-\beta_1}$-component such that the new secret isogeny $\varphi_1 = \varphi_2 \circ \kappa_2$ where $\varphi_2 : E_2 \to E$.



Let us denote $\widehat{\kappa_2\kappa_1}$ as η_2 and $\widetilde{\eta_2} : E_{start} \to C_2$ is the isogeny with kernel $\gamma_{start}(\ker\kappa_2\kappa_1)$ and for the ease of writing.

For each guess for κ_2 one computes

$$P_{c_2} = 2^{\alpha_2}\tilde{\eta_2}\gamma_{start}(P_0) \text{ and } Q_{c_2} = 2^{\alpha_2}\tilde{\eta_2}\gamma_{start}(Q_0)$$

Once these points are computed, checks whether the quotient of $C_2 \times E$ by the subgroup

$$\langle (P_{c_2}, 2^{\alpha_2}P), (Q_{c_2}, 2^{\alpha_2}Q) \rangle$$

is reducible or not.

By continuing in this way, one eventually retrieves all of φ by the composition of

$$E_0 \xrightarrow{\varphi_1} E_1 \xrightarrow{\varphi_2} E_2 \xrightarrow{\varphi_3} ..... \to E_{n-1} \xrightarrow{\varphi_n} E$$

4.1.2 Step sizes

One way of reducing the number of possible guesses in each iteration is by making the gaps between the consecutive integers $0, \beta_1, \beta_2, \beta_3, ..., \beta_r = b$ small as possible. In particular, the expected number of $(2,2)$-chains that need to be computed is about

$$\frac{1}{2}(3^{\beta_1} + 3^{\beta_2-\beta_1} + 3^{\beta_3-\beta_2} + ... + 3^{b-\beta_{r-1}})$$

Except on the last iteration where $\beta_r = b$, a necessary condition for each of the β_i is that $b - \beta_i$ is odd. Because, if $b - \beta_i > 0$ is even then

$$c_i = 2^{a-\alpha_i} - 3^{b-\beta_i} \equiv 3 \bmod 4$$

c_i must admit a prime factor that is congruent to 3 mod 4. Hence the best one can allow the sequence of $\{\beta_i\}$ to grow by steps of two. This makes the expected number of $(2,2)$-chains to be computed becomes about $9b/4$. Except in the cases of small β_i, the optimal estimate of $9b/4$ lies close to reality. Since as β_i grows, the number of candidates α_i's grows as well which leads to the increase in the probability of success as c_i is allowed to get smaller.

4.1.3 In terms of Bob's key

This is an attempt to rephrase the iteration in terms of Bob's secret key denoted as sk_{Bob}. SIDH comes with public generators P_B, Q_B of $E_0[3^b]$ and a secret isogeny φ (or the secret subgroup $\langle P_B + \mathrm{sk}_{Bob} Q_B \rangle$ which forms the $\ker\varphi$) is encoded as the integer

$$\mathrm{sk}_{Bob} \in [0, 3^b)$$

Let $\beta_0 = 0$ and expand

$$\mathrm{sk}_{Bob} = k_1 + k_2 3^{\beta_1} + ... + k_r 3^{\beta_{r-1}} \quad \text{where } k_i \in [0, 3^{\beta_i - \beta_{i-1}} - 1)$$

observe that for $i = 1$,

$$\ker\kappa_1 = \langle 3^{b-\beta_1} P_B + k_1 3^{b-\beta_1} Q_B \rangle$$

So the first iteration step to find the isogeny κ_1 amounts to

1. guessing k_1
2. computing the 3^{β_1}-isogeny $\tilde{\hat{\kappa_1}} : E_{start} \to C_1$ with the kernel as $\gamma_{start}(ker\kappa_1)$ where $\ker\kappa_1$ is as mentioned above.
3. computing the points $P_{c_1}, Q_{c_1} \in C_1$
4. checking whether the subgroup $\langle (P_{c_1}, 2^{\alpha_1} P), (Q_{c_1}, 2^{\alpha_1} Q) \rangle$ is reducible or not.

Once k_1 is found, proceed to determine k_2 by setting the kernel as

$$\ker\kappa_2 = \langle 3^{b-\beta_2} P_B + (k_1 + k_2 3^{\beta_1}) 3^{b-\beta_1} Q_B \rangle$$

and the process goes on. So the iteration finally determines the integer sk_{Bob} digit by digit. In addition, if all the gaps between β_i is two, then this results in determining one base-9 digit of sk_{Bob} at a time.

4.2 Some speed ups

The following are some methods for speeding up the iteration.

Choosing α_i as large as possible: For a given β_i, one usually choose an integer $\alpha_i \geq 0$ such that $c_i = 2^{a-\alpha_i} - 3^{b-\beta_i}$ is positive and has only prime factors which are congruent to 1 mod 4. But this is not a unique integer. Hence choosing a large α_i makes the length of $a - \alpha_i$-chain of (2,2)-isogenies smaller. Therefore it is efficient to choose larger α_i.

Using a precomputed table: A precomputed table which stores for all odd values $s \in \{1, 3, 5, ..., 239\}$, the smallest integer $t(s)$ such that $2^{t(s)} - 3^s$ has only prime factors which are congruent to 1 mod 4. It also stores the corresponding values for u and v. The table is available in uvtable.m. For every values of β_i such that $b - \beta_i$ is odd, check whether the corresponding value of $t(b - \beta_i) \leq a$. After choosing the right option, set $\alpha_i = a - t(b - \beta_i)$. This makes sure that, one can choose α_i as large as possible with u and v readily available without factoring.

Extending Bob's secret isogeny: Imagine a situation when some candidates of β_i does not admit an integer $\alpha_i \geq 0$ such that $2^{a-\alpha_i} - 3^{b-\beta_i}$ has prime factors only congruent to 1 mod 4. For instance, this happens when $b - \beta_i > 0$ is even. Suppose $\beta_i - 1$ does admit α_i such that all necessary conditions are satisfied. Then one can extend Bob's secret isogeny with an arbitrary 3-isogeny φ' such that $P' = \varphi'(P)$ and $Q' = \varphi'(Q)$. Now set $\varphi' \circ \varphi$ as the new secret isogeny , the relevant expression becomes $2^{a-\alpha_i} - 3^{b+1-\beta_i}$ and we already have the value for $\alpha_i \geq 0$ for which c_i is a product of prime factors that are congruent to 1 mod 4. One can use the attack to this extended isogeny for determining Bob's secret key.

This means that most step sizes drop from 2 to 1, by rephrasing it means that we are determining one base-3 digit of sk_{Bob} instead of base-9 in general case. The only possibly larger step occurs at the beginning of the iteration. Example: In case of SIKEp751, the smallest value β_1 such that c_1 satisfies the essential conditions is for $\beta_1 = 6$. This implies a costly start of the algorithm: out of 20.6 hours for breaking SIKEp751, about 14 hours are spent for determining the first 6 ternary digits of sk_{Bob}.

Remark 4.2.

1. If 2^a is smaller than 3^b, then it is more efficient to attack Alice's private key instead of Bob using 3^b torsion points and chain of $(3, 3)$-isogeny.

2. There is possibility that the randomly chosen isogeny φ' matches with the dual of the last degree-3 component of φ. This can lead to create false positives and clueless about the correct guess since the wrong guesses are also at distance $3^{b-\beta_i}$ from E. So if multiple guesses pass the test, all needs to do is change the φ', then one can identified the dual direction too. if this happens, it will not affect the correctness of β_1 since it doesn't depend on φ', but will be discovered when trying to determine the ternary digit at $\beta_2 = \beta_1 + 1$.

4.3 Complexity of the Algorithm

Modulo the factorization of polynomially many natural numbers which only depend on a and b, this attack runs in heuristic polynomial time on a classical computer.

The complexity of the attack is mainly dependent on the computation of the isogeny $\gamma : E_0 \to C$ of degree $c = 2^a - 3^b$.

Landau's classical theorem regarding sums of two squares

Let $b(n)$ be the characteristic function of integers that are representable as a sum of two squares and let

$$B(x) = \sum_{n \leq x} b(n)$$

be the number of such integers up to x. Landau's Theorem gives an asymptotic formula for $B(x)$:

$$B(x) = K\frac{x}{\sqrt{\log x}} + O(\frac{x}{\log^{3/2} x}), \quad x \to \infty$$

where $K \approx 0.764$ is the Landau–Ramanujan constant. More generally, for any integer $n \neq -k^2$, let $B_n(x)$ be the number of positive integers less than or equal to x of the form $u^2 + nv^2$, it was also shown that $B_4(x)$ is asymptotic to

$$\frac{0.5731..}{\sqrt{\log x}}x$$

We can use this to estimate the probability that our strategy succeeds in constructing an isogeny $\gamma : E_0 \to C$ of degree $c = 2^a - 3^b$: it is about $\approx \frac{0.6884}{\sqrt{a}}$.

The first iteration of our key recovery algorithm is where we choose β_1 such that there exists an α_1 for which $c_1 = 2^{a-\alpha_1} - 3^{b-\beta_1}$ is of the form $u^2 + 4v^2$. Observe that the first iteration dominates the overall runtime, because the expression can be recycled in the remaining iterations by extending Bob's secret isogeny. In view of Landau's theorem, it is expected that one should try in the order of $\sqrt{a}$ pairs (α_1, β_1) before we succeed. So the smallest β_1 is expected to be of magnitude $a^{1/4}$. While this is good enough for breaking the concrete parameter sets of SIKE, the asymptotic runtime is $L_p(1/4)$ rather than polynomial.

To achieve a polynomial time complexity, the attack is extended from sums of squares to more general quadratic forms and hope that there is a number $n \leq a$ such that c_1 can be written as $u^2 + nv^2$. Once such a decomposition is found, there is a polynomial time construction of an isogeny from E_{start} to an elliptic curve possessing an endomorphism $\sqrt{n}\mathbf{i}$ satisfying $\sqrt{n}\mathbf{i} \circ \sqrt{n}\mathbf{i} = -[n]$. This endomorphism can be transferred into a desired degree c isogeny γ.

Once γ is constructed, the algorithm proceeds by computing the corresponding $(a-\alpha_1)$ chain of $(2,2)$-isogenies. This mainly includes the Richelot isogenies between Jacobians of genus-2 curves. The formula consists of computing the determinant of a 3×3 matrix in each step. The computational complexity of finding the determinant of a $n \times n$ matrix is the same as matrix multiplication up to a constant. By a fast known algorithm, it is $O(n^3)$. In this case, we are always computing the determinant of a 3×3 matrix which have arithmetic operation complexity $O(1)$. For determinant, the order of magnitude of the bit complexity is different from that of arithmetic, as it depends on the length of the inputs. Given the inputs for the algorithm, it has complexity polynomial in $\log(p)$.

4.4 Generalizations

The proposed algorithm can also admit some generalizations which make one to use it even some of the initial inputs of the algorithm are not available explicitly.

Arbitrary torsion

As it is mentioned in the last remark, there is no theoretical obstruction to attack Alice's key instead of Bob especially in the situation when 2^a is considerably smaller than 3^b which makes none of the c_i to be positive. So in this case, one will be computing the chains of $(3,3)$-isogenies instead of $(2,2)$-isogenies.

This is doable using the machinery available in the literature *Descent via (3,3)-isogeny on Jacobians of genus 2 curves* by N. Bruin, E.V. Flynn, D. Testa.

It gives a parametrization of curves C of genus 2 with a maximal isotropic $(\mathbb{Z}/3\mathbb{Z})^2$ in $J_C[3]$, where J is the Jacobian variety of C, and develop the theory required to perform descent via $(3,3)$-isogeny. Consider curves C of genus 2 over a field k of characteristic not 2 or 3, that have special structure in the 3-torsion of their Jacobians J_C . In particular, we consider the situation where $J_C[3]$ contains a group χ of order 9. Such a curve C can be given by a model of the form

$$y^2 = F(x) = G(x)^2 + \lambda H(x)^3$$

where $G(x)$ is cubic and $H(x)$ is quadratic in x. It can be shown that the Weil pairing of the 3-torsion points can be easily expressed in terms of the corresponding polynomials $G(x)$ and $H(x)$ and this allows to show that the subgroup $\chi \subset J_C[3]$ of size 9 is maximally isotropic with respect to the Weil pairing. Therefore J_C/χ is a principally polarized surface and then descent via $(3,3)$-isogeny. The formula are practical and

one can recover Alice's private key bit per bit.

The attack will proceeds by computing $(3,3)$-isogenies of length at most b and then using a test "$\Delta = 0$" which plays a similar role as δ in the Richelot isogeny formula to verify whether the final $(3,3)$-isogeny splits or not.

In general, one can attack SIDH when set up use arbitrary primes l_A and l_B instead of just 2 and 3. This changes nothing but now one needs to compute chains of (l,l)-isogenies for primes $l \geq 5$. For isogenies between Jacobians of genus-2 curves, the formula is more involved than those to compute $(2,2)$ and $(3,3)$-isogenies, but they are polynomial in l, so practically enough to complete the attack. Away from $l = 2, 3$, a straightforward decision algorithm to verify whether the codomain of the Jacobian of a genus-2 curve result in a product of elliptic curves is not available. So this can be done by either compute an (l,l)-isogeny to a Jacobian and see if the theta constants fails to form a genus-2 curve or by writing down a system of equations expressing that the domain Jacobian is "(l,l)-split" and verify whether the system is consistent or not.

Other base curves with a known path to E_{start}

In practice, SIDH have a base curve $E_0 = E_{start}$ where

$$E_{start} : y^2 = x^3 + x \text{ or } E_{start} : y^2 = x^3 + 6x^2 + x$$

But with the currently known ways of generating supersingular elliptic curves, every publicly generated alternative to E_0 comes with a known path to E_{start}. The KLPT algorithm can convert this path into a required isogeny for the intial input $\tau : E_0 \to E_{start}$ of degree 3^β for some $\beta \geq 0$. Now one is in the position to start the attack by using glue and split method by constructing the auxiliary isogeny γ explicitly.

4.4.1 Base curves without a known path to E_{start}

In the case of unavailability of information about $\mathrm{End}(E_0)$, then letting γ to emanate from E rather than E_0 leading to consider $\gamma \circ \varphi : E_0 \to C$, an isogeny of degree $c3^b$ makes sense and it can be used to apply the decision criteria by checking whether or not the subgroup

$$\langle (P_0, x\gamma(P), (Q_0, x\gamma(Q)) \rangle \subset E_0 \times C$$

is reducible, with x be a multiplicative inverse of 3^b modula 2^a.

Even in the absence of a known path to E_{start}, there are situations when one can possibly construct the auxiliary isogeny γ. For instance, if $c = 2^a - 3^b$ is smooth.

Construction of γ: Let $c = l_1 l_2 ... l_s$ be the prime factorization of c and for each $i = 1, 2, ..., s$, let r_i denote the multiplicative order of $-\mathfrak{p}$ modulo l_i. Observe that since $\#E(\mathbb{F}_{p^2}) = (p+1)^2$ and End(E) is an order, the $\mathfrak{p}^2-$ Frobenius map acts as $[-\mathfrak{p}]$. Hence one can find a non-trivial point in $E_0[l_1] \subset E_0(\mathbb{F}_{p^{2r_1}})$ and the subgroup it generates is defined over $\mathbb{F}_{p^2}$. Using this subgroup as the kernel of an $\mathbb{F}_{p^2}$-rational isogeny $\gamma_1 : E_0 \to C_1$ of degree l_1 which can be computed and evaluated using formula of Velu type. By continuing this process, we eventually obtain $\gamma : E_0 \to C$ as a composition of $\gamma_s \circ \gamma_{s-1} \circ .... \circ \gamma_1$ where each γ_i is an $\mathbb{F}_{p^2}$ rational l_i-isogeny.

Using this γ in the decision criteria and key recovery algorithm is in the same way as in a general case. To begin with, choose a smallest integer $\beta \geq 1$ for which there exists an integer $\alpha \geq 0$ such that

$$c = 2^{a-\alpha} - 3^{b-\beta}$$

is smooth. Then for the each guess for κ_1-the first degree 3^β component of the secret isogeny φ, run the first iteration step to check whether or not there exist a degree $3^{b-\beta}$- isogeny $\varphi_1 : \kappa_1(E_0) \to E$ which maps $2^\alpha \kappa_1(P_0)$ to $2^\alpha P$ and $2^\alpha \kappa_1(Q_0)$ to $2^\alpha Q$ to find the right option.

Once κ_1 is found, proceed to find the next component of φ by steps of degree 3. Since smoothness is a rare event, it make sense to recycle the expression $c = 2^{a-\alpha} - 3^{b-\beta}$ all along. Hence the auxiliary isogeny γ is also recycled again i.e it has to be computed once, including pushing through points. It works by extending γ by an extra degree 3-isogeny $\varphi' : C \to E'$ and to find the option for κ_2, the decision is made by checking whether or not there is a degree $c3^{b-\beta}$-isogeny mapping $2^\alpha \kappa_2 \kappa_1(P_0)$ to $2^\alpha \varphi' \gamma(P)$ and $2^\alpha \kappa_2 \kappa_1(Q_0)$ to $2^\alpha \varphi' \gamma(Q)$. By continuing this iteration will give the entire isogeny chain.

Remark 4.3. **An attack on SIDH with arbitrary starting curve**:
Given the same source of inspiration, this way of generalizing the attack when there is no known path to E_{start} is quite similar to the attack proposed by Luciano Maino and Chole Martindale [2].

To recover the secret isogeny with an input of; let $A = l_A^a$ and $B = l_B^b$ be two coprime integers, two supersingular elliptic curves E_0 and E_A over $\mathbb{F}_{p^2}$ connected by an unknown degree-A- isogeny $\varphi_A : E_0 \to E_A$, a basis $\{P_B, Q_B\}$ of $E_0[B]$, a basis $\{P_A, Q_A\}$ of $E_0[A]$, the image points $\varphi_A(P_B), \varphi_A(Q_B)$.

The core idea behind their attack is to construct an elliptic curve E and an isogeny $\varphi_f : E \to E_0$ with $f = B - A$ be smooth and a polarized isogeny Φ originating from

the abelian surface $E \times E_A$ such that one of its components reveals the dual of the secret isogeny φ_A. A brief version of the algorithm can be given as:

1. Compute integers e, j, f, i such that e is small and smooth, $0 \leq j \leq b, f$ is smooth and positive, i is small, $(Al_A^{-i})^{-1} = c \bmod eBl_B^{-j}$ and $eBl_B^{-j} = f + Al_A^{-i}$. Set $A' = Al_A^{-i}$ and $B' = eBl_B^{-j}$ for ease of notation.

2. Compute a curve f-isogenous to E_0 by the isogeny φ_f.

3. Compute a basis $\{P_{eB'}, Q_{eB'}\}$ of $E[eB']$ such that $[e]P_{eB'} = [l_B^j]\hat{\varphi_f}(P_B)$ and $[e]Q_{eB'} = [l_B^j]\hat{\varphi_f}(Q_B)$.

4. Choose a guess $\varphi_{l_A^i}$ for the last i steps of φ_A and let $\varphi' : E_0 \to E'$ be the corresponding first $a - i$ steps of φ_A.

$$\begin{array}{ccccc} E_0 & \xrightarrow{\varphi'} & E' & \xrightarrow{\varphi_{l_A^i}} & E_A \\ {\scriptstyle\varphi_f}\uparrow & \nearrow_{\varphi} & & & \\ E & & & & \end{array}$$

5. Choose $R, S \in E'[eB']$ such that

$$[e]R = [l_A^{-i} f l_B^j]\hat{\varphi}_{l_A^i} \circ \varphi_A(P_B)$$
$$[e]S = [l_A^{-i} f l_B^j]\hat{\varphi}_{l_A^i} \circ \varphi_A(Q_B)$$

   R,S are a guess for the images of $\varphi(P_{eB'}), \varphi(Q_{eB'})$ respectively.

6. Compute a (eB', eB')-isogeny Φ_{guess} with domain $E \times E'$ and kernel;

$$\ker \Phi_{guess} = \text{Graph}(c\varphi|_{E[eB']}) = \langle (P_{eB'}, cR), (Q_{eB'}, cS) \rangle.$$

   Check if the codomain splits, if not take a new guess for $(\varphi_{l_A^i}, R, S)$.

7. Choose a basis $\{P, Q\}$ of $E'[A']$ and compute the images of $\hat{\varphi}'$ by evaluating Φ on $E'[A']$.

8. Compute $\ker(\varphi') = \langle \hat{\varphi}'(P), \hat{\varphi}'(Q) \rangle$ and return $\varphi_{l_A^i} \circ \varphi'$.

The existence of Φ with a kernel given by the graph of $c\varphi|_{E[B]}$ which is maximally isotropic with respect to the Weil pairing of $E \times E_A$ is proven by employing Kani's reducibility criteria. The complexity of this attack is mainly determined by the cost of computation of the cofactor isogeny φ_f and how smooth we require f to be.

So in summary the whole idea is as soon as one is able to find a small integer $\beta \geq 1$ such that there exist an α together with $c = 2^{a-\alpha} - 3^{b-\beta}$ is smooth, then the attack applies. Since finding c to be smooth is an optimistic goal, at least this might lower the security level of certain parameter sets.

There are some ways to make up the leeway of our attack

1. By extending the secret isogeny $\varphi : E_0 \to E$ by an arbitrary isogeny $\epsilon : E \to F$ of some smooth degree e and then working with $\epsilon \circ \varphi : E_0 \to F$ allows to easily find a smooth integer of the form $c = 2^{a-\alpha} - e3^{b-\beta}$ to construct the auxiliary isogeny $\gamma : F \to C$.

2. Let H be a genus-2 curve over $\mathbb{F}_{p^2}$ with a superspecial Jacobian J and let d be an integer. By any algorithm if one can efficiently solve to find if there exist a (d, d)-isogeny $\psi : J \to A$ such that A is a product of elliptic curves for a fixed d, this will also create more leeway.
   Indeed, this allows us to work with expression of the form $c = d2^{a-\alpha} - e3^{b-\beta}$. Each step consists of computing a $(2^{a-\alpha}, 2^{a-\alpha})$-isogeny and then checking the final Jacobian is (d, d)-split which likely to be more efficient by using the above algorithm.

Bibliography

[1] Castryck, Wouter, and Thomas Decru. *An efficient key recovery attack on SIDH (preliminary version).* Cryptology ePrint Archive (2022): Paper-2022.

[2] Maino, Luciano, and Chloe Martindale. *An attack on SIDH with arbitrary starting curve.* Cryptology ePrint Archive (2022).

[3] Edixhoven, Bas, Gerard Van der Geer, and Ben Moonen. *Abelian varieties.* Preprint (2012): 331.

[4] Silverman, Joseph H. *The arithmetic of elliptic curves.* Vol. 106. New York: Springer, 2009.

[5] Kani, Ernst. *The number of curves of genus two with elliptic differentials.* (1997): 93-122.

[6] De Feo, Luca, David Jao, and Jérôme Plût. *Towards quantum-resistant cryptosystems from supersingular elliptic curve isogenies.* Journal of Mathematical Cryptology 8, no. 3 (2014): 209-247.

[7] Bernstein, Daniel J. *Introduction to post-quantum cryptography.* In Post-quantum cryptography, pp. 1-14. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009.

[8] Howe, Everett W., Franck Leprévost, and Bjorn Poonen. *Large torsion subgroups of split Jacobians of curves of genus two or three.* (2000): 315-364.

[9] Smith, Benjamin Andrew. *Explicit endomorphisms and correspondences.* (2005).

[10] Galbraith, Steven D., Christophe Petit, and Javier Silva. *Identification protocols and signature schemes based on supersingular isogeny problems.* In Advances in Cryptology–ASIACRYPT 2017: 23rd International Conference on the Theory and Applications of Cryptology and Information Security, 2017.