Supersingular isogeny Diffie-Hellman

Gaurish Korpal

January 21, 2022

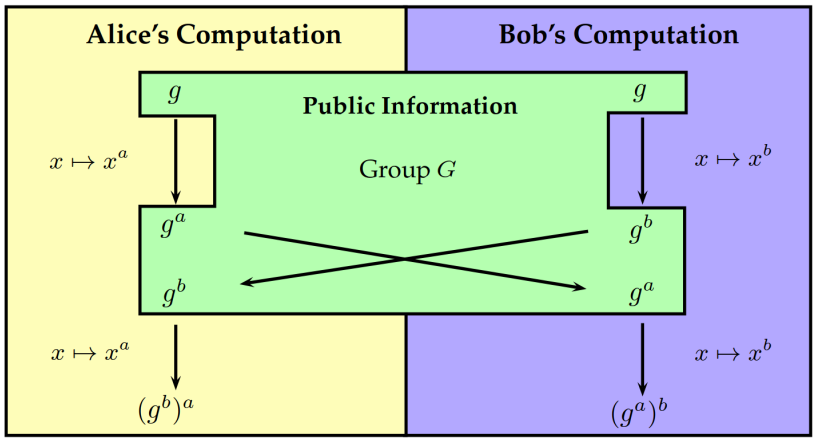
Abstract

In this seminar we will discuss the mathematics involved in the working of a post-quantum cryptographic protocol based on isogenies between supersingular elliptic curves. Some familiarity with the arithmetic properties of elliptic curves is assumed (for example, see my RTG presentation slides from Fall 2020: `https://bit.ly/3qtYzgT`).

1 Introduction

The security of the public-key cryptography techniques like RSA (traditional online payments) and ECC (blockchain and cryptocurrency) depend on the computational hardness of prime factorization and discrete-logarithm problem, respectively. However, with the advancement in quantum computer development, these problems can be solved in polynomial-time using Shor's algorithm[1]. Therefore, there is a need to develop quantum-safe cryptography techniques which can be deployed using the current digital computers[2] before the quantum computers arrive[3]. A lot of new protocols have been proposed as part of NIST's Post-Quantum Cryptography Standardization[4], and Supersingular Isogeny Key Encapsulation[5] (SIKE) is one of them. The isogeny-based cryptography is a kind of elliptic-curve cryptography, whose security relies on (various incarnations of) the problem of finding an explicit isogeny between two given isogenous supersingular elliptic curves over a finite field $\mathbb{F}_q$. Currently, quantum computers do not seem to make the isogeny-finding problem substantially easier. In this seminar we will look at the mathematics involved in the implementation of SIKE.

2 Diffie-Hellman key exchange



(a) A schematic version of the Diffie-Hellman protocol, emphasizing the public information (in green) and the private information of Alice and Bob (yellow and blue respectively), for $G = \langle g \rangle$. [Urbanik]

$$\begin{array}{ccc} g & \xrightarrow{x \mapsto x^a} & g^a \\ \downarrow{\scriptstyle x \mapsto x^b} & & \downarrow{\scriptstyle x \mapsto x^b} \\ g^b & \xrightarrow{x \mapsto x^a} & g^{ab} \end{array}$$

(b) Diffie-Hellman is actually a kind of commutative diagram.

[1] P. W. Shor, "Algorithms for quantum computation: discrete logarithms and factoring," Proceedings 35th Annual Symposium on Foundations of Computer Science, 1994, pp. 124-134, `https://doi.org/10.1109/SFCS.1994.365700`

[2] Hence we will focus on finding a computationally secure alternative rather than investing resources on the unconditionally secure quantum cryptography.

[3] Patrick Howell O'Neill. The US is worried that hackers are stealing data today so quantum computers can crack it in a decade. MIT Technology Review, November 2021. `https://www.technologyreview.com/2021/11/03/1039171/hackers-quantum-computers-us-homeland-security-cryptography/`

[4] `https://csrc.nist.gov/projects/post-quantum-cryptography`

[5] `https://sike.org/`

The protocol's security depends on the model chosen for the group G. For instance, if $G = \mathbb{Z}/n\mathbb{Z}$ with $g = 1$ then Alice and Bob will be sending a and b over the open channel and the protocol is trivially broken. Choosing a different generator g will make no difference since we can invert $g \mod n$ to get $a = g^{-1}(ag) \pmod n$ from the public information ag. Examples of better groups are $G = (\mathbb{Z}/p\mathbb{Z})^\times$ and $G = E(\mathbb{F}_q)$ where E is an elliptic curve over a finite field $\mathbb{F}_q$. The best known attacks against Diffie-Hellman (discrete logarithm) over these groups are non-trivial, and achieve only sub-exponential (index-calculus) and exponential (Pollard rho) time complexity, for $G = (\mathbb{Z}/p\mathbb{Z})^\times$ and $G = E(\mathbb{F}_q)$ respectively. Diffie-Hellman is broken by Quantum Computers on general grounds, irrespective of the particular implementation chosen.

3 Supersingular elliptic curves

Let E_1 and E_2 be elliptic curves. An *isogeny* from E_1 to E_2 is a morphism $\phi : E_1 \to E_2$ satisfying $\phi(O_{E_1}) = O_{E_2}$. In fact, every isogeny is a group homomorphism [AEC, Theorem III.4.8]. The set of isogenies $\mathrm{Hom}(E_1, E_2)$ from E_1 to E_2 form a group under addition where the sum of two isogenies is defined by $(\phi + \psi)(P) = \phi(P) \oplus \psi(P)$. The group $\mathrm{Hom}(E_1, E_2)$ is a torsion-free $\mathbb{Z}$-module [AEC, Proposition III.4.2(b)]. Let E be an elliptic curve and Φ be a finite subgroup of E. Then there is a unique elliptic curve E' and a separable isogeny $\phi : E \to E'$ satisfying $\ker(\phi) = \Phi$ [AEC, Proposition III.4.12]. The elliptic curve E' is often denoted by the quotient E/Φ. Given an elliptic curve E and subgroup Φ, Vélu's formulae give a recepie to explicitly write equations for the curve $E' = E/\Phi$ and isogeny $\phi : E \to E'$ with $\deg(\phi) = \#\Phi$ [MPKC, Theorem 25.1.6].

If $E_1 = E_2 = E$, then $\mathrm{Hom}(E_1, E_2) = \mathrm{End}(E)$ is a ring whose multiplication is given by composition defined as $(\phi\psi)(P) = \phi(\psi(P))$. $\mathrm{End}(E)$ is a (not necessarily commutative) ring of characteristic zero with no zero divisors [AEC, Proposition III.4.2(c)]. For each $m \in \mathbb{Z}$ we define the multiplication-by-m isogeny as

$$[m] : E \to E$$
$$P \mapsto \begin{cases} \underbrace{P \oplus P \oplus \cdots \oplus P}_{m \text{ times}} & \text{if } m > 0 \\ O & \text{if } m = 0 \\ [-m](-P) & \text{if } m < 0 \end{cases}$$

Let E/K be an elliptic curve and $m \in \mathbb{Z}$ with $m \neq 0$. Then $[m] : E \to E$ is nonconstant (surjective) on $E(\bar{K})$ [AEC, Proposition III.4.2(a)].

Let $\phi : E_1 \to E_2$ be a nonconstant isogeny such that[6] $\deg(\phi) = m$. Then there exists a unique isogeny $\hat{\phi} : E_2 \to E_1$ satisfying $\hat{\phi} \circ \phi = [m]$. The $\hat{\phi}$ obtained above is called the dual isogeny to ϕ. This assumes that $\phi \neq [0]$. If $\phi = [0]$ then we set $\hat{\phi} = [0]$. The degree map $\deg : \mathrm{Hom}(E_1, E_2) \to \mathbb{Z}$ is a positive definite quadratic form [AEC, Corollary III.6.3].

Let E/K be an elliptic curve with $m \in \mathbb{Z}_{\geq 1}$. Then the m-torsion subgroup of E, denoted by $E[m]$, is the set of points of E of order m, i.e.

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

Then we have [AEC, Corollary III.6.4]

1. If $m \neq 0$ in K, i.e. if either $\mathrm{char}(K) = 0$ or $\mathrm{char}(K) \nmid m$, then
$$E[m] \cong \mathbb{Z}/m\mathbb{Z} \times Z/m\mathbb{Z}$$
Thus $E[m]$ is a free $\mathbb{Z}/m\mathbb{Z}$-module of rank two.
2. If $\mathrm{char}(K) = p > 0$, then one of the following is true:
   (a) $E[p^d] = \{O\}$ for all $d = 1, 2, 3, \ldots$
   (b) $E[p^d] \cong \mathbb{Z}/p^d\mathbb{Z}$ for all $d = 1, 2, 3, \ldots$

Let K be a field of characteristic p, and E/K be an elliptic curve. Then E is called *supersingular* if the map $[p] : E \to E$ is purely inseparable and $j(E) \in \mathbb{F}_{p^2}$. Therefore, there are only finitely many supersingular elliptic curves for a given p. In fact, for finite field $\mathbb{F}_q$ of characteristic p [AEC, Theorem V.4.1]

[6] degree of a rational map $\phi : C_1 \to C_2$ defined over K. If ϕ is constant then $\deg(\phi) = 0$, otherwise we have $\deg(\phi) = [K(C_1) : \phi^* K(C_2)] < \infty$ with $\phi^* : K(C_2) \to K(C_1)$ defined as $\phi^* f = f \circ \phi$.

1. If $p = 2$, then $E : y^2 + y = x^3$ is the only supersingular elliptic curve over $\bar{\mathbb{F}}_2$.
2. If $p = 3$, then $E : y^2 = x(x-1)(x+1)$ is the only supersingular elliptic curve over $\bar{\mathbb{F}}_3$.
3. If $p \geq 5$, then

$$\#\text{ supersingular elliptic curves over } \bar{\mathbb{F}}_q = \left\lfloor \frac{p}{12} \right\rfloor + \begin{cases} 0 & \text{if } p \equiv 1 \pmod{12} \\ 1 & \text{if } p \equiv \pm 5 \pmod{12} \\ 2 & \text{if } p \equiv 11 \pmod{12} \end{cases}$$

Let p be a prime, and let E be a supersingular curve defined over a finite field $\mathbb{F}_q$ with $q = p^n$ elements. Let $t \in \mathbb{Z}$ be the trace of the Frobenius endomorphism of $E/\mathbb{F}_q$, i.e. $\#E(\mathbb{F}_q) = q - t + 1$ such that $|t| \leq 2\sqrt{q}$ [AEC, Theorem V.1.1, V.2.3.1]. The group structure of $E(\mathbb{F}_q)$ is one of the following [MPKC, Theorem 9.10.13]:

1. if $t = \pm 2\sqrt{q}$, then $E(\mathbb{F}_q) \cong \mathbb{Z}/(\sqrt{q} \mp 1)\mathbb{Z} \times \mathbb{Z}/(\sqrt{q} \mp 1)\mathbb{Z}$
2. if $t = \pm\sqrt{q}$ or $t = \pm\sqrt{pq}$ (for $p = 2, 3$), then $E(\mathbb{F}_q)$ is cyclic.
3. if $t = 0$, then $E(\mathbb{F}_q)$ is either cyclic or isomorphic to $\mathbb{Z}/((q+1)/2)\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$

Recall that two elliptic curves are isomorphic over $\bar{K}$ iff they both have the same j-invariant [AEC, Proposition III.1.4(b)]. Moreover, if $E_1/\mathbb{F}_q$ and $E_2/\mathbb{F}_q$ are two elliptic curves over a finite field, then E_1 and E_2 are isogenous over $\mathbb{F}_q$ iff $\#E_1(\mathbb{F}_q) = \#E_2(\mathbb{F}_q)$ [AEC, Exercise V.5.4] [MPKC, Theorem 9.7.4]. However, if E_1 and E_2 are ordinary (not supersingular) isogenous elliptic curves over $\mathbb{F}_q$ with $j(E_1) = j(E_2)$ then they are isomorphic over $\mathbb{F}_q$ [MPKC, Lemma 9.11.13].

Equivalently, E is supersingular iff $\mathrm{End}(E)$ is an order[7] in a quaternion algebra [AEC, Theorem V.3.1(a)]. This definition is useful for understanding the implementation of an other isogeny-based protocol called CSIDH. In general, computing isogenies between elliptic curves is very closely related to computations in the endomorphism ring of those curves [QA, Corollary 42.3.7]. Hence the larger and more complicated the endomorphism rings of the curves E and E/Φ are, the more difficult it will be for a potential attacker to discover the isogeny $\phi : E \to E/\Phi$. Now, since supersingular elliptic curves have a particularly large (and non-commutative) endomorphism ring and their quotient is also a supersingular curve, we choose to use supersingular curves for isogeny-based cryptography protocols.

4 Walking through the protocol

Now let's briefly discuss the toy example by Craig Costello [Costello]. Let $p = 431 = 2^4 3^3 - 1$. Thus $p \equiv 3 \pmod 4$ and we can choose $\mathbb{F}_{p^2} = \mathbb{F}_p(i)$ with $i^2 + 1 = 0$. In this case there are $\lfloor p/12 \rfloor + 2 = 37$ supersingular j-invariants in $\mathbb{F}_p^2$.

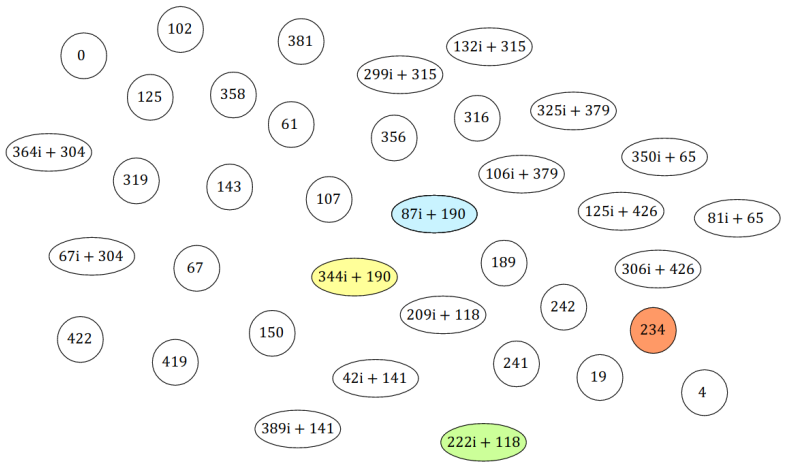


Figure 2: The set of 37 supersingular j-invariants in $\mathbb{F}_{431^2}$ [Costello]

[7] Given a ring A which is a finite-dimensional algebra over the field $\mathbb{Q}$, an order $\mathcal{O}$ of A is a subring of A that (1) spans A over $\mathbb{Q}$ and is a $\mathbb{Z}$-lattice in A.

Note that for any elliptic curve E over $\mathbb{F}_{p^2}$ with $\#E(\mathbb{F}_{p^2}) = (p+1)^2$ we will get the trace of Frobenius $t = -2 \cdot p = -2\sqrt{q}$. Therefore, if E is also supersingular then we will have $E(\mathbb{F}_{p^2}) \cong \mathbb{Z}/(p+1)\mathbb{Z} \times \mathbb{Z}/(p+1)\mathbb{Z} = E[p+1] = E[2^4 3^3]$. Such a supersingular elliptic curve is isomorphic to a curve in Montgomery or twisted Edward form [JFP, §4.3]. In this example, we will work with the ones which can we written in Montgomery form

$$E : y^2 = x^3 + ax^2 + x$$

where $a \in \mathbb{F}_{p^2}$. The main reason is that they facilitate very efficient x-only arithmetic, i.e. maps that ignore the y-coordinates entirely.

Let's begin with the public starting curve

$$E : y^2 = x^3 + (329i + 423)x^2 + x, \qquad j(E) = 87i + 190$$

Next, we choose any four public basis points:

$$E[2^4] = \langle P_A, Q_A \rangle \quad \text{with} \quad \begin{cases} P_A = (100i + 248, 304i + 199) \\ Q_A = (426i + 394, 51i + 79) \end{cases}$$

$$E[3^3] = \langle P_B, Q_B \rangle \quad \text{with} \quad \begin{cases} P_B = (358i + 275, 410i + 104) \\ Q_B = (20i + 185, 281i + 239) \end{cases}$$

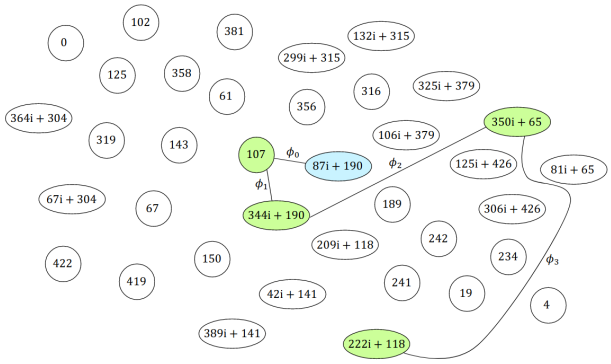
Alice chooses the secret $k_A = 11$ from $\mathbb{Z}/16\mathbb{Z}$. From this the secret generator of the kernel is computed

$$R_A = P_A + [k_A]Q_A = (271i + 79, 153i + 430)$$

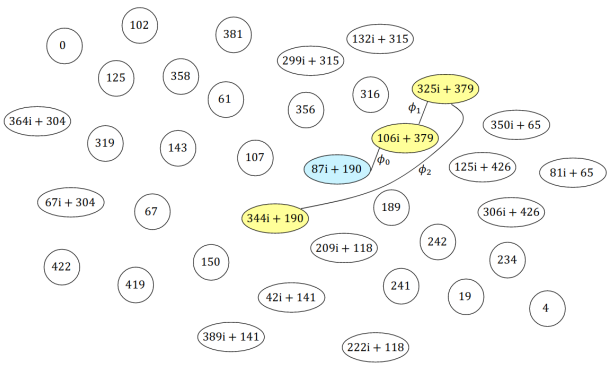
Similarly, Bob chooses the secret $k_B = 2$ from $\mathbb{Z}/27\mathbb{Z}$. From this the secret generator of the kernel is computed

$$R_B = P_B + [k_B]Q_B = (122i + 309, 291i + 364)$$

Note that R_A is a point of order 16 on E. Alice's secret 2^4-isogeny is the composition of the four 2-isogenies obtained using a combination of the point doubling operation and Vélu's formula for 2-isogeny. Similarly, R_B is a point of order 27 on E. Therefore, Bob's secret 3^3-isogeny is the compostion of three 3-isogenies a combination of the point tripling operation and Vélu's formula for 3-isogeny.

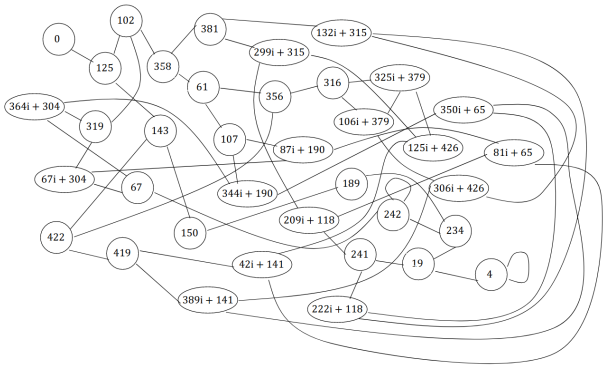


(a) Alice starts at the public curve corresponding to $j = 87i + 190$, her secret key is the 2^4-isogeny $\phi_A = (\phi_3 \circ \phi_2 \circ \phi_1 \circ \phi_0)$, and the destination node $222i + 118$ becomes part of her public key. [Costello]

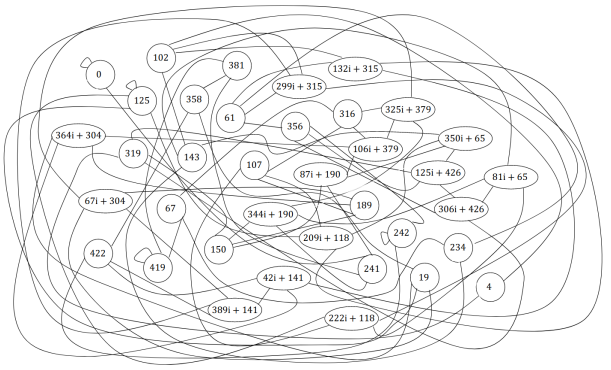


(b) Bob starts at the public curve corresponding to $j = 87i + 190$, his secret key is the 3^3-isogeny $\phi_B = (\phi_2 \circ \phi_1 \circ \phi_0)$, and the destination node $344i + 190$ becomes part of his public key. [Costello]

Recall that isogenous elliptic curves over $\mathbb{F}_q$ have the same number of points over $\mathbb{F}_q$. Therefore, since $E(\mathbb{F}_{p^2}) = E[2^4 3^3]$ we have 2-isogney and 3-isogeny graphs for Alice and Bob respectively. For all $j \notin \{0, 4, 242\}$, there are exactly 3 edges connecting a given node to other 2-isogenous j-invariants. Similarly, for all $j \notin \{0, 4, 125, 242\}$, there are exactly 4 edges connecting a given node to other 3-isogenous j-invariants. Moreover, there arem't any directions on the arrows since for any edge from $j(E)$ to $j(E')$ corresponding to an isogeny $\phi : E \to E'$, the dual isogeny gives an edge from $j(E')$ back to $j(E)$.

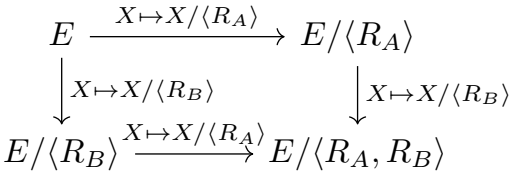


(a) Alice is walking on this 2-isogeny graph. [Costello]

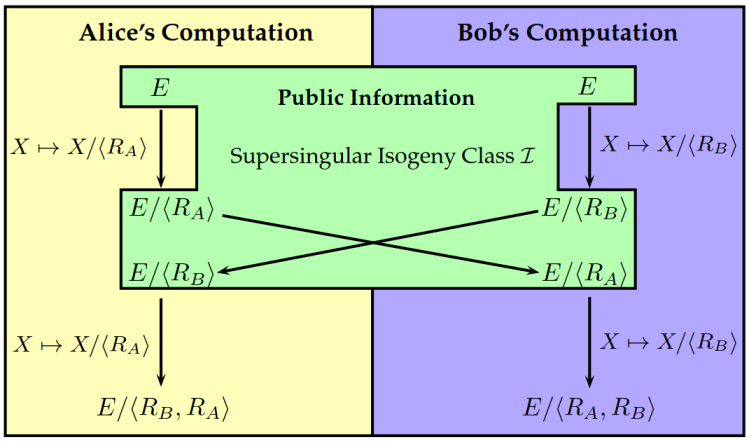


(b) Bob is walking on this 3-isogeny graph. [Costello]

Now if we try to mimic Diffie-Hellman key exchange using these isogeny-graphs, we will get the following scheme:



(a) Diffie-Hellman inspired intuitive commutative diagram using Vélu's formulae.



(b) A "wishful thinking" analogue of the Diffie-Hellman protocol schematic diagram we saw earlier. [Urbanik]

Now, for Alice to be able to compute the map $E/\langle R_B \rangle \mapsto (E/\langle R_B \rangle)/\langle R_A \rangle = E/\langle R_A, R_B \rangle$, i.e. $E/\langle R_B \rangle \mapsto (E/\langle R_B \rangle)/\langle \phi_B(R_A) \rangle$, she really needs to know the point $\phi_B(R_A)$. Similarly, Bob needs to know the point $\phi_A(R_B)$. Bob assists her by computing $\phi_B(P_A)$ and $\phi_B(Q_A)$ and sending the result over the public channel. Using this information, Alice can compute $\phi_B(R_A) = \phi_B(P_A) + [k_A]\phi_B(Q_A)$. The case for Bobs computation is analogous.

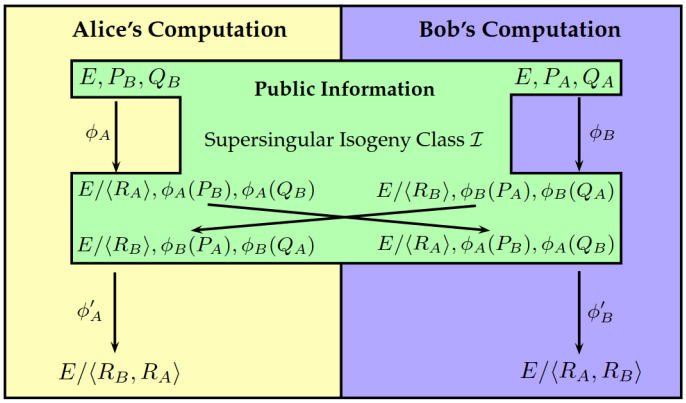


Figure 6: The SIDH protocol in practice. [Urbanik]

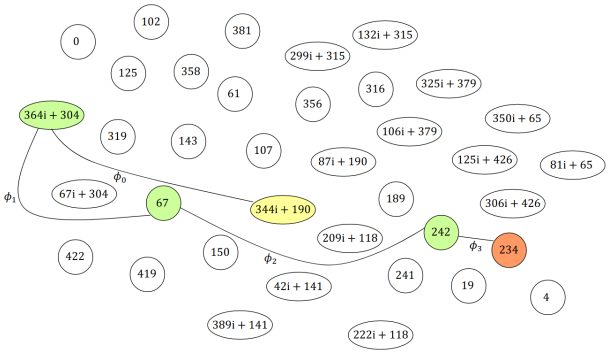
Therefore, we have the public key of Alice and Bob as follows:

$$\begin{aligned} PK_A &= (E/\langle R_A\rangle, \phi_A(P_B), \phi_A(Q_B)) \\ &= (423i + 179, (142i + 183, 119i + 360), (220i + 314, 289i + 10)) \end{aligned}$$

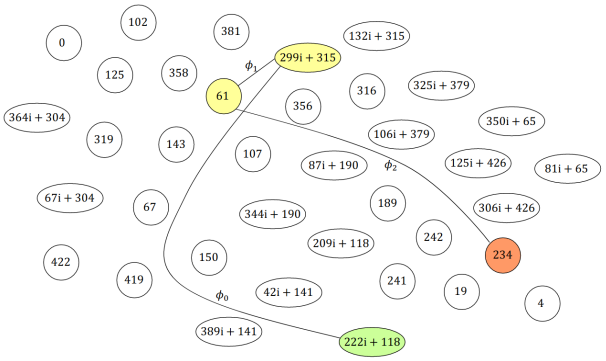
$$\begin{aligned} PK_B &= (E/\langle R_B\rangle, \phi_B(P_A), \phi_B(Q_A)) \\ &= (273i + 76, (187i + 226, 43i + 360), (325i + 415, 322i + 254)) \end{aligned}$$

where the value corresponding to the quotient is the coefficient defining the elliptic curve in Montgomery form.

Now Alice computes the shared secret by performing the analogous sequence of operations as during key generation, this time starting with $E/\langle R_B\rangle$ and $\phi_B(R_A)$. The only difference is that she no longer needs to move any basis points through the isogeny, saving some computation because it is only the destination curve that she needs. Similarly, Bob proceeds exactly as he did during key generation, with the exception of moving Alice's basis points through the isogeny. Alice and Bob end up with isomorphic curves $(E/\langle R_A\rangle)/\langle\phi_A(R_B)\rangle \cong (E/\langle R_B\rangle)/\langle\phi_B(R_A)\rangle$, they may take their shared secret s to be their j-invariant [JFP, §3.2].



(a) Alice starts at the public curve corresponding to $j = 344i + 190$. The shared secret is destination node 234 of the 2^4-isogeny $\phi'_A = (\phi'_3 \circ \phi'_2 \circ \phi'_1 \circ \phi'_0)$. [Costello]



(b) Bob starts at the public curve corresponding to $j = 222i + 118$. The shared secret is destination node 234 of the 3^3-isogeny $\phi'_B = (\phi'_2 \circ \phi'_1 \circ \phi'_0)$. [Costello]

References

[AEC] Joseph H. Silverman. "The Arithmetic of Elliptic Curves" (2nd edition). Springer-Verlag, New York, 2009. https://doi.org/10.1007/978-0-387-09494-6

[JFP] David Jao, Luca De Feo, and Jérôme Plût. Towards quantum-resistant cryptosystems from supersingular elliptic curve isogenies. An extended version of the paper by D. Jao and L. De Feo published in PQCrypto 2011. https://ia.cr/2011/506

[Urbanik] David Urbanik. A Frindly Introduction to Supersingular Isogeny Diffie-Hellman. Lecture notes for Undergraduate Research Seminar, University of Waterloo, July 2016. https://www.math.toronto.edu/dburbani/work/friendlysidh.pdf

[MPKC] Steven Galbraith. "Mathematics of Public Key Cryptography" (Extended Version 2.0), October 2018. https://www.math.auckland.ac.nz/~sgal018/crypto-book/crypto-book.html

[Costello] Craig Costello. Supersingular isogeny key exchange for beginners. Lecture notes for Microsoft Research Webinar, May 2020. https://ia.cr/2019/1321

[QA] John Voight. "Quaternion Algebras" (1st edition). Springer, Cham, 2021. https://doi.org/10.1007/978-3-030-56694-4