How to not break SIDH

Chloe Martindale and Lorenz Panny

chloemartindale@gmail.com, lorenz@yx7.cc

Department of Mathematics and Computer Science,
Technische Universiteit Eindhoven, The Netherlands

Abstract. We give a number of approaches which, to a newcomer, may seem like natural ways to attack the SIDH/SIKE protocol, and explain why each of these approaches seems to fail, at least with the specific setup and parameters of SIKE. Our aim is to save some time for others who are looking to assess the security of SIDH/SIKE. We include methods that fail to attack the pure isogeny problem, namely: looking at the $\mathbb{F}_p$-subgraph, lifting to characteristic zero, and using Weil restrictions. We also include methods that fail to make use of the public 2-power and 3-power torsion points, namely: interpolation techniques, any purely group-theoretic approaches, and constructing an endomorphism à la Petit to exploit the auxiliary points, but with balanced parameters.

Keywords: Isogenies, SIDH, cryptanalysis, negative results.

1 Introduction

Isogeny-based cryptography is a relatively new approach to post-quantum key exchange and other, more advanced (pre- and post-quantum) cryptographic constructions. Isogeny-based key exchange initially attracted attention due to the relatively tiny key sizes, while at the same time offering decent performance. In addition, it has by now crystallized that isogenies may actually be suitable for functionality that no other known post-quantum construction offers (efficiently), such as non-interactive key exchange.

An *isogeny* is a certain kind of map between two elliptic curves (or more generally, abelian varieties) that preserves these objects' structural properties: They are *group homomorphisms* which are given by *rational maps*; more precisely, an isogeny is a surjective morphism of elliptic curves (or abelian varieties) that preserves the identity.

The historically first practical isogeny-based key exchange is this paper's topic of interest: *Supersingular Isogeny Diffie–Hellman* (SIDH), conceived by Jao

[*] Author list in alphabetical order; see https://www.ams.org/profession/leaders/culture/CultureStatement04.pdf. This work was supported in part by the Commission of the European Communities through the Horizon 2020 program under project number 643161 (ECRYPT-NET) and CHIST-ERA USEIT under the Netherlands Organisation for Scientific Research project number 651.002.004. Date of this document: 2019.05.24.

and De Feo in 2011 [18], is first and foremost an ephemeral Diffie–Hellman-like key exchange. Unfortunately, it seems impossible to efficiently determine whether a public key was generated honestly; this leads to an active reaction attack which recovers a static private key in a linear (in the key size) number of queries [16]. Based on this observation, SIDH was later transformed into *SIKE* [17], a key-encapsulation mechanism (KEM) which is currently a second-round contestant in NIST's call for post-quantum cryptographic constructions [27]. In SIKE, one party (the server) can use a static key, while the other party generates a new ephemeral key pair for every connection. The construction is generally the same as SIDH, except that as part of his side of the key exchange, Bob encrypts his private key with the shared secret and sends it to Alice, who can then verify that the public key matches what one would get from Bob's alleged private key when following the protocol honestly. If Alice performs this check before doing anything else with the shared secret, she can be sure not to leak any information to dishonest clients: Bob only learns whether he was honest or not, but he is probably already aware of that.

This paper summarizes some of our and others' fruitless attempts to cryptanalyze SIDH, including a discussion of the reasons why they failed. We hope that this will be useful to other (in particular, novice) researchers in the field of isogeny-based cryptography: In the past, we have observed a tendency among practitioners to rediscover, and sink time into, some of the ideas outlined in the following. Ideally, this paper will provide a shortcut for those poor souls, allowing them to skip past some of the approaches doomed to fail. Finally, we strongly believe that publishing negative results can be valuable: One person's useless observation may be another person's missing link.

Finally, note that we do expect the ideas outlined in the following to strike experienced readers as naïve or foolish. This is by design: Documenting the insight to be gained while debunking — in hindsight — flawed ideas is exactly the point of this paper. *'Trivial' is but another word for 'we understood it'.*

Acknowledgements. The negative results presented in the paper are the result of discussions with many other researchers. We have tried to acknowledge all specific discussions in the relevant subsections. We would like to especially thank Tanja Lange for useful discussions regarding almost every part, if not every part, of this paper, as well as Dan Bernstein, Dan Boneh, Steven Galbraith, Ben Smith, and Fre Vercauteren for many insightful discussions.

2 Preliminaries

In this section, we give an account of the SIDH construction, introduce the problems it poses to cryptanalysts, and finally summarize the most important mathematical properties of the objects of interest.

2.1 The SIDH key-exchange protocol [18]

The core idea in isogeny-based key exchange is to compose two random walks on an isogeny graph of elliptic curves in such a way that the end node of both ways of composing is the same. However, the graph used in SIDH is chaotic — it does not carry a computationally useful structure regular enough to support the evident Diffie–Hellman-style key exchange depicted in Figure 1.

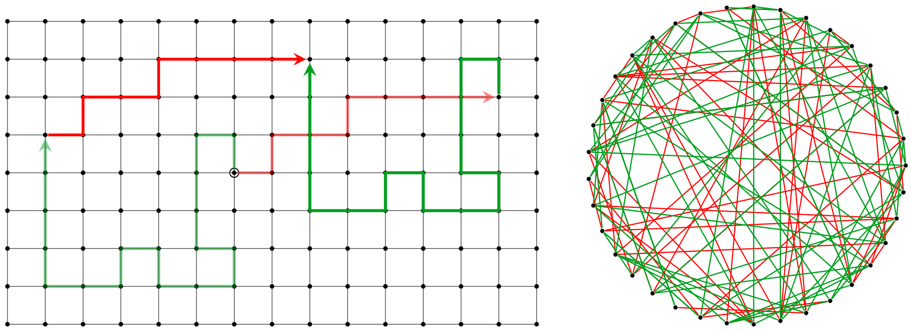
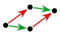


Figure 1. Left: Diffie–Hellman on a (too) structured graph. Right: The supersingular $\{2, 3\}$-isogeny graph over $\mathbb{F}_{431^2}$.

This creates a serious correctness challenge for key-exchange schemes trying to make use of this graph. The resolution of this problem is the core contribution of SIDH: By sending extra information (so-called 'auxiliary points') that helps Alice and Bob orient themselves when walking from the other party's public key node, they are able complete the DH 'diamond'  to obtain a shared secret.

Recall the following fundamental result [34, Prop. III.4.12]:

Lemma 1. *Let E be an elliptic curve and H a finite subgroup of E. Then there exists an elliptic curve E/H and a separable isogeny $\varphi_H \colon E \longrightarrow E/H$ whose kernel is H. The codomain E/H and isogeny φ_H are unique up to isomorphism.*

Parameters. The main parameter in SIDH is a large prime p of the form $p = \ell_A^{n_A} \cdot \ell_B^{n_B} \cdot f - 1$, where ℓ_A, ℓ_B are small distinct primes (typically $2, 3$) and f is a small cofactor (often 1) not divisible by ℓ_A or ℓ_B.

Other parameters are: a supersingular elliptic curve $E_0/\mathbb{F}_p$,[1] a basis (P_A, Q_A) of $E_0[\ell_A^{n_A}]$, and a basis (P_B, Q_B) of $E_0[\ell_B^{n_B}]$. Typically, $E_0 \colon y^2 = x^3 + x$ is used.

Note that the choice of p and E_0 implies that P_A, Q_A, P_B, Q_B are all defined over $\mathbb{F}_{p^2}$, since $E_0(\mathbb{F}_{p^2}) \cong \mathbb{Z}/(p+1) \times \mathbb{Z}/(p+1)$.

[1] In principle, it is not required that E_0 be defined over $\mathbb{F}_p$, but this is beneficial for a variety of reasons. However, there are some reasons to be concerned about special curves like the common choice $j = 1728$; see Section 4.3.

We refer to the curves used in SIDH as 'SIDH curves'. It turns out that these curves actually form a complete set of representatives of *all* isomorphism classes of supersingular elliptic curves over $\overline{\mathbb{F}_p}$.

Keys. Alice's secret key is an integer $a \in \{0, ..., \ell_A^{n_A} - 1\}$, defining the cyclic subgroup $A = \langle P_A + [a]Q_A \rangle \leq E_0[\ell_A^{n_A}]$.

Her public key is the curve E_0/A together with the images $\varphi_A(P_B), \varphi_A(Q_B)$ of Bob's public basis under her (secret) isogeny $\varphi_A \colon E_0 \to E_0/A$.

Bob follows the same process: his secret key is an integer $b \in \{0, ..., \ell_B^{n_B} - 1\}$ defining the cyclic subgroup $B = \langle P_B + [b]Q_B \rangle \leq E_0[\ell_B^{n_B}]$, and his public key is the tuple $(E_0/B, \varphi_B(P_A), \varphi_B(Q_A))$.

Key exchange. Bob takes Alice's public key $(E_0/A, \varphi_A(P_B), \varphi_A(Q_B))$ and uses the points contained in it to *shift* his secret $B \leq E_0[\ell_B^{n_B}]$ to E_0/A: He obtains

$$B' := \varphi_A(B) = \langle \varphi_A(P_B) + [b]\varphi_A(Q_B) \rangle \leq (E_0/A)[\ell_B^{n_B}]\,.$$

This allows him to compute the shared secret $(E_0/A)/B' \cong E_0/\langle A, B \rangle$.

Alice proceeds in exactly the same way: she computes $A' := \varphi_B(A)$ to obtain the shared secret $(E_0/B)/A' \cong E_0/\langle A, B \rangle$.

2.2 Basic observations

Rational points. Tate's theorem [37] implies that E_A and E_B have the same number of points as E_0, that is, $(p+1)^2$. Even stronger, [40, Theorem 4.4] shows that all SIDH curves E have isomorphic groups of $\mathbb{F}_{p^2}$-rational points:

$$E(\mathbb{F}_{p^2}) \;\cong\; \mathbb{Z}/(p+1) \times \mathbb{Z}/(p+1)\,.$$

Among other things, this (together with the smoothness of $p+1$) implies that the Pohlig–Hellman algorithm can compute discrete logarithms in $E(\mathbb{F}_{p^2})$ in polynomial time, and very efficiently in practice. Similarly, the generalization to 'two-dimensional discrete logarithms'—or in other words, decomposing a point in $E(\mathbb{F}_{p^2})$ over a basis of the group of rational points—is easy [35, Algorithm 9.3]. Therefore, the information $\varphi_A(P_B), \varphi_A(Q_B)$ and $\varphi_B(P_A), \varphi_B(Q_A)$ that Alice and Bob transmit reveals much more than just the action of the secret on mere two points: it encodes the action of φ_A resp. φ_B on the *entire* $\ell_B^{n_B}$- resp. $\ell_A^{n_A}$-torsion.

The graph structure. As mentioned before, the set of (isomorphism classes of) SIDH curves together with (a subset of) the rational isogenies between them can be viewed as a graph, a very useful viewpoint for understanding and arguing about isogeny-based cryptosystems. For example, for every finite set $S \subseteq \mathbb{Z}_{\geq 2}$, one obtains an S-isogeny graph where the edges are isogenies whose degree is in S; an important special case is $S = \{\ell\}$ where ℓ is a (typically small) prime. One can prove [14] that (up to isomorphism) there are $\lfloor p/12 \rfloor + \varepsilon$ supersingular elliptic curves defined over $\overline{\mathbb{F}_p}$, where $\varepsilon \in \{0,1,2\}$.[2] It turns out that all of these isomorphism classes have a representative defined over $\mathbb{F}_{p^2}$, hence the SIDH protocol actually works on the graph of *all* supersingular elliptic curves defined over characteristic-p fields.

Moreover, the ℓ-isogeny graph is always connected (for $p \nmid \ell$), and it has excellent mixing properties [30,18]: Any two nodes are expected to be connected via only $O(\log_\ell p)$ steps in the ℓ-isogeny graph, that is, an $\ell^{O(\log_\ell p)}$-isogeny. By counting, it is clear that one cannot hope for faster mixing: Since the ℓ-isogeny graph is $O(\ell)$-regular, there are at most $O(\ell^d)$ nodes at distance $\leq d$ from any given point in the graph. Setting $d \in \Omega(\log_\ell p)$ makes sure one can at least hope to reach all $\Theta(p)$ nodes within d steps, and the theory guarantees that this is indeed true. More careful handling of the constants in the relevant mixing bounds shows that the leading coefficient of the $O(\log_\ell p)$ is in fact a small constant (< 6 for reasonably-sized p), hence the SIDH shared secret is close to uniformly random in the supersingular isogeny graph. On the other hand, this is clearly not true for the public keys, which (by counting) lie in a negligibly small subset whose density is only $O(1/\sqrt{p})$.

Endomorphism rings. It is a classical result of Deuring [11] that the (full) endomorphism ring of a supersingular elliptic curve defined over $\overline{\mathbb{F}_p}$ is (isomorphic to) a maximal order in the quaternion algebra $B_{p,\infty}$ ramified at p and ∞. In the SIDH setting,[3] this means there exists a ring isomorphism from the endomorphism algebra $\mathrm{End}^\circ(E) = \mathrm{End}(E) \otimes_{\mathbb{Z}} \mathbb{Q}$ to the $\mathbb{Q}$-algebra $B_{p,\infty} = \mathbb{Q} \oplus i\mathbb{Q} \oplus j\mathbb{Q} \oplus ij\mathbb{Q}$ with multiplication rules $i^2 = -1$, $j^2 = -p$, and $ij = -ji$. The endomorphism ring $\mathrm{End}(E)$ is thus generated by four linearly independent elements of $B_{p,\infty}$ which span a maximal proper subring with respect to inclusion. The most prominent example is the SIDH starting curve $E_0\colon y^2 = x^3 + x$: Its endomorphism ring is generated as a ring by the endomorphisms ι and $(\iota+\pi)/2$, where ι is an automorphism of order 4 given by $\iota\colon (x,y) \mapsto (-x, \sqrt{-1}\cdot y)$, and $\pi\colon (x,y) \to (x^p, y^p)$ is the p-power Frobenius endomorphism.[4] Hence a $\mathbb{Z}$-basis of $\mathrm{End}(E_0)$ is given by $\langle 1, \iota, \frac{\iota+\pi}{2}, \frac{1+\iota\pi}{2} \rangle$. Note that one can in principle, although there are usually computational hurdles, express the endomorphisms of any other supersingular elliptic curve over $\overline{\mathbb{F}_p}$ with respect to this basis: Fixing an ℓ-isogeny $\psi\colon E_0 \to E$, we get an injective ring homomorphism

$$\mathrm{End}(E) \hookrightarrow \mathrm{End}^\circ(E_0) \cong B_{p,\infty},\ \alpha \mapsto \psi\alpha\widehat{\psi}/\ell. \tag{1}$$

Notice that evaluating an endomorphism given in this representation requires first computing an elliptic-curve point division by ℓ, which typically lies in a

[2] In the SIDH setting, where $p \equiv 11 \pmod{12}$, we have $\varepsilon = 2$.

[3] The technical condition here is $p \equiv 3 \pmod 4$; the other cases are slightly different but not harder in principle.

[4] To see why $(\iota+\pi)/2$ is an (integral) endomorphism of E_0, note that the affine 2-torsion points of E_0 are all of the form $(\xi, 0)$ where $\xi^3 + \xi = 0$, hence $\xi \in \{0, \pm\sqrt{-1}\}$. Since $\xi^p = -\xi$, we have $(\iota+\pi)(\xi,0) = (-\xi,0) + (\xi^p,0) = [2](-\xi,0) = \infty$.

field extension of degree $\Omega(\ell)$, hence special care needs to be taken to make sure this is feasible: for instance, choose ℓ to be powersmooth [15, Algorithm 5].

Also note that $\operatorname{End}(E)$ has many commutative subrings, the most important example being $\mathbb{Z}[\pi]$ when E is defined over $\mathbb{F}_p$. In principle, an efficient commutative subring can give rise to a subexponential quantum attack [4], although it seems just as hard to find *an* endomorphism as to break the scheme in the first place. Therefore the only known example of this idea being useful is $\mathbb{Z}[\pi]$. It does mean, however, that finding an isogeny to a curve defined over $\mathbb{F}_p$ can lead to a subexponential quantum attack; cf. Section 3.1.

Not only is the endomorphism ring isomorphic to a maximal quaternion order, but this so-called *Deuring correspondence* also works in the other direction: there is a bijection between the set of supersingular elliptic curves over $\overline{\mathbb{F}_p}$, up to isomorphism, and the set of 'oriented' maximal orders in $B_{p,\infty}$ [38, Section 42.4]. Simply put, this means for every maximal order $\mathcal{O} \subseteq B_{p,\infty}$ there is a set $\{j, j'\} \subseteq \mathbb{F}_{p^2}$ such that curves with j-invariant j or j' have endomorphism ring $\mathcal{O}$; furthermore, we have $j' = j^p$, hence there is either one such curve, which can be defined over $\mathbb{F}_p$, or the two curves are both defined over $\mathbb{F}_{p^2}$ and Galois conjugates of each other.

Moreover, this correspondence is categorical: Fixing a supersingular elliptic curve E_0 as a base object, every ℓ-isogeny $\alpha\colon E_0 \to E$ corresponds to a left[5] ideal $\mathfrak{a} \subseteq \operatorname{End}(E_0)$ of norm ℓ, and vice-versa (up to post-composition with isomorphisms) [38, Section 42.3]. The codomain E is determined up to isomorphism by the left-ideal *class* of $\mathfrak{a}$, hence finding different representatives of an ideal class corresponds to finding different isogenies between two fixed curves. Notably, given a left ideal $\mathfrak{a} \subseteq E_0$, it is easy to find the endomorphism ring of the image curve of the corresponding isogeny: Under the embedding $\operatorname{End}(E_0) \hookrightarrow B_{p,\infty}$ given in (1), it is isomorphic to another maximal order of $B_{p,\infty}$, and in fact, it turns out that the right order is the adequately named *right order*

$$\mathcal{O}_R(\mathfrak{a}) = \{\, r \in B_{p,\infty} \mid \mathfrak{a} r \subseteq \mathfrak{a} \,\}\,.$$

It may suggest itself at first that this correspondence will be very useful as an attack tool against SIDH. However, it seems that one simply cannot efficiently transcend into this alternate, equivalent reality: All known approaches to compute the endomorphism ring of a given curve essentially go through first finding an isogeny to either another curve with known endomorphism ring (such that one can compute the right order as above), or to itself [21].

2.3 Attack avenues against SIDH

The obvious way to attack SIDH is to try to recover one of the secret isogenies φ_A, φ_B from the public information. (We will often, without loss of generality, silently assume that we are attacking Alice's key.) A priori, it may seem like

[5] Since conjugation swaps the role of left- and right-multiplication, everything can equivalently be phrased in terms of right ideals.

one requires one of the actual secret isogenies; however, Galbraith–Petit–Shani–Ti have demonstrated that *any* isogeny ψ between E_0 and one of $\{E_A, E_B\}$ is enough to recover the *right* isogeny and therefore break the system [16]. The reduction makes use of the fact that the secret isogenies in SIDH are relatively 'short' compared to a 'random' isogeny between two given curves: There are $\Theta(\sqrt{p})$ different secrets, while the graph size is $\Theta(p)$, hence only an exponentially small fraction of SIDH curves can be reached from the starting curve by isogenies shorter than the secret keys. This observation is combined with the fact that isogenies from E_0 correspond to left ideals of $\mathrm{End}(E_0)$, and isogeny *codomains* correspond to left-ideal *classes* (see Section 2.2): The reduction first finds the ideal defining the known isogeny $\psi\colon E_0 \to E_A$, then employs lattice-basis reduction to compute an equivalent ideal of small norm. Except for rare cases of bad luck, this small-norm ideal corresponds to the secret isogeny φ_A. The 'pure' problem of finding an isogeny between E_0 and a given SIDH curve is discussed in Section 3.

The isogeny-finding problem does not capture the full power of an attacker in SIDH. In addition to the target curve, attackers also see the action of the secret isogeny on a coprime torsion subgroup, represented by the action on a few points that span said subgroup. These auxiliary points are the main innovation of SIDH, and the new setting they enable is the reason for SIDH's improved quantum security over other isogeny-based key exchanges [9,32,3], but the additional information that Alice and Bob disclose may also be worrisome: Petit has obtained cryptanalysis results on modified variants of SIDH using these extra points [29]. (Un)fortunately, it seems like there is little hope for his approach to apply to the original, balanced parameters; see Section 4.3. Other potential (but fruitless) approaches based on the extra points are outlined in Section 4.

Finally, note that analogously to the classical Diffie–Hellman setting, there is of course also the potential for an attack that obtains the shared secret *without* first recovering one party's secret key. Similar to the classical case, we are not aware of any ideas to attack SIDH from this direction.

3 Failed attempts to attack the pure isogeny problem

The *pure isogeny problem* for supersingular elliptic curves is:

> Given supersingular E and $E'/\mathbb{F}_{p^2}$, optionally with the guarantee that E and E' are ℓ^n-isogenous for some ℓ^n, compute an isogeny $\phi\colon E \to E'$.

We refer to this as the 'pure' isogeny problem because the hardness assumption on which SIDH is based features a stronger attacker: they also have knowledge of the images of some points under the isogenies φ_A, φ_B in addition to just the domain and the codomains. Moreover, recall from Section 2.3 that it is sufficient to recover *an* isogeny between E_0 and one of E_A, E_B; the correct isogeny can then (usually) be found by employing ideal-based techniques.

The best known classical or quantum attack to find an isogeny $E_0 \to E_A$ in the SIDH setting is essentially a generic approach searching for Alice's secret isogeny φ_A: compute and store random walks of length $n_A/2$ in the ℓ_A-isogeny graph starting from E_0 and E_A until two of them 'meet in the middle'; this algorithm takes time $O(p^{1/4})$ as Alice's isogeny from E_0 to E_A has degree approximately $p^{1/2}$. In practice, the memory cost of this algorithm is prohibitively high, so parallel versions of van Oorschot–Wiener's collision search algorithm with almost the same theoretical time complexity but much better time-space tradeoffs and hence superior real-world performance, are considered to be the best known attack against SIDH/SIKE [1,8]. Note that Tani's $O(p^{1/6})$ quantum algorithm [36] for the claw-finding problem is deemed unlikely to outperform the classical algorithm of van Oorschot–Wiener:

> Our conclusion is that an adversary with enough quantum memory to run Tani's algorithm with the query-optimal parameters could break SIKE faster by using the classical control hardware to run van Oorschot–Wiener. [20]

3.1 Finding the $\mathbb{F}_p$-subgraph

The idea of using the $\mathbb{F}_p$-subgraph to get a better classical attack on the pure isogeny problem was first studied by Delfs and Galbraith [10]. Biasse, Jao, and Sankar [2] later applied the same ideas to construct a more efficient quantum algorithm. The (other) attempts at exploiting the $\mathbb{F}_p$-subgraph presented here have certainly been considered by many people, but not written down as it has not (yet?) led to an improved attack on SIDH.

Trying to find a path to a curve in the $\mathbb{F}_p$-subgraph turns out to be common theme in attempts at attacking SIDH, so we now discuss the consequences such an algorithm would have.

Definition 2. *Let S be set of nodes in the SIDH ℓ-isogeny graph G, and let $S' \subseteq S$ be the subset of those nodes that are defined over $\mathbb{F}_p$. We define the* $\mathbb{F}_p$-subgraph *of G to be the full subgraph of G with nodes from S'.*

Fundamentally, the $\mathbb{F}_p$-subgraph forms a distinguished subset of the full isogeny graph that is easily recognizable once we have found it, and it is also easy to identify those edges that go to another node inside this subgraph.

Delfs–Galbraith use this observation to split the problem of finding an ℓ-isogeny between two arbitrary curves E, E' into two smaller subproblems: finding a path from both E and E' to curves defined over $\mathbb{F}_p$, and then connecting these two curves by an isogeny inside the subgraph. The composition of these three isogenies forms an isogeny $E \to E'$.

In total, one can show that there are approximately $\sqrt{p}$ supersingular elliptic curves defined over $\mathbb{F}_p$. The $\mathbb{F}_p$-subgraph G' of the SIDH ℓ-isogeny graph, with ℓ a prime, is either (if ℓ odd) a disjoint union of cycles of the same length, or (if $\ell = 2$) such a union of cycles with one single extra leaf 'hanging down' from each node in the cycles. The components of these graphs are known as a *volcanoes*, and we call the set of non-leaf nodes the *surface*.

Note that this implies that the surface subgraph is 2-regular, hence using a single ℓ-isogeny $\mathbb{F}_p$-subgraph leads to a time complexity of $\Theta(\sqrt{p})$ for either finding a path between two given nodes or determining that they do not lie in the same component. Using multiple ℓ yields an improvement, though: One can show that subexponentially many ℓ are sufficient to connect all nodes, and (under GRH) that random walks on this combined graph mix quickly [19]. Thus the usual meet-in-the-middle techniques apply, reducing the time complexity of connecting two $\mathbb{F}_p$-subgraph curves to $\widetilde{O}(p^{1/4})$. Note how this is not better at attacking SIDH than the easier meet-in-the-middle attack outlined before; this is because the isogenies in SIDH are known to be particularly short, a property which cannot be exploited by the Delfs–Galbraith approach since almost none of the curves on the path are defined over $\mathbb{F}_p$.

Moreover, *finding* the $\mathbb{F}_p$-subgraph in the first place by brute force costs $\widetilde{O}(\sqrt{p})$: The density of that subgraph is roughly $1/\sqrt{p}$, hence random walks can be expected to find a curve defined over $\mathbb{F}_p$ after walking approximately a number of steps that is the reciprocal of this proportion, i.e., $\sqrt{p}$.

With respect to quantum attacks, similar problems apply: Once the $\mathbb{F}_p$-subgraph has been found, isogeny walks can be interpreted as a commutative class-group action of an imaginary quadratic number ring, and therefore two nodes can be connected using a subexponential-time hidden-shift quantum algorithm [22,23]. This was first applied to isogeny graphs of elliptic curves in [4]. However, there is still no known efficient quantum algorithm to find the $\mathbb{F}_p$-subgraph, hence this does not lead to an improved attack.

An $\mathbb{F}_p$-compass? As stated above, the main problem to solve is finding an isogeny to a curve defined over $\mathbb{F}_p$. The evident brute-force approach is not cheaper than breaking SIDH 'directly' using meet-in-the-middle or collision finding, and more sophisticated methods seem out of reach. For instance, one observation is that a curve at distance d from the $\mathbb{F}_p$-subgraph in the ℓ-isogeny graph has an endomorphism of degree $\ell^{2d}p$ given by walking to the $\mathbb{F}_p$-subgraph, applying Frobenius, and walking back. Why this may seem a promising approach for detecting the $\mathbb{F}_p$-subgraph, it runs into the same problems as always: Checking whether a curve has an endomorphism of a certain norm seems to boil down to simply trying to *find* that endomorphism, which is infeasible unless (here) the distance d to the $\mathbb{F}_p$-subgraph is already extremely small. We have seen many similar or equivalent, but equally fruitless, attempts in this direction come and go in the past. For example, if a curve E is close to the $\mathbb{F}_p$-subgraph, there is a short isogeny between E and its Galois conjugate $E^{(p)}$, but again there is no known way to detect that isogeny unless we are already close enough to find the $\mathbb{F}_p$-subgraph with a generic approach.

Other subrings? One way to interpret the $\mathbb{F}_p$-subgraph is as the subset of curves with a certain endomorphism of norm p, namely the p-power Frobenius endomorphism. Hence, one is implicitly looking for those supersingular elliptic curves whose endomorphism ring contains the Frobenius order $\mathbb{Z}[\pi]$, and in principle the same sort of subgraph exists for other commutative subrings, like for example $\mathbb{Z}[\iota]$, although in this case it only consists of the single node E_0.

Finding, for instance, a bigger commutative subring than $\mathbb{Z}[\pi]$ that is contained in almost all endomorphism rings in the graph would potentially allow to spend less time on searching for the associated subgraph, but still apply the subexponential quantum attack once it is found.

However, there are a number of problems associated with this approach, one fundamental in nature and the others (as usual) computational: The embedding $\mathrm{End}(E) \hookrightarrow B_{p,\infty}$ is highly *non-canonical*. This means that even if one was able to compute (subrings of) the endomorphism rings of two curves, there is still no way to tell how these rings are related under the embedding from (1). The usual strategy to deal with this problem in theory is to make sure the embeddings are always compatible when considering two isogenous curves, but without knowing an isogeny, this of course seems impossible to do in practice. This issue does not apply to $\mathbb{Z}[\pi]$ as, given a curve $E/\mathbb{F}_p$, the endomorphism π is always trivial to find (it is just $(x, y) \mapsto (x^p, y^p)$), and since (by definition) isogenies defined over $\mathbb{F}_p$ commute with π, we automatically have $\psi\pi\widehat{\psi}/(\deg\psi) = \pi\psi\widehat{\psi}/(\deg\psi) = \pi$ for all isogenies $\psi\colon E \to E'$ defined over $\mathbb{F}_p$. Therefore it is possible to identify a canonical subring of the endomorphism ring which is automatically compatible between different $\mathbb{F}_p$-isogenous curves.

The computational problems are the usual: It is not clear how to tell whether a given curve E has an endomorphism of a given norm and trace, it seems impossible to make sure these endomorphisms are compatible choices without first finding an isogeny between the two curves in question, and for the quantum part of the attack it must also be efficient to *evaluate* the endomorphisms on points.

3.2 Lifting to characteristic zero

It is relatively well-known that to an ordinary elliptic curve $E/\mathbb{F}_q$ one can canonically associate an elliptic curve $E'/\mathbb{Q}_q$[6] with the same endomorphism ring (viewed as an order in a quadratic number field) — this is normally referred to as the ‘canonical lift’ [24] [26, Appendix], and E is the (unique) *reduction* of E'.

It is possible to compute this lift, for example via Satoh’s algorithm [33], albeit not efficiently for large characteristic p. Furthermore, it is functorial — we can also lift (and reduce) isogenies. A natural question is:

> Given a supersingular elliptic curve $E/\mathbb{F}_{p^2}$ with endomorphism ring $\mathcal{O}$, is there a way to canonically construct an elliptic curve $E'/\mathbb{C}$ whose endomorphism ring is isomorphic to a (well-chosen) commutative subring of $\mathcal{O}$?

[6] The field $\mathbb{Q}_q$, which can be embedded into $\mathbb{C}$, is the fraction field of $\mathbb{Z}_q$, which is a finite extension of the p-adic integers $\mathbb{Z}_p$, which has as elements power series in p.

Suppose for the sake of argument that such a construction is efficiently computable and that we can also lift and reduce isogenies. Then to find a path between $E_1/\mathbb{F}_q$ and $E_2/\mathbb{F}_q$ we could first compute their canonical lifts $E_1'/\mathbb{Q}_q$ and $E_2'/\mathbb{Q}_q$ respectively and then compute an isogeny $E_1' \to E_2'$, which one could subsequently hope to reduce back to $\mathbb{F}_q$. As $\mathbb{Q}_q \hookrightarrow \mathbb{C}$, the lifts E_1' and E_2' can be viewed as complex elliptic curves. As a complex elliptic curve is nothing but a torus and an isogeny between two such curves is just a $\mathbb{C}$-linear map, one may hope to be able to easily compute an isogeny over $\mathbb{C}$ using some linear algebra.

Unfortunately, the computational methods for computing the lift of an ordinary elliptic curve $E/\mathbb{F}_q$, such as Satoh's [33], all exploit a known endomorphism τ on E — in their case τ is the Frobenius π — and construct an elliptic curve $E'/\mathbb{C}$ with endomorphism algebra $\mathrm{End}^\circ(E') \cong \mathbb{Q}(\tau)$.

For a generic supersingular elliptic curve $E/\mathbb{F}_{p^2}$, the only endomorphisms we know of are scalar multiplications, i.e., lie in $\mathbb{Z}$. (Recall that in the SIDH case the p^2-power Frobenius is just $[-p]$.) So even if we could lift E in a meaningful and computable way to $E'/\mathbb{C}$ while preserving a known endomorphism, we simply wouldn't know how to find that endomorphism in the first place (as usual).

Computing a path from a generic supersingular elliptic curve $E/\mathbb{F}_{p^2}$ to a curve defined over $\mathbb{F}_p$ would be helpful in this context, but then there would then be easier ways to proceed, see Section 3.1.

3.3 Weil restrictions

Acknowledgements. Some of the ideas in this section were discussed with participants of the Spontaneous Isogeny Day in Leuven in October 2018. We had particularly enlightening discussions on this topic with Wouter Castryck, Steven Galbraith, Joost Renes, Ben Smith, and Fre Vercauteren (alphabetical order).

To any (supersingular) elliptic curve $E/\mathbb{F}_{p^2}$, one can in a natural way associate a (supersingular) principally polarizable abelian surface[7] $W(E)/\mathbb{F}_p$ called the *Weil restriction*.[8] Modulo (many) technical details, the fundamental idea is to interpret the defining equation of E over $\mathbb{F}_{p^2}$ as a set of equations over $\mathbb{F}_p$ instead by plugging in, then splitting over, an $\mathbb{F}_p$-basis of $\mathbb{F}_{p^2}$. The Weil restriction is functorial: isogenies of elliptic curves defined over $\mathbb{F}_{p^2}$ restrict to isogenies of their Weil restrictions over $\mathbb{F}_p$. This means that the the isogeny graph of supersingular elliptic curves defined over $\mathbb{F}_{p^2}$ can be viewed as a subgraph of the isogeny graph of supersingular principally polarized abelian surfaces over $\mathbb{F}_p$.

The centre of the $\mathbb{F}_p$-rational endomorphism ring $\mathrm{End}_{\mathbb{F}_p}(A)$ of an abelian variety defined over $\mathbb{F}_p$ is is an order in $\mathbb{Q}(\pi)$, where π is the p-power Frobenius of A [37, Theorem 2].

One might hope that in fact $\mathrm{End}_{\mathbb{F}_p}(A) \otimes_{\mathbb{Z}} \mathbb{Q} = \mathbb{Q}(\pi)$, as happens when $\dim(A) = 1$. We considered what the consequences of this might be: Assume that

[7] To read more about principally polarized abelian varieties, see [13, Chapter 11].

[8] To read more about Weil restrictions, see [12].

for the Weil restriction $W(E_A)$ of E_A (Alice's public key), the $\mathbb{F}_p$-rational endomorphism ring is commutative. For all but finitely many primes ℓ, we then expect that the (ℓ,ℓ)-isogeny[9] graph of supersingular principally polarized abelian surfaces defined over $\mathbb{F}_p$ is a disjoint union of cycles, as justified at the end of this section. If there is a list $\ell_1,\ldots,\ell_n$ such that the connected component of the union of the $(\ell_1,\ell_1),\ldots,(\ell_n,\ell_n)$-isogeny graphs contains $W(E_0)$ and $W(E_A)$, then the problem of finding a path from $W(E_0)$ to $W(E_A)$ can be viewed as a *hidden shift* problem, for which, if the individual steps in the path, i.e., isogenies, can be efficiently computed, there is a subexponential quantum algorithm due to Kuperberg [22,23].

Any hope? We need the probability of $W(E_0)$ and $W(E_A)$ being in the same connected component C of the union of the $(\ell_1,\ell_1),\ldots,(\ell_n,\ell_n)$-isogeny graphs to be high, which can only happen if C contains the Weil restrictions of almost all the supersingular elliptic curves defined over $\mathbb{F}_{p^2}$.

We expect (as justified at the end of this section) that the (ℓ_i,ℓ_i)-isogeny graphs will be the disjoint union of cycles of length $O(\sqrt{p})$. There exist $\Theta(p)$ (Weil restrictions of) supersingular elliptic curves defined over $\mathbb{F}_{p^2}$, so to have any chance of C covering almost all of these, we would need to take n to be at least $\Omega(\sqrt{p})$.

Currently we can't compute (ℓ,ℓ)-isogenies efficiently enough unless $\ell = 2$,[10] so we assume for the sake of argument that the complexity of a somewhat optimized algorithm to do this would scale at least as badly as Vélu's formulas for elliptic curves. That is, we assume that the evaluation of an (ℓ,ℓ)-isogeny takes time $\Omega(\ell)$ or $\Omega(\ell^2)$. Since we need to take at least $\Omega(\sqrt{p})$ different primes ℓ, it is then definitely not true that 'the individual steps in the path, i.e. isogenies, can be efficiently computed'.

More ideas? We considered two variations on this idea:

1. Instead of hoping that $W(E_A)$ is in the same connected component is $W(E_0)$, hope that it is in the same connected component of the Weil restriction of some curve defined over $\mathbb{F}_p$. Approximately one in $\sqrt{p}$ (Weil restrictions of) elliptic curves over $\mathbb{F}_{p^2}$ are (Weil restrictions of) elliptic curves over $\mathbb{F}_p$, so looking at one cycle of length $O(\sqrt{p})$, i.e., just one (ℓ,ℓ)-isogeny graph, might be enough.
   However, since we don't know which curve over $\mathbb{F}_p$ we're looking for, it seems impossible to phrase this as a hidden shift problem, so Kuperberg's algorithm doesn't apply.
2. Recall that we assume that $\mathrm{End}_{\mathbb{F}_p}(W(E_A))$ is an order in $\mathbb{Q}(\pi)$, where π is the p-power Frobenius on $W(E_A)$. We explain below that $\pi = \zeta_8\sqrt{p}$, where ζ_8 is an eighth root of unity, and that we then expect that the application of an

[9] To read more about (ℓ,ℓ)-isogenies, see [6].

[10] To read more about computing (2,2)-isogenies efficiently, see [7].

(ℓ,ℓ)-isogeny to a supersingular principally polarized abelian surface $A/\mathbb{F}_p$ can be viewed as the action of an ideal in the class group $\mathrm{cl}(\mathcal{O}_{\mathbb{Q}(\zeta_8\sqrt{p})})$ (the ℓ are chosen so that $\mathcal{O}_{\mathbb{Q}(\zeta_8\sqrt{p})} = \mathbb{Z}[\zeta_8\sqrt{p}]$ locally at ℓ). The reason that we expect the cycles in the (ℓ,ℓ)-isogeny graph to have length approximately $\sqrt{p}$ comes from this action — this is (heuristically) the size of this class group. However $\mathcal{O}_{\mathbb{Q}(\zeta_8\sqrt{p})}$ is *not* the largest commutative subring of $\mathrm{End}_{\overline{\mathbb{F}}_p}(A)$ (locally at ℓ): Since Frobenius commutes with every endomorphism, we could add another endomorphism to get a rank-4 $\mathbb{Z}$-module, the class group of which is likely to have a higher class number. But this is of course equivalent to finding non-obvious endomorphisms, which, if we could do, would lead to a much easier way of attacking SIDH, as explained in Section 2.2.

A couple of handwavy mathematical details. As stated above, we expect that, under the assumption that $\mathrm{End}_{\mathbb{F}_p}(W(E_A))$ is commutative: For all but finitely many primes ℓ, the (ℓ,ℓ)-isogeny graph of supersingular principally polarized abelian surfaces defined over $\mathbb{F}_p$ is a disjoint union of cycles. We also conjectured that the cycles have length $\Omega(\sqrt{p})$ (subject to some heuristics). We briefly justify our expectations here.

Suppose that ℓ does not divide the index $\left[\mathcal{O}_{\mathbb{Q}(\pi)} : \mathbb{Z}[\pi]\right]$, where π is the p-power Frobenius on $W(E_A)$. Since under our assumptions for any supersingular abelian surface over $A/\mathbb{F}_p$ with commutative $\mathbb{F}_p$-rational endomorphism ring we have that

$$\mathbb{Z}[\pi] \subseteq \mathrm{End}_{\mathbb{F}_p}(A) \subseteq \mathcal{O}_{\mathbb{Q}(\pi)},$$

it follows that every supersingular abelian surface over $A/\mathbb{F}_p$ has endomorphism ring $\mathcal{O}_{\mathbb{Q}(\pi)}$ locally at ℓ. An isogeny of abelian surfaces is uniquely determined by its kernel (just like with elliptic curves). In particular, if I is an ideal of $\mathrm{End}_{\mathbb{F}_p}(A)$ then we define f_I to be the isogeny from A with kernel

$$\bigcap_{\alpha \in I} \ker(\alpha).$$

Following exactly the same proof strategy as for elliptic curves, it is believable that the class group of $\mathcal{O}_{\mathbb{Q}(\pi)}$ acts on the set of supersingular abelian surfaces over $\mathbb{F}_p$ with endomorphism ring $\mathcal{O}_{\mathbb{Q}(\pi)}$ via

$$I * E = f_I(E).$$

Going one step further, we suppose for the sake of argument that horizontal (ℓ,ℓ)-isogenies even come from the action of an ideal $\mathfrak{l}$ such that $\ell\mathcal{O}_{\mathbb{Q}(\sqrt{-p})} = \mathfrak{l}\bar{\mathfrak{l}}$. If, as in the elliptic curve case, the results for supersingular abelian surfaces over a prime field turn out to be analogous to results for ordinary abelian surfaces, then such an ideal would send a supersingular abelian surface $A/\mathbb{F}_p$ with endomorphism ring $\mathcal{O}_{\mathbb{Q}(\pi)}$ equipped with a principal polarization $\zeta\colon A \to A^\vee$ to a supersingular abelian surface $f_{\mathfrak{l}}(A)/\mathbb{F}_p$ with endomorphism ring $\mathcal{O}_{\mathbb{Q}(\pi)}$ equipped with a principal polarization $\ell\zeta$. The analogous result for the ordinary case that we refer to here is [25, Proposition 3.6.1].

If all of this holds, then the (ℓ,ℓ)-isogeny graph of any prime ℓ not dividing $\big[\mathcal{O}_{\mathbb{Q}(\pi)} : \mathbb{Z}[\pi]\big]$ that splits in $\mathcal{O}_{\mathbb{Q}(\pi)}$ is a cycle. Suppose that $\ell\mathcal{O}_{\mathbb{Q}(\pi)} = \mathfrak{l}\bar{\mathfrak{l}}$. Then the length of the cycle is given by the order of $[\mathfrak{l}]$ in $\mathrm{cl}(\mathcal{O}_{\mathbb{Q}(\pi)})$.

Furthermore, by a theorem of Manin and Oort [28, p. 116], the Frobenius π equals $\zeta\sqrt{p}$, where ζ is a root of unity. By a theorem of Tate [37, Theorem 2], our assumption that the endomorphism algebra $B = \mathrm{End}_{\mathbb{F}_p}(W(E_A)) \otimes_{\mathbb{Z}} \mathbb{Q}$ is commutative is equivalent to saying that $[B : \mathbb{Q}] = 4$, so $\zeta = \zeta_8$ is in fact an eighth root of unity, and the characteristic polynomial of Frobenius is $x^4 - p^2$. According to standard class group heuristics [5], we expect that $\mathrm{cl}(\mathcal{O}_{\mathbb{Q}(\zeta_8\sqrt{p})})$ is cyclic or almost cyclic, and has order $\Omega(\sqrt{p})$ — hence the (ℓ,ℓ)-isogeny graph, where ℓ satisfies all of the conditions above, is heuristically speaking the disjoint union of cycles of length approximately $\sqrt{p}$.

4 Failed attack attempts that use the auxiliary points

The attacker has more information available than just two isogenous curves: They also get the action of Alice's and Bob's secret isogenies φ_A resp. φ_B on the $\ell_B^{n_B}$- resp. $\ell_A^{n_A}$-torsion. We focus on the problem of recovering the secret from a public key. Without loss of generality, suppose that $\ell_A^{n_A} < \ell_B^{n_B}$ and we are attacking Alice's public key $(E_A, \varphi_A(PB), \varphi_A(QB))$.

First, note that the extra information defines the secret isogeny uniquely: Consider two distinct d-isogenies $\phi, \psi\colon E \to E'$ with the same action on the m-torsion. Then $\ker(\phi - \psi) \supseteq E[m]$, hence $\deg(\phi - \psi) \geq \#\ker(\phi - \psi) \geq m^2$. On the other hand, Lemma V.1.2 of [34] implies $\deg(\phi - \psi) \leq 4d$. Combining these bounds yields $m^2 \leq 4d$. In SIDH, this implies that an $\ell_A^{n_A}$-isogeny is uniquely defined by its action on the $\ell_B^{n_B}$-torsion unless the parameters are highly unbalanced. However, no efficient way to make use of this information is known.

4.1 Interpolation problems

By definition, isogenies are rational maps, hence it is clear that given enough inputs and outputs, one can in principle recover the coefficients of that rational map [39, Section 5.8]. One can show [31, Proposition 1] that in the SIDH setting, the isogeny φ_A can be written as

$$\varphi_A\colon (x, y) \longmapsto \big(f(x),\, c_0 y \cdot f'(x)\big)$$

for some rational map $f \in \mathbb{F}_{p^2}(x)$ of degree $\ell_A^{n_A}$ and a constant $c_0 \in \mathbb{F}_q$. Therefore, being given the action of φ_A, and thereby f, on 'enough' points, one might hope to recover f and thus Alice's secret isogeny φ_A.

However, this is computationally infeasible: Even printing the *result* of the interpolation takes time linear in the degree, which in SIDH is exponentially large (in the bit length of the involved objects). One might wonder whether it is possible to *evaluate* the function while reconstructing it, thus circumventing the exponentially big output, but all known ways to do (polynomial or rational) interpolation still take time at least linear in the degree. The only conceivable way to succeed with this approach would be to reconstruct the rational map while *at the same time* rewriting it as a *composition* of rational maps, such that each of these maps has a degree polynomially small in ℓ_A. While there are of course methods to decompose polynomials and rational maps into a composition of smaller-degree maps, these algorithms require first storing the input in full.

Generally, the approach of rational-function interpolation seems similar in spirit to the interpolation idea in the next section, except that so far we haven't made any use of the group structure underlying the rational maps in question. Since we've been working with less than all the available structure, it seems reasonable to assume that this approach is fundamentally inferior to the ideas in the next sections.

4.2 Group-theoretic approaches

Perhaps the most obvious idea to make use of the auxiliary points is to try to extrapolate the known action of φ_A on the $\ell_B^{n_B}$-torsion to a bigger torsion subgroup to subsequently recover (part of) the secret.

Unfortunately, it is evident that purely group-theoretic methods are doomed to fail: Let $\gcd(m, \ell_B^{n_B}) = 1$. By the structure theorem of finite abelian groups, the $\ell_B^{n_B}$- and m-torsion subgroups of an elliptic curve are *independent*; i.e., there are simply no nontrivial relations between points of ℓ_B-power order and points of order m in the curve group. (In other words, the $\ell_B^{n_B}m$-torsion subgroup is an internal direct product of the $\ell_B^{n_B}$- and the m-torsion.) Perhaps a reliable extrapolation is too much to ask for, but it seems that even obtaining *any* information about the action on the ℓ_A-torsion with success probability (non-negligibly) better than random guessing seems infeasible. In a sense, this is remarkable, since elliptic curves are also equipped with a *geometric* structure, and many purely group-theoretical morphisms defined on elliptic curve groups do not come from an isogenies, i.e., do not respect the geometric structure. However, nobody has yet discovered an efficient way to exploit this.

An effective Tate's theorem? Rather than extrapolating to a coprime torsion subgroup, one may instead attempt to lift the action of φ_A on the $\ell_B^{n_B}$-torsion to a *higher* ℓ_B-power torsion subgroup. In the limit, this lifting process would yield the action of φ_A on the ℓ_B-adic Tate modules $T_{\ell_B}(E_0)$.[11] Write $\ell = \ell_B$.

If one knew how to do the lifting step, this observation may inspire hope: It is known [34, Theorem 7.7] that the natural map

$$\operatorname{Hom}_{\mathbb{F}_{p^2}}(E_0, E_A) \otimes_{\mathbb{Z}} \mathbb{Z}_\ell \longrightarrow \operatorname{Hom}_{\mathbb{F}_{p^2}}(T_\ell(E_0), T_\ell(E_A))$$

is an isomorphism of $\mathbb{Z}_\ell$-modules, hence the action of an isogeny defined over $\mathbb{F}_{p^2}$ on a sufficiently high ℓ^k-torsion completely determines the map. While this is

[11] The functor T_ℓ is defined as the inverse limit $T_\ell(E) = \varprojlim_n E[\ell^n]$ under the evident restriction maps $[\ell]\colon E[\ell^{n+1}] \twoheadrightarrow E[\ell^n]$; see for instance [34, Section III.7].

Note that if $\ell \neq 0$ in the field of definition of the curve E, then $T_\ell(E) \cong \mathbb{Z}_\ell \times \mathbb{Z}_\ell$.

an abstract result, Petit [29] found a way to turn this into an efficient algorithm assuming k grows big enough; see Section 4.3.

However, in any case, it seems that similar obstacles as in the previous section (extrapolating to another torsion subgroup) apply: Group-theoretically, the action on $E[\ell^k]$ can be lifted to an action on $E[\ell^{k+1}]$ in ℓ^4 different ways. Also taking into account the known information about the degree (coprime to ℓ), this expansion factor shrinks slightly,[12] but there still is no hope to learn anything about the action on the ℓ^∞-torsion without making use of the geometry of the underlying elliptic curve.

4.3 Constructing endomorphisms to exploit the auxiliary points

Acknowledgements. The ideas in this section are all based on Petit's paper [29], and in particular are the result of discussions with Dan Bernstein (who showed us the technique used below to estimate the expected size of solutions), Tanja Lange, and Christophe Petit (alphabetical order).

Recall that in two-party SIDH

$$\ell_A^{n_A} \approx \ell_B^{n_B} \approx \sqrt{p}\,,$$

corresponding to Alice's and Bob's secret isogenies having approximately the same degree. Petit [29] shows how to construct an endomorphism on E_A if instead

$$\ell_B^{n_B} \gg \ell_A^{n_A}\,,$$

such that the capability to evaluate this endomorphism on the $\ell_B^{n_B}$-torsion — which is granted to the attacker in the SIDH setting by means of the auxiliary points (see Section 2.2) — allows one to reconstruct Alice's secret isogeny.

Petit's attack. Following the notation of Section 2.2, let π be the p-power Frobenius on E_0 and let ι be the order-4 automorphism $(x, y) \mapsto (-x, \sqrt{-1}\cdot y)$ on E_0. Then for any $a, b, c \in \mathbb{Z}$, we have an endomorphism $a\iota\pi + b\pi + c\iota \in \operatorname{End}(E_0)$, and using the (unknown) $\ell_A^{n_A}$-isogeny $\varphi_A \colon E_0 \to E_A$ we can, for every $d \in \mathbb{Z}$, define the endomorphism

$$\alpha = \varphi_A(a\iota\pi + b\pi + c\iota)\widehat{\varphi_A} + d \quad \in \operatorname{End}(E_A)$$

of degree (or equivalently, norm)[13]

$$\deg(\alpha) = \ell_A^{2n_A}pa^2 + \ell_A^{2n_A}pb^2 + \ell_A^{2n_A}c^2 + d^2.$$

[12] The expansion factor is smaller, but still significant, for *endo*morphisms with known degree and trace: Forcing the characteristic polynomial limits the amount of choice. Concretely, there are ℓ^2 different ways to lift a known action on the ℓ^n-torsion to the ℓ^{n+1}-torsion while satisfying a given characteristic polynomial $\chi \bmod \ell^{n+1}$.

[13] A reader comparing this with the formula given in [29, p. 15] may wonder where q has gone, but the norm of this specific endomorphism ι is $q = 1$.

Of course, since the attacker does not know φ_A, they cannot compute α directly. However, writing $N_1 = \ell_A^{n_A}$ and $N_2 = \ell_B^{n_B}$, Petit gives conditions under which one can efficiently find $a, b, c, d \in \mathbb{Z}$ such that

$$N_1^2 pa^2 + N_1^2 pb^2 + N_1^2 c^2 + d^2 = eN_2\,, \tag{2}$$

where e is a small cofactor controlling the remaining amount of brute-force work the attacker has to do. If $N_2 = \ell_B^{n_B}$ is big enough relative to $N_1 = \ell_A^{n_A}$, then $\ker \alpha$, and subsequently the secret $\ker \varphi_A$, can be recovered from the action of φ_A on the $\ell_B^{n_B}$-torsion in polynomial time.

Any hope for $\boldsymbol{\ell_A^{n_A} \approx \ell_B^{n_B} \approx \sqrt{p}}$? We can heuristically estimate the expected size of solutions to (2) as follows. Suppose we want to count solutions with $e \leq M$ for some fixed bound M. Since all the terms in (2) are nonnegative, they cannot be bigger than the right-hand side $\approx M\sqrt{p}$. Hence

$$a, b \lessapprox \sqrt{M} \cdot p^{-3/4}\,; \qquad c \lessapprox \sqrt{M} \cdot p^{-1/4}\,; \qquad d \lessapprox \sqrt{M} \cdot p^{1/4}\,.$$

This means the total number of possible assignments for the variables a, b, c, d, e is approximately

$$M^3 p^{-3/2}\,.$$

Assuming (wrongly, but for the sake of a rough estimate) that for each such assignment, the left- and right-hand side of (2) are uniformly random nonnegative integers upper bounded by $\approx M\sqrt{p}$, the expected number of solutions with $e \leq M$ is seen to be about

$$\frac{M^3 p^{-3/2}}{M\sqrt{p}} = M^2 p^{-2}\,,$$

implying that one needs to increase M to approximately p before a solution can be expected. This means that the smallest expected solution to (2) features the undesirable property $e \approx p$, which means that in this case, Petit's attack performs much worse than simply applying one of the known graph-walking attacks from Section 3 directly. We can therefore conclude that at least heuristically, it seems extremely unlikely that Petit's attack can possibly apply to the actual, balanced SIDH parameters.

References

1. Gora Adj, Daniel Cervantes-Vázquez, Jesús-Javier Chi-Domínguez, Alfred Menezes, and Francisco Rodríguez-Henríquez. On the cost of computing isogenies between supersingular elliptic curves. In *SAC*, volume 11349 of *LNCS*, pages 322–343. Springer, 2018. https://ia.cr/2018/313.
2. Jean-François Biasse, David Jao, and Anirudh Sankar. A quantum algorithm for computing isogenies between supersingular elliptic curves. In *INDOCRYPT*, volume 8885 of *LNCS*, pages 428–442. Springer, 2014.

3. Wouter Castryck, Tanja Lange, Chloe Martindale, Lorenz Panny, and Joost Renes. CSIDH: an efficient post-quantum commutative group action. In *ASIACRYPT (3)*, volume 11274 of *LNCS*, pages 395–427. Springer, 2018. https://ia.cr/2018/383.
4. Andrew M. Childs, David Jao, and Vladimir Soukharev. Constructing elliptic curve isogenies in quantum subexponential time. *J. Mathematical Cryptology*, 8(1):1–29, 2014. https://arxiv.org/abs/1012.4019.
5. Henri Cohen and Hendrik W. Lenstra, Jr. Heuristics on class groups of number fields. In *Number Theory Noordwijkerhout 1983*, pages 33–62. Springer, 1984.
6. Romain Cosset and Damien Robert. Computing (ℓ, ℓ)-isogenies in polynomial time on Jacobians of genus 2 curves. *Math. Comp.*, 84(294):1953–1975, 2015.
7. Craig Costello. Computing supersingular isogenies on Kummer surfaces. In *ASIACRYPT (3)*, volume 11274 of *LNCS*, pages 428–456. Springer, 2018. https://ia.cr/2018/850.
8. Craig Costello, Patrick Longa, Michael Naehrig, Joost Renes, and Fernando Virdia. Improved classical cryptanalysis of the computational supersingular isogeny problem, 2019. IACR Cryptology ePrint Archive 2019/298. https://ia.cr/2019/298.
9. Jean-Marc Couveignes. Hard homogeneous spaces, 2006. IACR Cryptology ePrint Archive 2006/291. https://ia.cr/2006/291.
10. Christina Delfs and Steven D. Galbraith. Computing isogenies between supersingular elliptic curves over $\mathbb{F}_p$. *Des. Codes Cryptography*, 78(2):425–440, 2016. https://arxiv.org/abs/1310.7789.
11. Max Deuring. Die Typen der Multiplikatorenringe elliptischer Funktionenkörper. *Abhandlungen aus dem Mathematischen Seminar der Universität Hamburg*, 14:197–272, 1941.
12. Claus Diem and Niko Naumann. On the structure of Weil restrictions of abelian varieties. *J. Ramanujan Math. Soc.*, 18(2):153–174, 2003. https://arxiv.org/abs/math/0504359.
13. Bas Edixhoven, Gerard van der Geer, and Ben Moonen. *Abelian varieties.* 2007. Book in preparation. http://gerard.vdgeer.net/AV.pdf.
14. Martin Eichler. Über die Idealklassenzahl total definiter Quaternionenalgebren. *Mathematische Zeitschrift*, 43(1):102–109, December 1938.
15. Kirsten Eisenträger, Sean Hallgren, Kristin E. Lauter, Travis Morrison, and Christophe Petit. Supersingular isogeny graphs and endomorphism rings: Reductions and solutions. In *EUROCRYPT (3)*, volume 10822 of *LNCS*, pages 329–368. Springer, 2018. https://ia.cr/2018/371.
16. Steven D. Galbraith, Christophe Petit, Barak Shani, and Yan Bo Ti. On the security of supersingular isogeny cryptosystems. In *ASIACRYPT (1)*, volume 10031 of *LNCS*, pages 63–91, 2016. https://ia.cr/2016/859.
17. David Jao, Reza Azarderakhsh, Matthew Campagna, Craig Costello, Luca De Feo, Basil Hess, Amir Jalali, Brian Koziel, Brian LaMacchia, Patrick Longa, Michael Naehrig, Joost Renes, Vladimir Soukharev, and David Urbanik. SIKE. Submission to [27]. http://sike.org.
18. David Jao and Luca De Feo. Towards quantum-resistant cryptosystems from supersingular elliptic curve isogenies. In *PQCrypto*, volume 7071 of *LNCS*, pages 19–34. Springer, 2011. https://ia.cr/2011/506.
19. David Jao, Stephen D. Miller, and Ramarathnam Venkatesan. Ramanujan graphs and the random reducibility of discrete log on isogenous elliptic curves. *Journal of Number Theory*, 129(6):1491–1504, 2009. https://arxiv.org/abs/0811.0647.
20. Samuel Jaques and John Schanck. Quantum cryptanalysis in the RAM model: claw finding attacks on SIKE. 2019. IACR Cryptology ePrint Archive 2019/103. https://ia.cr/2006/291.

21. David Kohel. *Endomorphism rings of elliptic curves over finite fields*. PhD thesis, University of California, Berkeley, 1996. http://echidna.maths.usyd.edu.au/kohel/pub/thesis.pdf.
22. Greg Kuperberg. A subexponential-time quantum algorithm for the dihedral hidden subgroup problem. *SIAM J. Comput.*, 35(1):170–188, 2005. https://arxiv.org/abs/quant-ph/0302112.
23. Greg Kuperberg. Another subexponential-time quantum algorithm for the dihedral hidden subgroup problem. In *TQC*, volume 22 of *LIPIcs*, pages 20–34. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2013. https://arxiv.org/abs/1112.3333.
24. Jonathan Lubin, Jean-Pierre Serre, and John Tate. Elliptic curves and formal groups, 1964. Lecture notes. http://ma.utexas.edu/users/voloch/lst.html.
25. Chloe Martindale. *Isogeny graphs, modular polynomials, and applications*. PhD thesis, Universiteit Leiden and Université de Bordeaux, 2018. http://martindale.info/research/Thesis.pdf.
26. William Messing. *The crystals associated to Barsotti–Tate groups: with applications to abelian schemes*, volume 264 of *Lecture Notes in Mathematics*. Springer, 1972.
27. National Institute of Standards and Technology. Post-quantum cryptography standardization, December 2016. https://csrc.nist.gov/Projects/Post-Quantum-Cryptography/Post-Quantum-Cryptography-Standardization.
28. Frans Oort. Subvarieties of moduli spaces. *Invent. Math.*, 24:95–119, 1974.
29. Christophe Petit. Faster algorithms for isogeny problems using torsion point images. In *ASIACRYPT (2)*, volume 10625 of *LNCS*, pages 330–353. Springer, 2017. https://ia.cr/2017/571.
30. Arnold K. Pizer. Ramanujan graphs and Hecke operators. *Bull. Amer. Math. Soc. (N.S.)*, 23(1):127–137, July 1990.
31. Joost Renes. Computing isogenies between Montgomery curves using the action of $(0, 0)$. In *PQCrypto*, volume 10786 of *LNCS*, pages 229–247. Springer, 2018. https://ia.cr/2017/1198.
32. Alexander Rostovtsev and Anton Stolbunov. Public-key cryptosystem based on isogenies, 2006. IACR Cryptology ePrint Archive 2006/145. https://ia.cr/2006/145.
33. Takakazu Satoh. The canonical lift of an ordinary elliptic curve over a finite field and its point counting. *J. Ramanujan Math. Soc.*, 15(4):247–270, 2000.
34. Joseph H. Silverman. *The arithmetic of elliptic curves*. Number 106 in Graduate Texts in Mathematics. Springer, 2nd edition, 2009.
35. Andrew V. Sutherland. *Order computations in generic groups*. PhD thesis, Massachusetts Institute of Technology, 2007. https://groups.csail.mit.edu/cis/theses/sutherland-phd.pdf.
36. Seiichiro Tani. An improved claw finding algorithm using quantum walk. In *MFCS*, volume 4708 of *LNCS*, pages 536–547. Springer, 2007. https://arxiv.org/abs/0708.2584.
37. John Tate. Endomorphisms of abelian varieties over finite fields. *Inventiones mathematicae*, 2(2):134–144, 1966.
38. John Voight. *Quaternion algebras*. July 2018. Book in preparation. https://math.dartmouth.edu/~jvoight/quat-book.pdf.
39. Joachim von zur Gathen and Jürgen Gerhard. *Modern Computer Algebra*. Cambridge University Press, 3rd edition, 2013.
40. Lawrence C. Washington. *Elliptic curves: Number theory and cryptography*. Discrete Mathematics and its Applications. Chapman & Hall/CRC, 2nd edition, 2008.