A mathematical look at the Supersingular Isogeny Diffie Hellman key exchange

Author: Maaike Heijdenrijk
Student number: S4528301

Supervisor: Dr. Wieb Bosma
Second reader: Dr. Bernd Souvignier

June 30, 2018
Abstract

In this thesis the paper of Jao and De Feo [3] is analysed and partly summarised. The main contribution to this paper is the mathematical proof of most of the mathematical claims made in the paper about isogenies and isogeny graphs, and a elaborate explanation of the commutativity of the supersingular isogeny Diffie Hellman key exchange.
1. Introduction

As long as written word has existed, people have used encryption to encode messages they did not want anyone else to read. Starting with ‘easy’ encryption like Caesar shift, to modern-day encryption like AES that is currently implemented in our computers. It has been a cat-and-mouse game for centuries, with people trying to invent new inventive ways to hide their messages, and other persons doing their best to intercept and decrypt those. Modern day security systems on computers are based on the principle that anyone who has full knowledge of the cryptographic system used still shall not be able to decode any of the encoded messages. This is also known as Kerckhoff’s principle. A way to achieve this is using so called trapdoor functions. The idea is to find a certain mathematical problem that is hard to solve —like finding the discrete logarithm of a value— unless you have a certain key that makes the problem easier. When computers get more advanced and acquire more computing power, these problems need to get more advanced to still be secure.

The newest threat to computer safety is quantum computing. All over the world academia and companies are trying to build a quantum computer, a machine that will be able to break lots of encryption currently on computers. Since the current quantum computers are just in developing phase, and are by far not able to do any calculations needed to break today’s security, researchers have time to develop new, quantum-resistant cryptosystems. In this thesis one of the suggestions for the new cryptography standard will be discussed, a system using isogenies between supersingular elliptic curves. This paper will highlight the mathematics behind these curves and discuss the working of the actual cryptographic scheme. In the first part a brief introduction to cryptography, mainly focused on public-key cryptography and Diffie-Helmann key exchanges is given. Then there will be a mathematical background on elliptic curves and isogenies. This knowledge will then be applied to explain the Supersingular Isogeny Diffie-Hellman Key Exchange protocol (SIDH), which is a public key protocol. Lastly the security of the SIDH protocol is discussed, explaining why it is a quantum-computer secure cryptographic protocol.
2. Cryptography

There are two kinds of cryptosystems: symmetric and asymmetric. In symmetric systems the same key (the secret key) is used to encrypt and decrypt a message. Data manipulation in symmetric systems is faster than asymmetric systems as they generally use shorter key lengths. Asymmetric systems use a public key to encrypt a message and a private key to decrypt it. Asymmetric systems are also called public key systems. All the information in this chapter can be found in [18], unless stated otherwise. This book is also advised for anyone who wants further reading on the topic.

2.1 Public key cryptography

Public key cryptography, also called asymmetric cryptography, is the name for all the cryptography systems that make use of two different types of keys: public keys, that anyone is allowed to access, and private keys, only known by the owner of the key. Public key cryptographic systems use mathematically hard problems, like discrete logarithm. The goal of the cryptographic system can be either authentication, where the persons with the private key can authenticate who they are, or encryption, where only people knowing the private key can decipher a certain message. The big advantage of public key cryptography is that a secure channel of communication can be established using an ‘unsafe’ communication channel first. This means that even though all the information — the public keys — that two persons send to each other is intercepted by some adversary, the adversary will not be able to decode the message sent. This is because the protocol uses mathematically hard problems that are infeasible to crack to someone knowing only the public keys. The disadvantage of public key cryptography is that it uses a lot of computation to encode and decode a message, and is therefore not really suitable for exchanging large amounts of information.

In symmetric cryptography the same key is used for encryption and decryption. This key needs to be kept absolutely secure by the sender and the receiver. This system can only be established using secure communication, and there is no way of exchanging the key using symmetric cryptography. Its advantage is that it uses less complex algorithms, and therefore to decode or encode a message you need less space and fewer computational operations. This makes it better suited for sending long pieces of information than public key cryptography. A normal way of securing a message, therefore, is sending a message encrypted by a symmetric key, and along with it send the symmetric key, and encrypting that symmetric key using asymmetric key cryptography. This tactic takes advantage of the strong security of asymmetric cryptography for sending information over a secure channel, while a long message can be sent and read without it taking too many calculations. We will now discuss what security means when used in discussions of cryptography, and see how it is decided which cryptographic protocols are going to be implemented in computers.
2.2 Security

Security is the measure for how much effort it costs someone who is not supposed to be able to read the message to crack an encoded message. More exact, security can be expressed in bits. For example, 128-bit security means that it would take a hostile computer -on average- $2^{128}$ elementary operations to ‘crack’ the cryptographic scheme. Cracking a cryptographic protocol means obtaining the key to decipher the message. 128-bit quantum security would mean that it would take a quantum computer $2^{128}$ elementary operations to crack the cryptographic protocol. As computers get more advanced and faster, the need for more bit security rises. This is the reason that since the invention of computers, there has always been a need for updated and more secure cryptographic schemes. Currently, 96-bit security is safe, and 128-bit security is safe in the near future. Safe in this context means that it will take even the fastest computers (such as the ones used by the NSA) years to crack the protocol. How long exactly can never be said, since computers get faster every day, and it also depends on how much computational power can be accessed. How secure a cryptographic scheme is depends also on the length of the key used. For instance, for ordinary Diffie Hellman, a key with length of 300 bits can be factored in a few hours using your own laptop with software freely available online, while currently it is expected that no one can retrieve a key with a length of more than 2000 bits.

Sometimes a cryptographic system is less secure than first thought. As long as a scheme is used, there will be people —both in academia as with other —interests, trying to break the security. This means that they will try to find an attack on the system that needs less operations to break it than the security advertises. When a cryptographic scheme can be attacked in so few operations that it is no longer seen as secure, it is called 'broken.' This does not immediately mean that the system cannot be used anymore, since even the decreased security can still be safe enough.

**Definition 2.2.1** (Big-O-notation). Given a function $g : \mathbb{R} \rightarrow \mathbb{R}$, a function $f : \mathbb{R} \rightarrow \mathbb{R}$ has order $g$ if there exists a constant $c \in \mathbb{R}_{>0}$ and an $x_0 \in \mathbb{R}$ such that for all $x > x_0$, $|f(x)| \leq cg(x)$. This is denoted as $f \in O(g)$. This is called the *big-O-notation*.

In particular, if a function $f$ is a polynomial with several terms, the order of $f$ will be the term with the highest polynomial, omitting any constant. So for instance, if $f = 4x^4 + 6x - 5$, $f \in O(x^4)$.

In computer science, this notation is used to describe the amount of 'time' a polynomial requires. Here 'time' indicates the amount of elementary steps needed for a computer to complete the algorithm. So for instance a protocol can have a key length $n$. and it can be solved in $O(n^4)$ time.

2.3 NIST-competition

Cryptographic schemes that are implemented in computers are not chosen randomly. The authority in this field is the United States National Institute of Standards and Technology (NIST) [14]. When the time is right to start thinking about a new, improved, cryptographic scheme, they start a competition. People from all over the world can then send in their created protocols. In a competition that will last a couple of years, people from all over the world can analyse and try to break these protocols. Also the practical aspects will be reviewed, such as
the amount of time and memory implementing a certain scheme costs, and how easy it can be implemented in computers or other electronic devices. The winner of the competition will become the new standard of that type of cryptographic scheme, although it must be noted that lots of companies keep working with older standards, because the risk of being attacked does not weigh up to the cost of implementing the new system, for them. This way some banks for instance still work with the older triple DES or DES symmetric key system, instead of the more secure AES system.

Currently, NIST is in the middle of the competition to find one or more quantum-resistant public-key cryptographic algorithms. The application deadline was November 30th of 2017. There were 69 admissible applicants, and they are currently all under evaluation. One of the applications (Jao et al.) is a further developed version of the cryptographic scheme in this paper. Up until the writing of this thesis, no attacks that break this scheme are known.

2.4 Diffie-Hellman Key Exchange

The Diffie Hellman key exchange formed in the 1970’s, together with RSA integer factoring, the beginning of public key cryptography. Below is an explanation of the original Diffie Hellman key exchange, and a more advanced version, using elliptic curves. These schemes eventually lead to the creation of the Supersingular Isogeny Diffie Hellman protocol (SIDH). Assume there are two persons, Alice and Bob, who want to create a private session key together. The Diffie Hellman Key Exchange then works as follows. The letters in red will mark private keys, the ones in green public keys.

Alice and Bob agree on a public prime $p$, and a public base number $g$. They then separately choose numbers $a, b$ smaller than $p$, with a primitive root modulo $p$. They then calculate $A = g^a \mod p$ respectively $B = g^b \mod p$. Alice sends $A$ to Bob, Bob sends $B$ to Alice. Alice then calculates $F = B^a = (g^b)^a \mod p$, Bob acts mutatis mutandis. Since exponentiation is a commutative operation, Bob also will obtain $F$, thus making $F$ their shared private key. Retrieving the private keys using only all the public information, requires taking the discrete log of a value. This is for any intruder obtaining all the public information ($p, g, A, B$) with even the fastest computers at the moment an infeasible task, $a, b$ or $F$, given that $p$ is a large prime (at least 1400 digits). An example of the Diffie Hellman key exchange is given below.

Example 2.4.1. Bob and Alice agree on public prime $p = 23$ and base $g = 5$. First Alice calculates her private key $a = 4$, Bob chooses $b = 3$ Alice computes $A = 5^4 \mod 23 = 4$, Bob finds $B = 5^3 \mod 23 = 10$. They exchange $A$ and $B$. Alice then calculates $B^a \mod p = 18$, Bob computes in the same way $A^b \mod p = 18$, their private key.

2.4.1 Elliptic curve Diffie-Hellman

In the beginning of this century another version of the Diffie Hellman key exchange protocol has been introduced, a protocol using elliptic curves. The idea is that an elliptic curve $E$ over field $K$ is known, and a point $P = \langle x, y \rangle$ that generates a subgroup of $E[K]$. Then Alice and Bob both calculate their secret $Q_A = [a]P$ and $Q_B = [b]P$ and share these. Multiplying by $a$, respectively $b$, provides them with their secret shared key $[ab]P$. Retrieving $a$ from only $Q_A$ and the given parameters is a harder problem than solving a discrete logarithm in a cyclic group. This protocol is described below in Figure 2.1 using a commutative diagram.
2.5 Quantum cryptography

A normal computer is basically a large calculator. Operations are done by changing bits from 0 to 1 or the other way around. A quantum computer does not make use of normal bits, but so-called qubits. A qubit can be seen as a unit vector in the complex plane, that is, it is a linear combination of two vectors \(|0\rangle\) and \(|1\rangle\), denoted as

\[
\phi = \alpha|0\rangle + \beta|1\rangle.
\]

However, when measuring a qubit, we will not find it in this superposition state, but we will find the value 0 or 1, with probabilities \(|\alpha^2|\) and \(|\beta^2|\). Since probabilities sum to one, this gives \(|\alpha^2| + |\beta^2| = 1\).

The principle of quantum computing is that a quantum computer will not do sequential computations, but rather compare states and then ‘fall’ in the correct state when measured. This way, if only one variable has to be measured, this can theoretically be measured in one go, instead of \(n\) attempts. For some applications, even when the quantum computer needs lots of preparation time, this will speed up the process of measurement up to exponentially faster. This is for instance the case for a big part of modern day cryptographic protocols.

An example of a quantum algorithm is the Shor algorithm. With this algorithm a quantum computer can solve any discrete logarithm problem or integer factorisation problem in polynomial time, meaning that they can be efficiently solved using a quantum computer. This makes the Diffie Hellman and the Elliptic curve Diffie Hellman schemes vulnerable to quantum attacks. As far as we know so far, the Shor algorithm cannot be applied on the supersingular Isogeny Diffie Hellman scheme, since this scheme uses both different isogenies and different elliptic curves, making it a problem in not one but two variables, and also since it has no underlying abelian group structure.

Currently, quantum computers are only in a developing phase, and are not even close to be as useful as even normal computers are nowadays. It is not even completely clear if it will be possible to build quantum computers that can actually execute quantum algorithms in a way that they are faster than normal computer algorithms. The European Union expects the first fully functioning quantum computers to be working around 2035 [4]. This gives the world...
lots of time to develop post-quantum cryptography, making sure that when these computers arrive, they will not break all cryptography, as is a doomsday scenario used in modern day fiction. Also, until all people will have access to quantum computers, current symmetric-key cryptography is quantum computer resistant. So only public key cryptography would require an update when quantum computers arrive. The NIST-competition therefore only focuses on public key cryptographic protocols.

In the rest of the paper we will discuss the mathematical aspects of the cryptographic scheme of the above mentioned paper of Jao and De Feo. The more technical aspects like security and implementation will not be (deeply) discussed in this thesis, for more information on those topics see [3] and [24].
3. Elliptic curves and isogenies

This section contains basic information on elliptic curves and isogenies. For further reading on the topic, see the book of J. Silverman [17]. All definitions and theorems stated can be found there. A few definitions mentioned in this section will be defined exactly in the next chapter. This is to make this chapter more easy to read.

3.1 The basics

Definition 3.1.1. An elliptic curve is defined as a curve of genus one with a base point \( O \).

Since this definition does not make a lot of sense yet, we will first define a more easy to work with definition. Later in this thesis we will come back to the formal definition.

Definition 3.1.2. Given a field \( K \), the projective n-space \( \mathbb{P}^n(K) \) over \( K \) is defined as the set of \( n+1 \)-tuples \((x_0,...,x_n)\) \( \in \mathbb{A}^{n+1} \), with the \( x_i \)'s not all simultaneously zero. We define an equivalence relation

\[
(x_0,...,x_n) \sim (y_0,...,y_n)
\]

if there exists a \( \lambda \in \bar{K}^* \) such that, for all \( i \), \( x_i = \lambda y_i \). We will write \((x_0 : ... : x_n)\) for the equivalence class of \((x_0,...,x_n)\).

Definition 3.1.3. An elliptic curve over a field \( K \) with \( \text{Char}(K) \neq 2, 3 \) is defined as the set of solutions in the projective plane \( \mathbb{P}^2(K) \) of the Weierstrass equation

\[
Y^2Z = X^3 + aXZ^2 + bZ^3. \tag{3.1}
\]

The curve should also be non-singular (See Definitions 3.1.4 or 4.1.11 for the definition of non-singularity).

The only \( K \)-rational point on the equation with \( Z = 0 \) is \((0 : 1 : 0)\). We will call this point the point at infinity \( O \).

All the other \( K \)-rational points have \( ZZ \neq 0 \) so we can write the Weierstrass equation in affine form: Define \( x = X/Z \) and \( y = Y/Z \) and write (3.1) as

\[
y^2 = x^3 + ax + b,
\]

together with the point of infinity \( O \).

The points on an elliptic curve form a group. We will now define the group law on the curve. Given an elliptic curve \( E \), and two points \( P = (x_1,y_1) \) and \( Q = (x_2,y_2) \), define \( P + Q \) as follows:
• $P + \mathcal{O} = \mathcal{O} + P = P$,
• if $x_1 = x_2$ and $y_1 = -y_2$ then $P + Q = \mathcal{O}$,
• for $x_1 \neq x_2$ define $\lambda = \frac{y_2 - y_1}{x_2 - x_1}$,
• for $x_1 = x_2$, $y_1 = y_2$ and $y_1 \neq 0$, define $\lambda = \frac{3x_1^2 + a}{2y_1}$,
• for $x_1 = x_2$, $y_1 = y_2$ and $y_1 = 0$, $P + Q = \mathcal{O}$.

Then the point $P + Q = (x_3, y_3)$ is given by
\[ x_3 = \lambda^2 - x_1 - x_2 \]
\[ y_3 = \lambda x_3 - y_1 + \lambda x_1. \]

A graphic representation of the group law of an elliptic curve defined over $\mathbb{R}$ is shown in Figure 3.1.

![Figure 3.1: Addition of two points on an elliptic curve](image)

We now give two important definitions regarding elliptic curves, the discriminant and the \textit{j-invariant}.

\textbf{Definition 3.1.4.} The \textit{discriminant} $\Delta$ of an elliptic curve $E$ is given by $\Delta = -16(4a^3 + 27b^2)$. A curve is non-singular when $\Delta \neq 0$.

The \textit{j-invariant} of an elliptic curve $E : y^2 = x^3 + ax + b$ defined over a field $K$ is defined as
\[ j(E) = j(a, b) = \frac{1728}{4a^3 + 27b^2}. \]
Two curves have the same $j$-invariant if and only if they are isomorphic over the algebraic closure $\bar{K}$ of $K$.

3.2 Isogenies

In this section isogenies are discussed, a special kind of morphism between elliptic curves. In the SIDH protocol, two persons will create their unique isogenies. Isogenies of a specific kind, so called isogenies of smooth degree between supersingular curves, turn out to be especially suited for this task.

**Definition 3.2.1.** Given two elliptic curves $E_1$ and $E_2$, defined over the same field $K$, an *isogeny* $\phi : E_1 \rightarrow E_2$ is a surjective group morphism. An isogeny has a finite kernel, and also maps $O_1$ to $O_2$. We say that two elliptic curves are *isogenous over $K$* if there exists an isogeny between them.

An isogeny from a curve to itself is called an *endomorphism*. The most important example hereof is multiplication by $m$:

$$[m] : P \mapsto [m]P$$

The kernel of this endomorphism is the $m$-torsion group $E[m]$:

**Definition 3.2.2.** The *$m$-torsion subgroup* of an elliptic curve $E$, denoted $E[m]$ is the set of points of $E$ of order $m$:

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

**Theorem 3.2.3.** Given $E[\ell]$, the $\ell$-torsion group of an elliptic curve $E$ defined over $\overline{\mathbb{F}}_q$, take $\text{char} \, \overline{\mathbb{F}}_q = p$. For $p \nmid \ell$, $E[\ell] \cong \mathbb{Z}/\ell\mathbb{Z} \oplus \mathbb{Z}/\ell\mathbb{Z}$. In particular, this means that the torsion subgroup can be represented by two points on the elliptic curve.

*Proof.* See [17, III.6.4].

Taking all the endomorphisms of a group, together with multiplication-by-0, we get a ring under addition and composition, the *endomorphism ring*. This ring will be denoted by $\text{End}(E)$.

**Definition 3.2.4.** Two isogenies

$$\phi_1, \phi_2 : E \rightarrow E'$$

are called *isomorphic* if there exists an invertible $\alpha \in \text{End}(E')$ such that

$$\phi_2 = \alpha \phi_1.$$

We will now look further into the endomorphism ring of elliptic curves. We will show that $\text{End}(E)$ can have only one out of three forms. For that, we first need some definitions.

**Definition 3.2.5.** An *algebra $A$ over a field $K$*, usually just called an *algebra*, is a set $A$ which is both a ring and a vector space over a field $K$, such that $(\lambda a)b = a(\lambda b) = \lambda(ab)$ for all $\lambda \in K$ and $a, b \in A$. 

11
Definition 3.2.6. A quaternion algebra is an algebra of the form
\[ A = \mathbb{Q} + \mathbb{Q}\alpha + \mathbb{Q}\beta + \mathbb{Q}\alpha\beta \]
with multiplication rules \( \alpha^2, \beta^2 \in \mathbb{Q}, \alpha^2, \beta^2 < 0, \beta\alpha = -\alpha\beta. \)

Definition 3.2.7. An order of a ring \( A \) is a subring \( O \) such that \( A \) is a ring which is a finite-dimensional algebra over \( \mathbb{Q} \), and additively, \( O \) is a free abelian group generated by a basis for \( A \) over \( \mathbb{Q} \).

Theorem 3.2.8. Let \( E \) be an elliptic curve defined over a field \( k \) of characteristic \( p \). The ring \( \text{End}(E) \) is isomorphic to one of the following:

- \( \mathbb{Z} \), if and only if \( p = 0 \);
- An order in an imaginary quadratic field;
- An order of a quaternion algebra.

If \( \text{End}(E) \) of an elliptic curve over a finite field \( K \) is equal to a quaternion algebra, \( E \) is called supersingular. Otherwise, \( E \) is called ordinary.

Proof. See [17, III.9.4].

Supersingular curves form the basic of the SIDH protocol. In normal elliptic curve cryptography they are not used, since the discrete logarithm problem is relatively easy for these curves. Throughout the next chapters it will be shown that for isogeny based cryptography they are more useful than ordinary elliptic curves.

One of the main characteristics of supersingular elliptic curves is that their endomorphism ring is not commutative. This makes it harder to create a commutative diagram like other Diffie Hellman key exchanges have. But using some smart techniques, it will still be possible to create one.

For the SIDH protocol, the next theorems are essential. The proofs will be given in the next chapter, since a lot more mathematical background is required to give the proofs.

Definition 3.2.9. Let \( \phi : E_1 \rightarrow E_2 \) be a map between elliptic curves over field \( K \). The function \( \phi \) is separable, (insaperable, purely inseparable) if the function field extension \( K(E_1) = \phi*K(E_2) \) is separable as a field extension.

Definition 3.2.10. Let \( \phi : E_1 \rightarrow E_2 \) be a map between elliptic curves over field \( K \). The degree of \( \phi \) is defined as 0 if \( \phi \) is constant, and for all other maps \( \phi \) the degree is given by
\[ \deg(\phi) = [K(E_1) : \phi*K(E_2)]. \]

Theorem 3.2.11. If \( \phi \) is separable, \( \deg(\phi) = \# \ker(\phi) \).

Later in the text it is shown that all the isogenies we will work with are separable, so that we can take \( \deg(\phi) = \# \ker(\phi) \) as definition for the degree of \( \phi \). The next theorem provides us with a very useful tool, that we can determine every separable isogeny by its kernel.
**Theorem 3.2.12.** Let $E$ be an elliptic curve, and let $A$ be a finite subgroup of $E$. Then there is a unique elliptic curve $E'$ and a separable isogeny
\[ \phi : E \rightarrow E' \]
such that
\[ \ker(\phi) = A. \]
$E'$ can thus also be denoted as $E' = E/A$. 
4. Mathematical foundation

The goal of this chapter is to provide a mathematical background needed to prove the last two theorems of the previous chapter, and also to give an overview of the mathematics behind elliptic curves. The theorems are crucial to proving the commutativity of the SIDH protocol. While the theorems appear simple the proofs use lots of interesting features of algebraic geometry. This chapter will treat a lot of mathematics very briefly. For more background information, it is again advised to read [17]. The book is taken as reference for all the theorems and definitions in this chapter, unless stated otherwise.

Take from now on $K$ to be a perfect field, that is, every algebraic extension of $K$ is separable. This means that every irreducible polynomial over the field $K$ will have distinct roots. (like for instance if $K$ is a finite field, or its degree is 0). Define $\bar{K}$ to be the algebraic closure of $K$.

4.1 Defining a curve

An elliptic curve is a specific kind of curve. But while we can imagine what a curve is in a vector space of $\mathbb{R}$, we do not yet have an exact definition of what a curve is over any field. This first section will give this formal definition.

Definition 4.1.1. The affine $n$-space (over $K$) is the set of $n$-tuples

$$A^n = A^n(\bar{K}) = \{ P = (x_1, x_2, ..., x_n) : x_i \in \bar{K} \}.$$  

Definition 4.1.2. The set of $K$-rational points is defined as the set of $n$-tuples

$$A^n(K) = \{ P = (x_1, ..., x_n) : x_i \in K \}.$$  

Definition 4.1.3. Take $I$ to be an ideal in the polynomial ring $\bar{K}[X] = \bar{K}[X_1, ..., X_n]$. An algebraic set is defined as follows:

$$V_I = \{ P = (x_1, ..., x_n) \in A^n : \forall f \in I f(P) = 0 \}.$$  

Definition 4.1.4. If $V$ is an algebraic set, the ideal of $V$, defined in $\bar{K}[X]$, is given by

$$I(V) = \{ f \in \bar{K}[X] : \forall P \in V f(P) = 0 \}.$$  

We say an algebraic set is defined over $K$ if its ideal $I(V)$ can be generated by polynomials in $K[X]$. 

14
Definition 4.1.5. When $V$ is defined over $K$, the set of $K$-rational points of an algebraic set $V$ is the set $V(K) = V \cap \mathbb{A}^n(K)$.

Definition 4.1.6. An algebraic set $V$ is called an (affine) variety if $I(V)$ is a prime ideal in $\overline{K}[X]$. We write $V/K$ for a variety $V$ defined over $K$. The affine coordinate ring of $V/K$ is defined by

$$K[V] = \overline{K}[X]/I(V/K).$$

Its quotient field is denoted by $K(V)$ and is called the function field of $V/K$.

The transcendence degree of a field extension $L/K$ is defined as the largest cardinality of an algebraically independent subset of $L$ over $K$. Concretely, this means for a field extension $L = K(X_1,...,X_n)$ that the transcendence degree of $L/K$ equals $n$.

The dimension of a variety $V$, $\dim(V)$, is the transcendence degree of $\overline{K}(V)$ over $\overline{K}$.

Definition 4.1.7. A polynomial $f \in \overline{K}[X] = \overline{K}[X_0,...,X_n]$ is homogeneous of degree $d$ if

$$\forall \lambda \in \overline{K} \ f(\lambda X_0,...,\lambda X_n) = \lambda^d f(x_0,...,x_n).$$

An ideal $I \in \overline{K}[X]$ is homogeneous if it is generated by homogeneous polynomials.

Definition 4.1.8. A projective algebraic set is any set of the form $V_I$ for a homogeneous ideal $I$. If $V$ is a projective algebraic set, the homogeneous ideal of $V$ is the ideal of $\overline{K}$ generated by

$$\{f \in \overline{K} : f \text{ is homogeneous and } \forall P \in V \ f(P) = 0\}.$$

Definition 4.1.9. A projective algebraic set is called a (algebraic) variety if its homogeneous ideal $I(V)$ is a prime ideal in $\overline{K}[X]$.

With all these new definitions it is finally possible to give a definition of what a curve is: A curve is a projective variety of dimension one. A curve $C$ specified over a field $K$ is denoted as $C/K$.

For each point $P \in V$, define a maximal ideal $M_P$ of $\overline{K}[V]$ as follows:

$$M_P = \{f \in \overline{K}[V] : f(P) = 0\}.$$

Given a projective variety $V$ and a point $P \in V$, the local ring of $V$ at $P$, denoted as $\overline{K}[V]_P$, is defined as the localisation of $\overline{K}[V]$ at $M_P$:

$$\overline{K}[V]_P = \{F \in \overline{K}(V) : F = f/g \text{ for some } f, g \in \overline{K}[V] \text{ with } g(P) \neq 0\}.$$ 

A function $F \in \overline{K}(V)$ is called regular or defined at $P$ if it is in $\overline{K}[V]_P$.

Definition 4.1.10. Let $V_1$ and $V_2 \subset \mathbb{P}$ be projective varieties. A rational map from $V_1$ to $V_2$ is a map

$$\phi : V_1 \to V_2, \quad \phi = [f_0,...,f_n].$$
The functions \( f_0, \ldots, f_n \in \bar{K}(V_1) \) have the property that for every point \( P \in V_1 \) — provided that all \( f_i \) are regular at \( P \),

\[
\phi(P) = [f_0(P), \ldots, f_n(P)] \in V_2.
\]

A rational map \( \phi = [f_0, \ldots, f_n] : V_1 \to V_2 \) is regular if there is a function \( g \in \bar{K}(V_1) \) such that

- each \( gf_i \) is regular at \( P \);
- there is some \( i \) for which \( (gf_i)(P) \neq 0 \).

A rational map that is regular at every point is called a morphism.

**Definition 4.1.11.** Let \( V \) be a variety, \( P \in V \), and \( f_1, \ldots, f_m \in \bar{K}[X] \) generators for \( I(V) \). Then \( V \) is nonsingular or smooth at \( P \) if the matrix of partial differentials

\[
\left( \frac{\partial f_i}{\partial X_j}(P) \right)_{0 \leq i \leq m, 0 \leq j \leq n}
\]

has rank \( n - \dim(V) \). We say \( V \) is smooth if \( V \) is nonsingular at every point.

**Remark 4.1.12.** In Definition 3.1.4 it is stated that for elliptic curves, nonsingularity means having a discriminant that does not equal zero. In [17, Proposition II.1.4a.i] it is proven that these definitions are indeed the same.

### 4.2 Ramification and degree

**Definition 4.2.1.** Let \( C \) be a curve and \( P \in C \) a smooth point. Then the valuation on \( \bar{K}[C]_P \) is given by

\[
\text{ord}_P : \bar{K}[C]_P \to \{0, 1, 2, \ldots\} \cup \{\infty\},
\]

\[
\text{ord}_P(f) = \sup\{d \in \mathbb{Z} : f \in M^d_P\}.
\]

We can use \( \text{ord}_P(f/g) = \text{ord}_P(f) - \text{ord}_P(g) \) to extend \( \text{ord}_P \) to \( \bar{K}(C) \). This gives

\[
\text{ord}_P : \bar{K}(C) \to \mathbb{Z} \cup \{\infty\}.
\]

For a function \( f \in \bar{K}(C) \), \( \text{ord}_P(f) \) is called the order of \( f \) at \( P \). A function \( t \in \bar{K}(C) \) with \( \text{ord}_P(t) = 1 \), meaning \( t \) generates \( M_p \), is called a uniformizer for \( C \) at \( P \).

**Definition 4.2.2.** Let \( C_1/K \) and \( C_2/K \) be curves and take \( \phi : C_1 \to C_2 \) a nonconstant rational map over \( K \). Then composition with \( \phi \) induces an injection of function fields, fixing \( K \),

\[
\phi^* : K(C_2) \to K(C_1), \quad \phi^* f = f \circ \phi.
\]

**Theorem 4.2.3.** Let \( C_1 \) and \( C_2 \) be curves over a field \( K \). Let \( \phi : C_1 \to C_2 \) be a nonconstant map defined over \( K \). Then \( K(C_1) \) is a finite extension of \( \phi^*(K(C_2)) \) and, given \( K \subset K(C_1) \) a subfield of finite index containing \( K \), there exists a smooth curve \( C' \) over \( K \), unique up to \( K \)-isomorphism, and a nonconstant map \( \phi : C_1 \to C' \) defined over \( K \) such that \( \phi^* K(C') = K \).

**Proof.** See [6, II.6.8] for the first part, and for the second part see [6, I.6.12].
The next definition was already given at the end of the previous chapter. Here it is given again, since we finally have all the prior knowledge to fully understand its meaning.

**Definition 4.2.4.** Take \( \phi \) as in Definition 4.2.2 above, defined over a field \( K \). If \( \phi \) is constant, the degree of \( \phi \) is defined as 0. Otherwise, \( \phi \) is called a finite map and the degree of \( \phi \) is defined as

\[
\text{deg}(\phi) = [K(C_1) : \phi^*(K(C_2))].
\]

\( \phi \) is called separable, inseparable or purely inseparable if the field extension \( K(C_1)/\phi^*(K(C_2)) \) is separable, respectively inseparable. The degrees of the inseparable and the separable part of a map \( \phi \) are denoted by \( \text{deg}_i(\phi) \), respectively \( \text{deg}_s(\phi) \).

**Definition 4.2.5.** The norm \( N_{K_1/K_2} \) of a field \( K_1 \) into a field \( K_2 \), where \( K_1 \) is a finite field extension of \( K_2 \), is a mapping that sends an element \( \alpha \in K_1 \) to the element \( N_{K_1/K_2}(\alpha) \), that is the determinant of the matrix of the \( K_2 \) -linear mapping \( K_1 \to K_1 \) mapping \( x \in K_1 \) to \( \alpha x \). \( N_{K_1/K_2}(\alpha) \) is called the norm of \( \alpha \) \[10\].

**Definition 4.2.6.** Given the maps of curves \( \phi \) and \( \phi^* \) defined as in Definition 4.2.2, we can define a map \( \phi_\ast \) in the other direction:

\[
\phi_\ast : K(C_1) \to K(C_2),
\]

defined by

\[
\phi_\ast = (\phi^*)^{-1} \circ N_{K(C_1)/\phi^*(K(C_2))}.
\]

This is well defined since according to theorem 4.2.3 \( K(C_1) \) is a finite extension of \( \phi^*(K(C_2)) \).

**Definition 4.2.7.** Take \( \phi : C_1 \to C_2 \) a nonconstant map of smooth curves. Take a point \( P \in C_1 \) The ramification index of \( \phi \) at \( P \), denoted \( e_{\phi}(P) \) equals \( \text{ord}_P(\phi^*t_{\phi(P)}) \), with \( t_{\phi(P)} \in K(C_2) \) a uniformizer at \( \phi(P) \). \( \phi \) is called unramified at \( P \) if the ramification index at \( P \) is 1. If this is the case for all points of \( C_1 \), \( \phi \) is called unramified.

**Theorem 4.2.8.** Let \( \phi : C_1 \to C_2 \) be a nonconstant map of smooth curves. Then for every \( Q \in C_2 \)

\[
\sum_{P \in \phi^{-1}(Q)} e_{\phi} = \text{deg}(\phi),
\]

and for all but finitely many \( Q \in C_2 \)

\[
\#\phi^{-1}(Q) = \text{deg}_s(\phi).
\]

**Proof.** See [6, II.6.8 and II.6.9]. \( \square \)

4.3 Genus

As stated in the beginning of the previous chapter, an elliptic curve is a curve with genus one. In this section it will be explained what the definition of genus is.

**Definition 4.3.1.** The divisor group of a curve \( C \), denoted \( \text{Div}(C) \), is the free abelian group generated by the points of \( C \). Thus a divisor \( D \in \text{Div}(C) \) equals

\[
D = \sum_{P \in C} n_P(P),
\]
with \( n_P \in \mathbb{Z} \) and \( n_P = 0 \) for almost all \( P \in C \).

The **degree of** \( D \) is defined by

\[
\deg(D) = \sum_{P \in P} n_P.
\]

The **divisors of degree 0** form a subgroup of \( \text{Div}(C) \), denoted by

\[
\text{Div}^0(C) = \{ D \in \text{Div}(C) : \deg(D) = 0 \}.
\]

Assume the curve \( C \) is smooth, and let \( f \in \bar{K}(C)^* \). Then we define \( \text{div}(f) \) as follows:

\[
\text{div}(f) = \sum_{P \in C} \text{ord}_P(f).
\]

A divisor \( D \in \text{Div}(C) \) is **principal** if it has the form \( D = \text{div}(f) \) for some \( f \in \bar{K}(C)^* \).

The **divisor class group**, or Picard group, denoted \( \text{Pic}(C) \), is the quotient of \( \text{Div}(C) \) by its subgroup of principal divisors. Similarly we define the **degree-0 part of the divisor class group** of \( C \) to be the quotient of \( \text{Div}^0(C) \) by the subgroup of principal divisors, and denote this group by \( \text{Pic}^0(C) \).

It is also possible to define differentials of curves, or more precise, the vector space of differential forms on a curve.

**Definition 4.3.2.** Let \( C \) be a curve. The **space of differential forms** on \( C \), denoted \( \Omega_C \), is the \( \bar{K} \)-vector space generated by symbols of the form \( dx \) for \( x \in \bar{K}(C) \) with the relations

1. \( \forall \ x, y \in \bar{K}(C) \quad d(x + y) = dx + dy; \)
2. \( \forall \ x, y \in \bar{K}(C) \quad d(xy) = xdy + ydx; \)
3. \( \forall \ a \in \bar{K} \quad da = 0. \)

**Theorem 4.3.3.** Let \( C \) be a curve, \( P \in C \) and \( t \in \bar{K}(C) \) be a uniformizer at \( P \). Then for every for every \( \omega \in \Omega_C \) there exists a unique function \( g \in K(C) \), depending on \( \omega \) satisfying \( \omega = gdt \). Also, taking \( \omega \neq 0 \), the quantity \( \text{ord}_P(\omega/dt) \) depends only on \( \omega \) and \( P \) and is independent of the choice of the uniformizer \( t \). \( \text{ord}_P(\omega/dt) \) is called the **order of** \( \omega \) **at** \( P \) and is denoted by \( \text{ord}_P(\omega) \).

**Proof.** See [17, II.4.3].

**Definition 4.3.4.** Take \( \omega \in \Omega_C \). The **divisor associated to** \( \omega \) is

\[
\text{div}(\omega) = \sum_{P \in C} \text{ord}_P(\omega) \in \text{Div}(C).
\]

The **canonical divisor class on** \( C \) is the image in \( \text{Pic}(C) \) of \( \text{div}(\omega) \) for any nonzero differential \( \omega \in \Omega_C \). Any divisor in this class is called a **canonical divisor**.

**Definition 4.3.5.** A divisor \( D = \sum n_P(P) \) is **positive**, denoted by \( D \geq 0 \), if \( n_P \geq 0 \) for all \( P \in C \). We write

\[
D_1 \geq D_2
\]

to indicate that \( D_1 - D_2 \) is positive for two divisors \( D_1, D_2 \in \text{Div}(C) \).
Definition 4.3.6. Take $D \in \text{Div}(C)$. We define the set of functions

$$\mathcal{L}(D) = \{ f \in \bar{K}(C)^* : \text{div}(f) \geq -D \} \cup \{0\}.$$  

The set $\mathcal{L}(D)$ is a finite-dimensional $\bar{K}$-vector space [17, Theorem 5.12], and its dimension is denoted by

$$\ell(D) = \dim_{\bar{K}} \mathcal{L}(D).$$

Theorem 4.3.7 (Riemann-Roch). Given a smooth curve $C$ and a canonical divisor $K_C$ on $C$, there is an integer $g \geq 0$, called the genus of $C$, such that for every divisor $D \in \text{Div}(C)$

$$\ell(D) - \ell(K_C - D) = \deg(D) - g + 1.$$  

Proof. See [6, IV §1]. \hfill \square

Corollary 4.3.8. $\deg K_C = 2g - 2$.

Proof. Applying Theorem 4.3.7 with $D = 0$, and knowing that $\ell(0) = 1$, gives us that $\ell(K_C) = g$. Then applying Theorem 4.3.7 again, this time with $D = K_C$, gives us the proof needed. \hfill \square

The Riemann-Roch theorem gives us a really important definition, the definition of genus. Now we finally have the full definition of an elliptic curve. It can be shown that this definition of an elliptic curve and the one given in 3.1.3 are in fact equivalent, as is shown in [17].

Theorem 4.3.9 (Hurwitz). Let $\phi : C_1 \longrightarrow C_2$ be a nonconstant separable map of smooth curves with genera $g_1$ respectively $g_2$. Then

$$2g_1 - 2 \geq (\deg(\phi))(2g_2 - 2) + \sum_{P \in C_1} (e_{\phi}(P) - 1),$$

and equality only holds if and only if one of the following cases holds:

1. $\text{char}(K) = 0$;
2. $\text{char}(K) = p > 0$ and $p$ does not divide $e_{\phi}(P)$ for all $P \in C_1$.

Proof. See [17] II.5.9. \hfill \square

Remark 4.3.10. Note that if $\phi$ is unramified, meaning that $\forall P \in C_1 \ e_{\phi}(P) = 1$, then the equality always holds and also the sum in the right part of the equation vanishes, leaving the formula to be:

$$2g_1 - 2 = (\deg(\phi))(2g_2 - 2).$$

Definition 4.3.11. Take $E$ to be an elliptic curve over a field $K$, and take a point $Q \in E$. We define a translation-by-$Q$ map as follows:

$$\tau_Q : E \longrightarrow E, \quad P \mapsto P + Q.$$

Clearly, the map $\tau_Q$ is an isomorphism of varieties. We can now also define the map $\tau_Q^*$ that $\tau_Q$ induces on the function field $\bar{K}(E_1)$, as in 4.2.2. Since $P + Q$ is different for every $P \in E$ it follows that $\tau_Q^*$ is an isomorphism, and therefore an automorphism.
Theorem 4.3.12. Let \( \phi : E_1 \rightarrow E_2 \) be an isogeny, and take \( P, Q \in E_1 \). Then
\[
\phi(P + Q) = \phi(P) + \phi(Q).
\]

Proof. The case where \( \phi(P) = 0 \) for all \( P \in E_1 \) is clearly true. Otherwise, \( \phi \) is a finite map. It is stated in [17, II.3.7] that \( \phi \) then induces a homomorphism
\[
\phi_* : \text{Pic}^0(E_1) \rightarrow \text{Pic}^0(E_2)
\]
defined by
\[
\phi_*(\text{class of } \sum n_i(P_i)) = \text{class of } \sum n_i(\phi(P_i)).
\]
Also, it is proven in [17, II.3.4] that there exist group isomorphisms
\[
\kappa_i : E_i \rightarrow \text{Pic}^0(E_1)
\]
\[
P \mapsto \text{class of } (P) - (O).
\]
And since \( \phi(O) = 0 \) because \( \phi \) is an isogeny, this gives us the following commutative diagram
\[
\begin{array}{ccc}
E_1 & \overset{\cong}{\rightarrow} & \text{Pic}^0(E_1) \\
\downarrow{\phi} & & \downarrow{\phi_*} \\
E_2 & \overset{\cong}{\rightarrow} & \text{Pic}^0(E_2)
\end{array}
\]
Since \( \kappa_1, \kappa_2 \) and \( \phi_* \) are all group homomorphisms and \( \kappa_2 \) is an isomorphism and therefore injective, it follows that \( \phi \) is also a homomorphism.

4.3.1 The Frobenius map

Let \( K \) be a field, and \( \text{char}(K) = p > 0 \). Define \( q = p^r \). For a polynomial \( f \in K[X] \), define \( f^{(q)} \) as the polynomial obtained by raising each coefficient of \( f \) to the \( q^{th} \) power. For any curve \( C \) over \( K \) we can define a new curve \( C^{(q)} \), defined as the curve whose homogeneous ideal is given by
\[
I(C^{(q)}) = \text{ideal generated by } \{ f^{(q)} : f \in I(C) \}.
\]
There is a natural map \( \phi : C \rightarrow C^{(q)} \), called the \( q^{th} \)-power Frobenius morphism, defined by
\[
\phi[x_0, ..., x_n] = [x_0^q, ..., x_n^q],
\]
for the proof that \( \phi \) maps \( C \) to \( C^{(q)} \), see [17, II.2].

Theorem 4.3.13. Every map \( \psi : C_1 \rightarrow C_2 \) of smooth curves over a field \( K \) with \( \text{char}(K) > 0 \) factors as
\[
C_1 \overset{\phi}{\rightarrow} C_1^{(q)} \overset{\lambda}{\rightarrow} C_2,
\]
where \( q = \text{deg}_1(\psi) \), the map \( \phi \) is the \( q^{th} \)-power Frobenius morphism, and the map \( \lambda \) is separable.

Proof. See [17, II.2.12].
In particular this means that every map between two curves can be factored into a separable map and a Frobenius morphism. Since the Frobenius morphism is an endomorphism, we can represent the Frobenius map using a matrix. The trace of this matrix is called the Frobenius trace.

**Theorem 4.3.14** (Hasse). Take $E$ an elliptic curve over a field $\mathbb{F}_q$. Then

\[
\#E(\mathbb{F}_q) = q + 1 - t, \tag{4.1}
\]

with $t$ the Frobenius trace and $|t| \leq 2\sqrt{q}$.

*Proof.* See [19, 8.1].

**Theorem 4.3.15.** An elliptic curve $E \backslash \mathbb{F}_q$, is supersingular if and only if the Frobenius trace equals 0 mod $p$.

*Proof.* See [19, 14.2].

**Corollary 4.3.16.** An elliptic curve $E \backslash \mathbb{F}_p$, with $p$ a prime and $p > 3$, is supersingular if and only if $\#E(\mathbb{F}_q) = p + 1$.

*Proof.* By Theorem 4.3.14 the Frobenius trace is smaller than $2\sqrt{(p)}$, so for $p > 3$ this says that the Frobenius trace of a supersingular elliptic curve equals 0. Thus Equation 4.1 for these supersingular elliptic curves becomes

\[
\#E(\mathbb{F}) = q + 1.
\]

### 4.4 Kernel and subgroups

In this section the two theorems that were stated without proof in the previous chapter are finally proven.

**Theorem 4.4.1.** Given a nonzero isogeny $\phi : E_1 \rightarrow E_2$. Assume $\phi$ is separable. Then

\[
\# \ker(\phi) = \deg(\phi).
\]

*Proof.* From theorem [4.2.8] we know that

\[
\#\phi^{-1}(Q) = \deg_s(\phi) \quad \text{for all but finitely many } Q \in E_2.
\]

But since $\phi$ is surjective, we can take for any $Q, Q' \in E_2$ an $R \in E_1$ such that $\phi(R) = Q' - Q$, and given that $\phi$ is a homomorphism gives us a one-to-one correspondence

\[
\phi^{-1}(Q) \rightarrow \phi^{-1}(Q') \quad P \mapsto P + R.
\]

This gives that the degree of $\phi$ is the same for all $Q \in E_2$, so

\[
\#\phi^{-1}(Q) = \deg_s(\phi) \quad \text{for all } Q \in E_2.
\]

And since we required that $\phi$ is separable, we get that

\[
\deg_s(\phi) = \deg(\phi), \quad \text{giving for } Q = O:\n\deg(\phi) = \#\phi^{-1}(O) = \# \ker(\phi).
\]

\[\Box\]
Theorem 4.4.2. Given an elliptic curve $E$ and a finite subgroup $A$ of $E$, there is a unique elliptic curve $E'$ and a unique separable isogeny $\phi : E \to E'$ such that $\ker(\phi) = A$.

Proof. As we’ve seen, every point $Q \in A$, gives rise to an automorphism $\tau_Q^* \in K(E)$. Take $\tilde{K}(E)^A$ to be the subfield of $\tilde{K}(E)$ defined as all the points of $\tilde{K}(E)$ that remain invariant under applying the automorphisms $\tau_Q^*$, defined in Definition 4.3.11. Then $\tilde{K}(E)$ is a Galois extension of $\tilde{K}(E)^A$, with Galois group isomorphic to $A \cong \mathbb{Z}/2\mathbb{Z}$. The field $\tilde{K}(E)^A$ has transcendence degree one over $\tilde{K}$, since $A$, and therefore the Galois extension, is finite. Thus there exist a unique smooth curve $C$ over $\tilde{K}$ and a finite morphism

$$\phi : E \to C$$

satisfying $\phi^* \tilde{K}(C) = \tilde{K}(E)^A$.

This is a separable morphism since the extension

$$\tilde{K}(E)^A/\phi^* \tilde{K}(C) = \tilde{K}(E)^A/\tilde{K}(E)^A$$

is an algebraic extension, and because $\tilde{K}$ is a perfect field this extension is separable.

To prove the theorem we need to show that $\phi$ is the isogeny we are looking for, and that $C$ is an elliptic curve.

This will be done in two steps. The first step is to show that $\phi$ is unramified. If that is proven we can apply the Hurwitz genus formula as given in remark 4.3.10

$$2\text{genus}(E) - 2 = \deg(\phi)(\text{genus}(C) - 2).$$

This then gives that $C$ has genus one, and therefore is an elliptic curve. Since $\phi$ has a finite kernel, is an isogeny and the theorem will be proven.

So now we need only to prove that $\phi$ is unramified.

Take a point $P \in E$ and $Q \in A$. Then for every function $f \in \tilde{K}(C)$ we have

$$f(\phi(P + Q)) = f(\phi(\tau_Q(P))) = (\phi \circ \tau_Q)^* f(P) = \phi^* f(P) = f(\phi(P)).$$

The first equality follow from the definition of $\tau_Q$. The second and last equality follow from the definition of the map between function field as given in Definition 4.2.2. The third equality follows the fact that $Q \in A = \ker(\phi)$, so $\phi(P + Q) = \phi(P) + \phi(Q) = \phi(P)$.

Since $f$ was chosen randomly, it follows that $\phi(P + Q) = \phi(P)$. Then if we take a point $T \in C$, and choose a point $P \in E$ such that $\phi(P) = T$, we get

$$\phi^{-1}(T) \supset \{P + Q : Q \in A\}.$$  

Also, we know from theorem 4.2.8 that

$$\#\phi^{-1}(T) \leq \deg(\phi) = \#A.$$  \hspace{1cm} (4.2)

Theorem 4.2.8 also gives that Equation 4.2 is an equality if and only if $\phi$ is unramified at all points in the inverse image $\phi^{-1}(T)$. And since the points $P + Q$ are distinct as $Q$ ranges over the elements of $A$, we get that $\#\{P + Q : Q \in A\} = \#A$, so indeed equation 4.2 is an equality and therefore is $\phi$ unramified at $\phi^{-1}(T)$. But since $T$ was arbitrary, $\phi$ is unramified. So, as stated, we can apply Hurwitz genus formula as described above and can conclude that indeed $\phi$ is the unique required isogeny, and that $C = E'$ is the unique elliptic curve $\phi$ is mapping to.  

In the last part of this chapter some mathematical background on endomorphism rings is given.
4.5 Endomorphism ring of supersingular curves

**Theorem 4.5.1.** Two curves $E_1$ and $E_2$ defined over a field $K$ are isogenous if only if their endomorphism algebras $\text{End}(E_1) \otimes \mathbb{Q}$ and $\text{End}(E_2) \otimes \mathbb{Q}$ are isomorphic.

**Proof.** For a general version of the proof, see [21]. Here a rougher sketch of the proof is given, as given in [9]. Take an isogeny $\phi : E_1 \to E_2$ of degree $m$ and its dual isogeny $\hat{\phi}$, with $E_1$ and $E_2$ defined over a field $K$. For $\psi \in \text{End}(E_1)$ we have a $\mathbb{Z}$-module homomorphism $\text{End}(E_1) \to \text{End}(E_2)$ given by $\psi \mapsto \phi \psi \hat{\phi}$.

However, when $\deg(\phi) \neq 1$, this is not a ring homomorphism. To correct this, take an elliptic curve $E$ isogenous to $E_1$ and $E_2$, and set $K = \text{End}(E) \otimes \mathbb{Z} \mathbb{Q}$. For any isogeny $\phi_i : E_i \to E$ of degree $m$ there is a ring homomorphism $\text{End}(E_i) \to K$ given by $\psi \mapsto \hat{\phi}_i \psi \hat{\phi}_i \otimes \mathbb{Z} m^{-1}$.

From this it follows that $\text{End}(E_i) \otimes \mathbb{Z} \mathbb{Q} \cong K$ for all elliptic curves $E_i$ isogenous to $E$ over $K$. \hfill \Box

**Theorem 4.5.2.** All supersingular curves defined over $\overline{\mathbb{F}}_p$ are in the same isogeny class.

**Proof.** See [12]. \hfill \Box

**Theorem 4.5.3.** There is a bijection between left ideals of the endomorphism ring of an elliptic curve $E$, and isogenies with $E$ as starting curve.

**Proof.** For a more detailed proof using equivalence of categories, see [22, Chapter 42], [23, Theorem 4.5] or [9, Theorem 5.3].

The idea of the proof is that there is, for each pair $(E', \phi)$, with $\phi : E \to E'$ an isogeny, a left $\text{End}(E)$-ideal $I$, given by $I = \text{hom}(E', E)\phi$, and in the same way for every left $\text{End}(E)$-ideal $I$ there is an isogeny $\phi$, constructed as follows: For $I \neq (0)$, define $E[I]$ to be $E[I] = \bigcap_{\alpha \in I} \ker(\alpha)$.

Then by Theorem 4.4.2 there is a unique isogeny $\phi_I : E \to E/E[I]$ with kernel $E[I]$. \hfill \Box

The last theorems combined give that up to isomorphism, all supersingular elliptic curves have the same endomorphism algebra, and that we can represent isogenies by left ideal classes of the endomorphism ring. Since the endomorphism ring of a supersingular elliptic curve is not commutative, it shows that taking isogenies between supersingular isogenies is not commutative either.
5. The Supersingular Isogeny Diffie Hellman Key Exchange protocol

Now that we have the sufficient mathematical background and enough information about Diffie Hellmann key exchanges, we can finally explain the SIDH protocol. The goal of the SIDH protocol, like any other key exchange, is for two persons, Bob and Alice, to form a shared secret key together. In the process of creating this key it must be infeasible for an adversary, Eve, to get access to information that will lead her to this secret key.

The first step is constructing a supersingular elliptic curve $E_0$ over a field $\mathbb{F}_q$, with $q = p^2$. $p$ is a prime of the form $p = \ell_A^e \ell_B^f - 1$, with $f$ a cofactor to make $p$ prime. The curve will have cardinality $(\ell_A^e \ell_B^f)^2$. Usually $\ell_A$ is taken to be 2, and $\ell_B$ to be 3.

Besides the curve $E_0$, the public parameters also include points, $P_A$ and $P_B$, $Q_A$ and $Q_B$, such that $\langle P_A, Q_A \rangle = E_0[\ell_A^e]$ and $\langle P_B, Q_B \rangle = E_0[\ell_B^f]$, the points are chosen to generate the $\ell_A^e$- and $\ell_B^f$-torsion groups.

Alice then chooses her secret integers $m_A$ and $n_A$, not both divisible by $\ell_A^e$, so that $R_A = [m_A]P_A + [n_A]Q_A$ has order $\ell_A^e$. She computes an isogeny (it is possible to compute a specific isogeny using for instance Vélu’s formulas [13]) $\phi_A : E_0 \rightarrow E_A$ with $E_A \cong E_0/\langle R_A \rangle$, as according to Theorem [4.4.2] every subgroup $A$ of $E_0$ gives way to a unique curve $E_0/A$ and a unique isogeny between them. Bob acts mutatis mutandis.

Then Alice sends her public keys; curve $E_A$, and the points $\phi_A(P_B), \phi_A(Q_B)$ to Bob. The isogeny $\phi_A$ and the point $R_A$ stay private. She receives from Bob the points $\phi_B(P_A), \phi_B(Q_A)$ and the curve $E_B \cong E_0/\langle R_B \rangle$. She then computes an isogeny $\phi_A' : E_B \rightarrow E_{AB}$ with kernel $[m_a]\phi_B(P_A) + [n_A]\phi_B(Q_A)$. Bob also proceeds in the same way to generate $E_{BA} \cong E_{AB}$.

They can then use the common $j$-invariant of $E_{AB}$ as a secret shared key, since isomorphic curves have the same $j$-invariant.

The SIDH protocol is also explained in schematic form in Figure [5.1]
Key exchange protocol

Public parameters:

\[E_0, p, \ell_A, \ell_B, P_A, P_B, Q_A, Q_B\]

Alice

\[R_A = [m_A]P_A + [n_A]Q_A\]
\[\phi_A : E \to E_A = E/\langle R_A \rangle\]

Bob

\[R_B = [m_B]P_B + [n_B]Q_B\]
\[\phi_B : E \to E_B = E/\langle R_B \rangle\]

\[E_A, \phi_A(P_B), \phi_A(Q_B)\] \[E_B, \phi_B(P_A), \phi_B(Q_A)\]

\[E_{AB} = E_B/\langle [m_A]\phi_B(P_A), [n_A]\phi_B(Q_A) \rangle\] \[E_{BA} = E_A/\langle [m_B]\phi_A(P_B), [n_B]\phi_A(Q_B) \rangle\]

Output: \(j(E_{AB})\) \[\text{Output: } j(E_{BA})\]

Figure 5.1: The SIDH protocol explained schematically.

The SIDH protocol is more complicated than other Diffie Hellman key exchanges. Instead of Bob and Alice just sending each other their curves \(E_A\) and \(E_B\), auxiliary points are also included as parameters. The reason for this is that taking isogenies between supersingular curves is not a commutative action. If the protocol were to use ordinary elliptic curves instead of supersingular ones, as proposed by [16], an easier scheme would have sufficed. But as explained in Section 7.3, ordinary curve Diffie Hellman is not really suitable for post-quantum cryptography.

The reason that the SIDH protocol needs auxiliary points is because the endomorphism ring is not commutative, but instead, as we have seen in Theorem 4.5.1, an order of a quaternion algebra. Also, Theorem 4.5.3 showed the bijection between the left ideals of the endomorphism ring of elliptic curves and isogenies between elliptic curves. This together gives that taking isogenies, under composition, is not commutative for supersingular isogenies. So it would not suffice for Alice and Bob just to exchange their curves \(E_A\) and \(E_B\) to get a commutative diagram. This is the reason it took researchers five years to develop the SIDH protocol after the ordinary isogeny Diffie Hellman key exchange was already developed.

In the theorem below the proof is given that with these extra auxiliary points, the key exchange does work commutatively.
5.1 Commutativity

Theorem 5.1.1. The curves $E_{AB}$ and $E_{BA}$ are isomorphic.

Proof. According to theorem 4.4.2 the curve $E_{AB}$ is isomorphic to

$$E_{AB} \cong E_B / \langle [m_a] \phi_B(P_A) + [n_A] \phi_B(Q_A) \rangle.$$

Theorem 4.3.12 gives

$$[m_a] \phi_B(P_A) + [n_A] \phi_B(Q_A) = \phi_B([m_a] P_A) + \phi_B([n_A] Q_A) = \phi_B([m_a] P_A + [n_A] Q_A) = \phi_B(R_A).$$

Therefore

$$E_{AB} \cong E_B / \langle \phi_B(R_A) \rangle.$$

In the same way it follows that

$$E_B \cong E_0 / \langle R_B \rangle.$$

Therefore we have that the kernel of

$$(\phi'_A \circ \phi_B) : E_0 \longrightarrow E_{AB} \quad \text{is} \quad \langle R_A, R_B \rangle.$$  

And since

$$E_{BA} \cong E_A / \langle [m_b] \phi_A(P_B) + [n_B] \phi_A(Q_B) \rangle = E_A / \langle \phi_A(R_B) \rangle,$$

and $E_A = E_0 / \langle R_A \rangle$, this gives for the kernel of

$$(\phi'_B \circ \phi_A) : E_0 \longrightarrow E_{BA}$$

(which is again an isogeny since both $\phi'_B$ and $\phi_A$ are surjective, making $\phi'_B \circ \phi_A$ a surjective homomorphism between elliptic curves, and therefore an isogeny) that

$$\ker((\phi'_B \circ \phi_A) : E_0 \longrightarrow E_{BA}) = \langle R_B, R_A \rangle.$$  

And since $\langle R_B, R_A \rangle = \langle R_A, R_B \rangle$, it follows that the kernels of $\phi'_B \circ \phi_A$ and $\phi'_A \circ \phi_B$ are the same. The only thing left to prove is that the kernels are subgroups of $E_0$, then it follows from Theorem 4.4.2 that $\phi'_B \circ \phi_A = \phi'_A \circ \phi_B$, and that $E_0 / \langle R_A, R_B \rangle$ again is an elliptic curve. And since

$$\text{order}(R_A) = \ell^{e_A} \neq \ell^{e_B} = \text{order}(R_B),$$

$\langle R_A \rangle + \langle R_B \rangle = \langle R_A, R_B \rangle$ is a subgroup of $E_0$. This proves that $E_{AB} \cong E_{BA}$. \qed
6. Isogeny graphs

We now know how the SIDH protocol works. There is however another way to describe how Alice and Bob can create their shared secret key, using isogeny graphs. The benefit of describing the SIDH protocol this way is that it gives more insight in the security of the protocol, and also gives a nice graphical image on how the protocol works. First we will give some background information.

6.1 Dual isogenies

Definition 6.1.1. Let $\phi : E_1 \to E_2$ be an isogeny. The dual isogeny is the unique isogeny $\phi^* : E_2 \to E_1$, such that $\phi^* \circ \phi = \phi \circ \phi^* = [m]$, with $m = \deg(\phi)$.

Theorem 6.1.2. Every isogeny has a dual isogeny.

Proof. See [17, III.6.1].

Since every isogeny has a dual isogeny, we can see the property of being isogenous as an equivalence relation on the set of $\overline{\mathbb{F}}_q$-isomorphism classes of elliptic curves defined over $\mathbb{F}_q$. For transitivity note that isogenies are surjective, so the composition of two isogenies is again an isogeny. Theorem 4.5.1 shows that two elliptic curves can only be isogenous if their endomorphism ring has the same structure, giving in particular that supersingular elliptic curves are isogenous only to other supersingular elliptic curves.

6.2 Isogeny graphs

Definition 6.2.1. An isogeny graph is a graph with as nodes the $j$-invariants of isogenous curves, and as edges the isomorphism classes of isogenies between them.

Since every isogeny has a dual isogeny of the same degree, the isogeny graph is undirected. Isogeny graphs are usually drawn with only isogenies of one specific degree as edges. A graph with only isogenies of degree $\ell$ is called an isogeny graph of degree $\ell$.

Theorem 6.2.2. All supersingular $j$-invariants of curves in $\overline{\mathbb{F}}$ are defined over $\mathbb{F}_{p^2}$.

Proof. See [17, V.3.1].
Remark 6.2.3. This theorem also says that all supersingular elliptic curves over a field $\mathbb{F}_p$ are not only defined over $\bar{\mathbb{F}}_p$, but more specifically over $\mathbb{F}_{p^2}$.

Theorem 6.2.2 gives us that, up to isomorphism, there are only a finite supersingular curves in an isogeny class. This way it is possible to represent the isogeny class using a finite graph. It also gives us that, if we are looking for points on a curve $E$, they will be defined over $\mathbb{F}_{p^2}$ and not in a random field (of high order) in the algebraic closure of $\mathbb{F}_p$, making it relatively easy to explicitly describe points on a curve.

Theorem 6.2.4. In a supersingular isogeny class over a field $\bar{\mathbb{F}}_p$, with $p > 3$, there are

$$\left\lfloor \frac{p}{12} \right\rfloor + \begin{cases} 0, & \text{if } p = 1 \text{ mod } 12 \\ 1, & \text{if } p = 5, 7 \text{ mod } 12 \\ 2, & \text{if } p = 11 \text{ mod } 12 \end{cases}$$

different isomorphism classes of supersingular curves.

Proof. [17] V.4.1c].

Example 6.2.5. An example of an isogeny graph of degree 3 over $\bar{\mathbb{F}}_{97}$ is given below [2].

![Figure 6.1: Isogeny graph of degree 3 over $\bar{\mathbb{F}}_{97}$](image)

Definition 6.2.6. The adjacency matrix of a finite graph is a matrix that shows whether or not two vertices of a graph are adjacent. For a graph $G = (V, E)$, the adjacency matrix is a square $|V| \times |V|$ matrix $A$ such that it counts the edges between vertices; $A_{ij}$ is one when there is one edge from vertex $i$ to vertex $j$, zero when there is no edge.

Definition 6.2.7. A connected $d$-regular graph $G = (V, E)$ has the Ramanujan property, and is then called a Ramanujan graph, if the eigenvalues of the corresponding adjacency matrix have the following property:

If for $\lambda_0 \geq \lambda_1 \geq ... \geq \lambda_{n-1}$, with $n = |V|$, there exists a $\lambda_i$ such that $|\lambda_i| < d$, and given $\lambda(G) = \max_{|\lambda_i| < d} |\lambda_i|$, then $|\lambda(G)| < 2\sqrt{d - 1}$. 28
**Theorem 6.2.8.** The graph of supersingular curves in $\mathbb{F}_p$ with $\ell$-isogenies is connected, $\ell + 1$ regular and has the Ramanujan property.

*Proof.* The fact that the graph is $\ell + 1$ regular follows from the fact that $E[\ell] \cong \mathbb{Z}_\ell \oplus \mathbb{Z}_\ell$ has $\ell + 1$ subgroups of size $\ell$, and that every subgroup is the kernel of a unique isogeny with the degree of the isogeny equal to the size of the kernel. For the proof of the other parts of the theorem, see [15, Theorem 1].

**Theorem 6.2.9.** Every supersingular separable isogeny between elliptic curves defined over $\mathbb{F}$ of degree greater than 1 can be factored into a composition of isogenies of prime degree between elliptic curves defined over $\mathbb{F}_{p^2}$.

*Proof.* Take an isogeny $\phi : E \rightarrow E'$, with $E$ and $E'$. According to Theorem [4.4.1] the cardinality of the kernel of $\phi$ equals its degree, and since $\phi$ is an isogeny, its kernel is finite. Now calculate all the subgroups of $\ker(\phi)$. Theorem [4.4.2] says that with a subgroup of an isogeny as kernel, we can create a unique isogeny mapping to a unique new elliptic curve. Label the subgroups $\{H_1, \ldots, H_n\}$. Then start with taking $H_1$ as the kernel for our first isomorphism $\phi_1$, mapping to curve $E_1$. Now calculate $H_2/H_1$. If this is the trivial group, continue to the next subgroup and repeat. If it is not trivial, $H_2/H_1 = H_2$, since they are both of prime degree. Take $H_2$ as kernel for the new isogeny $\phi_2 : E_1 \rightarrow E_2$. Repeat this process until $H_n$ is reached. Composing all these $\phi_i$ gives a map with the same kernel as $\phi$. Therefore the composition of all these $\phi_i$ equals $\phi$, so it is possible to write a separable isogeny as a composition of isogenies of prime degree. And since all the isogenies used are between supersingular elliptic curves used are supersingular and in the same isogeny graph as $E$ and $E'$, they are all defined over $\mathbb{F}_{p^2}$.

Now we can represent the SIDH protocol by Alice and Bob taking a random walk in the isogeny graphs of curves isogenous to $E_0$. Alice first takes a walk in the $\ell_A$-graph, Bob in the $\ell_B$-graph, and for the second step they switch graphs. Since an isogeny can be split into isogenies of prime degrees, finding an isogeny of degree $\ell_A^{e_A}$ is the same as taking a walk of length $e_A$ in the $\ell_A$-isogeny graph, without backtracking.

As we know, Alice computes an isogeny of degree $\ell_A^{e_A}$, since that is the cardinality of the kernel. This is a separable isogeny. We can thus split the isogeny into $e_A$ isogenies of degree $\ell_A$. All these isogenies will be in the same isogeny graph. Alice thus takes a walk of length $e_A$. An adversary will have to find, given two points — or isomorphism classes, in the graph, the shortest path of length $e_A$ between them. This path will correspond to Alice’s secret subgroup of $E[\ell_A^{e_A}]$. Finding this shortest path is for an outstander infeasable due to certain properties of the graph, explained in the section below, providing the security for the SIDH scheme.
7. Security

In the beginning of this thesis the claim was stated that this Isogeny based Diffie Hellman scheme was quantum-secure. In this part we will shed some light on why that is the case, and also explain the difference between using isogenies between supersingular and ordinary curves, since the first is —by the best of our knowledge— quantum resistant, whilst for the latter already exists a quantum algorithm partly breaking the quantum security.

As said in the previous section, we can explain the hardness of the SIDH protocol using isogeny graphs. We will use the property that a supersingular isogeny graph has the Ramanujan property.

7.1 Expander graphs

A graph is said be an expander graph, or have the expander property if the graph has really high connectivit properties, meaning that a lot of edges have to be removed to make the graph unconnected. There exist multiple formal definitions of expander graphs, here we will give one. For further reference, see [11]. This paper is also the reference for all definitions and theorems in this section.

**Definition 7.1.1.** Take a graph $X$ with set of vertices $V$. The boundary of a subset $Y$, denoted $\partial Y$, is the set of vertices that are connected to at least one vertex of $Y$, but are not in $Y$.

**Definition 7.1.2.** Take $0 < \epsilon \in \mathbb{R}$ a finite graph $X = (V,E)$, with $|V| = n$. $X$ is an $\epsilon$-expander if for every subset $Y$ of $V$ with

$$|Y| \leq |V|/2, \quad |\partial Y| \geq \epsilon|Y|.$$ 

The largest $\epsilon$ for which $X$ is an $\epsilon$-expander is denoted as $\epsilon(X)$.

A $k$-regular graph $X$ is called an expander graph if it is an $\epsilon$-expander graph for a certain $\epsilon > 0$.

Recall from the Definition 6.2.6 section that every graph has an adjacency matrix $A$, with eigenvalues $\lambda_0 \geq \ldots \geq \lambda_{n-1}$.

**Theorem 7.1.3.** For $k$-regular graphs, being an $\epsilon$-expander is equivalent to having a spectral gap $\lambda_1 < k - \epsilon'$ for a certain $\epsilon' > 0$.

We know that $\lambda_1 < k$, and we know from Definition 6.2.7 that for a Ramanujan graph $\lambda(X) \leq 2\sqrt{k-1}$. So we can apply Theorem 7.1.3 to Ramanujan graphs, thus Ramanujan graphs have the expander property. The main benefit of expander graphs is that they have nice rapid mixing properties, like the one stated in the theorem below, that states that a random walk in a graph can end with high probability at any subgroup of the graph.
Theorem 7.1.4. Let $G$ be a regular graph of degree $k$ on $h$ vertices. Suppose that the
eigenvalue of any nonconstant eigenvector satisfies the bound $|\lambda| \leq c$ for some $c < k$. Let $S$
be any subset of the vertices of $G$, and $x$ be any vertex in $G$. Then a random walk of length
of at least $\frac{\log 2h/|S|^{1/2}}{\log k/c}$ starting from $x$ will land in $S$ with probability at least $\frac{|S|}{2h} = \frac{|S|}{2|G|}$.


Combining the Ramanujan property of supersingular isogeny graphs en the properties of
Ramanujan graphs, we can conclude that a supersingular isogeny graph has the expander
property, meaning that a random walk will become ‘random’ really fast.

7.2 Security of the SIDH protocol

The cryptographical problem related to the SIDH can be stated as follows.

Problem 7.2.1. Given two isogenous elliptic curves $E, E'$ over a finite field $K$, find an isogeny
$\phi : E \rightarrow E'$ of smooth degree.

This problem is known to be a hard problem to solve, since finding the secret isogeny is the
same as finding the shortest random walk between two points in an isogeny graph, and we
now know that if that walk is sufficiently large, the endpoint of the walk is so random that
finding the shortest path between the two points is really difficult.
The fastest known non-quantum attack to the SIDH scheme is the generic Claw attack. For
an adversary, only the starting point and the end point of the random walk are given, and
the degree of the isogeny, and therefore the length of the path. The claw algorithm works as
follows: Given a graph with starting point $E$, end point $E'$ and a random walk of length $\ell^e$,
calculate for the starting point all random walks of length $\ell^e \sqrt{2}$ and store these walks and their
end points $E_i$. Then calculate for the end point random walks of length $\ell^e \sqrt{2}$, until the end
point of one of these walks matches a point $E_i$. This path is —almost certain— the shortest
walk between $E$ and $E'$, and the required random walk. The security of the SIDH protocol
then becomes $O(p^{1/4})$.\footnote{This level of security actually makes the SIDH protocol not the best public key protocol, if you look at
key sizes and time needed to complete the key exchange. SIDH therefore serves best only as a quantum secure
protocol, where it —to the best of our knowledge — can compete with all the other proposed cryptographic
schemes.}
The fastest known quantum attack is the quantum claw attack described in [20]. The quantum
security of the SIDH protocol becomes $O(p^{1/6})$.

7.3 Isogenies over ordinary elliptic curves

The question may have arosen why only supersingular elliptic curves are used, and not ordi-
nary curves as well. There is in fact, an isogeny based Diffie Hellman protocol using ordinary
curves, created by Stolbunov [16]. The ordinary isogeny Diffie Hellman protocol is based on
constructing an isogeny between two ordinary elliptic curves with the same endomor-
phism ring. As explained earlier, there is a bijection between isogenies and ideals in the endomor-
phism ring. For ordinary elliptic curves, the endomorphism ring is commutative. Using this
commutativity, researchers [1] were able to use a quantum algorithm to break the protocol
in supexponential time. Even though the time needed to break the protocol was still quite long, the scheme was declared inapplicable for post-quantum cryptographic uses. This was because running the key exchange on a computer already took a really long time compared to other post-quantum protocols, and with weakened security there was no practical use for it anymore.
Bibliography


