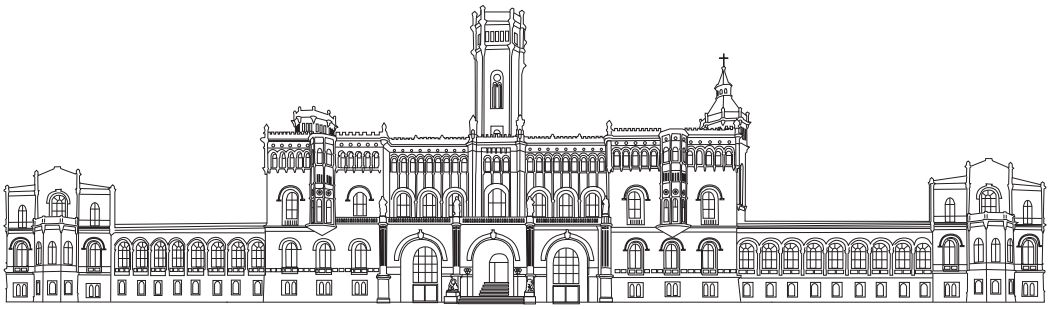

Leibniz Universität Hannover

Fakultät für Elektrotechnik und Informatik
Institut für Theoretische Informatik

Elliptische Kurven in der Kryptographie

Bachelorarbeit

eingereicht von

Jan Eberhardt

am 29. März 2016

Erstprüfer:	Prof. Dr. Heribert Vollmer
Zweitprüfer:	Dr. Arne Meier
Betreuer:	M.Sc. Anselm Haak

Erklärung

Hiermit versichere ich, dass ich diese Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe.

Inhaltsverzeichnis

1 Einleitung

In einer immer mehr vernetzten Welt, in der Informationen und deren Austausch zunehmend an Bedeutung gewinnt, ist auch der Schutz dieser Informationen sehr wichtig. Im digitalen Zeitalter ist die Verschlüsselung von Daten daher eine Sache von hoher Notwendigkeit. Die meisten Menschen nutzen Verschlüsselungsverfahren unbewusst, während sie im Internet agieren, ohne darüber nachdenken zu müssen. Dass sicherer Informationsaustausch im Internet heute problemlos möglich ist, ist dem Fortschritt in der modernen Kryptographie zu verdanken. Maßgeblich für die Verschlüsselung von Daten auf den Transportwegen im Internet ist die asymmetrische (Public-Key-) Kryptographie. Diese wurde ab Anfang der 1970er Jahre entwickelt und ist somit noch eine ziemlich junge Entdeckung.

Diese Arbeit soll eine Einführung in elliptische Kurven geben und orientiert sich inhaltlich hauptsächlich an dem Buch *Elliptische Kurven in der Kryptographie* von Annette Werner ([We]). Nach einer kurzen Übersicht über die Public-Key-Kryptographie und einiger Verfahren aus diesem Bereich, sollen die mathematischen Grundlagen für elliptische Kurven besprochen werden. Dabei werden elementare Kenntnisse aus der Algebra vorausgesetzt. Eine Zusammenfassung dazu liefert [We] Kapitel 6. Nachdem die mathematischen Grundlagen gelegt sind, werden elliptische Kurven eingeführt und ein Gruppengesetz auf diesen definiert. Abschließend beschäftigen wir uns mit der praktischen Nutzung von elliptischen Kurven in der Kryptographie. Außerdem wurde eine Prinzipimplementierung vorgenommen, die in Kapitel 5 vorgestellt wird.

1.1 Public-Key-Kryptographie

In der Zeit vor dem Internet wurden geheime Nachrichten stets mit ein und demselben Schlüssel ver- und entschlüsselt. Verschlüsselungsverfahren, die auf diesem Prinzip basieren werden als symmetrisch bezeichnet. Um hier einen sicheren Austausch von Informationen zu ermöglichen, muss jeder Kommunikationspartner die Informationen, die er übermitteln will, mit einem symmetrischen Verschlüsselungsalgorithmus verschlüsseln. Dazu wird ein Schlüssel benötigt, der sowohl dem Sender als auch dem Empfänger einer Nachricht vorliegen muss, um sie zu ver- bzw. entschlüsseln. Dieser geheime Schlüssel muss daher zuvor über einen sicheren Kanal ausgetauscht werden. Im Internet ist eine solche Lösung jedoch nicht praktikabel, da meist große Entfernungen zwischen den Kommunikationspartnern liegen. Niemand möchte vor dem verschlüsselten Abrufen einer Webseite zunächst einen Schlüssel über einen sicheren Weg austauschen müssen, zum Beispiel per Post. Es musste also eine Lösung gefunden werden, um eine sichere Verbindung über einen unsicheren Übertragungsweg initiieren zu können.

Ein Verfahren zum Schlüsselaustausch über einen unsicheren Verbindungskanal veröffentlichten Martin E. Hellman und Whitfield Diffie 1976[1], den sogenannten Diffie-Hellman-Schlüsselaustausch. Dieses Verfahren wird auch heute noch bei fast allen

[1] *New Directions in Cryptography.* In: IEEE Transactions on Information Theory. 22, Nr. 6, 1976, S. 644–654

HTTPS-Verbindungen im Internet verwendet. Damit ist es möglich symmetrische Verschlüsselung (meistens AES[2]) im Internet beim sicheren Zugriff auf Webseiten zu verwenden. Wenig später wurde ein Algorithmus entwickelt, der auf den Überlegungen von Hellman und Diffie aufbaute und das erste asymmetrische Verschlüsselungsverfahren darstellte. Das Verfahren ist nach seinen Entwicklern R. L. Rivest, A. Shamir und L. Adleman benannt (RSA). Bei einem asymmetrischen Verschlüsselungsverfahren gibt es nicht nur einen Schlüssel, mit dem ver- und entschlüsselt wird, sondern zwei separate Schlüssel, die als Paar fungieren. Der sogenannte Public-Key wird dabei zum verschlüsseln verwendet. Eine damit verschlüsselte Nachricht kann nur mit dem dazugehörigen Private-Key entschlüsselt werden. Deswegen nennt man asymmetrische Kryptographie auch Public-Key-Kryptographie. Wie der Name schon sagt, muss der private Schlüssel geheim bleiben, um die Sicherheit des Kryptosystems zu gewährleisten. Der Public-Key kann hingegen öffentlich zugänglich sein und kann zum Beispiel auf einem öffentlichen Schlüsselserver hinterlegt werden. Die Sicherheit solcher Public-Key-Kryptosysteme basiert auf der Schwierigkeit durch die Kenntnis eines Public-Keys den entsprechenden Private-Key berechnen zu können. Dazu müssen im Allgemeinen schwierige mathematische Probleme gelöst werden.

Im Folgenden sollen die RSA-Verschlüsselung, der Diffie-Hellman-Schlüsselaustausch und die EL-Gamal-Verschlüsselung kurz in ihrer mathematisch, algorithmischen Form dargestellt werden, um einen Überblick über die wichtigsten Verfahren in der Public-Key-Kryptographie zu geben. Die beiden Kommunikationspartner, die verschlüsselt kommunizieren wollen, werden in der Literatur über Kryptographie traditionell Bob und Alice genannt. Daher werden wir diese Konvention in den nachfolgenden Beschreibungen der Kryptographie-Verfahren ebenfalls übernehmen.

1.2 RSA-Verschlüsselung

Dieser Abschnitt soll die Idee der RSA-Verschlüsselung in kurzer Form vermitteln. Eine ausführliche Beschreibung findet sich zum Beispiel in [Bu], 9.3.

Damit Bob und Alice das RSA-Verfahren nutzen können müssen sie zunächst ein Schlüsselpaar generieren. Dazu wählt Bob zwei (verschiedene) große Primzahlen p, q und berechnet $n = p \cdot q$. Dann wählt er noch eine zufällige Zahl e zwischen 1 und $\varphi(n)$ wobei

$$\varphi(n) = (p-1) \cdot (q-1)$$

ist und e teilerfremd zu $\varphi(n)$ sein muss. φ ist hier die Eulersche φ-Funktion. Das Paar (n, e) ist Bobs öffentlicher Schlüssel. Um einen dazu passenden privaten Schlüssel zu erzeugen berechnet er $d \in \{1, \ldots, \varphi(n)\}$, sodass $e \cdot d \equiv 1 \mod \varphi(n)$ mithilfe des erweiterten Euklidischen Algorithmus. Denn dadurch erhält man die Bézout-Koeffizienten in der Formel

$$de + y\varphi(n) = 1.$$

Nun hat Bob sein Schlüsselpaar erzeugt und kann Alice seinen öffentlichen Schlüssel auf beliebigem Wege übergeben. Alice ist nun in der Lage Bob eine verschlüsselte Nachricht

[2] Advanced Encryption Standard

zu senden. Dafür benutzt sie die für Bob spezifische Funktion

$$f_B : \mathbb{Z}/n\mathbb{Z} \to \mathbb{Z}/n\mathbb{Z}$$
$$x \mapsto x^e.$$

Hier ist x die Nachricht, die verschlüsselt werden soll. Damit man eine beliebige Nachricht mit f_B verschlüsseln kann muss sie zunächst als Element von $\mathbb{Z}/n\mathbb{Z}$ dargestellt werden und falls nötig auch in mehrere Teile aufgeteilt werden. Diese Codierung muss Bob natürlich auch bekannt sein, damit er sie später rückgängig machen kann.

Wenn Bob nun die verschlüsselte Nachricht x^e erhalten hat, kann er sie mit seinem privaten Schlüssel d entschlüsseln und zwar indem er $(x^e)^d = x^{ed}$ berechnet. So kommt er an die ursprüngliche Nachricht x, denn da $ed \equiv 1 \mod \varphi(n)$ ist gibt es eine ganze Zahl a mit $ed = 1 + a \cdot \varphi(n)$. Jetzt gibt es zwei Fälle zu unterscheiden:

1) x und p sind teilerfremd. Dann folgt aus dem kleinen Satz von Fermat, dass

$$x^{ed} = x \cdot x^{a\varphi(n)} \equiv x \mod p$$

da $\varphi(p) = p - 1$ also $\varphi(p)$ ein Teiler von $\varphi(n)$ ist.

2) p teilt x. Dann gilt $x^{ed} \equiv x \mod p$ (beide Seiten sind 0 mod p)

Dasselbe gilt für q und x ($x^{ed} \equiv x \mod q$) und dadurch kann man mit dem chinesischen Restsatz folgern, dass $x^{ed} \equiv x \mod n$. So erhält Bob die ursprüngliche Nachricht x zurück.

Dieses Verfahren ist besonders geeignet für die Verschlüsselung im Internet (per HTTPS), weil sowohl die Verschlüsselung $f_B(x)$ als auch die Entschlüsselung $x^e \mapsto (x^e)^d$ relativ leicht, also schnell zu berechnen sind. Aus $f_B(x) = x^e$ jedoch ohne Kenntnis von d wieder x zu berechnen ist allerdings ein schwieriges mathematisches Problem. Alternativ kann ein Angreifer versuchen, aus dem öffentlichen Schlüssel (n, e) den privaten Schlüssel d zu berechnen und damit $f_B(x)$ zu entschlüsseln. Dazu muss er die Primfaktoren p und q von n ermitteln, also das Faktorisierungsproblem lösen. Wenn die Primfaktoren bekannt sind, kann d ebenso wie Bob es gemacht hat mit dem erweiterten Euklidischen Algorithmus berechnet werden. Die theoretische Sicherheit des RSA-Verfahrens hängt also entscheidend von der Wahl von p und q ab. Sie müssen so groß gewählt werden, dass man sie in der Praxis nicht mit vertretbarem Aufwand mithilfe bekannter Faktorisierungsalgorithmen ermitteln kann. Somit ist die Sicherheit von RSA nicht bewiesen, sondern man verlässt sich in der Praxis auf eine ausreichend große Abschätzung bei der Wahl von p und q.[3] Offen bleibt allerdings, ob zukünftig bessere Faktorisierungsalgorithmen entwickelt werden, die RSA praktisch unbrauchbar machen können.

[3] Die aktuelle Empfehlung des BSI für die Schlüssellänge beim Einsatz von RSA, ist mindestens Zahlen in der Größenordnung von 2000 Bit (mehr als 600 Dezimalstellen) oder besser $\geq$ 3000 Bit zu verwenden. [bsi]

1.3 Diskreter Logarithmus

Für die folgenden Verfahren der Public-Key-Kryptographie ist das sogenannte Diskreter-Logarithmus-Problem von großer Bedeutung. Daher werden wir es hier einführen und zwar im Kontext einer abelschen Gruppe G mit additiver Operation $(P, Q) \mapsto P + Q$ und neutralem Element 0. Für ein gegebenes Element $P \in G$ sei n die Ordnung der von P erzeugten zyklischen Untergruppe

$$\langle P \rangle = \{kP \mid k \in \mathbb{Z}\}.$$

Es ist also $nP = 0$. Jetzt definieren wir eine Funktion

$$\begin{aligned} f : \mathbb{Z}/n\mathbb{Z} &\to \langle P \rangle \\ k \bmod n &\mapsto kP. \end{aligned}$$

Das **Diskreter-Logarithmus-Problem** (DL-Problem oder DLP) lautet nun wie folgt: *Bestimme zu den gegebenen Daten G, $P \in G$, $n = ord(P)$ und $Q \in \langle P \rangle$ das Element k in $\mathbb{Z}/n\mathbb{Z}$ mit*

$$Q = kP.$$

Gesucht wird also eine Umkehrfunktion zu $k \mapsto kP$, daher die Bezeichnung Logarithmus, welche sich aus den altgriechischen Worten lógos (Verständnis, Lehre, Verhältnis) und arithmós (Zahl) zusammensetzt. Anschaulicher ist es bei einer multiplikativen Gruppenoperation, wo dann die Umkehrfunktion zu $k \mapsto p^k$ gesucht werden würde, so wie man es sonst aus vielen Bereichen der Mathematik kennt (z.B. $\ln(x)$ als Umkehrfunktion für e^x). Bei elliptischen Kurven ist die additive Definition der Gruppenoperation allerdings üblicher.

In kryptographischen Verfahren, die auf dem DL-Problem basieren ist es von großer Wichtigkeit geeignete Gruppen zu wählen, sodass das DL-Problem schwer zu lösen ist. Dazu ist es auch wichtig, dass $n = |\langle P \rangle|$ ausreichend groß ist, da man sonst mit einem Brute-Force-Ansatz angreifen kann, indem man alle Möglichkeiten $(P, 2P, 3P, \dots)$ durchprobiert. (Im statistischen Mittel findet man k beim Ausprobieren schon nach der Hälfte der Schritte.)

1.4 Diffie-Hellman-Schlüsselaustausch

Wie oben erwähnt, dient dieses Verfahren zum Austausch, oder besser gesagt zur Verständigung auf einen Schlüssel über einen öffentlichen Kanal, der dann für ein symmetrisches Verschlüsselungsverfahren verwendet werden kann. Dafür seien wie beim DL-Problem zunächst die Gruppe G, $P \in G$ und $n = |\langle P \rangle|$ allgemein bekannt.

Wenn Alice und Bob sich auf einen geheimen Schlüssel verständigen wollen, führen sie folgende Schritte durch:

1) Alice wählt zufällig eine ganze Zahl $d_A \in \{1, 2, \ldots, n-1\}$ und schickt das Gruppenelement $d_A P$ an Bob.

2) Bob wählt seinerseits zufällig eine Zahl $d_B \in \{1, 2, \ldots, n-1\}$ und schickt das Gruppenelement $d_B P$ an Alice.

3) Alice berechnet mit ihrer Zahl d_A das Element $d_A(d_B P) = (d_A d_B)P$; Bob berechnet $d_B(d_A P) = (d_B d_A)P = (d_A d_B)P$.

Das Element $d_A d_B P$, dass Alice und Bob beide errechnet haben dient jetzt als gemeinsamer Schlüssel. Beide Kommunikationspartner brauchen dafür nicht die geheime Zahl des jeweils anderen (d_A bzw. d_B) zu kennen. Und genau darin liegt die Schwierigkeit für einen Angreifer, der die Kommunikation zwischen Alice und Bob zwar abhören könnte und somit in Besitz von $d_A P$ und $d_B P$ gelangen könnte, aber nun das sogenannte Diffie-Hellman-Problem (DH-Problem) lösen müsste. Dies wird folgendermaßen definiert:

Diffie-Hellman-Problem: *Berechne zu zwei Elementen kP und lP in $\langle P \rangle$ das Element klP in $\langle P \rangle$.*

Es besteht nun ein Zusammenhang zwischen dem DL- und DH-Problem. Offensichtlich kann jemand, der das DL-Problem in der Gruppe G lösen kann auch das DH-Problem lösen, weil er l und k bestimmen kann. Ob auch die Umkehrung gilt, ist aber nicht bekannt: Wenn in einer Gruppe das DL-Problem schwer zu lösen ist kann man nicht daraus schließen, dass darin auch das DH-Problem schwer zu lösen ist.

1.5 ElGamal-Verschlüsselung

Dieses Public-Key-Verfahren wurde 1985 von T. ElGamal vorgestellt. Die ElGamal-Verschlüsselung hat starke Ähnlichkeiten mit dem DH-Schlüsselaustausch, und verwendet dieselben grundlegenden Daten, nämlich G, n und ein $P \in G$. Die Schlüsselerzeugung ist denkbar einfach. Um ein Schlüsselpaar zu erzeugen wählt man ein $d \in \{1, 2, \ldots, n-1\}$ und erzeugt damit dP. d ist dann der private Schlüssel und dP der Öffentliche. Wenn Alice nun Bob eine verschlüsselte Nachricht m zusenden will, braucht sie Bobs öffentlichen Schlüssel. Außerdem ist es bei der ElGamal-Verschlüsselung wichtig, das die Nachricht m mit einem Element aus G identifiziert werden kann. Dafür kann es notwendig sein, dass die Nachricht in geeigneter Weise aufgeteilt wird und die Teile einzeln verschlüsselt werden.

Um m zu verschlüsseln sind folgende Schritte nötig:

1) Alice wählt zufällig eine ganze Zahl $k \in \{1, 2, \ldots, n-1\}$ und berechnet $Q = kP$. Mit Bobs öffentlichem Schlüssel $d_B P$ berechnet sie $R = k(d_B P) + m$.

2) Dann schickt sie das Paar (Q,R) an Bob.

3) Bob nimmt seinen privaten Schlüssel d_B, um $d_B Q = d_B kP$ zu berechnen. Nun kann er die Nachricht m ermitteln, indem er $R - d_B Q = kd_B P + m - d_B kP = m$ ausrechnet.

Wie sähe ein mathematischer Angriff auf dieses Verfahren aus? Ein Angreifer, der die Kommunikation abhören konnte, hat nun insgesamt folgende Daten vorliegen: Bobs öffentlichen Schlüssel $d_B P$ und die abgefangenen Daten $Q = kP$ und $R = kd_B P + m$. Um an m zu gelangen, muss er $kd_B P$ berechnen, also ein DH-Problem lösen können.

Das k, was Alice zufällig wählt, sollte zur Sicherheit bei jeder Nachricht neu gewählt werden. Andernfalls tut sich eine Schwachstelle auf. Wenn nämlich zwei verschiedene Nachrichten abgefangen werden, also etwa (Q, $R_1 = kd_B P + m_1$) und (Q, $R_2 = kd_B P + m_2$), kann $R_1 - R_2 = m_1 - m_2$ berechnet werden. Falls ein Angreifer also m_1 bereits bekannt ist, so erhält er auch m_2 sehr leicht.

1.6 Geeignete Gruppen

Die dargestellten Verfahren aus der Public-Key-Kryptographie (außer RSA) basieren also auf dem DL-Problem. Die Sicherheit der Verfahren hängt wie schon angedeutet maßgeblich von der Wahl der Gruppe ab, in der ein DL- bzw. DH-Problem gelöst werden muss. In der Praxis benötigen wir eine ausreichend große, endliche abelsche Gruppe, in der das DL-Problem schwer zu lösen ist. Eine schlechte Wahl wäre zum Beispiel die additive Gruppe innerhalb eines endlichen Körpers $\mathbb{F}_q$. Haben wir hier zu einem $P \in \mathbb{F}_q$ ein $Q = kP$, so ist entweder $P = Q = 0$ oder $k = \frac{Q}{P}$. Der diskrete Logarithmus lässt sich hier also mit einfacher Division im Körper $\mathbb{F}_q$ berechnen. Die multiplikative Einheitengruppe $\mathbb{F}_q^\times$ ist wiederum besser geeignet. Doch auch hier sind Angriffe möglich, durch Algorithmen, die speziell für diese Gruppe entworfen sind. Dazu später mehr.

Seit 1985 werden elliptische Kurven bzw. gewisse Gruppen über den Punktemengen $E(\mathbb{F}_q)$ einer elliptischen Kurve als Kandidaten für verschiedene Public-Key-Verfahren untersucht. Ein Vorteil von Elliptic Curve (ECC)-Verfahren ist, dass manche DL-Problem-Algorithmen für $\mathbb{F}_q^\times$ nicht auf den EC-Gruppen funktionieren, was sie widerstandsfähiger gegen Angriffe macht.

Im Kapitel 4 gehen wir noch genauer auf die praktischen Vorteile von elliptischen Kurven in der Kryptographie ein. Im nächsten Kapitel werden wir die mathematischen Grundlagen legen, um dann die elliptischen Kurven einführen zu können.

2 Mathematische Grundlagen

Um elliptische Kurven verstehen zu können, müssen wir uns zunächst mit affinen und projektiven zweidimensionalen Räumen bzw. Ebenen und den Kurven innerhalb dieser beschäftigen. Es besteht ein Zusammenhang zwischen affinen und projektiven Ebenen, aber es ist wichtig, den Unterschied zwischen beiden und die jeweils spezifischen Eigenschaften zu erkennen. Die Definitionen in diesem Kapitel sind, wenn nicht anders angegeben, in leicht veränderter Form aus [We] entnommen.

2.1 Affine Ebenen

Affine Ebenen werden in [Ka-Ki] 5.3 als Paar $(\mathcal{A}, \mathcal{G})$ definiert, wobei $\mathcal{A}$ eine nicht-leere Menge ist und $\mathcal{G}$ eine Menge von Teilmengen von $\mathcal{A}$ ($\mathcal{G} \subseteq \mathrm{Pot}(\mathcal{A})$). Außerdem müssen folgende Bedingungen erfüllt sein:

A1) Zu je zwei Elementen $a, b \in \mathcal{A}$ mit $a \neq b$ existiert genau ein $G \in \mathcal{G}$ mit $a, b \in G$; wir schreiben $\overline{a, b}$ für dieses G (die Gerade, die durch die beiden Punkte a, b definiert ist).

A2) (**Parallelenaxiom**) Zu $G \in \mathcal{G}$ und $a \in A \setminus G$ existiert genau ein $G' \in \mathcal{G}$ mit $a \in G'$ und $G \cap G' = \varnothing$.

A3) Es gibt drei Punkte $a, b, c \in \mathcal{A}$ mit $c \notin \overline{a, b}$.

$\mathcal{A}$ bezeichnen wir dann als Punktemenge und $\mathcal{G}$ als Geradenmenge. Ein anschauliches und intuitives Beispiel ist die normale Anschauungsebene, mit der man schon in der Schulmathematik zu tun hat. Hier ist $\mathcal{A} = \mathbb{R}^2$ und $\mathcal{G}$ besteht aus allen möglichen Geraden darin. Allgemeiner kann man über beliebige Körper F affine Ebenen definieren mit $\mathcal{A} = F^2$ und

$$\mathcal{G} := \{a + Fb \mid a, b \in F^2, b \neq 0\}$$

Zusammen bezeichnen wir die affine Ebene $(\mathcal{A}, \mathcal{G})$ in diesem Fall auch als $\mathbb{A}^2(F)$.

2.1.1 Affine Kurven

Eine Kurve in einer affinen Ebene definieren wir als Nullstellenmenge eines Polynoms in zwei Variablen.

Definition 2.1 *Es sei f ein Polynom in zwei Variablen x und y mit Koeffizienten $\gamma_{\mu,\nu}$ im Körper F:*

$$f(x, y) = \sum_{i,j \geq 0} \gamma_{i,j} x^i y^j,$$

wobei nur endlich viele der $\gamma_{\mu,\nu}$ gleich null sind. Weiterhin sei $f \neq 0$. Dann nennen wir die Nullstellenmenge

$$C_f(F) := \{(a, b) \in F \times F \mid f(a, b) = 0\}$$

affine ebene Kurve.

Oft schreiben wir auch kurz *affine Kurve*, da wir im Rahmen dieser Arbeit keine höherdimensionalen Kurven betrachten.

Als Beispiel betrachten wir das Polynom f:

$$f(x, y) := y^2 - x^3 - x$$

mit Koeffizienten im endlichen Körper $\mathbb{F}_p$ für eine Primzahl p. Für verschiedene p erhalten wir unterschiedliche Lösungen für die Gleichung

$$y^2 = x^3 + x, \tag{2.1}$$

welche den Nullstellen von f entsprechen. Zunächst halten wir fest, dass $(0, 0)$ immer eine Lösung ist; egal für welches p. Ob ein bestimmter Punkt in $\mathcal{A}(\mathbb{F}_p)$ ein Element der Nullstellenmenge $C_f(\mathbb{F}_p)$ ist, finden wir heraus, indem wir prüfen, ob im entsprechenden Körper die Auswertung von $x^3 + x$ für ein bestimmtes x ein Quadrat im Körper ist.

Für $p = 3$ ergibt sich $C_f(F) = \{(0, 0), (2, 1), (2, 2)\}$. Denn für $x = 1$ ergibt sich $1^3 + 1 = 2$. In $\mathbb{F}_3$ ist 2 jedoch kein Quadrat (es gibt kein $b \in \mathbb{F}_3$ mit $b^2 = 2$), was man hier durch ausprobieren schnell herausbekommt. In größeren Körpern hilft auch das quadratische Reziprozitätsgesetz (siehe [We] 6.3.5) bei der Bestimmung der Punkte. Bleibt noch $x = 2$: $2^3 + 2 = 10 \equiv 1 \mod 3$. Diesen Wert bekommen wir ebenfalls beim Quadrieren von 1 und 2 in $\mathbb{F}_3$ ($1^2 = 1$ bzw. $2^2 = 4 \equiv 1 \mod 3$).

Um später elliptische Kurven definieren zu können, benötigen wir noch den Begriff der *(Nicht-)Singularität* einer Kurve. Dafür betrachten wir zu einer Kurve $C_f(F)$ zusätzlich die Kurve $C_f(\overline{F})$ über dem algebraischen Abschluss von F. Sowohl die Koeffizienten von f, als auch die Lösungen für f können dann in dem F enthaltenden Körper $\overline{F}$ liegen. Es ist also

$$C_f(F) \subseteq C_f(\overline{F}).$$

Wir definieren nun:

Definition 2.2 *i) Die ebene affine Kurve $C_f(F)$ heißt singulär in dem Punkt $(a, b) \in C_f(F)$, falls beide partiellen Ableitungen von f in (a, b) null sind. Das heißt, dass (a, b) ein Punkt in $\mathbb{A}^2(F)$ ist für den gilt $f(a, b) = 0$, $\frac{\partial f}{\partial x}(a, b) = 0$ und $\frac{\partial f}{\partial y}(a, b) = 0$.*

ii) $C_f(F)$ heißt nicht-singulär, falls die Kurve $C_f(\overline{F})$ in keinem Punkt (a,b) singulär ist. Mit anderen Worten, es gibt keinen Punkt $(a, b) \in \mathbb{A}^2(\overline{F})$, in dem die drei Polynome $f, \frac{\partial f}{\partial x}$ und $\frac{\partial f}{\partial y}$ gleichzeitig null sind.

$C_f(F)$ kann also singulär sein, obwohl sie keine singulären Punkte enthält, sondern nur die übergeordnete Kurve $C_f(\overline{F})$. Als Beispiel soll uns die Kurve $C_f(\mathbb{R})$ über $\mathbb{R}$ für

$$f(x, y) := y^2 - x^4 - 2x^2 - 1$$

dienen. Die partiellen Ableitungen sind

$$\frac{\partial f}{\partial x} = -4x(x^2 + 1) \text{ und } \frac{\partial f}{\partial y} = 2y.$$

Im Reellen haben die drei Polynome $f, \frac{\partial f}{\partial x}$ und $\frac{\partial f}{\partial y}$ keine gemeinsame Nullstelle. In den komplexen Zahlen $\mathbb{C}$, dem algebraischen Abschluss von $\mathbb{R}$, finden sich jedoch $(i, 0)$ und $(-i, 0)$ als singuläre Punkte von f, sodass $C_f(\mathbb{R})$ keine nicht-singuläre Kurve ist.

2.2 Projektive Ebenen

Wie affine Ebenen sind projektive Ebenen definiert als ein Mengenpaar, bestehend aus der Menge $\mathcal{P}$ und dem Teilmengensystem $\mathcal{G}$ von $\mathcal{P}$, d.h. $\mathcal{G} \subseteq \mathrm{Pot}(\mathcal{P})$ ([Ka-Ki] 13.1). Dieses Paar heißt projektive Ebene, wenn folgende Eigenschaften gegeben sind:

P1) Zu je zwei Punkten $P, Q \in \mathcal{P}$ mit $P \neq Q$ existiert genau eine Gerade $G \in \mathcal{G}$ mit $P, Q \in G$.

P2) Für je zwei Geraden $G, H \in \mathcal{G}$ gilt $|G \cap H| = 1$.

P3) Es gibt vier Punkte, von denen drei nicht auf einer Geraden liegen.

Die Definitionen von affinen und projektiven Ebenen unterscheiden sich hauptsächlich im zweiten Punkt. Der Unterschied lässt sich auf die einfache Aussage herunterbrechen, dass es im Affinen parallele Geraden gibt, im Projektiven jedoch nicht. Hier schneiden sich je zwei Geraden immer in genau einem Punkt. Wenn wir später ein Gruppengesetz auf elliptischen Kurven definieren, benötigen wir dazu genau diese Eigenschaft einer projektiven Ebene.

Wir wollen uns zunächst die affinen Kurven, d.h. die Nullstellen von Polynomen in zwei Variablen, genauer ansehen und anschließend zeigen, dass die daraus hervorgehende Definition einer projektiven Ebene den obigen Anforderungen entspricht. So können wir die projektive Ebene als Erweiterung zur affinen Ebene betrachten. Dies wird sich auch bei späteren Anwendungen als nützlich erweisen. Wie wir bereits gesehen haben, ist die affine Kurve $C_f(F)$ für unser Beispielpolynom $f(x, y) = y^2 - x^3 - x$ als Lösungsmenge der Gleichung

$$y^2 = x^3 + x \tag{2.2}$$

definiert. Um von der affinen Kurve $C_f(F)$ zu einer projektiven Kurve zu gelangen müssen wir uns die Lösungen von (2.2) genauer anschauen und sie ergänzen. Sei $(a, b) \in \mathbb{A}^2(F)$ eine Lösung von (2.2); es gilt also $b^2 = a^3 + a$. Jetzt nehmen wir eine beliebe Zahl $c \neq 0$ in F und definieren $a' = ac$ und $b' = bc$. Dann gilt:

$$\left(\frac{b'}{c}\right)^2 = \left(\frac{a'}{c}\right)^3 + \left(\frac{a'}{c}\right)$$

Wenn man nun mit c^3 multipliziert, erhält man $b'^2 c = a'^3 + a' c^2$ und sieht, dass das Tripel $(a', b', c) \in F^3$ eine Lösung folgender Gleichung in drei Variablen ist:

$$Y^2 Z = X^3 + X Z^2 \tag{2.3}$$

Eine allgemeine Lösung $(a, b, c) \in F^3$ von (2.3) lässt sich zu einer Lösung von (2.2) umformen wenn c ungleich 0 ist. Denn wenn man bei $b^2 c = a^3 + ac^2$ wieder durch c^3 teilt,

erhält man eine Lösung von (2.2) (nämlich $\left(\frac{a}{c}, \frac{b}{c}\right)$). Ist anderenfalls $c = 0$, so ergibt sich $0 = a^3$, also muss auch $a = 0$ sein und b darf beliebige Werte annehmen. Offensichtlich findet die Lösung von (2.3) in diesem Fall keine Entsprechung in (2.2).

Wenn man die Lösungen so beschreibt, bemerken wir, dass für ein beliebiges $t \neq 0$ aus F zu jeder Lösung (a, b, c) von (2.3) auch (ta, tb, tc) eine Lösung ist. Denn es ist

$$\begin{aligned} & (t^2b^2)tc = t^3a^3 + ta(t^2c^2) \\ \iff\ & t^3(b^2)c = t^3(a^3 + ac^2) \\ \iff\ & b^2c = a^3 + ac^2 \end{aligned}$$

Das liegt an der Homogenität von (2.3). Allgemein definieren wir:

Definition 2.3 *Es sei g ein Polynom in drei Variablen X, Y und Z über F. Dann heißt g* homogen *vom Grad d, falls gilt:*

$$g(X, Y, Z) = \sum_{i,j,k \geq 0} \gamma_{i,j,k} X^i Y^j Z^k$$

mit Koeffizienten $\gamma_{i,j,k}$, die nicht alle null sind, und für die $i + j + k = d$ ist, wenn $\gamma_{i,j,k} \neq 0$ ist.

Kommen wir zurück zur Betrachtung von Lösungen von (2.2), die aus der erweiterten Gleichung (2.3) kommen. Wenn c ungleich 0 ist, so ist auch $tc \neq 0$. Und da $\left(\frac{a}{c}, \frac{b}{c}\right) = \left(\frac{ta}{tc}, \frac{tb}{tc}\right)$ ist, wird klar, dass jede Lösung (ta, tb, tc) von (2.3) zu derselben Lösung von (2.2) führt. Deswegen ist es sinnvoll diese Vielfachen zu identifizieren.

Definition 2.4 *i) Wir nennen (a, b, c) und (a', b', c') aus $F \times F \times F$ äquivalent und schreiben $(a, b, c) \sim (a', b', c')$, falls es ein $t \in F \setminus \{0\}$ gibt mit*

$$a = ta', \quad b = tb' \text{ und } c = tc'.$$

ii) Wir definieren den zweidimensionalen projektiven Raum, bzw. die projektive Ebene $\mathbb{P}^2(F)$ als den Quotienten von $F \times F \times F \setminus \{(0, 0, 0)\}$ nach der Äquivalenzrelation $\sim$:

$$\mathbb{P}^2(F) := (F \times F \times F \setminus \{(0, 0, 0)\}) / \sim$$

Die Punkte in der projektiven Ebene sind also Äquivalenzklassen, die durch ein Tripel $(a, b, c) \neq (0, 0, 0)$ beschrieben werden können. Die Tripel sind deren Repräsentanten und werden im Folgenden mit $[a : b : c]$ bezeichnet. Es gilt $[a : b : c] = [a' : b' : c']$ genau dann, wenn $a = ta', b = tb'$ und $c = tc'$ für ein $t \neq 0$ ist.

Wie lassen sich nun Geraden in der projektiven Ebene darstellen? Wir definieren wie folgt.

Definition 2.5 *Ist $g \in F[X,Y,Z]$ ein homogenes Polynom vom Grad 1, also*

$$g(X,Y,Z) = \alpha X + \beta Y + \gamma Z,$$

für α, β und γ in F, die nicht alle gleichzeitig null sind, so nennen wir die Nullstellenmenge von g projektive Gerade und schreiben dafür $L(\alpha, \beta, \gamma)$.

Nun können wir zeigen, dass die so definierte projektive Ebene und ihre projektiven Geraden die Bedingungen aus der obigen Definition (P1, P2, P3) erfüllen.

Lemma 2.6 *$\mathbb{P}^2(F)$ ist eine projektive Ebene.*

Beweis: (nach [We] S. 33 f.)

P1) Es seien $P = [a_1 : b_1 : c_1]$ und $Q = [a_2 : b_2 : c_2]$ zwei verschiedene Punkte aus $\mathbb{P}^2(F)$. Um die gemeinsame projektive Gerade zu finden, auf der die beiden Punkte liegen, müssen wir ein Tripel $(\alpha, \beta, \gamma) \neq (0,0,0)$ finden, dass die Gleichungen

$$\alpha a_1 + \beta b_1 + \gamma c_1 = 0 \text{ und } \alpha a_2 + \beta b_2 + \gamma c_2 = 0$$

erfüllt. Daraus lässt sich die Koeffizientenmatrix $\begin{pmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \end{pmatrix}$ ableiten. Da $P \neq Q$ ist, sind die beiden Zeilen der Matrix linear unabhängig. Somit hat die Matrix den Rang 2 und es bleibt nach der Dimensionsformel für lineare Abbildungen ein eindimensionaler Unterraum als Lösungsraum in F^3 übrig. Dieser wird durch einen Punkt in $\mathbb{P}^2(F)$ bzw. ein Tripel $(\alpha, \beta, \gamma) \neq (0,0,0)$ definiert, sodass $P \in L(\alpha, \beta, \gamma)$ und $Q \in L(\alpha, \beta, \gamma)$. Jedes andere Tripel $(\alpha', \beta', \gamma')$, dass die obigen Gleichungen erfüllt, muss ein Vielfaches von (α, β, γ) sein (also den gleichen Punkt in $\mathbb{P}^2(F)$ beschreiben), da der Lösungsraum sonst nicht mehr eindimensional wäre.

P2) Betrachten wir nun zwei verschiedene projektive Geraden $G_1 := L(\alpha_1, \beta_1, \gamma_1)$ und $G_2 := L(\alpha_2, \beta_2, \gamma_2)$. Wir können wieder eine Matrix $\begin{pmatrix} \alpha_1 & \beta_1 & \gamma_1 \\ \alpha_2 & \beta_2 & \gamma_2 \end{pmatrix}$ mit Rang 2 angeben, denn da G_1 und G_2 verschieden sind, sind $(\alpha_1, \beta_1, \gamma_1)$ und $(\alpha_2, \beta_2, \gamma_2)$ linear unabhängig. Der Kern der Matrix ist also eindimensional und enthalte den Vektor $\begin{pmatrix} a \\ b \\ c \end{pmatrix} \neq 0$, welcher mit dem projektiven Punkt $[a : b : c]$ identifiziert werden kann. Dies ist genau der einzige Schnittpunkt von G_1 und G_2.

P3) Die vier Punkte $[1:0:0], [0:1:0], [0:0:1]$ und $[1:1:1]$ erfüllen die Bedingung.

□

Um zu sehen, dass die projektive Ebene eine Erweiterung der affinen Ebene ist, betrachten wir folgende Abbildung:

$$i : \mathbb{A}^2(F) \to \mathbb{P}^2(F)$$
$$(a, b) \mapsto [a : b : 1]$$

Man sieht leicht, dass i injektiv ist, denn für $i(a, b) = i(a', b')$, also $[a : b : 1] = [a' : b' : 1]$, gibt es ein $t \in F$ mit $a = ta', b = tb'$ und $1 = t1$. Dafür kommt nur $t = 1$ in Frage und somit folgt wiederum $(a, b) = (a', b')$. Wir können i also als Einbettung von $\mathbb{A}^2(F)$ in $\mathbb{P}^2(F)$ verstehen. Jetzt stellt sich die Frage, welche Punkte noch in $\mathbb{P}^2(F)$ enthalten sind, die nicht schon in der affinen Ebene liegen. Alle Punkte $[a : b : c]$ können für $c \neq 0$ als $[\frac{a}{c} : \frac{b}{c} : 1]$ geschrieben werden und somit mit $i\left(\frac{a}{c}, \frac{b}{c}\right)$ identifiziert werden. Andersherum kann auch kein Punkt $[a : b : 0] = i(a', b')$ sein. Neben dem Bild von i enthält die projektive Ebene $\mathbb{P}^2(F)$ also genau die Punkte $[a : b : 0]$, wobei a und b nicht beide 0 sein dürfen. Man bezeichnet sie auch als „unendlich ferne Punkte“.

2.2.1 Graphische Veranschaulichung

Nun wollen wir diesen Begriff und den Zusammenhang zwischen projektiven und affinen Ebenen veranschaulichen. Dazu machen wir uns folgendes klar: Damit die Bedingung erfüllt werden kann, dass sich je zwei Geraden in der projektiven Ebene in genau einem Punkt schneiden (P2), müssen die Geraden, die in der affinen Ebene parallel sind (also keinen Schnittpunkt haben), einen zusätzlichen Punkt im „Unendlichen“ bekommen. In diesem unendlich fernen Punkt schneiden sich dann diese Geraden. Eine gute Vorstellung davon liefert der Blick auf Eisenbahnschienen, welche sich am Horizont zu berühren scheinen (schematische Darstellung in Abbildung 1). Für jede Geradenrichtung gibt es solch einen Punkt. Alle unendlich fernen Punkte zusammengenommen bilden die sogenannte unendlich ferne Gerade.

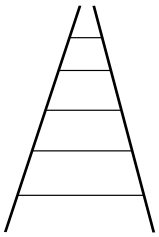

Abbildung 1: Schienen (schematisch)

Räumliche Darstellung

Räumlich veranschaulichen lässt sich die projektive Ebene am Beispiel der reellen Zahlen. Die hier erläuterte Veranschaulichung orientiert sich frei an [Rü], wobei die Grafiken selbst erstellt wurden.

Die Abbildung i kann als Anheben der affinen Ebene (genauer $\mathbb{A}^2(\mathbb{R})$) um eins nach oben im dreidimensionalen Raum $\mathbb{R}^3$ verstanden werden. In den nachfolgenden Abbildungen ist sie als braune Gitterfläche dargestellt. Jeder Punkt $(a, b, 1)$ in der angehobenen Ebene definiert mit dem Punkt $(0, 0, 0)$ eine eindeutige Ursprungsgerade (siehe Abbildung 2). Ebenso definiert eine Gerade in der affinen Ebene , welche durch zwei Punkte $(a_1, b_1, 1)$ und $(a_2, b_2, 1)$ verläuft, eine Ursprungsebene durch diese beiden Punkte (Abbildung 3). Auf diese Weise lassen sich Punkte in der projektiven Ebene als Ursprungsgeraden im $\mathbb{R}^3$ und projektive Geraden als Ursprungsebenen im $\mathbb{R}^3$ verstehen.

Wenn sich nun zwei Geraden in der affinen Ebene schneiden und einen Schnittpunkt S definieren, so schneiden sich die zwei zugehörigen Ursprungsebenen genau in der Geraden, die durch $(0, 0, 0)$ und den Schnittpunkt S verläuft (Abbildung 4). Wie verhält es sich mit den Geraden, die in der affinen Ebene parallel sind? Auch hier schneiden sich die entsprechenden Ursprungsebenen und definieren eine Schnittgerade. Diese hat aber keinen Schnittpunkt mit der affinen Ebene sondern liegt ganz in der x,y-Ebene ($z = 0$), wie in Abbildung 5 zu sehen ist. So bekommen alle Parallelen aus der affinen Ebene einen eindeutigen Schnittpunkt im Projektiven. Durch Hinzunahme dieser Schnittpunkte zu der affinen Ebene entsteht die projektive Ebene. Wenn man nun alle Ursprungsgeraden im $\mathbb{R}^3$ betrachtet, sieht man, dass sie sich in zwei Kategorien einteilen lassen. Solche, die einen Schnittpunkt mit $\mathbb{A}^2(\mathbb{R})$ haben und solche, die keinen haben. Die Ursprungsgeraden ohne Schnittpunkt mit $\mathbb{A}^2(\mathbb{R})$ sind die unendlich fernen Punkte, die zusammen die unendlich ferne Gerade in $\mathbb{P}^2(\mathbb{R})$ bilden (hier repräsentiert durch die x,y-Ebene).

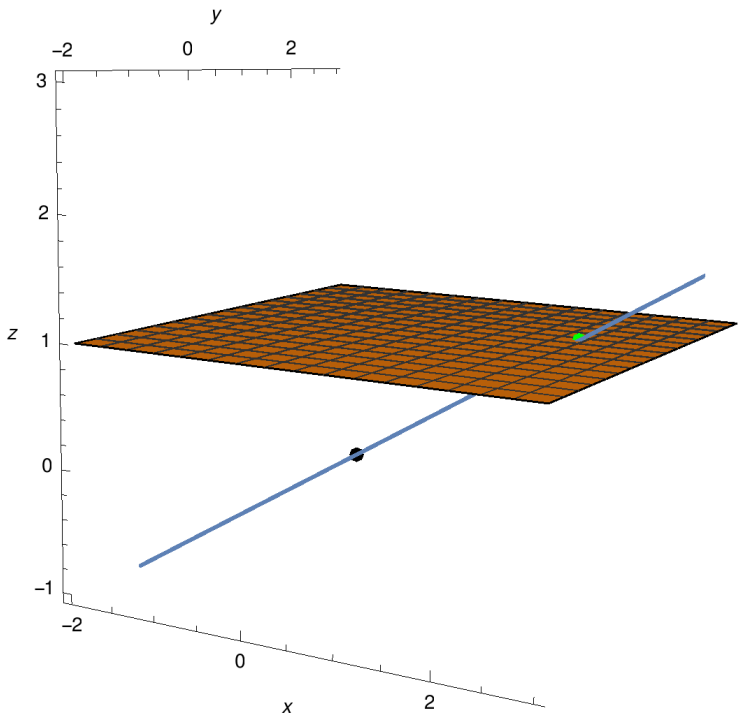

Abbildung 2: Projektiver Punkt als Ursprungsgerade

Zusammenfassend lässt sich sagen, dass die projektive Ebene durch Hinzunahme einer Geraden entsteht, die mit den Parallelen der affinen Ebene Schnittpunkte bildet. Dieser Umstand wird sich später auch bei den Berechnungen auf elliptischen Kurven als nützlich erweisen.

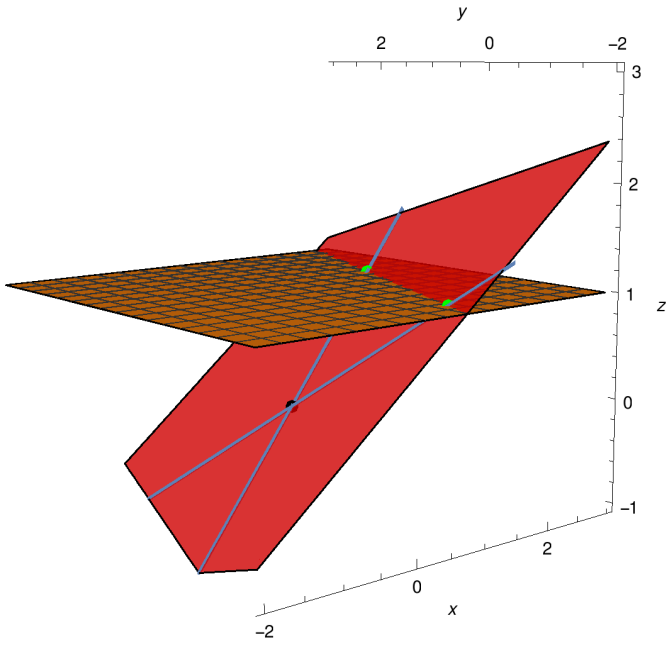

Abbildung 3: Projektive Gerade als Ursprungsebene

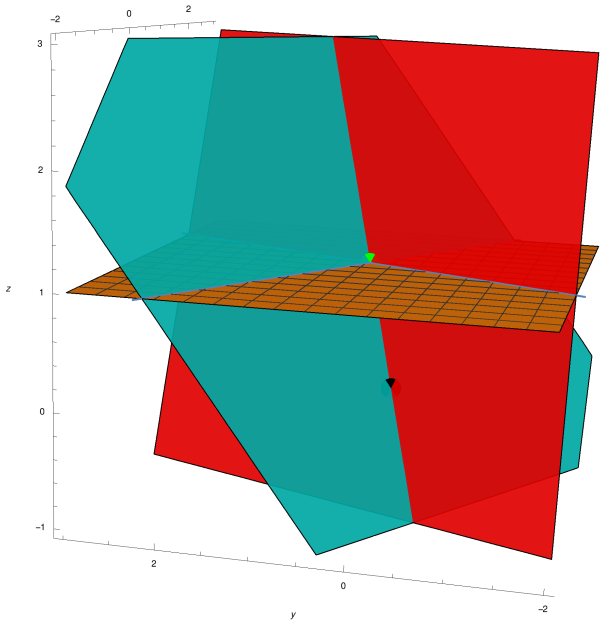

Abbildung 4: Projektive Geraden, die sich in $\mathbb{A}^2(F)$ schneiden

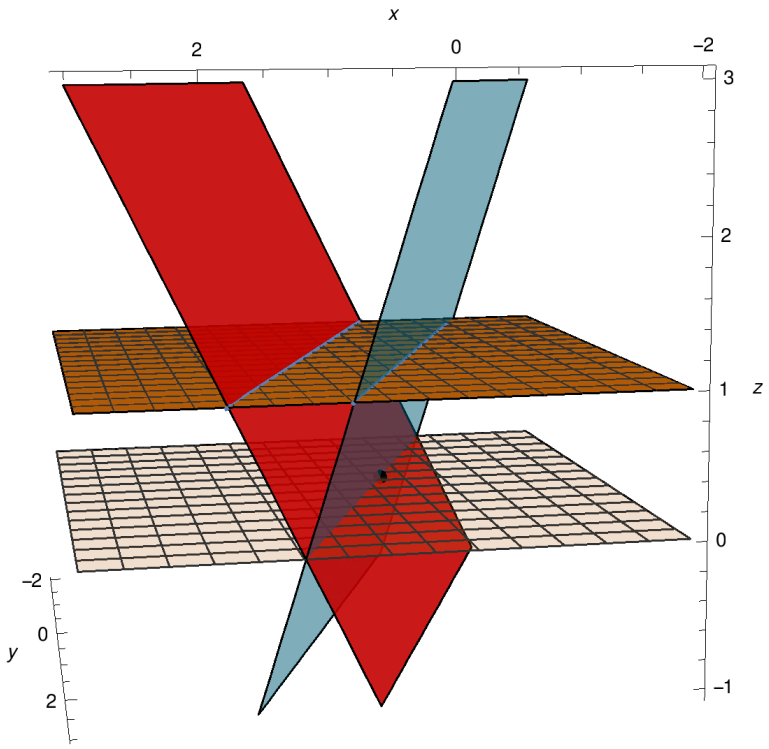

Abbildung 5: Projektive Geraden, die sich im „Unendlichen“ schneiden

2.2.2 Projektive Kurven

Nachdem wir uns mit der projektiven Ebene vertraut gemacht haben definieren wir nun analog zur Definition affiner Kurven projektive Kurven als Nullstellenmenge bestimmter Polynome:

Definition 2.7 *Sei g ein homogenes Polynom in $F[X,Y,Z]$. Dann bezeichnen wir die Menge der Nullstellen von g in $\mathbb{P}^2(F)$ als $C_g(F)$:*

$$C_g(F) = \left\{ [a:b:c] \in \mathbb{P}^2(F) \mid g(a,b,c) = 0 \right\}.$$

Jede solche Nullstellenmenge $C_g(F)$ nennen wir eine projektive ebene Kurve.

Auch hier schreiben wir kurz *projektive Kurve*. Die geforderte Homogenität des Polynoms g sorgt für die Korrektheit bezüglich der projektiven Punkte, denn es gilt folgendes Lemma:

Lemma 2.8 *Ist $g \in F[X,Y,Z]$ ein homogenes Polynom vom Grad d, so gilt für alle a,b,c aus F und $t \in F \setminus \{0\}$:*

$$g(a,b,c) = 0 \iff g(ta,tb,tc) = 0.$$

Somit ist es egal, wie wir einen Punkt auf $C_g(F)$ in $\mathbb{P}^2(F)$ schreiben. So wie die projektive Ebene als Erweiterung der affinen Ebene betrachtet werden kann, sind auch projektive Kurven eine Erweiterung ihrer affinen Verwandten. Dazu sehen wir uns nochmal das Beispiel von oben an:

$$f(x, y) = y^2 - x^3 - x, \tag{2.4}$$

die zugehörige affine Kurve

$$C_f(F) = \{(a, b) \in F^2 \mid b^2 = a^3 + a\}$$

und die projektive Kurve $C_g(F)$ mit

$$g(X, Y, Z) = Y^2Z - X^3 - XZ^2. \tag{2.5}$$

Wir haben gesehen, dass jeder Punkt auf $C_f(F)$, also jede Nullstelle von (2.4) auch ein Punkt auf $C_g(F)$ ist (als Nullstelle von (2.5)). Für $(a, b) \in C_f(F)$ ist nämlich $[ac : bc : c] \in C_g(F)$, wobei $c \neq 0$ beliebig ist. Den projektiven Punkt $[ac : bc : c]$ können wir auch als $[a : b : 1]$ schreiben und bemerken, dass Punkte dieser Art genau dem Bild der Abbildung $i : \mathbb{A}^2(F) \to \mathbb{P}^2(F)$ entsprechen. Die Punkte $[a : b : c] \in C_g(F)$ auf der projektiven Kurve können also mit den Punkten aus der affinen Ebene identifiziert werden wenn $c \neq 0$ ist. Ist c hingegen gleich 0, so muss a ebenfalls 0 sein und b darf beliebige Werte annehmen. Dies ist der Punkt $[0 : b : 0] = [0 : 1 : 0]$. Die projektive Kurve $C_g(F)$ setzt sich somit folgendermaßen zusammen:

$$C_g(F) = i(C_f(F)) \cup \{[0 : 1 : 0]\}$$

Unter i wird $C_f(F)$ in $C_g(F)$ eingebettet. Der umgekehrte Vorgang, also das Schneiden von $C_g(F)$ mit $\mathbb{A}^2(F)$, wird auch als Übergang zu affinen Koordinaten bezeichnet.

Allgemein lässt sich aus einer affinen Kurve eine projektive konstruieren, indem man dem Polynom der affinen Kurve eine zusätzliche Variable hinzufügt, und die einzelnen Monome so ergänzt, dass ein homogenes Polynom in drei Variablen entsteht. Umgekehrt kommen wir von der Gleichung einer projektiven Kurve zu der entsprechenden affinen Kurvengleichung indem wir $Z = 1$ setzen:

Proposition 2.9 *Sei $f \neq 0$ ein beliebiges Polynom in $F[x, y]$ von Grad d. Für*

$$f(x, y) = \sum_{i,j \geq 0} \gamma_{i,j} x^i y^j \text{ ist dann}$$
$$g(X, Y, Z) = \sum_{i,j \geq 0, i+j \leq d} \gamma_{i,j} X^i Y^j Z^{d-i-j}$$

homogen vom Grad d und erfüllt $g(a, b, 1) = f(a, b)$ für alle $(a, b) \in \mathbb{A}^2(F)$.

Als letzten Schritt, bevor wir elliptische Kurven definieren können, benötigen wir noch die Definition singulärer Punkte auf projektiven Kurven, welche analog zur Definition ihrer affinen Verwandten ist.

Definition 2.10 *Sei g ein homogenes Polynom in $F[X,Y,Z]$ vom Grad d.*

i) Die projektive Kurve $C_g(F)$ heißt singulär im Punkt $P = [a : b : c] \in C_g(F)$, falls alle partiellen Ableitungen von g in P gleich null sind, d.h.

$$\frac{\partial g}{\partial X}(a,b,c) = \frac{\partial g}{\partial Y}(a,b,c) = \frac{\partial g}{\partial Z}(a,b,c) = 0$$

ii) $C_g(F)$ heißt nicht-singulär, falls $C_g(\overline{F})$ keinen singulären Punkt enthält.

Außerdem besteht folgender Zusammenhang zu affinen Kurven.

Lemma 2.11 *Sei für ein homogenes Polynom $g(X,Y,Z)$ vom Grad d das im Sinne von Proposition 2.9 zugehörige Polynom $f(x,y)$. Für jeden Punkt $P \in C_g(F)$ gilt: Falls $P = i(Q)$ in $i(\mathbb{A}^2(F))$ liegt, so ist $C_g(F)$ singulär in P genau dann, wenn die affine Kurve $C_f(F)$ singulär in Q ist.*

(Beweis, siehe [We] S.21)

3 Elliptische Kurven

Nachdem wir nun alle nötigen Vorüberlegungen getroffen haben, können wir uns den elliptischen Kurven zuwenden. Es sind bestimmte projektive Kurven, die besondere Eigenschaften haben, welche es ermöglichen ein Gruppengesetz auf ihren Punktemengen zu definieren. Dazu später mehr; zunächst definieren wir angelehnt an die Definitionen in [We] 2.3:

Definition 3.1 *Eine elliptische Kurve ist eine nicht-singuläre projektive Kurve $C_g(F)$, wobei g ein homogenes Polynom vom Grad drei der folgenden Gestalt ist:*

$$g(X,Y,Z) = Y^2Z + a_1XYZ + a_3YZ^2 - X^3 - a_2X^2Z - a_4XZ^2 - a_6Z^3$$

mit a_1, a_2, a_3, a_4 und $a_6 \in F$.

Die Gleichung

$$Y^2Z + a_1XYZ + a_3YZ^2 = X^3 + a_2X^2Z + a_4XZ^2 + a_6Z^3, \tag{3.1}$$

deren Lösungen den Punkten auf einer elliptischen Kurve entsprechen, wird Weierstraß-Gleichung genannt. Die Bezeichnung der Koeffizienten a_i hat historische Gründe und wird so einheitlich in der Literatur über elliptische Kurven verwendet.

Das Beispiel $C_g(F)$ mit $g(X,Y,Z) = Y^2Z - X^3 - XZ^2$ im vorigen Kapitel ist eine elliptische Kurve mit den Koeffizienten $a_4 = 1$ und $a_i = 0$ für $i \in \{1,2,3,6\}$. Dort haben wir gesehen, dass $C_g(F)$ aus der affinen Kurve $C_f(F)$ hervor geht, unter Hinzunahme eines zusätzlichen unendlich fernen Punktes ($[0:1:0]$). Wie man leicht nachrechnen kann, gilt das im Allgemeinen für elliptische Kurven. Alle Punkte $[a:b:c]$ einer elliptischen Kurve $C_g(F)$ sind für $c \neq 0$ durch die Abbildung i mit Punkten aus der affinen Ebene $\mathbb{A}^2(F)$ identifizierbar, da $C_g(F)$ insbesondere eine projektive Kurve ist. Setzen wir nun $P = [u:v:0] \in \mathbb{P}^2(F)$ in die Weierstraß-Gleichung ein, erhalten wir $u^3 = 0$. Also ist $v \neq 0$ und

$$P = [0:v:0] = [0:1:0].$$

Aufgrund dessen zeichnet man den Punkt $[0:1:0]$ aus und gibt ihm die Bezeichnung $\mathcal{O}$. Damit gilt für jede elliptische Kurve $C_g(F)$,

$$C_g(F) = i(C_f(F)) \cup \{\mathcal{O}\},$$

wobei $C_f(F)$ die zugehörige affine Kurve mit

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

ist (vgl. Proposition 2.9).

3.1 Singularitäts-Kriterium

Nun sind nicht alle projektiven Kurven $C_g(F)$ der Form (3.1) elliptische Kurven. Sie müssen nach Definition 3.1 nicht-singulär sein, dürfen also keine singulären Punkte im algebraischen Abschluss des zugrunde liegenden Körpers enthalten. Zunächst findet man beim Punkt $\mathcal{O}$, dass $\frac{\partial g}{\partial Z}(\mathcal{O}) = 1$ ist. Somit ist $\mathcal{O}$ auf allen elliptischen Kurven nicht-singulär. Da dies der einzige nicht-affine Punkt auf der Kurve ist, können wir uns auf die Betrachtung der zugehörigen affinen Kurve $C_f(F)$ beschränken (siehe Lemma 2.11).

Die Singularität einer elliptischen Kurve wird durch die Wahl der Koeffizienten a_i und des jeweiligen Körpers F, über den die Kurve definiert ist, bestimmt und hängt mit der Diskriminante des Kurven-Polynoms g, bzw. f zusammen. Doch zunächst wollen wir die etwas sperrige Weierstraß-Gleichung umformen und so die nachfolgenden Berechnungen vereinfachen. Stellt man bestimmte Anforderungen an den Körper F, lassen sich Bijektionen für eine elliptische Kurve $C_g(F)$ (mit g wie in Definition 3.1) finden. Hier nennen wir die beiden wichtigsten:

1. Falls die Charakteristik[4] von F ungleich 2 ist, so ist die Abbildung

$$\begin{aligned} \Phi : \mathbb{P}^2(F) &\to \mathbb{P}^2(F) \\ [r : s : t] &\mapsto [r : s + \frac{a_1}{2}r + \frac{a_3}{2}t : t] \end{aligned}$$

   bijektiv und es gilt

$$\Phi(C_g(F)) = C_{h_1}(F)$$

   mit $h_1(X, Y, Z) = Y^2Z - X^3 - \frac{1}{4}b_2X^2Z - \frac{1}{2}b_4XZ^2 - \frac{1}{4}b_6Z^3$, mit neuen Koeffizienten $b_2 = a_1^2 + 4a_2$, $b_4 = 2a_4 + a_1a_3$ und $b_6 = a_3^2 + 4a_6$. $C_{h_1}(F)$ ist ebenfalls eine elliptische Kurve.

2. Falls die Charakteristik von F ungleich 2 und ungleich 3 ist, so ist die Abbildung

$$\begin{aligned} \Psi : \mathbb{P}^2(F) &\to \mathbb{P}^2(F) \\ [r : s : t] &\mapsto [36r + 3b_2t : 216s : t] \end{aligned}$$

   bijektiv und es gilt

$$\Psi(C_{h_1}(F)) = C_{h_2}(F)$$

   mit $h_2(X, Y, Z) = Y^2Z - X^3 + 27c_4XZ^2 + 54c_6Z^3$, mit neuen Koeffizienten $c_4 = b_2^2 - 24b_4$ und $c_6 = -b_2^3 + 36b_2b_4 - 216b_6$. $C_{h_2}(F)$ ist ebenfalls eine elliptische Kurve.

Den Beweis findet man z.B. in [We] 2.3.

Für $\mathrm{char}(F) \neq 2, 3$ erhalten wir durch die Verknüpfung von Φ und Ψ die sogenannte kurze Weierstraß-Gleichung

$$Y^2Z = X^3 + aXZ^2 + bZ^3 \tag{3.2}$$

[4] Entweder 0, oder bei endlichen Körpern die kleinste natürliche Zahl $n > 1$ mit $\sum_{i=1}^{n} 1 = 0$, n prim

und in ihrer affinen Variante $y^2 = x^3 + ax + b$. Wir werden uns im Folgenden auf diese einfache Form beschränken, da in kryptographischen Anwendungen oft mit Körpern von Charakteristik ungleich zwei und drei gearbeitet wird.

Jetzt wollen wir untersuchen, unter welchen Bedingungen eine Kurve der Form (3.2) nicht-singulär, also eine elliptische Kurve ist. Dabei hilft uns folgende Proposition:

Proposition 3.2 *Sei $g(X, Y, Z) = Y^2Z - X^3 - aXZ^2 - bZ^3$ ein kurzes Weierstraß-Polynom und F ein Körper mit char(F) $\neq 2, 3$. Dann ist die Kurve $C_g(F)$ nicht-singulär genau dann, wenn die Diskriminante $\Delta = -4a^3 - 27b^2$ ungleich null ist.*

Beweis: Wie wir bereits gesehen haben, ist $C_g(F)$ genau dann singulär, wenn $C_f(F)$ singulär ist, für $f(x, y) = y^2 - x^3 - ax - b$. Bei einem singulären Punkt $P = (r, s)$ auf der Kurve $C_f(F)$ müssen beide partiellen Ableitungen

$$\frac{\partial f}{\partial x} = -3x^2 - a \text{ und } \frac{\partial f}{\partial y} = 2y$$

null sein. Es müssen also folgende Gleichungen gelten

$$s^2 - r^3 - ar - b = 0,\ -3r^2 - a = 0 \text{ und } 2s = 0$$

Daraus folgt sofort, dass $s = 0$ sein muss und wir können uns auf die Betrachtung des Polynoms

$$h(x) = x^3 + ax + b$$

beschränken. Der Punkt $(r, 0)$ ist genau dann singulär, wenn $h(r) = 0$ und $h'(r) = \frac{\partial h}{\partial x}(r) = 0$ ist. Da wir h über dem algebraischen Abschluss $\overline{F}$ betrachten, zerfallen h und h' in Linearfaktoren:

$$h(x) = (x - \alpha_1)(x - \alpha_2)(x - \alpha_3)$$
$$h'(x) = (x - \alpha_2)(x - \alpha_3) + (x - \alpha_1)(x - \alpha_3) + (x - \alpha_1)(x - \alpha_2)$$

wobei $\alpha_i \in \overline{F}$ die Nullstellen von h sind. Da r Nullstelle von h und h' sein soll, können wir ohne Einschränkung $r = \alpha_1$ setzen. Damit fallen beim Einsetzen von r in $h'(x)$ die letzten beiden Terme weg und es bleibt $h'(r) = (r - \alpha_2)(r - \alpha_3)$. h' ist als Ableitung von h von Grad 2 und hat somit genau 2 Nullstellen in $\overline{F}$. Da r, wie gefordert, auch Nullstelle von h' ist, müssen zwei der α_i gleich sein. Mit anderen Worten, wenn r eine Nullstelle von h und dessen Ableitung ist, muss h eine doppelte Nullstelle haben. Diese Eigenschaft kann mithilfe der Diskriminante eines Polynoms bestimmt werden.
Für $h(x) = (x - \alpha_1)(x - \alpha_2)(x - \alpha_3)$ ist sie definiert als

$$D_h = (\alpha_1 - \alpha_2)^2(\alpha_1 - \alpha_3)^2(\alpha_2 - \alpha_3)^2.$$

Hier sieht man sofort, dass die Diskriminante null wird, wenn h eine doppelte Nullstelle hat. Die Umkehrung gilt ebenso. Durch Umformen kann man die Diskriminante auch durch die Koeffizienten eines Polynoms ausdrücken (siehe [Bo] S. 160) und erhält

$$D_h = \Delta = -4a^3 - 27b^2.^\dagger$$

[†] Manchmal wird Δ auch mit einem Normierungsfaktor (in diesem Fall 16) definiert (siehe [Bu] 14.2.1). Da er aber keine Rolle dabei spielt, ob $\Delta = 0$ ist oder nicht, kann er hier auch ignoriert werden.

Die Diskriminante ist somit ein einfaches Kriterium für die Bestimmung der (Nicht-)-Singularität einer Kurve. Von nun an werden wir elliptische Kurven auch mit $E(F)$ bezeichnen und meinen damit eine Kurve mit kurzer Weierstraß-Gleichung wie in (3.2); damit sind g bzw. f implizit gegeben.

3.2 Tangenten und Schnittgeraden

Wenn eine Kurve nicht-singulär ist, sorgt dies vereinfacht gesagt, dafür, dass man an jeden Punkt der Kurve eine Tangente anlegen kann. Diese Eigenschaft ist essentiell, um ein Gruppengesetz auf solchen Kurven definieren zu können. Bevor wir das tun, schauen wir uns die Definition von Tangenten an elliptischen Kurven und die Vielfachheit der Schnittpunkte mit ihnen an.

Definition 3.3 *Sei $C_g(F)$ eine elliptische Kurve und $P = [a : b : c]$ ein Punkt auf $E(F)$. Die projektive Gerade*

$$L\left(\frac{\partial g}{\partial X}(a,b,c), \frac{\partial g}{\partial Y}(a,b,c), \frac{\partial g}{\partial Z}(a,b,c)\right)$$

heißt Tangente in P an $E(F)$.

Tangenten an affinen Punkten kann man sich bildlich gut vorstellen (siehe auch Abbildung 6) und bedürfen keiner weiteren Veranschaulichung. Wie sieht nun aber eine Tangente an einem unendlich fernen Punkt aus? Wir betrachten die Tangente an Punkt $\mathcal{O}$ auf einer beliebigen elliptischen Kurve $E(F)$. Dazu setzen wir $\mathcal{O} = [0 : 1 : 0]$ in die Tangentengleichung ein und erhalten

$$\begin{aligned} &L\left(\frac{\partial g}{\partial X}(0,1,0), \frac{\partial g}{\partial Y}(0,1,0), \frac{\partial g}{\partial Z}(0,1,0)\right) \\ &= L\,(0,0,1) \end{aligned}$$

Damit erhalten wir die Geradengleichung der Tangente: $Z = 0$ (vgl. Definition 2.5). Wir sehen also, dass die Tangente in $\mathcal{O}$ an $E(F)$ die unendlich ferne Gerade ist, auf der auch alle anderen unendlich fernen Punkte ($[a : b : 0]$) liegen. Als nächstes wollen wir uns das Schnittverhalten von projektiven Geraden und elliptischen Kurven genauer anschauen.

Definition 3.4 *Sei $L(\alpha, \beta, \gamma)$ eine projektive Gerade und $E(F)$ eine elliptische Kurve. Wir fixieren einen Punkt $P = [a : b : c] \in L(\alpha, \beta, \gamma)$ und wählen einen beliebigen weiteren Punkt $P' = [a' : b' : c'] \in L(\alpha, \beta, \gamma)$. Dann ist die Vielfachheit, mit der sich $L(\alpha, \beta, \gamma)$ und $E(F)$ in P schneiden, definiert als die Nullstellenordnung in 0 des Polynoms*

$$\psi(t) = g(a + ta', b + tb', c + tc').$$

Wir bezeichnen sie mit $m(P, L(\alpha, \beta, \gamma), E(F))$.

Um die Nullstellenordnung in 0 bestimmen zu können betrachten wir das Polynom ψ in der Form

$$\psi(t) = w_0 + w_1 t + w_2 t^2 + \dots w_l t^l.$$

Der kleinste Index $j \in \{0, \dots, l\}$ bei dem $w_j \neq 0$ ist, ist dann die Nullstellenordnung in 0 von ψ (vgl. [We] 6.5). Anders ausgedrückt, ist es das kleinste j bei der $\psi^{(j)}(0)$ nicht null ist. Sei beispielsweise $\psi(0) = 0,\ \psi'(0) = 0,\ \psi''(0) = 0$, aber $\psi'''(0) = w_3$, so ist die Nullstellenordnung in 0 von ψ gleich 3. Nun gilt $\psi(0) \neq 0$ genau dann, wenn $P \notin E(F)$ ist. Somit ist $m(P, L(\alpha, \beta, \gamma), E(F)) = 0$, für alle $P \notin (L(\alpha, \beta, \gamma) \cap E(F))$. Betrachten wir nun eine beliebige Tangente $L(\alpha, \beta, \gamma)$ an einer elliptischen Kurve $E(F)$ in Punkt $P \in E(F)$. Da der Berührungspunkt P auf der Kurve liegt, ist $\psi(0) = 0$ und die Vielfachheit von P bzgl. L und $E(F)$ ist somit bereits größer als 1. Die erste Ableitung von ψ wird in 0 ebenfalls null, da

$$\psi'(0) = \frac{\partial g}{\partial X}(a, b, c) \cdot a' + \frac{\partial g}{\partial Y}(a, b, c) \cdot b' + \frac{\partial g}{\partial Z}(a, b, c) \cdot c' = 0$$

und der Punkt P' ebenfalls auf der Tangente liegt (vgl. [We] S. 36).
Daher ist $m(P, L, E(F)) \geq 2$. Generell gilt folgender Satz:

Satz 3.5 *Für eine projektive Gerade L und eine elliptische Kurve E(F) gilt: Die Summe der Vielfachheiten*

$$\sum_{P \in \mathbb{P}^2(F)} m(P, L, E(F))$$

ist entweder 0,1 oder 3.

Für den Beweis sei auf [We] S. 37 ff. verwiesen. Mit diesem Satz ist die wesentliche Grundlage für das Gruppengesetz auf elliptischen Kurven gelegt, denn es lässt sich somit folgern:

Korollar 3.6 *Für eine elliptische Kurve E(F) gilt:*

i) Für zwei verschiedene Punkte P und Q auf E(F) sei L die projektive Gerade, die durch beide Punkte verläuft. Dann hat L noch einen dritten Schnittpunkt mit E(F).

ii) Die Tangente L an E(F) im Punkt P ∈ E(F) hat noch einen dritten Schnittpunkt, wenn man P doppelt zählt.

Die Anzahl der Schnittpunkte ergibt sich durch Aufsummierung der jeweiligen Vielfachheiten $m(P, L, E(F))$. Der erste Fall ist leicht ersichtlich, denn wenn es schon zwei verschiedene Schnittpunkte zwischen L und $E(F)$ gibt, sagt uns Satz 3.5, dass es noch einen dritten Schnittpunkt geben muss. Dieser kann wiederum von P und Q verschieden sein, wobei alle drei dann die Vielfachheit 1 hätten, oder aber einer der Schnittpunkte P und Q hat die Vielfachheit 2. Bei der Tangente an einen Punkt P wissen wir schon, dass die Vielfachheit dieses Schnittpunkts mindestens 2 sein muss. Und nach Satz 3.5 gibt es entweder einen weiteren Schnittpunkt $P' \neq P$ mit Vielfachheit 1, oder P ist selbst der „dritte“ Schnittpunkt mit Vielfachheit 3.

Diese Tatsachen sind auch gut an Darstellungen einer elliptischen Kurve in affinen Koordinaten über den reellen Zahlen ersichtlich (siehe Abbildung 6 und 7). Auch wenn wir uns dabei auf die affinen Koordinaten beschränken bleiben die Aussagen aus 3.6 unter gewissen Bedingungen gültig. Wenn nämlich alle drei Schnittpunkte schon im Affinen liegen, gibt es keinen weiteren mehr im Projektiven. Diese Fälle sehen wir in den eben genannten Abbildungen. Wenn es im Affinen nur zwei Schnittpunkte zwischen einer Geraden und der elliptischen Kurve gibt, bspw. bei einer Parallelen zur y-Achse, ist der dritte Schnittpunkt der unendlich ferne Punkt $\mathcal{O}$. Dies gilt für alle Geraden, die zur y-Achse parallel verlaufen (siehe [Ka-Ki] S. 224). Diese Gesetzmäßigkeiten führen uns unmittelbar zum Gruppengesetz auf elliptischen Kurven.

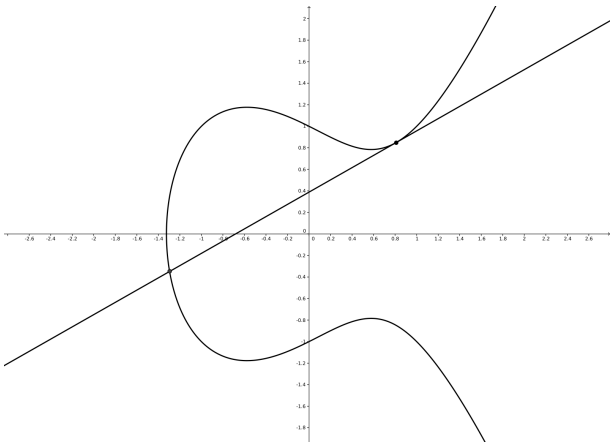

Abbildung 6: Die Tangente an $E(\mathbb{R})$ schneidet $E(\mathbb{R})$ in einem weiteren Punkt

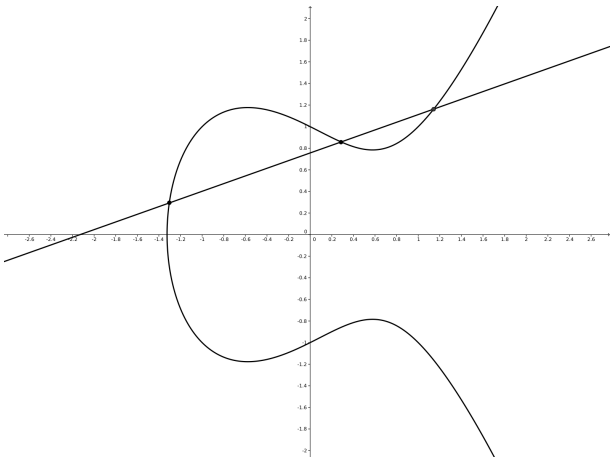

Abbildung 7: Eine Gerade durch zwei Punkte schneidet $E(\mathbb{R})$ in einem dritten Punkt

3.3 Gruppengesetz auf *E(F)*

Mit den obigen Vorbereitungen lässt sich ein Gruppengesetz auf elliptischen Kurven sehr gut geometrisch motivieren. Dazu führen wir eine Punktaddition ein, die sich in zwei Schritte aufgeteilt.

i) Für zwei verschiedene Punkte P und Q auf $E(F)$ legen wir zunächst eine projektive Gerade durch diese beiden Punkte $(\overline{PQ})$. Diese Gerade schneidet $E(F)$ in einem weiteren Punkt R.

ii) Dann legen wir wiederum eine projektive Gerade durch R und $\mathcal{O}$, deren dritter Schnittpunkt schließlich $P + Q$ ist.

Abbildung 8 macht dies sehr gut deutlich.

Im Fall $P = Q$ suchen wir den Punkt $P + P$. Dazu legen wir zunächst eine Tangente an P und betrachten wiederum den dritten Schnittpunkt (R) mit $E(F)$. R wird auch hier als dritter Schnittpunkt bezeichnet, da die Vielfachheit von P bereits mindestens 2 ist. $P + P$ ist nun wieder der dritte Schnittpunkt von $\overline{R\mathcal{O}}$.

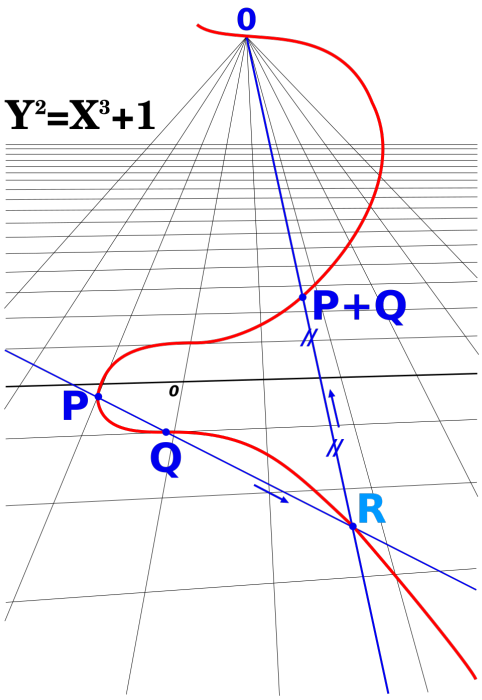

Abbildung 8: Das Gruppengesetz auf $E(\mathbb{R})$[5]

[5]Quelle: Jean Brette derivative work: Beao (talk) - Addition_on_cubic.jpg, CC BY 3.0, https://commons.wikimedia.org/w/index.php?curid=7981011

Wenn bereits R (der dritte Schnittpunkt von $\overline{PQ}$) gleich $\mathcal{O}$ sein sollte, legen wir im zweiten Schritt eine Tangente an $\mathcal{O}$. $P+Q$ ist dann der weitere Schnittpunkt dieser Tangente an $E(F)$. Dies ist, wie wir oben gesehen haben, wieder $\mathcal{O}$, denn die Tangente an $\mathcal{O}$ schneidet $E(F)$ in $\mathcal{O}$ mit Vielfachheit 3, also in keinem anderen Punkt. Auf der Kurve gegenüberliegende Punkte (bei Spiegelung an der x-Achse) haben diese Eigenschaft, denn eine Gerade durch zwei derartige Punkte ist eine Parallele zur y-Achse und hat somit den dritten Schnittpunkt $\mathcal{O}$.

Eine ähnliche Situation ergibt sich für $Q = \mathcal{O}$. Der dritte Schnittpunkt von $\overline{P\mathcal{O}}$ ist der P gegenüberliegende Punkt R (vgl. [We] S. 41). $P + \mathcal{O}$ ist also wieder P, denn die zweite Gerade (aus Schritt *ii)*) entspricht der Ersten ($\overline{P\mathcal{O}} = \overline{R\mathcal{O}}$), wobei hier P der dritte Schnittpunkt ist.

Wir sehen also, dass $P + \mathcal{O} = P$ und außerdem $\mathcal{O} + \mathcal{O} = \mathcal{O}$ ist. $\mathcal{O}$ hat anscheinend die Eigenschaft eines neutralen Elements. Dieser Schein trügt nicht, denn die so beschriebene Addition von Punkten auf $E(F)$ macht diese zu einer abelschen (kommutativen) Gruppe:

Satz 3.7 *Seien $P, Q, R \in E(F)$. Die Verknüpfung*

$$+ : (P, Q) \mapsto P + Q$$

bildet zusammen mit $E(F)$ eine abelsche Gruppe mit neutralem Element $\mathcal{O}$, sodass gilt

i) $P + \mathcal{O} = P$ *für alle* $P \in E(F)$

ii) *Für alle* $P \in E(F)$ *gibt es genau einen Punkt* $-P \in E(F)$ *mit* $P + (-P) = \mathcal{O}$

iii) $P + Q = Q + P$ *für alle* $P, Q \in E(F)$

iv) $(P + Q) + R = P + (Q + R)$

Beweisskizze:

i) Haben wir oben bereits gesehen.

ii) Zwei sich gegenüberliegende Punkte sind zueinander invers (siehe oben). Die Existenz und Eindeutigkeit der Inversen kann man sich dadurch erschließen, dass man beachtet, dass eine elliptische Kurve eine an der x-Achse gespiegelte kubische Funktion darstellt. Für einen affinen Punkt $(x, y) \in E(F)$ ist ebenfalls $(x, -y) \in E(F)$.

iii) Klar: Für die Definition einer Geraden durch zwei Punkte spielt die Reihenfolge keine Rolle.

iv) Der Beweis der Assoziativität ist eine schwierige Aufgabe und wird in [Kna] III.3 und tiefergehend in [Si] Prop. 3.4 erbracht.

Um die mehrfache Addition eines Punktes zu sich selbst auszudrücken, definieren wir der Vollständigkeit halber, das Skalarprodukt als

$$mP = \underbrace{P + \cdots + P}_{m} \quad \text{für } m > 0,$$
$$(-m)P = -(mP) \quad \text{für } m > 0 \text{ und}$$
$$0P = \mathcal{O}.$$

Für die Implementierung ist es notwendig die Punktaddition in Koordinaten anzugeben. Dabei können wir uns auf die affinen Koordinaten von $E(F)$ beschränken, denn der einzige nicht-affine Punkt auf $E(F)$ dient als neutrales Element. Somit ist die Addition von $\mathcal{O}$ mit beliebigen Punkten aus $E(F)$ bereits unabhängig von ihren Koordinaten definiert.

Satz 3.8 *Für eine elliptische Kurve $E(F)$, mit zugehöriger affiner Kurve $C_f(F)$ in kurzer Weierstraß-Gleichung (3.2), gelten folgende Formeln für die Punktaddition.*

i) Für $P_1 = (x_1, y_1) \in C_f(F)$ ist $-P_1 = (x_1, -y_1)$.

ii) Für $P_1 = (x_1, y_1)$ und $P_2 = (x_2, y_2)$ aus $C_f(F)$ mit $P_1 \neq -P_2$ ist $P_1 + P_2 = P_3 = (x_3, y_3)$, wobei

$$x_3 = \lambda^2 - x_1 - x_2 \text{ und } y_3 = \lambda(x_1 - x_3) - y_1 \text{ ist, mit}$$
$$\lambda = \begin{cases} \dfrac{y_2 - y_1}{x_2 - x_1}, & \text{falls } P_1 \neq P_2 \\ \dfrac{3x_1^2 + a}{2y_1}, & \text{falls } P_1 = P_2. \end{cases}$$

Diese Formeln ergeben sich durch Betrachtung der Geradengleichungen der jeweiligen Schnittgeraden und werden in [We] 2.3 ausführlich, auch für die allgemeine (lange) Weierstraß-Gleichung, bewiesen.

4 Praktische Anwendung von elliptischen Kurven in der Kryptographie

Elliptische Kurven und ihre Punktegruppen werden seid den 1980er Jahren auch im Bezug auf Kryptographie erforscht. Sie werden heutzutage weitläufig in der Public-Key-Kryptographie, zum Beispiel beim Diffie-Hellman-Verfahren, eingesetzt. Sie lassen sich generell bei Verfahren verwenden, die auf einer Gruppenstruktur aufbauen, insbesondere bei Verfahren, die auf dem Diskreter-Logarithmus-Problem (DL-Problem, bzw. DLP) basieren. Der Oberbegriff Elliptic-Curve-Cryptography (ECC) fasst alle Public-Key-Verfahren zusammen, die mit elliptischen Kurven arbeiten und auf dem DLP basieren. Hier spricht man auch genauer vom ECDLP.

Um die Sicherheit, der auf ECDLP basierenden Kryptosysteme, nicht zu gefährden, gibt es einige Anforderungen an die Kurven-Parameter und die Körper, über denen die elliptischen Kurven definiert sind. Entscheidend dabei ist, welche mathematischen Angriffsmöglichkeiten es jeweils gibt.

Zunächst sollte der verwendete endliche Körper ausreichend viele Elemente haben, da dadurch auch die Anzahl der Punkte auf einer elliptischen Kurve über diesen Körper bestimmt wird (siehe unten). In der Praxis, d.h. in den meisten offiziellen Standards[6] und Anwendungen von ECC, werden elliptische Kurven über endlichen Körpern $\mathbb{F}_p$ mit einer großen Primzahl p verwendet und als sicher angesehen. Prinzipiell ist die Sicherheit von ECC, also die Schwere der Lösbarkeit des DL-Problems in der Punktegruppe einer elliptischen Kurve, auch über Körpern mit nicht-primer Ordnung (etwa $GF(2^k)$) gewährleistet. Doch hierbei gibt es ein paar Aspekte zu beachten, die die Sicherheit des Kryptosystems beeinträchtigen können. Um solche möglichen Schwächen von vornherein zu vermeiden, verwendet man daher hauptsächlich Körper mit primer Ordnung (siehe dazu [saf]).

4.1 DL-Problem Algorithmen

Die Sicherheit eines Verschlüsselungsverfahrens, dass auf dem ECDL-Problem basiert, hängt maßgeblich davon ab, wie effizient die ECDLP-Algorithmen sind, die es für die verwendete Gruppe gibt.

Ein Hauptaspekt ist dabei die Anzahl der Punkte auf einer elliptischen Kurve, also die Gruppenordnung $\#E(\mathbb{F}_p)$. Bei ECC-Verfahren wird eine zyklische Untergruppe von $E(\mathbb{F}_p)$ verwendet. Somit ist die Ordnung n dieser Untergruppe ebenfalls relevant. Eine Schranke für $\#E(\mathbb{F}_p)$ kann mit dem Theorem von Hasse (vgl. [Bu] S. 256 oder [We] 3.2.1) bestimmt werden.

$$\#E(\mathbb{F}_p) = p + 1 - t \quad \text{mit } |t| \leq 2\sqrt{p}$$

Die Anzahl der Punkte von $E(\mathbb{F}_p)$ liegt also ungefähr bei p. Effiziente Verfahren zur genauen Bestimmung von $\#E(\mathbb{F}_p)$ sind in [We] 3 oder in [Mir] zu finden.

In [Bu] 11 werden die gängigsten Algorithmen zur Lösung des DL-Problems vorgestellt. Generische Verfahren, die prinzipiell auf beliebigen Gruppen funktionieren haben

[6] Zum Beispiel ANSI X9.62 (1999), NIST FIPS 186-2 (2000), oder Brainpool (2005).

eine Laufzeit von $O(\sqrt{|G|})$. Für multiplikative Gruppen endlicher Körper gibt es spezielle Index-Calculus-Verfahren zur Berechnung des diskreten Logarithmus. Diese Verfahren sind deutlich schneller als die generischen Verfahren. Ihre Laufzeit ist subexponentiell. Der derzeit effizienteste Index-Calculus-Algorithmus ist das Zahlkörpersieb mit der Laufzeit $L_p[\frac{1}{3}, (\frac{64}{9})^{\frac{1}{3}}]$[7] (vgl. [Bu] 11.7). Die Index-Calculus-Verfahren sind im Allgemeinen jedoch nicht zum Lösen von ECDLP einsetzbar. Daher sind beim Einsatz von ECC-Verfahren geringere Schlüssellängen nötig (siehe unten). Allerdings ist nicht klar, ob es spezielle DLP-Lösungsverfahren für Gruppen auf elliptischen Kurven gibt.

Es gibt zwar spezielle Verfahren zum Lösen des DLP für bestimmte Klassen von elliptischen Kurven (supersingulär[8], anomal[9]), wie in [We] 4.2 gezeigt, jedoch lassen sich die Kriterien dafür überprüfen und solche Kurven daher leicht vermeiden.

Der Pohlig-Hellman-Algorithmus ist einer der schnellsten Algorithmen zum Lösen von DLP. Er ist besonders effizient, wenn die Gruppenordnung n nicht prim ist und der größte Primteiler von n relativ klein ist (siehe [Bu] Theorem 11.5.4). Daher sollte man den Basispunkt, der die zyklische Untergruppe in $E(\mathbb{F}_p)$ erzeugt, bestenfalls so wählen, dass deren Ordnung n prim ist.

4.2 Geeignete elliptische Kurven

Auf der Webseite `savecurves.cr.yp.to` werden elliptische Kurven verschiedener Formen über diversen endlichen Körpern vorgestellt, die auf ihre Eignung für Kryptosysteme untersucht worden sind. Außerdem werden in den erwähnten Standards Kriterien für elliptische Kurven veröffentlicht. Das BSI gibt ebenfalls Richtlinien vor, die elliptische Kurven erfüllen sollten (siehe [bsi2]).

Einige Kriterien für sichere elliptische Kurven sollen hier genannt werden.

- Die Kurve darf nicht supersingulär sein.
- Die Kurve darf nicht anomal sein.
- Die verwendete Untergruppe $\langle P \rangle \leq E(\mathbb{F}_p)$ sollte von der Ordnung n, n prim, $\geq 2^{192}$ sein.
- Die kleinste Zahl t mit $n|(p^t - 1)$ sollte größer als 104 sein.

4.3 Praktische Vorteile von ECC

Die wichtigsten Vorteile von ECC-Verfahren, sind die gebotenen Alternativen zum RSA-Verfahren und Diffie-Hellman-Schlüsselaustausch. Beim RSA-Verfahren ist die Sicherheit nicht garantiert (siehe [Bu] 9.3.11). Hier bietet ECC einen besseren Überblick über mögliche Angriffsarten. Für ECC-Verfahren sind außerdem effizientere Implementierungen möglich, da die benötigten Schlüssellängen nicht so groß sind, wie etwa bei RSA.

[7] L-Notation: $L_p[\alpha, c] = e^{(c+o(1))(\ln p)^{\alpha}(\ln \ln p)^{1-\alpha}}$

[8] Eine elliptische Kurve heißt supersingulär genau dann, wenn $\#E(\mathbb{F}_q) \equiv 1 \bmod \text{char}(\mathbb{F}_q)$

[9] Eine elliptische Kurve heißt anomal genau dann, wenn $\#E(\mathbb{F}_p) = p$

Eine Übersicht der Schlüssellängen für ein etwa gleich großes Sicherheitsniveau zeigt Tabelle 1. Hier ist n wieder die Gruppenordnung der verwendeten (Unter-)Gruppe in der das DLP gelöst werden muss.

Symmetrische Schlüssellänge	ECC, Bitlänge von n	RSA, Bitlänge von p ($\mathbb{F}_p^\times$)
80	160	1024
112	224	2048
128	256	3072
160	320	7680
256	512	15360

Tabelle 1: Vergleich von Schlüssellängen für gleiches Sicherheitsniveau[10]

Wie man sieht, sind bei ECC-Verfahren deutlich kürzere Schlüssellängen erforderlich. Die Angaben lassen sich auch auf das Diffie-Hellman-Verfahren übertragen, bei der es zu der Variante, die eine Einheitengruppe eines endlichen Körpers verwendet, auch die Alternative gibt, die auf elliptischen Kurven basiert (ECDH). Die kurzen Schlüssellängen haben mehrere praktische Vorteile. Einerseits ist die geringere Bitzahl bei der Implementierung auf eingebetteten Systemen nützlich, wegen der unter Umständen stark begrenzen Resscourcen, andererseits sind die Berechnungen mit Zahlen geringerer Bitlänge generell deutlich schneller.

[8]Quelle: ECC Brainpool Standard: http://www.ecc-brainpool.org/download/Domain-parameters.pdf

5 Prinzipimplementierung

Im Rahmen dieser Bachelorarbeit wurde eine Prinzipimplementierung in Java angefertigt, die es ermöglicht mittels des ElGamal-Verfahrens auf elliptischen Kurven beliebige Dateien zu ver- und entschlüsseln. Dabei wurde auf ein modulares Designkonzept Wert gelegt, sodass Verschlüsselungsalgorithmen auf verschiedenen elliptischen Kurven zum Einsatz kommen können. Man kann das Tool leicht mit Algorithmen, die auf elliptischen Kurven arbeiten, erweitern. Diese Flexibilität hat allerdings zur Folge, dass das Tool nicht auf eine bestimmte Kurve optimiert ist. Außerdem wurden keine Performanceoptimierungen vorgenommen, weshalb das Tool nur bedingt für den praktischen Gebrauch geeignet ist. Darüber hinaus sollte man sich nicht auf die Sicherheit des Tools verlassen, da neben den mathematischen Anforderungen an elliptische Kurven auch hohe Anforderungen an die Implementierungssicherheit gestellt werden müssen, welche nicht Teil dieser Arbeit sind.

Konkret implementiert wurde, neben der Integer-Arithmetik von endlichen Körpern, das Gruppengesetz auf elliptischen Kurven mit kurzer Weierstraß-Gleichung und das Verschlüsselungsverfahren ElGamal in einer leicht angepassten Variante (siehe unten).

Die Implementierung der Gruppenoperationen auf elliptischen Kurven wurde mithilfe des Computer-Algebra-Systems PARI/GP[11] stichprobenartig überprüft.

ectool

Das Verschlüsselungstool *ectool* setzt die Verwendung von Oracle-Java 1.8 voraus. Es ist ein reines Kommandozeilenprogramm. Ruft man es ohne Parameter auf, wird eine knappe Hilfe ausgegeben, in der die verfügbaren Funktionen kurz skizziert sind.

- Schlüsselpaar generieren (Option -g)
- Verschlüsseln einer Datei mittels eines Private-Keys (Option -s und -f)
- Entschlüsseln einer Datei mittels eines Public-Keys (Option -p und -f)
- Einlesen von Kurvenparametern (Option -l)
- Ver-/Entschlüsseln einer Nachricht, welche direkt als Parameter angegeben wird (Option -m)

Bevor eine Nachricht, oder eine Datei ver- oder entschlüsselt werden kann, muss ein Schlüsselpaar generiert werden. Danach kann man zum Verschlüsseln jeweils einen öffentlichen Schlüssel und zum Entschlüsseln, einen privaten Schlüssel angeben.

Es können beliebige Kurven in kurzer Weierstraß-Form angegeben werden. Dazu ist es möglich eine Datei einzulesen, in der man folgende Parameter definieren kann. Dies ist jedoch nicht zwingend notwendig, da eine elliptische Kurve (Curve25519) bereits fest eingebaut ist. Wenn keine Datei mit Kurvenparametern angegeben wird, wird diese als Standardkurve verwendet.

[11] PARI/GP: http://pari.math.u-bordeaux.fr

- Definition eines endlichen Körpers $\mathbb{F}_p$ über die Angabe einer Primzahl p.
- Definition einer elliptischen Kurve in einfacher Weierstraß-Form ($y^2 = x^3 + ax + b$)
  - Ein Element a des Körpers $\mathbb{F}_p$
  - Ein Element b des Körpers $\mathbb{F}_p$
  - Ein Punkt ($P = (x, y)$ mit $x, y \in \mathbb{F}_p$) auf der elliptischen Kurve, der als Basispunkt für ECC-Algorithmen verwendet wird.
  - Die Ordnung der durch P erzeugten zyklischen Gruppe ($|\langle P \rangle| = n \in \mathbb{F}_p$)

All diese 6 Parameter werden zeilenweise als Hexadezimal-String (ohne 0x) angegeben. Dabei ist die Reihenfolge zu beachten. Viele Beispiele elliptischer Kurven, mit den entsprechenden Parametern, sind auf `safecurves.cr.yp.to` zu finden. Das Programm enthält keine Überprüfung der eingelesen Körper- und Kurvenparameter, etwa ob die angegebene Zahl p, die den endlichen Körper definiert, wirklich eine Primzahl ist. Es wird jedoch überprüft, ob die angegebene Kurve überhaupt eine elliptische Kurve ist (duch Berechnung der Diskriminante).

Zu beachten ist, dass Schlüsselpaare offensichtlich nur mit der Kurve verwendet werden können, für die sie generiert wurden.

5.1 Hashed-ElGamal

Beim klassischen ElGamal-Verschlüsselungsverfahren ist es notwendig, dass beliebige Nachrichten (Bytestrings) als Elemente der verwendeten Gruppe codiert und decodiert werden können (siehe Abschnitt 1.5). Bei der Verwendung einer Gruppe auf elliptischen Kurven muss also eine Nachricht als Punkt auf einer solchen eindeutig repräsentiert werden. Dieses Problem ist nicht so leicht zu lösen, vor allem wenn man dabei nicht die Sicherheit der Verschlüsselung gefährden will. Unter gewissen Bedingungen ist es möglich injektive Abbildungen zu elliptischen Kurven zu finden (siehe dazu zum Beispiel [FJT]). Außerdem kann man probabilistische Verfahren einsetzen, indem man die Nachricht so aufteilt, dass die einzelnen Teile als Zahlen des zugrunde liegenden Körpers repräsentiert werden können. Diese Werte setzt man dann als x-Werte in die Kurvengleichung ein und überprüft, ob es ein passendes y^2 im Körper gibt, womit man einen Punkt auf der Kurve gefunden hätte. Anderenfalls erhöht man den x-Wert sukzessive, bis ein Punkt gefunden ist. Um diesen Aufwand zu vermeiden, lässt sich das ElGamal-Verfahren auch leicht anpassen, sodass Nachrichten mit beliebigen Punkten verknüpft werden können. Dazu wird bei der Ver- und Entschlüsselung eine kryptographische Hashfunktion verwendet und die Nachricht mit dem Hashwert XOR-verknüpft. Dieses sogenannte Hashed-ElGamal-Verfahren wurde auch in der vorliegenden Implementierung eingesetzt (unter Verwendung von SHA-256 als Hashfunktion), da es flexibel auf elliptischen Kurven unterschiedlicher Größe einsetzbar ist. Eine Beweis-Skizze für die Sicherheit dieses Verfahrens findet sich in [Ro] (S. 77 ff.). Eine genaue Beschreibung des Verfahrens folgt nun.

Wie in Abschnitt 1.5 erklärt, erzeugen Alice und Bob jeweils ein Schlüsselpaar. Alice verschlüsselt mit Bobs offentlichen Schlüssel ($d_B P$) eine Nachricht m und sendet sie an Bob, der sie mit seinem privaten Schlüssel (d_B) entschlüsseln kann. Dabei ist P der Erzeuger der verwendeten (Unter-)Gruppe mit Ordnung n und H eine Hashfunktion mit

$$\mathsf{H} : \langle P \rangle \to \{0,1\}^m \qquad \text{bspw. } m = 256 \text{ bei SHA-256}$$

Der genaue Ablauf eines verschlüsselten Nachrichtenaustausches mittels Hashed-ElGamal verläuft nun nach diesem Schema:

Hashed ElGamal

Alice		**Bob**
$k \leftarrow_\$ \{1, \ldots, n-1\}$		
$Q \leftarrow kP$		
$R \leftarrow \mathsf{H}(k(d_B P)) \oplus m$		
	$\xrightarrow{(Q,R)}$	
		$S \leftarrow d_B Q = d_B(kP)$
		$m \leftarrow \mathsf{H}(S) \oplus R$

5.2 Implementierungen in der Praxis

In Implementierungen, die in der Praxis zum Einsatz kommen, werden diverse Mittel eingesetzt, um die Berechnungen auf den elliptischen Kurven effizienter zu machen. Viele Maßnahmen sind schon auf mathematischer Ebene anzusiedeln. Zum Beispiel kommen unterschiedliche Formen von Gleichungen der elliptischen Kurven zum Einsatz. Die allgemeine lange Weierstraß-Gleichung (3.1) lässt sich auf verschiedene Arten äquivalent umformen, wenn man gewisse Anforderungen an den Körper stellt, über den die elliptische Kurve definiert werden soll. Die Umformung zur kurzen Weierstraß-Gleichung wurde in Kapitel 3 gezeigt. Darüber hinaus gibt es noch weitere Formen von Gleichungen für elliptische Kurven. Die wichtigsten seien hier kurz genannt:

- Edwards-Kurve: $x^2 + y^2 = 1 + dx^2y^2$
- Montgomery-Kurve: $By^2 = x^3 + Ax^2 + x$

Ziel dieser Umformungen ist es, die Berechnung des additiven Gruppengesetzes auf der elliptischen Kurve zu beschleunigen, indem man die Zahl der benötigten Grundoperationen (Addition, Multiplikation, Division (Inversion in $\mathbb{F}_p$)) verringert. Im *Handbook of Elliptic and Hyperelliptic Curve Cryptography* ([Co-Fr]) werden die verschiedenen Verfahren vorgestellt und praktische Implementierungstipps gegeben.

6 Fazit und Ausblick

Elliptische Kurven eignen sich gut für die Verwendung in der Public-Key-Kryptographie, denn mit gewissen Anforderungen ist das Diskreter-Logarithmus-Problem auf ihren Punktegruppen schwer zu lösen. Die (heute) benötigten kleinen Schlüssellängen bei ECC machen es zu einer effizienten Alternative zu herkömmlichen Verfahren und bieten noch die Möglichkeit größere Schlüssellängen praktikabel zu nutzen. Die Forschung an elliptischen Kurven für den Einsatz in der Kryptographie ist noch nicht sehr alt und daher können in diesem Bereich noch interessante Ergebnisse erwartet werden. In den letzten Jahren wurden neue Formen elliptischer Kurven spezifiziert, um etwa die Berechnungen der Gruppenoperation zu optimieren und übersichtliche Kriterien für die Sicherheit von elliptischen Kurven zu erhalten.

Trotzdem ist ECC nicht die endgültig sichere Lösung für Public-Key-Kryptographie. Die Sicherheit von Public-Key-Verfahren wird heutzutage durch die Entwicklung von Quantencomputern bedroht. P. W. Shor hat gezeigt, dass mithilfe von Quantencomputern alle heute gängigen Public-Key-Verfahren, die auf dem DL-Problem basieren, leicht gelöst werden können ([Sh]). Heutzutage gibt es zwar noch keine Quantencomputer, die dafür praktisch nutzbar wären, da die Forschung in diesem Bereich aber immer weiter voranschreitet[12], ist dies langfristig ein Problem für die Public-Key-Kryptographie. Auf der Webseite pqcrypto.org, von Daniel J. Bernstein und Tanja Lange findet man aktuelle Forschungsergebnisse zu dem Thema Post-Quantum-Kryptographie.

[12] http://www.spiegel.de/netzwelt/web/google-und-nasa-praesentieren-ihren-quantencomputer-a-1066838.html

Literatur

[Bo] Bosch, S.: *Algebra*, Springer Verlag, 6. Auflage, 2006

[bsi] Bundesamts für Sicherheit in der Informationstechnik (BSI): https://www.bsi.bund.de/DE/Publikationen/TechnischeRichtlinien/tr02102/index_htm.html, BSI TR-02102-1, abgerufen am 28.03.2016

[bsi2] Bundesamts für Sicherheit in der Informationstechnik (BSI): https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Zertifizierung/Interpretationen/AIS_46_ECCGuide_e_pdf.pdf, abgerufen am 28.03.2016

[Bu] Buchmann, J. A.: *Einführung in die Kryptographie*, Springer Verlag, 5. Auflage, 2010

[Co-Fr] Cohen, H. und Frey, G., u.a.: *Handbook of Elliptic and Hyperelliptic Curve Cryptography*, Chapman & Hall/CRC, 2006

[FJT] Fouque, P.-A., Joux A. und Tibouchi, M.: *Injective Encodings to Elliptic Curves*, http://eprint.iacr.org/2013/373.pdf, abgerufen am 28.03.2016

[Ka-Ki] Karpfinger, C., Kiechle, Hubert, *Kryptologie - Algebraische Methoden und Algorithmen*, Vieweg+Teubner Verlag, 1. Auflage, 2010

[Kna] Knapp, A.: *Elliptic Curves, Mathematical Notes 40*, Princeton University Press, 1992

[Mir] Mirbach, A.: *Elliptische Kurven: Die Bestimmung ihrer Punktezahl und Anwendungen in der Kryptographie*, Verlaghaus Monsenstein und Vannerdat, 2003

[Ro] Rogaway, P.: *Advances in Cryptology* – CRYPTO 2011: 31st Annual Cryptology Conference, Santa Barbara, CA, USA, August 14-18, 2011, Proceedings, Springer Verlag, 2011

[Rü] Rück, H.-G., *Die projektive Ebene - Was sind unendlich ferne Punkte?*, http://www.mathematik.uni-kassel.de/TdM/media/rueck-proj-ebene.pdf, abgerufen am 28.03.2016

[saf] https://safecurves.cr.yp.to, abgerufen am 28.03.2016

[Sh] Shor, P. W.: *Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer*, SIAM Journal on Computing, 26:1484–1509, 1997

[Si] Silverman, J.H.: *The arithmetic of elliptic curves*, Graduate Texts in Mathematics 106, Springer Verlag, 1986

[We] Werner, A., *Elliptische Kurven in der Kryptographie*, Springer Verlag, 2002

Abbildungsverzeichnis