Criptografía y curvas elípticas. La curva de WhatsApp

Irene Ayerra Balduz

Trabajo de fin de grado en Matemáticas
Universidad de Zaragoza

Director del trabajo: Manuel Vázquez Lapuente
Directora del trabajo: Paz Jiménez Seral
13 de septiembre de 2018

Abstract

The aim of this study consists on studying the basis of the algebraic theory of elliptic curves, to later analyse its application in the encryption of messages.

In order to facilitate the understanding of the elliptic curves, a first chapter of basic notions related to the algebra of groups, rings, fields and extensions of fields has been developed.

For more than three centuries, elliptic curves have played a fundamental role in mathematics. In the Chapter 2, the operation that gives the elliptic curves of abelian group structure is developed. This unusual operation results in the elliptic curves having excepcional arithmetic properties. Next, the theory of elliptic curves on finite fields is developed, paying special attention to the structure and cardinality of the finite group generated. To finish this chapter, the arithmetic of the *Montgomery form* is analysed, since it is one of the best ways to express the elliptic curves to apply them in cryptography.

Even though it was not until about 30 years ago that the whole theory developed on elliptic curves began to be applied to cryptography, it is important to point out that more than 4000 years ago there were already methods of message encryption. In fact, the word cryptography comes from the Greek "kryptos", which means hidden, and "graphia", which means writing, and is defined as "the art of writing with a secret key or in an enigmatic way".

Currently, there are thousands of daily transactions and the world is electronically connected in a global way through the network. For this reason, it has been necessary to create mathematical algorithms that guarantee integrity, confidentiality, authenticity and non-repudiation. Hence, the Chapter 3 has been dedicated to the study of public key cryptography based on the *Discrete Logarithm Problem*. More specifically, it focuses on the *Diffie-Hellman key exchange* (DHKE).

In April 2016, WhatsApp began to encrypt its messages with the "*end to end encryption*" method. This protocol is based on the DHKE applied to the group generated by an elliptic curve on a finite field. WhatsApp chose Curve25519, which belongs to a particular family of Montgomery curves. The Chapter 4 includes the algebraic study of this family of curves and the analysis of the computational aspects of Curve25519, which bring to light that an exceptional curve has been found for message encryption. Finally, the encryption protocol used by WhatsApp is explained step by step.

Índice general

Capítulo 1

Introducción

La necesidad de comunicarse de forma secreta ha existido siempre, por lo que desde la antigüedad se han desarrollado distintas técnicas de cifrado de mensajes sobre todo entre diplomáticos, militares y gobernantes; basta remontarnos a los jeroglíficos empleados por los egipcios o a la escitala utilizada por los espartanos para encontrar las primeras muestras. Posteriormente, surgió uno de los cifrados más famosos, el cifrado César, que consistía en escribir el mensaje con un alfabeto que estaba formado por las letras del alfabeto latino pero desplazadas tres posiciones a la derecha. Avanzando en la historia nos encontramos cifrados cada vez más complejos como son la tabla de Polibio, el método de las frecuencias, los cifrados de transposición, el cifrado Alberti, el polialfabético de Vigenère o el de tipo Playfair; todos ellos constituyen la *criptografía clásica*. Con la llegada de la revolución industrial, la criptografía se fue mecanizando dando lugar a la aparición de los primeros rotores y máquinas de cifrado, que culminaron con la creación de la máquina Enigma, empleada por los alemanes en la Segunda Guerra Mundial. Se trata en todos los casos de métodos de *clave simétrica*, es decir, se emplea la misma clave para cifrar y descifrar.

Las primeras investigaciones académicas en el ámbito de la criptografía comenzaron a mediados de 1970. A inicios de 1980, con la llegada de la informática a las comunicaciones, diferentes empresas financieras y de telecomunicaciones ya introdujeron métodos criptográficos en sus sistemas hardware. Desde entonces, rápidamente, la criptografía ha ido creciendo a pasos agigantados extendiéndose a todos los ámbitos hasta tal punto que hoy en día es usada a diario por todos nosotros. Desde abrir la puerta del garaje con control remoto, hasta conectarse a una red wifi, como hacer compras por Internet o instalar una actualización de software son acciones que llevan implementados diferentes métodos de encriptación. Así, en un mundo en el que existen miles de transacciones diarias y que electrónicamente está conectado de manera global a través de la red, ha sido necesario crear algoritmos matemáticos que garanticen **integridad**, **confidencialidad**, **autenticidad** y **no repudio**, los cuales deben permitir además el intercambio de claves entre dos o más partes que nunca antes habían estado en contacto.

El primer algoritmo creado fue el de clave simétrica DES (*Data Encryption Standard*) en la década de los 70, el cual fue sustituido en 1999 por AES (*Advanced Encryption Standard*), un método también de clave simétrica muy rápido en el aspecto computacional. A pesar de ser muy eficientes computacionalmente, los criptosistemas de clave simétrica presentaban varios defectos como el gran número de claves que eran necesarias, la dificultad de distribuir éstas a través de canales inseguros o la falta de autenticidad del mensaje.

Paralelamente Whitfield Diffie, Martin Hellman y Ralph Merkle introdujeron en 1976 la *criptografía de clave pública*, también llamada *criptografía asimétrica* porque ya no se utilizaba la misma clave para cifrar y descifrar. Este tipo de algoritmos garantizan el resto de propiedades que los algoritmos simétricos no pueden garantizar: permiten establecer una clave a través de un canal inseguro, identificar a cada persona que interviene, encriptar mensajes y asegurar la integridad de los mismos, si bien

surge con ellos el problema de la autenticidad de las claves públicas. Todos emplean métodos matemáticos, aunque los más relevantes se basan en el problema de la factorización entera (*Algoritmo de Rivest-Shamir-Adleman*, RSA) y en el Problema del Logaritmo Discreto (*Intercambio de claves Diffie-Hellman* y *Elgamal*).

En este trabajo vamos a estudiar uno de los métodos de clave pública con más visión de futuro que hay en la actualidad, los criptosistemas basados en *curvas elípticas*.

Las curvas elípticas sobre cuerpos finitos generan una cantidad inagotable de grupos abelianos finitos que, escogidos adecuadamente, pueden ser aplicados a la computación. Sin embargo, la teoría de curvas elípticas presenta una complejidad matemática mucho mayor que la que hay detrás de los algoritmos de encriptación que ya existían (AES, RSA, ...), por lo que el capítulo 2 lo dedicamos totalmente al desarrollo de la teoría de curvas elípticas sobre cuerpos finitos.

En el capítulo 3 presentaremos el Problema del Logaritmo Discreto y el principal método criptográfico basado en él, el intercambio de claves Diffie-Hellman (DHKE). Con todo lo explicado hasta este punto ya seremos capaces de aplicar el Problema del Logaritmo Discreto a grupos generados por curvas elípticas y desarrollar así el sistema de encriptación DHKE pero con el grupo generado por una curva elíptica.

Un claro ejemplo de que el DHKE aplicado a curvas elípticas es muy eficiente y probablemente sea el criptosistema más utilizado dentro de unos años es que se trata del método de cifrado que se utiliza para las criptomonedas *bitcoin*, así como forma parte del protocolo de encriptación que desde 2016 emplea WhatsApp, la aplicación de mensajería que supera los mil millones de usuarios diarios e intercambia al día más de 60000 millones de mensajes, siendo la app más utilizada de todo el mundo para el intercambio de mensajes. La curva elíptica que emplea WhatsApp para el cifrado de mensajes es la llamada *Curva25519*. El capítulo 4 lo dedicaremos al estudio algebraico y computacional de esta curva con el fin de mostrar las propiedades que la hacen idónea para su uso en criptografía. Por último, presentaremos una versión reducida del protocolo completo de encriptación de WhatsApp.

Antes de comenzar con la teoría de curvas elípticas vamos a definir y enunciar algunos conceptos básicos del álgebra que serán fundamentales para seguir el texto.

1.1. Grupos

Definición 1.1. Un grupo G es un conjunto con una operación binaria interna asociativa, con elemento neutro e inverso para cada elemento. Se dice *abeliano* si la operación binaria interna es conmutativa.

Definición 1.2. Un subgrupo H de un grupo G es un subconjunto de G que con la operación restringida es también grupo. Lo denotamos $H \leq G$.

Proposición 1.1. *Sea G grupo, si H es un subconjunto no vacío de G, H es subgrupo de G si y solo si* $xy^{-1} \in H$ *para todo* $x, y \in H$.

Si G es un grupo finito, llamaremos *orden* o *cardinalidad* de G al número de elementos de G y lo denotamos mediante $|G|$.

Definición 1.3. Si H es subgrupo de G y $x \in G$ al conjunto $Hx = \{hx \mid h \in H\}$ se le llama *clase a derecha de x módulo H* (o *coclase*). Si $H \leq G$, llamaremos *índice de H en G* al número de clases a derecha y lo denotamos $|G : H|$.

Teorema 1.2 (Teorema de Lagrange)**.** *Sea G finito y H subgrupo de G. Se tiene*

$$|G| = |G : H||H|.$$

Demostración. Probamos primero que $Hx = Hy \Leftrightarrow xy^{-1} \in H$.$\Leftarrow$) Sea $h = xy^{-1} \in H \Rightarrow x = hy$. Si $h'x \in Hx \Rightarrow h'x = h'hy \in Hy$. Luego $Hx \subseteq Hy$. Del mismo modo $Hy \subseteq Hx$ y así $Hx = Hy$.$\Rightarrow$) Sea $x = hy$ para algún $h \in H \Rightarrow xy^{-1} = h \in H$.

Antes de continuar veamos que $Hx \cap Hy \neq \emptyset$ si y solo si $Hx = Hy$.$\Leftarrow$) Trivial por definición de intersección.$\Rightarrow$) Supongamos que tenemos $z = hx = h'y$ con $h, h' \in H \Rightarrow x = h^{-1}h'y$. Ahora $xy^{-1} = (h^{-1}h'y)y^{-1} = h^{-1}h' \in H$. Como $Hx = Hy \Leftrightarrow xy^{-1} \in H \Rightarrow Hx = Hy$. Luego las clases a derecha forman una partición de G en $|G : H|$ conjuntos distintos, todas ellas con el mismo cardinal $|H|$.

Queda probado que $|G| = |G : H||H|$. □

Del teorema de Lagrange deducimos que todo subgrupo tiene orden divisor del orden del grupo.

Teorema 1.3 (Pequeño Teorema de Fermat)**.** *Sea p primo tal que p no divide a b. Entonces*

$$b^{p-1} \equiv 1 \mod p.$$

1.1.1. Homomorfismos de grupos

Definición 1.4. Un *homomorfismo* entre dos grupos $(G, \cdot)$ y $(H, \cdot)$ es una aplicación $f : G \to H$ tal que $f(ab) = f(a)f(b)$ para todo $a, b \in G$. Un homomorfismo biyectivo se dice *isomorfismo*.

Proposición 1.4. *Sea f un homomorfismo entre dos grupos $(G, \cdot)$ y $(H, \cdot)$. Se tiene*
a) $f(1) = 1$ y para todo $a \in G$, $f(a^{-1}) = f(a)^{-1}$.
b) Se define $Kerf = \{a \in G \mid f(a) = 1\}$ y $Kerf \trianglelefteq G$.
c) Se define $Imf = f(G) = \{f(a) \mid a \in G\}$ y se tiene que $Imf \leq H$.
d) $Kerf = 1$ si y solo si f es inyectiva.

Teorema 1.5 (Teorema de isomorfía)**.** *Sea $f : G \to H$ homomorfismo. La aplicación*

$$\overline{f} : G/Kerf \to f(G)$$

dada por $\overline{f}(xKerf) = f(x)$, es un isomorfismo de grupos.

Demostración. Veamos que $\overline{f}$ está bien definida. Sea $Kerfx = Kerfy$, se tiene que $xy^{-1} \in Kerf$, luego $f(xy^{-1}) = 1$ pero $f(xy^{-1}) = f(x)f(y^{-1}) = f(x)f(y)^{-1}$ y así $f(x) = f(y)$. Comprobamos que $\overline{f}$ es homomorfismo: $\overline{f}(Kerfx\ Kerfy) = \overline{f}(Kerfxy) = f(xy) = f(x)f(y) = \overline{f}(Kerfx)\overline{f}(Kerfy)$. Veamos ahora que $\overline{f}$ es inyectiva. Sean $Kerfx$ y $Kerfy$ tales que $f(x) = \overline{f}(Kerfx) = \overline{f}(Kerfy) = f(y)$. Entonces, $f(x)f(y)^{-1} = 1$ y como $f(x)f(y)^{-1} = f(x)f(y^{-1}) = f(xy^{-1})$, $xy^{-1} \in Kerf$ y se tiene que $Kerfx = Kerfy$. Por último, probemos que $\overline{f}$ es sobre, es decir, que $\forall\, y \in f(G)\ \exists\, x \in G$ tal que $f(x) = y$. Sea $y = f(x) = \overline{f}(Kerfx)$, $\forall\, y \in f(G)\ \exists\, Kerfx \in G/Kerf$ tal que $\overline{f}(Kerfx) = y$. □

1.1.2. Grupos cíclicos

Se conviene que $x^0 = 1$, $x^n = x \cdot \ldots \cdot x$ (n-veces) y $x^{-n} = (x^{-1})^n$. Sea $x \in G$, al conjunto $\langle x \rangle = \{x^n \mid n \in \mathbb{Z}\}$ se le llama subgrupo generado por el elemento x.

Definición 1.5. Sea $x \in G$, si existe un natural n tal que $x^n = 1$, al menor tal n se le llama *orden* de x. Si no existe dicho n se dice que x no tiene orden finito.

Si x tiene orden n, entonces $x^{-1} = x^{n-1}$ y $\langle x \rangle = \{1, x, \ldots, x^{n-1}\}$.

Definición 1.6. Un grupo G se dice *cíclico* si existe $x \in G$ tal que $G = \langle x \rangle$. A los elementos x de orden máximo los llamamos *elementos primitivos* o *generadores*.

Como ejemplos sencillos de grupos cíclicos tenemos:

- $(\mathbb{Z}, +)$ y se tiene que $\mathbb{Z} = \langle 1 \rangle$.

- $(\mathbb{Z}_n, +)$ donde $\mathbb{Z}_n = \{0,\ 1,\ 2, \ldots,\ n-1\}$ y la operación de grupo denotada por "$+$" es la suma módulo n. Entonces $\mathbb{Z}_n = \langle 1 \rangle$.
- Si p primo, $(\mathbb{Z}_p^*, \cdot)$ es grupo cíclico, donde $\mathbb{Z}_p^* = \{1,\ 2, \ldots,\ p-1\}$ y la operación "$\cdot$" es el producto ordinario módulo p.

Proposición 1.6. *Todo subgrupo de un grupo cíclico es cíclico.*

La demostración es muy sencilla y no la desarrollamos.

Proposición 1.7. *Si G es cíclico con n elementos, tiene exactamente $\phi(n)$ generadores distintos.*

Proposición 1.8. *Si G es grupo cíclico de orden $n \in \mathbb{N}$ se tiene:*
a) Todo elemento tiene orden divisor de n.
b) Para todo d divisor de n existe un único subgrupo de orden d.

Proposición 1.9. *Todo grupo de orden primo es cíclico.*

Proposición 1.10. *Sean C_n y C_m cíclicos de órdenes n y m respectivamente, entonces $C_n \times C_m$ es abeliano de orden $n \cdot m$. Además,*
a) Si $mcd(n,m) = 1$, entonces $C_n \times C_m \simeq C_{n \cdot m}$.
b) Todo grupo abeliano finito es isomorfo a un producto de cíclicos.

1.2. Anillos, cuerpos y extensiones de cuerpos

Definición 1.7. Un *anillo* es un conjunto $A \neq \emptyset$ dotado de dos operaciones binarias internas "$+$" y "$\cdot$" verificando: A es un grupo abeliano con la operación "$+$" y para todo $a, b, c \in A$,

$$\begin{aligned} a \cdot (b \cdot c) &= (a \cdot b) \cdot c, \\ a \cdot (b + c) &= (a \cdot b) + (a \cdot c), \\ (a + b) \cdot c &= (a \cdot c) + (b \cdot c). \end{aligned}$$

Además el anillo A se dice *conmutativo* si $a \cdot b = b \cdot a$ para todo $a, b \in A$.

Definición 1.8. Un anillo A se dice con *identidad* si el producto posee elemento neutro. Este elemento es único y lo de notamos por 1_A.

Si A es anillo, se donota con $A[x]$ el anillo de polinomios sobre la indeterminada x con coeficientes en R. El grado de un polinomio $p(x)$, se denota mediante $grad(p(x))$.

Definición 1.9. Sean A, B anillos. Un *homomorfismo* de A en B es una aplicación $f : A \longrightarrow B$ tal que para todo $a, b \in A$ cumple que $f(a+b) = f(a) + f(b)$ y $f(ab) = f(a)f(b)$.

Es fácil ver que si $f : A \longrightarrow B$ es un homomorfismo de anillos $f(0_A) = 0_B$ y $f(-a) = -f(a)$. Si A y B son anillos con identidad, entonces $f(1_A) = 1_B$. Además, si el homomorfismo es inyectivo se dice *monomorfismo*.

Proposición 1.11. *Sea A anillo conmutativo y $a \in A$. La aplicación $e_a : A[x] \longrightarrow p(a)$ es un homomorfismo de anillos, conocido como homomorfismo evaluación.*

Definición 1.10. Un *cuerpo K* es un anillo conmutativo con identidad en el que todo elemento no nulo tiene inverso.

Definición 1.11. Sea p primo. Al anillo entero $\mathbb{Z}_p$ lo llamamos *cuerpo primo*. Notar que además de anillo es cuerpo finito.

Es fácil ver que $(\mathbb{Z}_p, +, \cdot)$ donde "$+$" denota la suma ordinaria modulo p y "$\cdot$" denota el producto ordinario modulo p, es un anillo conmutativo con elemento neutro $0_{\mathbb{Z}_p}$ y elemento identidad $1_{\mathbb{Z}_p}$. Además como p es primo, entonces $(\mathbb{Z}_p, +, \cdot)$ es cuerpo (cuerpo primo).

Proposición 1.12. *Si el polinomio $p(x)$ es irreducible en $K[x]$, entonces $K[x]/p(x)$ es cuerpo.*

Definición 1.12. Decimos que E/K es una *extensión de cuerpos* si K es un subcuerpo de E.

Es claro que E es un K-espacio vectorial. La extensión es finita si la dimensión de E como K-espacio vectorial es finita. Denotamos $dim_K(E) = |E : K|$. Si E/K es una extensión y $X \subseteq E$, es obvio que $K(X)$ es el menor subcuerpo de E que contiene a K y a X.

Proposición 1.13. *Supongamos que $p(x) \in K[x]$ es irreducible, entonces existe una extensión E/K tal que $p(x)$ tiene una raíz en E.*

Proposición 1.14. *Sea $p(x) \in K[x]$ y $a \in K$, entonces $p(a) = 0$ si y sólo si $x - a$ divide a $p(x)$.*

Proposición 1.15. *Sea K un subcuerpo de E y $f(x) \in K[x]$ con $f'(x) \neq 0$. Se tiene:*
a) Si $a \in E$, a es raíz múltiple de $f(x)$ si y sólo si $f(a) = f'(a) = 0$.
b) Si $(f(x), f'(x)) = 1$, entonces $f(x)$ no tiene raíces múltiples en E.
c) Si $f(x)$ es irreducible en $K[x]$, entonces todas las raíces de $f(x)$ son distintas.
d) Si la característica de K es p, los irreducibles de $K[x]$ no tienen raíces múltiples en ningún cuerpo extensión de K.

Definición 1.13. El cuerpo primo del cuerpo K es la intersección de todos los subcuerpos de K.

Proposición 1.16. *Sea F el cuerpo primo de K. Si la característica de K es cero, entonces F es isomorfo a $\mathbb{Q}$ y si la característica es p, entonces F es isomorfo a $\mathbb{Z}_p$.*

Teorema 1.17. *Sea K un cuerpo finito, entonces $|K| = p^n$ para algún primo p y algún entero n. Recíprocamente para todo primo p y entero positivo n, existe un único cuerpo, salvo isomorfismos, de p^n elementos.*

Al cuerpo de p^n elementos se le denota $GF(p^n)$.

Capítulo 2

Curvas elípticas

2.1. Definición

Definición 2.1. Una *ecuación de Weierstrass generalizada* sobre un cuerpo K es una ecuación de la forma

$$y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6, \tag{2.1}$$

con $a_1, a_2, a_3, a_4, a_6 \in K$.

Sea $K \times K$, recordemos que dada una función $F : K \times K \longrightarrow K$ diferenciable y una curva en implícitas $C = \{(x,y) \in K \times K \mid F(x,y) = 0_K\}$, un punto $(x_0, y_0) \in C$ se dice *singular* si $\partial F/\partial x\,(x_0, y_0) = \partial F/\partial y\,(x_0, y_0) = 0$. En otro caso se dice *no singular* o *regular*. Entonces, una ecuación generalizada de Weierstrass es no singular si para cada punto (x_0, y_0) con coordenadas en $K \times K$ satisfaciendo la ecuación, las derivadas parciales en ese punto $\partial E/\partial x\,(x_0, y_0)$ y $\partial E/\partial y\,(x_0, y_0)$ no se anulan simultáneamente.

Definición 2.2. Una curva elíptica E sobre un cuerpo K es el conjunto de puntos $(x,y) \in K \times K$ que satisfacen la ecuación (2.1) no singular. Además de un punto especial que se llama *punto en el infinito* y se denota P_∞.

Nota. El hecho de que la curva correspondiente a la ecuación (2.1) sea no singular equivale gráficamente a que no presenta auto-cortes o vértices. Esto será esencial para definir la operación de grupo, ya que es necesario que todos los puntos tengan una única tangente.

Si $car(K) \neq 2$, podemos simplificar la ecuación (2.1) con la transformación $y \to \frac{1}{2}(y - a_1x - a_3)$ obteniendo

$$E : y^2 = 4x^3 + b_2x^2 + 2b_4x + b_6, \tag{2.2}$$

donde $b_2 = a_1^2 + 4a_2$, $b_4 = 2a_4 + a_1a_3$ y $b_6 = a_3^2 + 4a_6$. Además definimos los siguientes coeficientes:

$$\begin{aligned} b_8 &= a_1^2a_6 + 4a_2a_6 - a_1a_3a_4 + a_2a_3^2 - a_4^2, \\ c_4 &= b_2^2 - 24b_4, \\ c_6 &= -b_2^3 + 36b_2b_4 - 216b_6. \end{aligned}$$

Definición 2.3. Se llama *discriminante* de la ecuación de Weierstrass a $\Delta = -b_2^2b_8 - 8b_4^3 - 27b_6^2 + 9b_2b_4b_6$.

Proposición 2.1. *La curva W dada por la ecuación* (2.1) *es no singular, es decir, es una curva elíptica, si y solo si $\Delta \neq 0$.*

Demostración. Supongamos que W es singular y que $P = (x_0, y_0)$ es un punto singular de W. Sustituyendo $x = x' + x_0$ e $y = y' + y_0$ los valores de Δ y c_4 no cambian, por lo que tenemos ahora una curva singular

en $(0,0)$. Como $a_6 = f(0,0)$, $a_4 = \partial f/\partial x\ (0,0)$ y $a_3 = \partial f/\partial y\ (0,0)$, tenemos que $a_6 = a_4 = a_3 = 0$. La ecuación implícita de la curva queda

$$f(x,y) = y^2 + a_1xy - a_2x^2 - x^3 = 0,$$

con $c_4 = (a_1^2 + 4a_2)^2$ y $\Delta = 0$. Queda probado que si la curva es singular, entonces su discriminante es nulo. Veamos ahora que si W es no singular, entonces $\Delta \neq 0$. Para simplificar los cálculos, asumimos que $car(K) \neq 2$ para tomar así la ecuación de Weierstrass reducida (2.2) en forma implícita

$$f(x,y) = 4x^3 + b_2x^2 + 2b_4x + b_6 - y^2 = 0$$

Entonces W es singular si y sólo si existe un punto $P = (x_0, y_0)$ tal que cumple $\partial f/\partial x\ (x_0,y_0) = \partial f/\partial y\ (x_0,y_0) = 0$, es decir, si satisface

$$2y_0 = 12x_0^2 + 2b_2x_0 + 2b_4 = 0.$$

Luego el punto singular debe ser de la forma $(x_0, 0)$ con $x_0 \in K$ y x_0 raíz doble de $4x^3 + b_2x^2 + 2b_4x + b_6 = 0$. Este polinomio cúbico tiene una raíz doble si y sólo si su discriminante $\Delta = 0$. □

Por último, notar que podemos simplificar la ecuación de Weierstrass todavía más ya que si $car(K) \neq 2,3$ con el cambio $(x, y) \to ((x - 3b_2)/36, y/108)$ obtenemos la siguiente ecuación más simple

$$y^2 = x^3 - 27c_4x - 54c_6.$$

Suponiendo entonces que $car(K) \neq 2,3$ podemos tomar para cualquier curva la *ecuación de Weierstrass simplificada*

$$y^2 = x^3 + ax + b, \tag{2.3}$$

con $\Delta = -16(4a^3 + 27b^2)$, de donde se deduce fácilmente que la condición de no singular equivale a $4a^3 + 27b^2 \neq 0$.

Nota. Si $car(K) = 2$ entonces podemos expresar la ecuación de Weierstrass como $y^2 + y = x^3 + cx + d$ con $c, d \in K$; mientras que si $car(K) = 3$ la ecuación de Weierstrass se reduce a $y^2 = x^3 + bx^2 + cx + d$ con $b, c, d \in K$.

A partir de ahora, salvo que se indique lo contrario, asumiremos que $car(K) \neq 2,3$ y utilizaremos la forma reducida de la ecuación de Weierstrass (2.3) con $4a^3 + 27b^2 \neq 0$, aunque todos los resultados son válidos para el caso general. Denotaremos la curva elíptica como E si no da lugar a confusión, o como $E(K)$ si es necesario remarcar el cuerpo sobre el que está definida.

2.2. Operación de grupo

Definimos una operación para los puntos de E que denotamos por $\oplus$. Para visualizarlo mejor vamos a considerar primero que $K = \mathbb{R}$, de manera que la curva elíptica es ahora una curva ordinaria en el plano.

El punto P_∞ es el elemento identidad del grupo. Sin embargo, la forma más natural de introducir dicho punto es a través del *plano proyectivo*. Recordemos que el plano proyectivo representa al conjunto de clases de equivalencia de la forma (X,Y,Z) (todas las componentes no pueden ser nulas a la vez) donde

$$(X_1,Y_1,Z_1) \sim (X_2,Y_2,Z_2) \Leftrightarrow \exists\ \lambda \in \overline{K}^* \text{ tal que } X_2 = \lambda X_1,\ Y_2 = \lambda Y_1 \text{ y } Z_2 = \lambda Z_1.$$

A cada clase de equivalencia la denominamos *punto proyectivo* y la identificamos con uno cualquiera de sus elementos. De esta manera si en un punto proyectivo $Z \neq 0$, entonces existe una única terna en su clase de equivalencia con la forma $(x,y,1)$, es decir, hay una correspondencia entre el punto proyectivo y el plano afín ($x = X/Z$, $y = Y/Z$). Así, podemos identificar el plano proyectivo como todos los puntos

(x,y) del plano afín más los puntos para los que $Z=0$, estos últimos constituyen la denominada *línea del infinito*. La ecuación (2.1) que está expresada en coordenadas afines la podemos representar ahora en coordenadas proyectivas, como

$$E: Y^2Z + a_1XYZ + a_3YZ^2 = X^3 + a_2X^2Z + a_4XZ^2 + a_6Z^3.$$

Vemos entonces que existe un único punto de E que satisface $Z=0$. Haciendo $Z=0$ en la ecuación anterior obtenemos $0=X^3$, de dónde $X=0$ y así la única clase de equivalencia posible con ambas coordenadas X y Z nulas es la $(0,1,0)$. Así $P_\infty=(0,1,0)$ en coordenadas proyectivas, por lo que es la intersección de la línea del infinito con el eje-y, de ahí que podamos colocar el punto P_∞ infinitamente lejos en el eje-y.

Sea E una curva elíptica general con $K=\mathbb{R}$ y sea ℓ una recta del plano. Entonces la recta ℓ y la curva E tienen tres puntos de intersección P, Q, R (estos tres puntos no tienen por qué ser distintos ya que puede ser que la recta ℓ sea tangente a E).

Operación de grupo. Dados dos puntos $P,Q\in E$, sea ℓ la recta que conecta P y Q (que denotamos como δ en el caso en que sea tangente a E, es decir, en el caso $P=Q$) y sea R el tercer punto de intersección de la recta ℓ con E. Llamamos ℓ' a la recta que conecta R y P_∞. Entonces, el punto $P\oplus Q$ es el punto en el que ℓ' corta con E en R, P_∞ y $P\oplus Q$.

Vamos a expresar más claramente la *suma de dos puntos* $(P\oplus Q)$ y *doblar un punto* $(P\oplus P)$ para la curva particular E dada por 2.3, además de hallar sus expresiones algebraicas. Dada la clara simetría respecto al eje-x de los puntos de E, para cada coordenada x_1 existen dos puntos en E: (x_1,y_1) y $(x_1,-y_1)$. Notar que la recta que conecta (x_1,y_1) y $(x_1,-y_1)$ pasa también, por definición, por P_∞.

- Dados $P,Q\in E$ con $P\neq\pm Q$, definimos la *suma de dos puntos* $P\oplus Q$ como el simétrico al punto de corte entre la recta que une P y Q y la curva E, como se muestra en la Figura 2.1. En el caso particular de que la recta que conecta P y Q sea tangente a la curva en P, definimos $P\oplus Q$ como el simétrico a P.

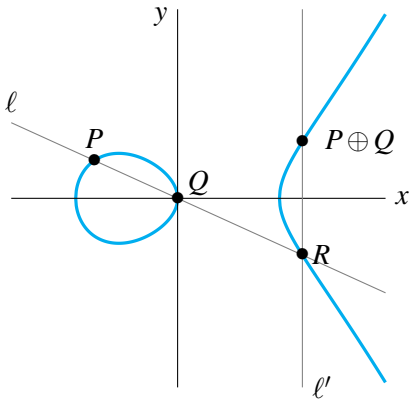

Figura 2.1: Suma de $P\oplus Q$ con $P\neq\pm Q$.

Sea $P=(x_1,y_1)$, $Q=(x_2,y_2)$ con $x_1\neq x_2$ y $P\oplus Q=(x_3,y_3)$ y sea $y=\alpha x+\beta$ la recta ℓ descrita anteriormente. Por definición, $\alpha=\frac{y_2-y_1}{x_2-x_1}$ y $\beta=y_1-\alpha x_1$. Un punto de ℓ está en la curva elíptica E si y sólo si $(\alpha x+\beta)^2=x^3+ax+b$, con lo que obtenemos la ecuación cúbica $x^3-(\alpha x+\beta)^2+ax+b=0$. Por lo que hay un punto de intersección para cada raíz de la ecuación cúbica, teniendo en cuenta que ya sabemos que x_1 y x_2 son raíces de la ecuación. Aplicando que la suma de raíces de un polinomio mónico es igual a menos el coeficiente de la segunda potencia más alta obtenemos que la tercera raíz de la ecuación cúbica es $x_3=\alpha^2-x_1-x_2$ y así $y_3=-(\alpha x_3+\beta)$.

$$P \oplus Q = (x_3, y_3) = \left(\left(\frac{y_2 - y_1}{x_2 - x_1} \right)^2 - x_1 - x_2,\ -y_1 + \left(\frac{y_2 - y_1}{x_2 - x_1} \right) (x_1 - x_3) \right) \tag{2.4}$$

- Dado P, definimos *doblar un punto* $P \oplus P$ como el simétrico al punto de corte entre la recta tangente a P, que la denotamos por δ, y la curva E, como se muestra en la Figura 2.2. En el caso concreto de que el punto P sea un punto de inflexión de la curva, definimos $P \oplus P$ como el simétrico a P.

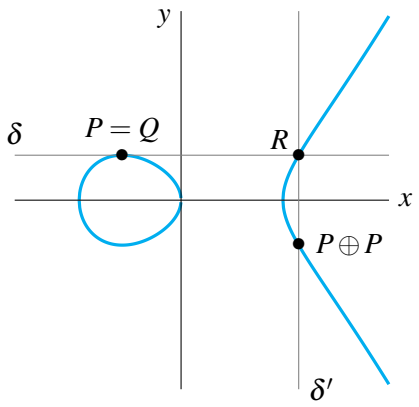

Figura 2.2: Doblar un punto P.

Buscamos ahora las coordenadas del punto $P \oplus P = (x_2, y_2)$ con $P = (x_1, y_1)$, $y_1 \neq 0$. Tenemos que $y = \alpha x + \beta$ es la recta δ donde $\alpha = \dfrac{dx}{dy}(x_1, x_2) = \dfrac{3x_1^2 + a}{2y_1}$. Del mismo modo que para la suma obtenemos

$$P \oplus P = (x_2, y_2) = \left(\left(\frac{3x_1^2 + a}{2y_1} \right)^2 - 2x_1,\ -y_1 + \left(\frac{3x_1^2 + a}{2y_1} \right) (x_1 - x_3) \right). \tag{2.5}$$

En el caso general, es decir, sea cual sea la curva elíptica E (ecuación generalizada, ecuación reducida,...), se tiene que E con la operación $\oplus$ forma un grupo abeliano.

Proposición 2.2. *Una curva elíptica E con la operación $\oplus$ forma un grupo abeliano.*

Demostración. Veamos en primer lugar que si la recta ℓ interseca con E en los puntos P, Q y R (no necesariamente distintos), entonces $(P \oplus Q) \oplus R = P_\infty$. Es claro por definición de la operación de grupo y se puede ver, en particular, en la Figura 2.1, ya que la recta tangente a E en el punto P_∞ corta con E en el punto P_∞ con multiplicidad 3. Notar que la operación definida es claramente conmutativa por construcción, ya que la recta que va de P a Q es la misma que va de Q a P. Veamos que $P \oplus P_\infty = P\ \forall P \in E$. Primero tomamos $Q = P_\infty$. Si construimos la recta ℓ que pasa por P y por P_∞ obtenemos una recta de la forma $x = c$ para alguna constante c. La recta ℓ' coincide con ℓ y se tiene que $\ell \cap E$ en los puntos P, P_∞ y R mientras que $\ell' \cap E$ en los puntos R, P_∞ y $P \oplus P_\infty$. Luego, $P \oplus P_\infty = P$. Probamos ahora que todo elemento tiene inverso. Sea $P \in E$, la recta que une P y P_∞ corta con E en otro punto R. Entonces $P_\infty = (P \oplus P_\infty) \oplus R = P \oplus R$. Así $\forall P \in E$ existe un punto de E que denotamos como $-P$ tal que $P \oplus (-P) = P_\infty$ (en particular, $-P_\infty = P_\infty$). El hecho de que sea asociativa no es para nada obvio, la comprobación puede realizarse utilizando las fórmulas caso por caso, lo que supone una demostración muy laboriosa. Una demostración más clara, utilizando el Teorema de Riemann-Roch, se puede encontrar en [9, pág. 66–67]. Entonces E es grupo abeliano con la operación definida y elemento identidad P_∞. □

Nota. La operación de grupo está formada por funciones racionales y estas fórmulas solo utilizan las cuatro operaciones algebraicas básicas. Remarcar que se cumple entonces que $P \oplus P_\infty = P\ \ \forall P \in E$, $P \oplus Q = Q \oplus P\ \ \forall P, Q \in E$ y el inverso de P lo denotamos $-P$ y $P \oplus (-P) = P_\infty$.

2.2.1. Multiplicación escalar

Dado $n \in \mathbb{Z}$ se tiene que, si n positivo,

$$\begin{aligned}[n]: E &\longrightarrow E \\ P &\longmapsto [n]P = \underbrace{P \oplus P \oplus \cdots \oplus P}_{n \text{ veces}}\end{aligned}$$

Si $n = 0$ se considera que $[0]P = P_\infty$ y si $n < 0$, entonces $[n]P = [-n](-P)$. A $[n]$ la denominamos *multiplicación escalar* en E.

Definición 2.4. Un punto $P = (x, y)$ de una curva elíptica E se dice que tiene *orden finito* si existe un entero n tal que $[n]P = P_\infty$. En tal caso, el *orden* de P es el menor entero positivo tal que $[n]P = P_\infty$.

Proposición 2.3. *Sea $P \in E$ y sea un entero n, entonces $[n](-P) = -[n]P$.*

Demostración. Procedemos por inducción sobre n. Si $n = 1$, $[1](-P) = -P = -[1]P$. Supongamos que se cumple para n, es decir, $[n](-P) = -[n]P$. Veamos si es cierto para $n+1$.

$$[n+1](-P) = [n](-P) \oplus [1](-P) = (-[n]P) \oplus (-[1]P) = -[n+1]P$$

□

2.3. Curvas elípticas definidas sobre $K = \mathbb{Z}_p$

En criptografía, necesitamos grupos abelianos finitos. Sin embargo, las curvas elípticas sobre los reales que hemos visto no forman un grupo abeliano finito, por lo que no se pueden utilizar de esta manera. Es necesario entonces que el cuerpo sobre el que definimos la curva sea finito, para que el grupo abeliano generado por la curva elíptica sobre dicho cuerpo sea también finito. Además, es muy común en criptografía el uso de curvas sobre *cuerpos primos finitos* ($K = \mathbb{Z}_p$ con p primo) que es en lo que nos centraremos en esta sección. Por seguridad computacional el primo p debe ser muy grande, por lo que no hay problema en asumir que $p > 3$ y tomar entonces para E la ecuación de la forma 2.3, dónde $a, b \in \mathbb{Z}_p$.

A partir de ahora vamos a considerar $K = \mathbb{Z}_p$ con $p > 3$ y aunque gráficamente no podemos expresar la operación de grupo como hemos hecho antes, sí que la tenemos expresada algebraicamente con las fórmulas obtenidas en 2.4 y 2.5, que para $K = \mathbb{Z}_p$ pasarían a ser:

- Suma de puntos

$$x_3 = \left(\frac{y_2 - y_1}{x_2 - x_1}\right)^2 - x_1 - x_2 \mod p \qquad y_3 = -y_1 + \left(\frac{y_2 - y_1}{x_2 - x_1}\right)(x_1 - x_3) \mod p$$

- Doblar un punto

$$x_2 = \left(\frac{3x_1^2 + a}{2y_1}\right)^2 - 2x_1 \mod p \qquad y_2 = -y_1 + \left(\frac{3x_1^2 + a}{2y_1}\right)(x_1 - x_3) \mod p$$

Para el caso en el que el primo p sea pequeño, sí que es posible obtener todos los elementos de la curva elíptica $E(\mathbb{Z}_p)$ siguiendo estos pasos que describimos a continuación:

1. Determinar los $x \in \mathbb{Z}_p$ tales que $z = x^3 + cx + d \mod p$ es cuadrado perfecto en $\mathbb{Z}_p$.
2. Buscar todos los $y \in \mathbb{Z}_p$ tales que $y^2 \equiv z \mod p$.

Llamamos *residuos cuadráticos* a los $x \in \mathbb{Z}_p^*$ que son cuadrados perfectos. De esta manera, los valores de z obtenidos con los dos pasos anteriores son $0 \in \mathbb{Z}_p$ y los residuos cuadráticos en $\mathbb{Z}_p^*$.

Proposición 2.4. *Un elemento $z \in \mathbb{Z}_p^*$ es un residuo cuadrático en $\mathbb{Z}_p^*$ si y sólo si $z^{\frac{p-1}{2}} \equiv 1$ mod p. Así, z es un cuadrado perfecto en $\mathbb{Z}_p$ si y sólo si $z = 0$ o $z^{\frac{p-1}{2}} \equiv 1$ mod p.*

Demostración. Dado $(\mathbb{Z}_p^*, \cdot)$ consideramos el siguiente homomorfismo de grupos

$$\begin{aligned} s:\ \mathbb{Z}_p^* &\longrightarrow \mathbb{Z}_p^* \\ y &\longmapsto y^2 \end{aligned}$$

Entonces $Kers = \{a \in \mathbb{Z}_p^* \mid s(a) = 1\} = \{a \in \mathbb{Z}_p^* \mid a^2 = 1\} = \{1, -1\}$, donde $-1 \equiv p-1 \bmod p$. Luego $|Kers| = 2$. Consideramos el conjunto de residuos cuadráticos $Q = \{z \in \mathbb{Z}_p^* \mid z = s(y),\ y \in \mathbb{Z}_p^*\}$. Aplicando el Teorema de isomorfía 1.5 nos queda

$$\begin{aligned} \overline{s}:\ \mathbb{Z}_p^*/Kers &\longrightarrow s(\mathbb{Z}_p^*) \\ xKers &\longmapsto s(x) \end{aligned}$$

Notar que $s(\mathbb{Z}_p^*) = \{y^2 \mid y \in \mathbb{Z}_p^*\} = \{z \in \mathbb{Z}_p^* \mid z = s(y), y \in \mathbb{Z}_p^*\} = Q$. Entonces

$$|Q| = |s(\mathbb{Z}_p^*)| = |\mathbb{Z}_p^*/Kers| = |\mathbb{Z}_p^*|/|Kers| = (p-1)/2.$$

Luego $|Q| = \frac{p-1}{2} = m$. Consideramos la función $g(x) = x^m - 1$. Dado $z \in Q$, entonces $z \equiv y^2 \bmod p$ para algún $y \in \mathbb{Z}_p^*$. Aplicando el Pequeño Teorema de Fermat

$$g(z) = z^m - 1 = y^{2m} - 1 = y^{p-1} - 1 \equiv 0 \mod \text{p}.$$

De donde concluimos que las m raíces de $g(x)$ son precisamente los m elementos de Q. □

Proposición 2.5. *Suponer $p \equiv 3$ mod 4. Si z es un residuo cuadrático en $\mathbb{Z}_p^*$, entonces $y \equiv z^{\frac{p+1}{4}}$ mod p es una raíz cuadrada de z en $\mathbb{Z}_p^*$. La otra raíz cuadrada de z en $\mathbb{Z}_p^*$ es $-y$.*

Demostración. Supongamos que z es un residuo cuadrático en $\mathbb{Z}_p^*$, es decir, $z \equiv y^2 \bmod p$, se sigue que $(z^{\frac{p+1}{4}})^2 = y^{p+1} = y^2 \equiv z \bmod p$. Por tanto, si $p \equiv 3 \bmod 4$ podemos encontrar una raíz cuadrada de z calculando $z^{\frac{p+1}{4}} \bmod p$. □

De esta manera si p es un primo pequeño, $p > 3$ y $p \equiv 3 \bmod 4$, podemos calcular los elementos de la curva elíptica $E(\mathbb{Z}_p)$ con $4a^3 + 27b^2 \neq 0$ como los elementos del conjunto

$$E(\mathbb{Z}_p) = \{(x, \pm y) \in \mathbb{Z}_p \times \mathbb{Z}_p \mid z = x^3 + ax + b \in Q,\ y = z^{\frac{p+1}{4}} \bmod \text{p}\} \cup \{(x,0) \mid x^3 + ax + b = 0\} \cup \{P_\infty\}.$$

2.3.1. Estructura general de las curvas elípticas sobre $\mathbb{Z}_p$

La estructura general de las curvas elípticas sobre $\mathbb{Z}_p$ es simple y está totalmente definida. El grupo abeliano obtenido no tiene por qué ser un grupo cíclico, pero sí se puede probar que siempre es isomorfo al producto de dos grupos cíclicos.

Teorema 2.6. *Sea E una curva elíptica sobre $\mathbb{Z}_p$ con $p > 3$ primo, entonces E es isomorfa al producto directo $\mathbb{Z}_{n_1} \times \mathbb{Z}_{n_2}$ de los grupos aditivos $\mathbb{Z}_{n_1}$ y $\mathbb{Z}_{n_2}$ para algunos enteros positivos n_1 y n_2 tales que $n_2 | n_1$ y $n_2 | (p-1)$.*

Una demostración de este teorema se puede encontrar en [2, pág. 111 y sig.].

2.3.2. Cardinalidad

Por último veamos que aunque el primo p sea muy grande y por tanto no sea posible construir todos los elementos de $E(\mathbb{Z}_p)$, sí que es posible acotar el orden de $E(\mathbb{Z}_p)$. Citamos entonces el siguiente teorema, del que podemos encontrar una demostración detallada en [9, pág. 131].

Teorema 2.7 (Teorema de Hasse)**.** *Sea E una curva elíptica sobre $\mathbb{Z}_p$, entonces*

$$p+1-2\sqrt{p} \leqslant |\, E(\mathbb{Z}_p)\,| \leqslant p+1+2\sqrt{p}.$$

2.4. Forma de Montgomery

Hemos visto que una curva elíptica E puede expresarse respecto a diferentes sistemas de coordenadas (afines, proyectivas,...) y ecuaciones, por lo que resulta de gran interés estudiar la rapidez de la operación de grupo (suma de puntos y doblar un punto) en diferentes sistemas de coordenadas y ecuaciones. Medimos esta rapidez según el número de operaciones elementales necesarias para calcular la suma de puntos y doblar un punto. Dentro de todos los tipos de sistemas de coordenadas y ecuaciones que podemos encontrar, cabe destacar la mejora en cuanto a rapidez de cálculo que aporta la forma de Montgomery.

Definición 2.5. Decimos que una curva elíptica E está en *forma de Montgomery* si la ecuación de la curva es del tipo

$$By^2 = x^3 + Ax^2 + x.$$

En ese caso denotaremos a la curva elíptica como E_M o como E si no da lugar a confusión.

Como la ecuación de la curva elíptica cambia, ya no sirven las expresiones algebraicas calculadas anteriormente para la operación de grupo. Siguiendo el mismo procedimiento, pero con esta nueva ecuación, llegamos a las expresiones algebraicas para el caso de una curva elíptica $E(\mathbb{Z}_p)$ en forma de Montgomery. Para ello, sean $P=(x_1,y_1)$ y $Q=(x_2,y_2)$ con $x_1 \neq x_2$. Entonces $P \oplus Q = (x_3,y_3)$ queda:

$$x_3 = \left(\frac{y_2-y_1}{x_2-x_1}\right)^2 - A - x_1 - x_2 \mod p, \qquad y_3 = \left(\frac{y_2-y_1}{x_2-x_1}\right)(x_1-x_3) - y_1 \mod p. \tag{2.6}$$

Sea ahora $P=(x_1,y_1)$ con $y_1 \neq 0$, entonces doblar un punto $P \oplus P = (x_2,y_2)$ tendría la expresión algebraica:

$$x_2 = \left(\frac{3x_1^2+2Ax_1+1}{2y_1}\right)^2 - A - 2x_1 \mod p, \qquad y_2 = \left(\frac{3x_1^2+2Ax_1+1}{2y_1}\right)(x_1-x_2) - y_1 \mod p. \tag{2.7}$$

Recordemos que $P \oplus P_\infty = P$ para todo $P \in E$ y en particular, $P_\infty \oplus P_\infty = P_\infty$.

La aritmética en esta forma es muy eficiente en cuanto a la computación sólo de la coordenada-x (ver [2, cap. 13, pág. 285] y [7, sec. 10.3.1]). Sea $P=(x_1,y_1) \in E_M$, en coordenadas proyectivas $P=(X_1,Y_1,Z_1)$ (la relación es $x_1 = X_1/Z_1$ y $y_1 = Y_1/Z_1$). Sea $[n]P = (X_n,Y_n,Z_n)$, entonces la suma $[n+m]P = [n]P \oplus [m]P$ viene dada por las siguientes fórmulas donde la coordenada Y_n no aparece.

- Suma de puntos ($n \neq m$)

$$X_{m+n} = Z_{m-n}((X_m - Z_m)(X_n + Z_n) + (X_m + Z_m)(X_n - Z_n))^2$$

$$Z_{m+n} = X_{m-n}((X_m - Z_m)(X_n + Z_n) + (X_m + Z_m)(X_n - Z_n))^2$$

- Doblar un punto ($n = m$)

$$4X_nZ_n = (X_n+Z_n)^2-(X_n-Z_n)^2$$

$$X_{2n} = (X_n+Z_n)^2(X_n-Z_n)^2$$

$$Z_{2n} = 4X_nZ_n((X_n-Z_n)^2+((A+2/4))(4X_nZ_n))$$

Vemos que la suma necesita 4 multiplicaciones elementales y 2 cuadrados; mientras que la operación doblar un punto necesita 3 multiplicaciones y 2 cuadrados. Aunque en algunos criptosistemas se podrá trabajar únicamente con la coordenada-x, en otros será necesario obtener también la coordenada-y.

$$y_n = \frac{Y_n}{Z_n} = \frac{(x_1x_n+1)(x_1+x_n+2A)-2A-(x_1-x_n)^2x_{n+1}}{2By_1}$$

donde $P=(x_1,y_1)$ y x_n y x_{n+1} son la coordenada-x afín de $[n]P$ y $[n+1]P$ respectivamente.

Esta idea para la multiplicación escalar se puede llevar también al caso de una curva elíptica dada por la ecuación de Weierstrass simplificada, aunque aumenta el número de multiplicaciones y cuadrados que se obtienen al sumar y doblar un punto.

2.4.1. Transformación a la forma de Montgomery

Cualquier curva en la forma de Montgomery se puede transformar a la forma simplificada de Weierstrass, basta hacer $a = 1/B^2 - A^2/3B^2$ y $b = -A^3/27B^3 - aA/3B$. El recíproco es falso.

Proposición 2.8. *Una curva en la forma simplificada de Weierstrass se puede transformar a la forma de Montgomery si y sólo si verifica:*

a) El polinomio x^3+ax+b tiene al menos una raíz $\alpha \in \mathbb{Z}_p$.

b) $3\alpha^2+a$ es un residuo cuadrático en $\mathbb{Z}_p$.

En ese caso, basta tomar $A = 3\alpha s$ y $B = s$, donde s es una raíz cuadrada de $(3\alpha^2+a)^{-1}$. Entonces, haciendo el cambio de variable $(x,y) \mapsto (x/s+\alpha, y/s)$ transformamos la ecuación de la curva a la forma de Montgomery. Este tipo de curvas satisfacen:

- El punto $(0,0)$ tiene orden 2.
- El orden de E_M es divisible por 4.

Ejemplo. Sea la curva elíptica E definida sobre $\mathbb{Z}_p$ con $p = 2003$ primo dada por la ecuación simplificada de Weierstrass

$$E : y^2 = x^3 + 1132x + 278.$$

Tomamos $\alpha = 1702 \in \mathbb{Z}_p$, $\alpha^3 \equiv 1947 \bmod p$ y $1132 \cdot 1702 \equiv 1781 \bmod p$. Entonces

$$\alpha^3 + 1132\alpha + 278 = 1947 + 1132(1702) + 278 = 1947 + 1781 + 278 = 4006 \equiv 0 \mod 2003.$$

$\alpha^2 \equiv 466 \bmod p$. Calculamos $\beta = 3\alpha^2 + a = 3(466) + 1132 = 2530 \equiv 527 \bmod 2003$ y comprobamos que es un residuo cuadrático en $\mathbb{Z}_p$ aplicando la proposición 2.4,

$$527^{\frac{2003-1}{2}} \equiv 1 \mod 2003.$$

Calculamos el inverso de β utilizando el algoritmo de Euclides extendido $\beta^{-1} = 992$. Tenemos que β^{-1} también es un residuo cuadrático en $\mathbb{Z}_p$. Para calcular una raíz cuadrada de β^{-1} aplicamos la proposición 2.5:

$$2003 \equiv 3 \bmod 4, \qquad 992^{\frac{2003+1}{4}} \equiv 899 \bmod 2003.$$

Luego $s = 899$ y tenemos $A = 3 \cdot 1702 \cdot 899 \equiv 1421 \bmod p$ y $B = 899$. Por último, construimos el isomorfismo $(x,y) \mapsto (899x+1702, 899y)$ que transforma los puntos de E en puntos de E_M,

$$E_M : 899y^2 = x^3 + 1421x^2 + x.$$

Nota. Los cálculos han sido realizados con Sage.

Capítulo 3

Criptografía de clave pública basada en el Problema del Logaritmo Discreto

3.1. Problema del Logaritmo Discreto

Definición 3.1. Problema del Logaritmo Discreto (PLD)
Sea (G, $\cdot$) grupo cíclico finito y sea $|G| = n$. Tomamos un elemento primitivo $\alpha \in G$ y otro elemento $\beta \in G$, el *Problema del Logaritmo Discreto* consiste en encontrar el entero x, con $1 \leq x \leq n$ tal que

$$\beta = \alpha \cdot \alpha \cdot \ldots \cdot \alpha = \alpha^x.$$

A este entero x lo llamamos *logaritmo discreto de* β *en base* α y lo denotamos

$$x = log_\alpha \beta.$$

Es claro que dicho entero x existe ya que α es un elemento primitivo. Además, notar que computar $\beta = \alpha^x$, es muy sencillo, pero invertir la aplicación resulta casi imposible debido a la aleatoriedad de la operación de grupo, ya que no esperamos encontrar el x tal que $\beta = \alpha^x$ hasta haber comprobado todas las posibilidades.

Nota. La dificultad varía según el grupo G escogido.

Ejemplo. PLD aplicado a $\mathbb{Z}_p^*$.
Sea p primo y sea el grupo cíclico finito $\mathbb{Z}_p^*$ de orden $p-1$. Dado un elemento primitivo $\alpha \in \mathbb{Z}_p^*$ y otro elemento $\beta \in \mathbb{Z}_p^*$, el PLD consiste en determinar el entero $1 \leq x \leq p-1$ tal que

$$\alpha^x \equiv \beta \bmod p.$$

Si el primo p es suficientemente grande, computar logaritmos discretos módulo p es muy complicado, mientras que la exponenciación $\alpha^x \equiv \beta \bmod p$ es muy sencilla. Sea $p = 11$ y $G = (\mathbb{Z}_{11}^*, \cdot)$ grupo cíclico finito con elemento primitivo $\alpha = 2$.

$\alpha^1 = 2$	$\alpha^3 = 8$	$\alpha^5 = 10$	$\alpha^7 = 7$	$\alpha^9 = 6$
$\alpha^2 = 4$	$\alpha^4 = 5$	$\alpha^6 = 9$	$\alpha^8 = 3$	$\alpha^{10} = 1$

Vemos que, efectivamente, como $\alpha = 2$ es un elemento primitivo, las potencias de α generan todos los elementos del grupo $\mathbb{Z}_{11}^*$, sin embargo la forma en la que van apareciendo dichos elementos es totalmente arbitraria. La dificultad de resolver el Problema del Logaritmo Discreto se basa entonces en la aleatoriedad entre el exponente y los elementos del grupo.

Actualmente no se conoce ningún algoritmo eficaz para resolver el PLD y se considera irresoluble siempre que se escojan adecuadamente el grupo G y α. Es importante, además, para posibilitar los procesos criptográficos, que el grupo G tenga una operación rápida.

3.2. Criptografía de clave pública

La idea general de la criptografía de clave pública es que en el proceso de cifrar-descifrar se utilizan dos claves, una conocida por todos, la *clave pública* y otra conocida únicamente por cada persona, la *clave privada*, de manera que cada persona dispone de un par (k_{pub}, k_{pr}), de ahí que se denomine también criptografía asimétrica.

Como la criptografía con clave simétrica (algoritmos DES y AES [8, cap. 3 y 4]) es computacionalmente más rápida y además no es aconsejable utilizar la misma clave en el cifrado masivo de mensajes, la criptografía de clave pública se debe complementar con el uso de una clave simétrica, por lo que es necesario utilizar una clave temporal y conjunta para las dos partes y que, por supuesto, no pueden conocer terceras partes. En mantener en secreto esa clave conjunta es donde interviene directamente la criptografía de clave pública.

El procedimiento de clave pública más popular y extendido en la actualidad es el RSA, muy empleado en la firma electrónica (certificados digitales) y en el intercambio seguro de claves (transporte de claves), donde se combina con un método simétrico, que suele ser AES, con el que se encripta. Podemos encontrar una descripción de su funcionamiento y de aspectos computacionales en [8, pág. 173–199]. Nosotros vamos a centrarnos en los sistemas criptográficos de clave pública basados en el Problema del Logaritmo Discreto, los cuales describimos a continuación de forma general y posteriormente los aplicaremos al caso de grupos de curvas elípticas.

3.2.1. Intercambio de claves Diffie-Hellman

El intercambio de claves Diffie-Hellman (DHKE) fue propuesto por Whitfield Diffie y Martin Hellman en 1976 con la publicación del artículo "New directions in Cryptography" [3], influenciados por el trabajo anterior de Ralph Merkle, siendo el primer método asimétrico publicado en la literatura. Este método, sencillo pero muy eficaz, permite establecer una clave común entre dos partes que se comunican a través de un canal inseguro (de ahí que todos los parámetros que se envían sean públicos), aplicando el Problema del Logaritmo Discreto. Con esa clave común establecida, las dos partes pueden cifrar y descifrar, es decir, esta clave funciona como clave simétrica.

Consideramos un grupo cíclico finito G y denotamos la operación del grupo $\cdot$, sea $|G| = n$ y $\alpha \in G$ tal que $G = \langle \alpha \rangle$. Notar que

$$k = (\alpha^x)^y = (\alpha^y)^x,$$

ya que la operación de grupo es conmutativa (todo grupo cíclico es abeliano). Ésta es la idea básica detrás del intercambio de claves Diffie-Hellman y el valor k es la clave de sesión que se establece entre las dos partes.

Para explicar el funcionamiento del método, suponemos que Alicia y Bob son las dos partes que quieren establecer una clave secreta común a través de un canal que siempre se considera inseguro. En primer lugar, se lleva a cabo la elección de los parámetros: escoger un grupo cíclico finito $(G, \cdot)$ y elegir $\alpha \in G$ tal que $G = \langle \alpha \rangle$ ($|G| = n$), siendo ambos parámetros públicos. A partir de aquí se genera una clave secreta conjunta empleando el siguiente método:

1. Alicia escoge su clave privada $a = k_{pr,A} < n$, a partir de la cual genera su clave pública $k_{pub,A} = \alpha^a$ y envía $k_{pub,A}$ a Bob.

2. Simultáneamente Bob escoge su clave privada $b = k_{pr,B} < n$, genera su clave pública $k_{pub,B} = \alpha^b$ y envía $k_{pub,B}$ a Alicia.

3. Alicia dispone de (G, α, a, α^b) y calcula $k_{AB} = (\alpha^b)^a$.

4. Bob dispone de (G, α, b, α^a) y calcula $k_{BA} = (\alpha^a)^b$.

Se prueba fácilmente que Alicia y Bob calculan, respectivamente, lo mismo.

$$k_{AB} = k^a_{pub,B} = (\alpha^b)^a = (\alpha)^{ab}$$
$$k_{BA} = k^b_{pub,A} = (\alpha^a)^b = (\alpha)^{ab}$$

Nota. No es necesario que G sea cíclico, podemos tomar G grupo finito cualquiera, $\alpha \in G$ y tomar el grupo cíclico $C = \langle \alpha \rangle$ de tal manera que escogemos $a, b \in C$.

Así $k_{AB} = k_{BA}$ es la clave común que emplearán Alicia y Bob para mantener una comunicación segura. Habitualmente esta clave establecida se utiliza como clave simétrica en algoritmos de tipo AES; para lo cual es suficiente con tomar los últimos 128 bits (en el caso de AES-128), o también es común aplicarle a k_{AB} una *función hash* (como SHA-1 [8, cap. 11]) y trabajar con el valor hash obtenido.

Es importante ver que en DHKE se utilizan ambos métodos: clave asimétrica y clave simétrica. El par (k_{pub}, k_{pr}) es asímetrico y se emplea para mantener en secreto la clave conjunta común k_{AB}, que es simétrica. Notar que el tamaño de la clave de cada persona sería $(k_{pr}, k_{pub}) = (a, \alpha^a)$.

Respecto a la seguridad del método, éste no es seguro contra ataques *activos* (denominados *man in the middle*), que permiten modificar o crear falsos mensajes llegando a ser eficaces en la suplantación de las partes. Sin embargo, contra ataques *pasivos*, en los que no interviene de forma activa una tercera parte, sino que ésta sólo dispone de los datos públicos, el DHKE es un problema computacional intratable si se escogen los parámetros adecuadamente.

3.3. Curvas elípticas en criptografía de clave pública

En los últimos 30 años se han desarrollado numerosas investigaciones sobre curvas elípticas y, concretamente, se han realizado múltiples estudios de éstas sobre diferentes cuerpos y coordenadas. Aunque a simple vista pueda parecer extraña su utilización en el ámbito de la criptografía, la introducción de curvas elípticas en aplicaciones criptográficas ([6] y [5]) fue debida a la **fácil construcción de un grupo** y a la **rapidez en el cálculo de su operación**, así como a la dificultad que supone encontrar un ataque eficiente contra el Problema del Logaritmo Discreto aplicado al grupo generado por éstas, hecho del que se está todavía muy lejos de encontrar una solución aceptable. Durante más de dos décadas se ha intentado atacar el PLD con muy poco éxito, aunque sí que se ha llegado a encontrar la solución a dicho problema para grupos de orden no suficientemente grande.

La criptografía con curvas elípticas proporciona un nivel de seguridad similar a RSA con la ventaja de que los operandos son mucho más pequeños (160-256 bits frente a 1024-3072 bits en RSA), lo que las convierte en candidatas perfectas para sustituir al famoso método RSA.

3.3.1. DHKE con curvas elípticas

Definición 3.2 (PLD aplicado a curvas elípticas)**.** Sea E una curva elíptica sobre $\mathbb{Z}_p$, dados dos puntos $Q, P \in E$ con $P \in \langle Q \rangle$, el Problema del Logaritmo Discreto en E consiste en encontrar el entero positivo x tal que $[x]Q = P$.

Para utilizar el grupo generado por una curva elíptica sobre $\mathbb{Z}_p$ para algún primo p como grupo para el criptosistema DHKE, es necesario que p sea extremadamente grande para que se trate de un método de encriptación seguro, hasta tal punto que se considera seguro si el grupo G generado contiene un subgrupo cíclico de orden al menos 2^{160}. Además, sólo es necesario encontrar un elemento de E que tenga orden suficientemente grande, y no es necesario en ningún caso hallar todos los elementos de E.

Recordemos que E es la curva elíptica formada por el conjunto de pares $(x, y) \in \mathbb{Z}_p \times \mathbb{Z}_p$ soluciones a la ecuación

$$y^2 = x^3 + ax + b \mod p,$$

donde $4a^3 + 27b^2 \neq 0 \mod p$ y el punto en el infinito P_∞.

Sea $Q \in E$ de orden primo q cercano a p y consideramos $\langle Q \rangle$, es decir, $q = |\langle Q \rangle|$. En el caso de que E sea grupo abeliano cíclico, tomamos $E = \langle Q \rangle$. Recordar que Q es público.

1. Alicia escoge su clave privada $a = k_{pr,A} < q$, a partir de la cual genera su clave pública $k_{pub,A} = [a]Q$ y envía $k_{pub,A}$ a Bob.
2. Simultáneamente Bob escoge su clave privada $b = k_{pr,B} < q$, genera su clave pública $k_{pub,B} = [b]Q$ y envía $k_{pub,B}$ a Alicia.
3. Alicia dispone de $(G, Q, a, b[Q])$ y calcula $k_{AB} = [a]([b]Q)$.
4. Bob dispone de $(G, Q, b, [a]Q)$ y calcula $k_{BA} = [b]([a]Q)$.

Se prueba fácilmente que Alicia y Bob calculan, respectivamente, lo mismo.

$$k_{AB} = [a]k_{pub,B} = [a]([b]Q) = [a][b]Q$$
$$k_{BA} = [b]k_{pub,A} = [b]([a]Q) = [a][b]Q$$

En este caso el tamaño de clave de cada persona sería $(k_{pr}, k_{pub}) = (a, [a]Q) = (a, (x, y))$ con a, x, $y \in \mathbb{Z}_p$.

3.3.2. Protocolo de firma con curvas elípticas

El protocolo de firma con curvas elípticas permite verificar la autenticidad de los mensajes transmitidos. Es decir, permite ver si el mensaje ha sido enviado realmente por la persona que dice haberlo enviado.

Seguimos con la misma notación que en la sección anterior. Sea m un mensaje y sea $Q \in E$ de orden primo q cercano a p tal que $C = \langle Q \rangle$, es decir, $q = |C|$. El par de claves escogido por Alicia sería $(k_{pr}, k_{pub}) = (a, T)$, con $T = [a]Q$. El protocolo de firma sería:

1. Tomar un entero e con $e < q$ como clave temporal $k_E = e$.
2. Calcular $R = [k_E]Q = [e]Q = (x_R, y_R)$.
3. Nos quedamos sólo con la coordenada-x, $r = x_R$.
4. Sea $h(\cdot)$ una función hash concreta. Aplicamos la función hash al mensaje m, $h(m)$ y calculamos $s \equiv (h(m) + r \cdot a)k_E^{-1} \mod q$.

De esta manera, la firma de un mensaje m consiste en el par de enteros (r, s). Para saber si el mensaje m ha sido enviado realmente por Alicia, Bob calcula, en primer lugar, los valores auxiliares $w \equiv s^{-1} \mod q$, $u_1 \equiv w \cdot h(m) \mod q$ y $u_2 \equiv w \cdot r \mod q$. A continuación computa $P = u_1 Q + u_2 T$ y escoge la coordenada-x, que la denotamos x_P. Entonces, si $x_P \equiv r \mod q$ la firma es válida; si $x_P \not\equiv r \mod q$ la firma no es válida y el mensaje no ha sido enviado por Alicia.

Comprobamos que este método funciona. Para ello, veamos cómo una firma (r, s), satisface la condición $r \equiv x_P \mod q$. Sea $s \equiv (h(m) + r \cdot a)k_E^{-1} \mod q$. Entonces

$$k_E \equiv s^{-1}h(m) + as^{-1}r \mod q.$$

Como $w \equiv s^{-1}$, $u_1 \equiv w \cdot h(m)$ y $u_2 \equiv w \cdot r$, tenemos que

$$k_E \equiv u_1 + au_2 \mod q.$$

Como $Q \in E$ genera un grupo cíclico de orden q, multiplicamos a ambos lados por Q.

$$[k_E]Q = [u_1 + au_2]Q = [u_1]Q + [au_2]Q = [u_1]Q + [u_2]([a]Q) = [u_1]Q + [u_2]T$$

$[k_E]Q$ es igual a $[u_1]Q + [u_2]T$ si se ha utilizado la clave privada correcta, es decir si Alicia utiliza su clave privada. Esta condición es la que se comprueba al verificar la firma, comparando la coordenada-x de $P = u_1 Q + u_2 T$ con la coordenada-x de $R = [k_E]Q$.

Capítulo 4

Curva25519, la curva elíptica de WhatsApp

La plataforma WhatsApp inició en abril de 2016 un proceso de cifrado de todos los mensajes entre sus usuarios de manera que sólo ellos pudieran descifrarlos. Es lo que denominaron *"cifrado de extremo a extremo"* y aparece indicado en nuestros teléfonos en forma de rótulo amarillo en la parte superior de la pantalla la primera vez que iniciamos conversación con cualquier contacto. Para este cifrado se eligió una curva elíptica de tal manera que sus claves fueran más pequeñas que lo usual en curvas elípticas, a la vez que fuera segura y de fácil cálculo, ya que todo el proceso de cifrar-descifrar se lleva a cabo en los teléfonos móviles. La curva escogida se conoce como *Curva25519*.

En este capítulo vamos a estudiar en profundidad la familia de curvas elípticas a la cual pertenece la utilizada en WhatsApp, tomando como artículo base [1]. La primera sección la dedicamos al estudio algebraico de estas curvas, mientras que en la siguiente sección construimos la Curva25519 y explicamos las ventajas computacionales que ofrece y, por tanto, la hacen idónea para su uso en criptografía. Por último, explicamos el protocolo de encriptación de WhatsApp.

4.1. Estudio algebraico de una familia de curvas en forma de Montgomery sobre $\mathbb{Z}_p$

En esta sección consideramos una familia especial de curvas elípticas sobre $\mathbb{Z}_p$, con $p \geq 5$ primo, dadas por la ecuación de Montgomery

$$y^2 = x^3 + Ax^2 + x,$$

donde $A^2 - 4$ no es residuo cuadrático módulo p. El discriminante es $\Delta = A^2 - 4 \neq 0$.

Notar que los residuos cuadráticos juegan un papel muy importante en curvas elípticas, ya que por definición de la ecuación de la curva, sólo están en E los puntos de coordenada x tales que $x^3 + Ax^2 + x$ es residuo cuadrático.

En la demostración de la proposición 2.4 hemos visto que en $\mathbb{Z}_p$ hay exactamente $(p-1)/2$ elementos que no son cuadrados perfectos. Por esa misma proposición sabemos que:

$$\begin{aligned}
&\alpha^{(p-1)/2} \equiv 1 \mod p \text{ si } \alpha \text{ es un residuo cuadrático en } \mathbb{Z}_p,\\
&\alpha^{(p-1)/2} \equiv -1 \mod p \text{ si } \alpha \text{ no es cuadrado perfecto en } \mathbb{Z}_p,\\
&\alpha^{(p-1)/2} \equiv 0 \text{ si } \alpha = 0.
\end{aligned}$$

Ahora fijamos un elemento $\lambda \in \mathbb{Z}_p$ que no sea residuo cuadrático, por ejemplo, el menor. Si α no es cuadrado perfecto en $\mathbb{Z}_p$, entonces α/λ es residuo cuadrático en $\mathbb{Z}_p$, porque

$$\left(\frac{\alpha}{\lambda}\right)^{(p-1)/2} \equiv (-1/-1) = 1 \mod p. \tag{4.1}$$

Sea el polinomio $p(x) = x^2 - \lambda$ irreducible en $\mathbb{Z}_p$, porque λ no es residuo cuadrático. Formamos L como sigue

$$L = \mathbb{Z}_p[x]/(x^2 - \lambda) = \{f(x) \in \mathbb{Z}_p[x] \mid grad(f(x)) < grad(p(x))\},$$

que es cuerpo con las operaciones de suma y producto de polinomios módulo $p(x)$, ya que $x^2 - \lambda$ es irreducible (proposición 1.12). Por definición, $L = \{a + bx \mid a, b \in \mathbb{Z}_p\}$. Con la identificación $a + bx = (a, b)$ tenemos que L es el conjunto

$$L = \{(a, b) \mid a, b \in \mathbb{Z}_p\}.$$

Y las operaciones de suma y producto con esta identificación pasan a ser:

$$(a, b) + (c, d) = (a + c, b + d), \qquad (a, b) \cdot (c, d) = (ac + \lambda bd, ad + bc).$$

El inverso de cada elemento del cuerpo $(a, b) \neq (0, 0)$ es de la forma $\left(\frac{a}{a^2 - \lambda b^2}, \frac{-b}{a^2 - \lambda b^2}\right)$. Notar que $(0, 1) \cdot (0, 1) = (\lambda, 0)$.

Sea t raíz del polinomio, es decir, $t^2 - \lambda = 0$, entonces $t = \sqrt{\lambda} \notin \mathbb{Z}_p$. Así, $\mathbb{Z}_p(\sqrt{\lambda}) = \{a + b\sqrt{\lambda} \mid a, b \in \mathbb{Z}_p\}$ es el menor subcuerpo de L que contiene a $\mathbb{Z}_p$ y a $\sqrt{\lambda}$. Luego tenemos que

$$\mathbb{Z}_p \subseteq \mathbb{Z}_p(\sqrt{\lambda}) \subseteq \mathbb{Z}_p[x]/(x^2 - \lambda) = L.$$

Es decir, $\mathbb{Z}_p(\sqrt{\lambda})/\mathbb{Z}_p$ es una extensión de cuerpo y el grado de la extensión es $|\mathbb{Z}_p(\sqrt{\lambda}) : \mathbb{Z}_p| = 2$. Del mismo modo, $L/\mathbb{Z}_p$ es una extensión de cuerpo y el grado de la extensión es $|L : \mathbb{Z}_p| = 2$. Luego, $\mathbb{Z}_p(\sqrt{\lambda}) = \mathbb{Z}_p[x]/(x^2 - \lambda) = L$.

Sea $GF(p^2)$ el cuerpo de Galois de p^2 elementos. Como $\mathbb{Z}_p(\sqrt{\lambda})$ es finito y $|\mathbb{Z}_p(\sqrt{\lambda})| = p^2$, entonces

$$\mathbb{Z}_p(\sqrt{\lambda}) = \mathbb{Z}_p[x]/(x^2 - \lambda) = GF(p^2).$$

Llamamos M al siguiente homomorfismo de anillos.

$$\begin{aligned} M : \mathbb{Z}_p &\longrightarrow \mathbb{Z}_p(\sqrt{\lambda}) \\ a &\longmapsto (a, 0) \end{aligned}$$

Sean $a, b \in \mathbb{Z}_p$,

$$\begin{aligned} M(a + b) = (a + b, 0) = (a, 0) + (b, 0) = M(a) + M(b), \\ M(ab) = (ab, 0) = (a, 0) \cdot (b, 0) = M(a) \cdot M(b). \end{aligned}$$

Además, por ser homomorfismo $M(0_{\mathbb{Z}_p}) = 0_{E(\mathbb{Z}_p(\sqrt{\lambda}))}$ y por ser ambos anillos con identidad $M(1_{\mathbb{Z}_p}) = 1_{E(\mathbb{Z}_p(\sqrt{\lambda}))}$. Por último, es claro que la aplicación es inyectiva. Podemos abreviar entonces $(a, 0) = a$ y en ese caso, $(0, 1) = \sqrt{\lambda}$.

Proposición 4.1. *Si $x^3 + Ax^2 + x = 0$, entonces $x = 0$.*

Demostración. En cualquier otro caso, si $x \neq 0$, tendríamos $A = -(x + x^3)/x^2$.

$$A^2 - 4 = \left(\frac{-(x + x^3)}{x^2}\right)^2 - 4 = \frac{1 + 2x^2 + x^4}{x^2} - 4 = \left(\frac{x^2 - 1}{x}\right)^2$$

Entonces $A^2 - 4 \bmod p$ sería residuo cuadrático en $\mathbb{Z}_p$, luego contradicción. □

Consideramos la curvas elípticas $E(\mathbb{Z}_p)$ y $E(\mathbb{Z}_p(\sqrt{\lambda}))$. Como consecuencia de la proposición anterior, los elementos $(a, 0) \notin E(\mathbb{Z}_p(\sqrt{\lambda}))$ si $a \neq 0$. Recordemos que las curvas elípticas $E(\mathbb{Z}_p)$ y $E(\mathbb{Z}_p(\sqrt{\lambda}))$ junto con la operación que hemos definido como $\oplus$, son grupos abelianos.

A continuación, demostramos un teorema que asegura el óptimo funcionamiento de esta familia de curvas elípticas, ya que nos permitirá trabajar sólo con la coordenada-x de los puntos de $E(\mathbb{Z}_p)$.

Teorema 4.2. *Sea $p \geq 5$ primo. Sea A entero tal que A^2-4 mod p no es residuo cuadrático en $\mathbb{Z}_p$. Sea la curva elíptica E sobre $\mathbb{Z}_p$ de ecuación*

$$y^2 = x^3 + Ax^2 + x.$$

Definimos la aplicación

$$\begin{aligned} X_0 : \ E(\mathbb{Z}_p(\sqrt{\lambda})) &\longrightarrow \ \mathbb{Z}_p(\sqrt{\lambda}) \\ (x,y) \quad &\longmapsto X_0(x,y) = x \\ P_\infty \quad &\longmapsto X_0(P_\infty) = 0 \end{aligned}$$

Entonces dado un entero n y dado $q \in \mathbb{Z}_p$, existe un único $s \in \mathbb{Z}_p$ tal que $X_0([n]Q) = s \ \forall Q \in E(\mathbb{Z}_p(\sqrt{\lambda}))$ tal que $X_0(Q) = q$.

Demostración. Vamos a probar que hay exactamente dos $Q \in E(\mathbb{Z}_p(\sqrt{\lambda}))$ tales que $X_0(Q) = q$, y ambos tienen el mismo valor de $X_0([n]Q) \in \mathbb{Z}_p$. Recordemos que $[n]$ representa la multiplicación escalar. Definimos $\alpha = q^3 + Aq^2 + q$ y analizamos los distintos casos dependiendo del valor de α.

Si $\alpha = 0$, por la proposición 4.1, $q = 0$. La única raíz cuadrada de 0 en $\mathbb{Z}_p$ es 0. Es más, en $\mathbb{Z}_p(\sqrt{\lambda})$ la única raíz de 0 es 0. El conjunto

$$\Omega = \{Q \in E(\mathbb{Z}_p(\sqrt{\lambda})) \mid X_0(Q) = q\} = \{Q \in E(\mathbb{Z}_p(\sqrt{\lambda})) \mid X_0(Q) = 0\} = \{(0,0),\ P_\infty\}$$

es subgrupo de $E(\mathbb{Z}_p(\sqrt{\lambda}))$, ya que hemos visto que

$$P_\infty \oplus P_\infty = P_\infty, \qquad (0,0) \oplus P_\infty = (0,0), \qquad (0,0) \oplus (0,0) = P_\infty.$$

Para cada $Q \in E(\mathbb{Z}_p(\sqrt{\lambda}))$ tal que $Q \in \Omega$, entonces $[n]Q \in \{(0,0),\ P_\infty\}$, por lo que $X_0([n]Q) = 0$.

Si α es un residuo cuadrático en $\mathbb{Z}_p$, tomamos $r \in \mathbb{Z}_p$ raíz cuadrada de α. Ahora $q \neq 0$. Las únicas raíces cuadradas de α en $\mathbb{Z}_p$ son $\pm r$. Es claro que $\pm r$ son también las únicas raíces de α en $\mathbb{Z}_p(\sqrt{\lambda})$. Entonces

$$\Omega = \{Q \in E(\mathbb{Z}_p(\sqrt{\lambda})) \mid X_0(Q) = q\} = \{(q,r), (q,-r)\}.$$

Sea $s = X_0([n](q,r))$. Como $(q,r) \in E(\mathbb{Z}_p) \leq E(\mathbb{Z}_p(\sqrt{\lambda}))$, entonces $[n](q,r) \in E(\mathbb{Z}_p)$. Luego $s \in \mathbb{Z}_p$. Por otra parte, aplicando la proposición 2.3,

$$X_0([n](q,-r)) = X_0([n](-(q,r))) = X_0(-[n](q,r)) = X_0(-(s,y')) = X_0(s,-y') = s \ \text{ con } y' \in \mathbb{Z}_p.$$

Queda probado que $X_0([n]Q) = s \ \forall Q \in E(\mathbb{Z}_p(\sqrt{\lambda}))$ tal que $Q \in \Omega$.

Si α no es residuo cuadrático en $\mathbb{Z}_p$, tomamos α/λ con λ el menor elemento de $\mathbb{Z}_p$ tal que λ no es residuo cuadrático en $\mathbb{Z}_p$. Hemos probado en (4.1) que α/λ es residuo cuadrático en $\mathbb{Z}_p$. Sea r raíz cuadrada de α/λ. Como $q \neq 0$, las únicas raíces cuadradas de $\alpha = q^3 + Aq^2 + q$ en $\mathbb{Z}_p(\sqrt{\lambda})$ son $\pm r\sqrt{\lambda}$. Entonces

$$\Omega = \{Q \in E(\mathbb{Z}_p(\sqrt{\lambda})) \mid X_0(Q) = q\} = \{(q, r\sqrt{\lambda}), (q, -r\sqrt{\lambda})\}.$$

Veamos ahora que el conjunto $\{P_\infty\} \cup \left(E(\mathbb{Z}_p(\sqrt{\lambda})) \cap (\mathbb{Z}_p \times \sqrt{\lambda}\mathbb{Z}_p)\right) = H$ es subgrupo de $E(\mathbb{Z}_p(\sqrt{\lambda}))$. Los elementos de H son los puntos de la curva $E(\mathbb{Z}_p(\sqrt{\lambda}))$ de la forma $(a, b\sqrt{\lambda})$ con $a, b \in \mathbb{Z}_p$. Recordamos las expresiones algebraicas calculadas para la suma de puntos (ver ecuación (2.6)) y doblar un punto (ver ecuación (2.7)). Para la suma de puntos, dados $P = (a, b\sqrt{\lambda})$, $Q = (c, d\sqrt{\lambda}) \in H$ con $a \neq c$, tenemos

$$\frac{d\sqrt{\lambda} - b\sqrt{\lambda}}{c-a} = \frac{(d-b)\sqrt{\lambda}}{c-a} = (c-a)^{-1}(d-b)\sqrt{\lambda} = u\sqrt{\lambda} \in \sqrt{\lambda}\mathbb{Z}_p$$

con $u = (c-a)^{-1}(d-b) \in \mathbb{Z}_p$. Entonces,

$$\beta = \left(\frac{d\sqrt{\lambda} - b\sqrt{\lambda}}{c-a}\right)^2 = (u\sqrt{\lambda})^2 = u^2\lambda \in \mathbb{Z}_p, \text{ ya que } \lambda \in \mathbb{Z}_p.$$

Así tenemos que $\beta - A - a - c \in \mathbb{Z}_p$. Luego la coordenada-$x$ de $P \oplus Q$ está en $\mathbb{Z}_p$. Para la operación doblar un punto, dado $P = (a, b\sqrt{\lambda}) \in H$ con $b\sqrt{\lambda} \neq 0$, tenemos

$$\frac{3a^2 + 2Aa + 1}{2b\sqrt{\lambda}} = \frac{(3a^2 + 2Aa + 1)\sqrt{\lambda}}{2b\lambda} = u\sqrt{\lambda} \in \sqrt{\lambda}\mathbb{Z}_p,$$

ya que $u = \dfrac{(3a^2 + 2Aa + 1)}{2b\lambda} = (3a^2 + 2Aa + 1)(2b\lambda)^{-1} \in \mathbb{Z}_p$, porque $\lambda \in \mathbb{Z}_p$. Entonces

$$\beta = \left(\frac{3a^2 + 2Aa + 1}{2b\sqrt{\lambda}}\right)^2 = (u\sqrt{\lambda})^2 = u^2\lambda \in \mathbb{Z}_p, \text{ ya que } \lambda \in \mathbb{Z}_p.$$

Así tenemos que $\beta - A - 2a \in \mathbb{Z}_p$. Luego la coordenada-$x$ de $P \oplus P$ está en $\mathbb{Z}_p$. Por tanto, H es subgrupo de $E(\mathbb{Z}_p(\sqrt{\lambda}))$. Sea $s = X_0([n](q, r\sqrt{\lambda}))$. Como $(q, r\sqrt{\lambda}) \in H$, entonces $[n](q, r\sqrt{\lambda}) \in H$. Así, $s \in \mathbb{Z}_p$. Por otro lado, aplicando la proposición 2.3 queda

$$[n](q, -r\sqrt{\lambda}) = [n](-(q, r\sqrt{\lambda})) = -[n](q, r\sqrt{\lambda}),$$

$$X_0([n](q, -r\sqrt{\lambda})) = X_0(-[n](q, r\sqrt{\lambda})) = X_0(-(s, y')) = X_0(s, -y') = s \text{ con } y' \in \mathbb{Z}_p(\sqrt{\lambda}).$$

Queda probado que $X_0([n]Q) = s \ \ \forall Q \in E(\mathbb{Z}_p(\sqrt{\lambda}))$ tal que $Q \in \Omega$. □

4.2. La curva de WhatsApp

En esta sección describimos los parámetros escogidos por WhatsApp para formar la Curva25519, utilizada en su protocolo de cifrado de mensajes. Esta curva pertenece a la familia de curvas descritas en la sección anterior.

Como primo p se ha tomado $2^{255} - 19$. Entonces $\mathbb{Z}_p = \mathbb{Z}_{2^{255}-19}$. Comprobando en Sage

$$2^{\frac{2^{255}-19-1}{2}} \equiv p - 1 \not\equiv 1 \mod 2^{255} - 19,$$

por la proposición 2.4, 2 no es residuo cuadrático en $\mathbb{Z}_p$. Así, tomamos $\lambda = 2$. Entonces $GF(p^2) = GF((2^{255} - 19)^2) = \mathbb{Z}_p(\sqrt{2})$. Se ha escogido $A = 486662$. Comprobamos, como antes, que $A^2 - 4$ mod p no es residuo cuadrático en $\mathbb{Z}_p$, es decir,

$$(A^2 - 4)^{\frac{2^{255}-19-1}{2}} \equiv p - 1 \not\equiv 1 \mod 2^{255} - 19.$$

Definimos la *Curva25519* como la curva elíptica E sobre $\mathbb{Z}_{2^{255}-19}$ formada por los puntos $(x,\ y) \in \mathbb{Z}_{2^{255}-19} \times \mathbb{Z}_{2^{255}-19}$ que satisfacen la ecuación

$$y^2 = x^3 + 486662x^2 + x,$$

junto con el punto P_∞.

Por último, para aplicar el protocolo DHKE, como se ha explicado en la sección 3.3.1, es necesario tomar un elemento $Q \in E(\mathbb{Z}_{2^{255}-19})$ de orden primo q cercano a p. A este punto Q lo llamamos *punto base*. Hemos probado que podemos trabajar sólo con la coordenada-x, por lo que no necesitamos el punto base sino la coordenada-x del punto base. En este caso, 9 es la coordenada-x del punto Q elegido. En cualquier caso, sabiendo la coordenada-x, podemos calcular la coordenada-y del punto Q como

$$9^3 + A9^2 + 9 = 9^3 + 486662 \cdot 9^2 + 9 = 39420360 = d.$$

Igual que antes, volvemos a comprobar que este valor obtenido es residuo cuadrático de $\mathbb{Z}_p$ viendo que

$$d^{\frac{2^{255}-19-1}{2}} \equiv 1 \mod 2^{255} - 19.$$

Entonces, la coordenada-y del punto base es la raíz cuadrada de d en $\mathbb{Z}_p$.

Este punto base no tiene orden igual a p, pero sí muy cercano a p. En concreto, tiene orden primo igual a $2^{252} + 27742317777372353535851937790883648493$. El punto Q no genera todos los puntos de la curva elíptica $E(\mathbb{Z}_p)$, pero sí la gran mayoría.

Una vez que tenemos la curva elíptica y el cuerpo, dado $n \in 2^{254} + 8\{0, 1, 2, 3, \ldots, 2^{251} - 1\}$ y $q \in \mathbb{Z}_p$, el teorema 4.2 nos garantiza que obtendremos $s \in \mathbb{Z}_p$, que es único.

4.2.1. Aspectos computacionales

Como ya hemos explicado, el PLD se considera irresoluble y, por tanto, los criptosistemas basados en él, también lo son. Para el caso del método de intercambio de claves Diffie-Hellman aplicado a curvas elípticas y en concreto a esta curva, ocurre lo mismo. Cualquier ataque específico es muchísimo más costoso que un ataque por fuerza bruta, el cual si se escogen adecuadamente los parámetros de la curva y el cuerpo resulta ineficaz. En el caso de la Curva25519 el tamaño del cuerpo primo utilizado es de 2^{255}, lo que lo convierte en un cuerpo primo suficientemente grande como para pasar cualquier ataque a su PLD. En [1, sección 3] podemos encontrar diferentes ataques genéricos y específicos contra esta curva. Para la Curva25519 se tienen las siguientes ventajas computacionales:

- **Velocidad de computación muy alta**: los valores registrados en diferentes procesadores muestran un nuevo record de velocidad en cuanto a criptosistemas Diffie-Hellman. Este dato es el más relevante, ya que justifica la gran eficiencia del criptosistema creado. Se puede mejorar la velocidad de computación fácilmente disminuyendo el tamaño, es decir, pasando por ejemplo a 160 bits; sin embargo, esto hace que disminuya considerablemente el nivel de seguridad del método. Con la Curva25519 conseguimos máxima velocidad de computación pero sin disminuir el nivel de seguridad.

  Esta mejora en la velocidad de computación se ha conseguido en parte por la elección de la curva en forma de Montgomery (2.4), que disminuye considerablemente el número de operaciones necesarias para calcular la operación de grupo. Aunque en este caso es necesario, para el mismo nivel de seguridad, un primo p mucho más grande que en otros casos porque $|E|$ es divisible por 4. Esto no importa ya que se gana mucho en velocidad de operaciones.

  Por otro lado, se ha conseguido mejorar la velocidad de cálculo gracias a la aritmética rápida módulo $p = 2^{255} - 19$, es decir, gracias a la elección del cuerpo $\mathbb{Z}_{2^{255}-19}$. Las curvas elípticas sobre cuerpos primos presentan muchas ventajas en la seguridad de los criptosistemas basados en ellas. Se ha escogido el primo $2^{255} - 19$ basándose en el criterio de que los primos cercanos a una potencia de 2 disminuyen el tiempo de cálculo en las operaciones del cuerpo, sin tener repercusión en la seguridad. Tomando un primo de longitud $32k$ bits, permitimos poder transmitir las claves públicas en palabras de 32 bits. Si $k = 8$ ya se consigue un alto nivel de seguridad, luego escogemos un primo de longitud 256 bits. Se han considerado varios números primos cercanos ($2^{255} + 95$, $2^{255} - 19$, $2^{255} - 31$,...) y se ha escogido $2^{255} - 19$ por ser $19 < 95, 31, \ldots$

- **Claves públicas y privadas pequeñas:** Las claves públicas y privadas en el DHKE de la Curva25519 son de 256 bits. Para cada usuario consideramos el par de claves $(k_{pr}, k_{pub}) = (a, c)$ con $a, c \in \mathbb{Z}_{2^{255}-19}$. La clave privada es un elemento de $\mathbb{Z}_{2^{255}-19}$, es decir, un entero de 256 bits. Ahora, la clave pública, gracias al teorema 4.2, disminuye de tamaño y ya no es un punto de E, sino la coordenada-x de un punto de E. Luego la clave pública es un entero de 256 bits. Se ha conseguido reducir la clave pública de 64 bytes a 32 bytes.

- No hay validación de las claves: cualquier cadena de 256 bits se acepta como clave pública. Normalmente en otros criptosistemas del tipo DHKE hay una fase de validación de la clave pública.

- Código muy pequeño: se ha conseguido un código a compilar para el DHKE de la Curva25519 muy pequeño, por lo que ocupa muy poco espacio en el procesador. Esto permite que la encriptación pueda llevarse a cabo en dispositivos como teléfonos móviles.

4.3. Protocolo simplificado de encriptación de WhatsApp

Para terminar, vamos a explicar el protocolo simplificado de encriptación de WhatsApp con la Curva25519. El software que WhatsApp utiliza es el propuesto por Daniel J. Bernstein, autor del artículo en el que se basa este capítulo del trabajo. Para una detallada explicación de este protocolo ver [4] y [10].

Antes de comenzar con el método vamos a definir la notación que usaremos: si $a, b \in \mathbb{Z}_{2^{255}-19}$ son números de 256 bits, $Curva25519(a, b)$ es la coordenada-x del punto $[a]P$, siendo $P \in E$ un punto de coordenada b en el eje de las x.

Para generar las claves privadas, cada usuario de WhatsApp dispone en su teléfono móvil de un procedimiento para encontrar números aleatorios de 256 bits. Una vez que tenemos el número a de 256 bits, es decir, la clave privada, calculamos la clave pública asociada a a como

$$Curva25519(a, 9).$$

En el momento de la instalación de WhatsApp en el teléfono, el usuario A construye automáticamente multitud de pares de claves, como hemos explicado en el párrafo anterior, y envía al servidor de WhatsApp las claves públicas. En ningún momento el servidor de WhatsApp tiene acceso a las claves privadas. La primera clave pública se denomina *clave de identificación* y el resto de claves públicas se conocen como *claves de un sólo uso*. Estas claves de un sólo uso se firman con la clave de identificación y se reponen cada cierto tiempo cuando sea necesario.

Denotamos como $I_A = (I_{A,pr}, I_{A,pub})$ y $U_A = (U_{A,pr}, U_{A,pub})$ a los pares de claves anteriores. A continuación, describimos el método de encriptación de mensajes en una conversación:

1. Cuando el usuario A quiere comunicarse con el usuario B por primera vez, A solicita al servidor $I_{B,pub}$ y $U_{B,pub}$. La clave pública $U_{B,pub}$, que es de un único uso, se borra entonces del servidor. El usuario A construye una clave efímera $E_A = (E_{A,pub}, E_{A,pr})$.

2. A construye una clave K_1 de tipo AES-256. Para ello, computa $Curva25519(E_{A,pr}, U_{B,pub})$ y aplica la función hash SHA al valor obtenido.

   $$K_1 = SHA(Curva25519(E_{A,pr}, U_{B,pub}))$$

   Notar que $Curva25519(E_{A,pr}, U_{B,pub})$ es la clave conjunta del método DHKE formada por la clave privada de A y la pública de B.

3. A cifra el mensaje m con el método AES-256 utilizando la clave K_1. Envía a B, en la cabecera del mensaje, la clave pública temporal $E_{A,pub}$ y $I_{A,pub}$.

4. De esta manera, B puede construir K_1 a partir de sus claves privadas y las claves públicas enviadas por A en la cabecera del mensaje. Con K_1 ya puede descifrar el mensaje, es decir, obtener m.

5. El usuario B borra la clave $U_{B,pub}$ ya utilizada por A.

Queda descrito el método de establecimiento de sesión. A partir de aquí, para el resto de mensajes entre A y B, ambos construyen sucesivas claves K_i mediante $K_i = SHA(K_{i-1})$.

El protocolo de *"cifrado de extremo a extremo"* no permite que WhatsApp ni terceras partes puedan descifrar mensaje alguno.

4.4. Conclusión

A modo de conclusión, señalar que con este trabajo he pretendido analizar las propiedades algebraicas de la Curva25519 para demostrar que es suficiente utilizar la coordenada-x para los cálculos necesarios para el cifrado de mensajes. Como ampliación del trabajo, sería interesante analizar más en profundidad por qué esta es la curva elegida y no otra, así como demostrar las fórmulas de Montgomery para la computación rápida de la coordenada-x, lo cual requiere un trabajo más extenso.

Bibliografía

[1] DANIEL J. BERNSTEIN, *Curve25519: new Diffie-Hellman speed records*, Yung, Moti; Dodis, Yevgeniv; Kiayias, Aggelos; et al., eds. Public Key Cryptography. Lecture Notes in Computer Science. 3958. New York: Springer. pp. 207-228.

[2] HENRI COHEN, GERHARD FREY, *Handbook of Elliptic and Hyperelliptic Curve Cryptography*, Taylor and Francis Group, LLC, Chapman and Hall/CRC, 2006.

[3] WHITFIELD DIFFIE, MARTIN HELLMAN, *New directions in Cryptography*, Vol. IT-22, pp. 644-654, Nov. 1976. `https://ee.stanford.edu/~hellman/publications/24.pdf`

[4] ANTONIO JUANO, *Criptografía y Seguridad en WhatsApp*, Trabajo Fin de Máster, Universidad Nacional de Educación a Distancia, 2016.

[5] NEAL KOBLITZ, *Elliptic curve cryptosystems*, Journal: Math. Comp. 48, pp. 203-209, 1987. `http://www.ams.org/journals/mcom/1987-48-177/S0025-5718-1987-0866109-5S0025-5718-1987-0866109-5.pdf`

[6] VICTOR S. MILLER, *Use of Elliptic Curves in Cryptography*, Advances in Cryptology - CRYPTO ′85 Proceedings, pp 417-426. Springer-Verlag, Berlin Heidelberg, 1985.

[7] PETER L. MONTGOMERY, SPEEDING THE POLLARD AND ELLIPTIC CURVE METHODS OF FACTORIZATION, Mathematics of Computation 48 (1987), 243-264.

[8] CHRISTOF PAAR, JAN PELZL, *Understanding Cryptography*, Springer-Verlag, Berlin Heidelberg, 2010.

[9] JOSEPH H. SILVERMAN, *The arithmetic of Eliptic Curves*, Springer-Verlag, New York Inc., Graduate Texts in Mathematics 106, 1986.

[10] WHATSAPP ENCRYPTION OVERVIEW, `https://www.whatsapp.com/security/`.