UNIVERSIDAD DE LA RIOJA

TRABAJO FIN DE ESTUDIOS

Título

Criptografía basada en curvas elípticas

Autor/es

ALBERTO CALLEJA PASCUAL

Director/es

JUAN LUIS VARONA MALUMBRES

Facultad

Facultad de Ciencia y Tecnología

Titulación

Grado en Matemáticas

Departamento

MATEMÁTICAS Y COMPUTACIÓN

Curso académico

2019-20

publicaciones.unirioja.es
E-mail: publicaciones@unirioja.es

UNIVERSIDAD DE LA RIOJA

Facultad de Ciencia y Tecnología

TRABAJO FIN DE GRADO

Grado en Matemáticas

Criptografía basada en curvas elípticas

Realizado por:

Alberto Calleja Pascual

Tutelado por:

Juan Luis Varona Malumbres

Logroño, junio de 2020

UNIVERSIDAD
DE LA RIOJA

Criptografía basada en curvas elípticas

Autor: Alberto Calleja Pascual

Tutor: Juan Luis Varona Malumbres

Grado en Matemáticas

Facultad de Ciencia y Tecnología
Universidad de La Rioja

Curso académico: 2019/2020

Índice general

Resumen/Abstract

Resumen

En el primer capítulo de este trabajo hablaremos sobre las curvas elípticas. Estudiaremos cómo definir estas curvas, observaremos sus propiedades, veremos la "ley de grupo" mediante la cual seremos capaces de construir un grupo abeliano a partir del conjunto de puntos de una curva elíptica, y concluiremos con algunos resultados sobre estas curvas que nos serán útiles en la criptografía.

En el segundo capítulo empezaremos estudiando algunos conceptos básicos de la criptografía y después veremos varios métodos y algoritmos criptográficos basados en las curvas elípticas, así como algunos ejemplos de uso de estos métodos. Al final del capítulo analizaremos la importancia de la computación cuántica en este tema.

Hemos dedicado después un capítulo a dos curvas elípticas concretas que se utilizan hoy en día, una de ellas para cifrar los mensajes de la plataforma WhatsApp y la otra en Bitcoin. En este capítulo vemos una aplicación real de la criptografía basada en curvas elípticas.

Finalmente hay un último capítulo con conclusiones personales. La bibliografía utilizada para realizar este trabajo aparece en las últimas páginas del trabajo.

Abstract

In the first chapter of this work we are going to talk about elliptic curves. We are going to study how elliptic curves are defined, observe their properties, see the "group law" that allow us to create an abelian group from the set of points of an elliptic curve, and conclude with some results that are going to be used in cryptography.

In the second chapter we are going to study some basic concepts of cryptography. After that we are going to see some cryptographic methods and algorithms based on elliptic curves and we are going to do some examples of this methods. At the end of the chapter we are going to analyze how quantum computers affect the elliptic curves cryptography.

We have dedicated a chapter to study two elliptic curves that are used nowadays, one of the curves is used to encrypt WhatsApp messages and the other one is used in Bitcoin. In this chapter we are going to see a real usage of elliptic curves cryptography.

Finally there is a chapter about my own conclusions. At the end of the work appears the bibliography used to do this work.

Capítulo 1

Curvas elípticas

En este capítulo vamos a conocer y estudiar la noción de **curva elíptica**, así como las propiedades de éstas y qué ventajas podemos sacar de ellas para su uso en la criptografía.

Podemos definir una curva elíptica como:

Definición 1.1. *Una **curva elíptica** es el conjunto de los pares de puntos (x, y) pertenecientes a un determinado cuerpo $\mathbb{F}$ que cumplen la siguiente ecuación:*

$$y^2 = x^3 + Ax + B \tag{1.1}$$

con A, B constantes en $\mathbb{F}$.

Generalmente, el cuerpo $\mathbb{F}$ será $\mathbb{Q}, \mathbb{R}, \mathbb{C}$, los cuerpos finitos $\mathbb{F}_p$ (esta notación se usa para hacer referencia a $\mathbb{Z}_p$ pensado como cuerpo), con p primo, o uno de los cuerpos finitos $\mathbb{F}_q$ con $q = p^k$. En todo caso, lo expresaremos diciendo que la curva está definida sobre $\mathbb{F}$. Usaremos, en estas primeras explicaciones, los números reales.

Nota 1.1. Hemos expresado la ecuación de la curva mediante la conocida como **ecuación de Weierstrass**; sin embargo, queremos recalcar que existen ecuaciones con otras formas que, de igual manera, determinan una curva elíptica.

Para comenzar a trabajar con las curvas elípticas explicaremos en primer lugar los espacios afines y los proyectivos, ambas nociones necesarias para proseguir con nuestro trabajo.

1.1. Espacio afín y espacio proyectivo

La motivación primera que encontramos para usar estos espacios no es otra que la observación de que, en ocasiones, para encontrar los puntos de intersección entre

dos curvas (definidas sobre un cuerpo) no basta utilizar una extensión algebraica para hallarlos. Por ejemplo, cuando tenemos dos rectas paralelas, intuimos que en el infinito deben tener un punto donde intersecan; sin embargo, cambiar de $\mathbb{R}$ a $\mathbb{C}$ no soluciona el problema.

Añadiendo esos puntos del infinito podríamos arreglar el problema, y a esta construcción del plano, junto con los puntos del infinito, se le conoce como **plano proyectivo**. No obstante, debemos ser cautelosos a la hora de añadir los puntos.

Usaremos los números reales por una cuestión de comodidad, pero estas nociones pueden igualmente ser usadas para otros cuerpos. Llamamos a $\mathbb{R}$ la recta afín, a $\mathbb{R}^2$ el plano afín, a $\mathbb{R}^3$ el espacio afín... Podemos generalizar llamando a $\mathbb{R}^n$ el n-espacio afín, y lo denotamos así:

$$A^n(\mathbb{R}) := \mathbb{R}^n = \{(a_1, \ldots, a_n) \mid a_i \in \mathbb{R}\}.$$

La definición tendrá sentido aunque cambiemos el cuerpo $\mathbb{R}$ por otro, como $\mathbb{Q}$.

Antes de explicar cómo añadir puntos a un espacio afín, es conveniente explicar una construcción que contendrá un espacio afín de dimensión dada, n, en un espacio proyectivo.

Queremos interpretar la noción de punto en el espacio proyectivo. Para ello debemos de tener en cuenta que un punto en el espacio proyectivo es representado mediante una relación de equivalencia. Nos interesa encontrar un representante de una clase de equivalencia que tenga sentido.

Definiremos primero, como un primer acercamiento a la idea de espacio proyectivo, la **recta proyectiva**. Consideramos el conjunto de vectores no nulos del plano afín, $A^2(\mathbb{R})$; éstos tendrán origen en $(0,0)$ y final en un punto arbitrario $(a,b) \neq (0,0)$ del plano afín. Tenemos así el conjunto

$$C := \{(a,b) \in A^2(\mathbb{R}) \mid (a,b) \neq (0,0)\}.$$

Sea ahora la relación de equivalencia entre los puntos $(a,b) \sim (c,d)$ si y solo si existe λ con $\lambda \in \mathbb{R} \setminus \{0\}$ tal que $(c,d) = (\lambda a, \lambda b)$. Teniendo en cuenta esta relación de equivalencia, el conjunto de puntos de C que está en la clase de equivalencia de (a,b), son los puntos que están en la recta que pasa por el origen y (a,b), es decir, las clases de equivalencia de C son las rectas que pasan por el origen. Busquemos el represante de cada clase de equivalencia; por un lado consideramos la recta $y = 1$; notemos ahora que las rectas que pasan por el origen y $b \neq 0$ (son distintas del eje x) cortarán a $y = 1$ en un único punto (a,b) y $(a,b) \sim (\frac{a}{b}, 1)$. De otra manera, sea $y_A = mx$ con $m \neq 0$, interseca a $y = 1$ en $(\frac{1}{m}, 1)$.

Así, toda recta que pasa por el origen determina un único punto $(r,1)$ en $y = 1$ y, recíprocamente, cada punto $(r,1)$ en $y = 1$ determina una recta que pasa por $(0,0)$, excepto al eje x. Esto implica una relación de correspondencia uno a uno entre los puntos r del plano afín y las rectas que pasan por el origen. Dada $(a : b)$

la clase de equivalencia del vector (a, b) de C, el conjunto de clases de equivalencia de C es

$$\{(a : b) \mid a, b \in \mathbb{R}, (a, b) \neq (0, 0)\} = \{(r : 1) \mid r \in \mathbb{R}\} \cup \{(1 : 0)\}.$$

Notar que el conjunto $\{(a : b) \mid a, b \in \mathbb{R}, (a, b) \neq (0, 0)\}$ contiene al eje x, ya que éste queda determinado por cualquier vector de la forma $(a, 0)$ con a distinto de 0 y además $(a, 0) \sim (1, 0)$, por lo tanto $(a : 0) = (1 : 0)$. Luego definimos la recta proyectiva como este conjunto, o sea,

$$P^1(\mathbb{R}) := \{(r : 1) \mid r \in A^1(\mathbb{R})\} \cup \{(1 : 0)\} = A^1(\mathbb{R}) \cup \{(1 : 0)\}.$$

Aquí vemos claramente cómo el plano afín está contenido en la recta proyectiva.

Nota 1.2. En este documento denotamos la clase de equivalencia como $(a : b)$. Sin embargo, en otros libros y documentos (por ejemplo en [8]) se denota como $[a, b]$.

Ahora, si generalizamos, podemos aumentar la dimensión; entonces, para el plano proyectivo $P^2(\mathbb{R})$ usaremos el espacio afín $A^3(\mathbb{R})$ y

$$P^2(\mathbb{R}) = \{(a : b : c) \mid a, b, c \in \mathbb{R}, (a, b, c) \neq (0, 0, 0)\}.$$

Notar que se cumple la igualdad

$$P^2(\mathbb{R}) = \{(a : b : 1) \mid (a, b) \in A^2(\mathbb{R})\} \cup \{(a : b : 0) \mid (a, b, 0) \neq (0, 0, 0)\},$$

donde $\{(a : b : 0) \mid (a, b, 0) \neq (0, 0, 0)\}$ es una copia de $P^1(\mathbb{R})$.

En general, si la dimensión es n y el cuerpo $\mathbb{F}$, definimos $P^n(\mathbb{F})$ como sigue. En

$$S := \{(a_1, \ldots, a_{n+1}) \in A^{n+1}(\mathbb{F}) \mid (a_1, \ldots, a_{n+1}) \neq (0, \ldots, 0)\}$$

decimos que dos puntos son equivalentes, $(a_1, \ldots, a_{n+1}) \sim (b_1, \ldots, b_{n+1})$, si y solo si existe $t \in \mathbb{R}$, $t \neq 0$, tal que $(b_1, \ldots, b_{n+1}) = t(a_1, \ldots, a_{n+1})$. Entonces, el n-plano proyectivo será

$$P^n(\mathbb{R}) := \{(a_1 : \cdots : a_{n+1}) \mid (a_1, \ldots, a_{n+1}) \in S\}.$$

Podemos expresar $P^n(\mathbb{R})$ como

$$P^n(\mathbb{R}) = \{(a_1 : \cdots : a_n : 1) \mid (a_1, \ldots, a_n) \in A^n(\mathbb{F})\} \cup X,$$

con X el conjunto

$$X = \{(a_1 : \cdots : a_n : 0) \mid (a_1, \ldots, a_n, 0) \neq (0, \ldots, 0)\}.$$

El primer conjunto es igual a $A^n(\mathbb{F})$ y el segundo, al que hemos denotado X, es igual al hiperplano proyectivo o $(n-1)$-espacio proyectivo.

1.2. Curvas en el plano afín y proyectivo

Como ya hemos comentado en la sección 1.1, una de las motivaciones para el uso de los espacios proyectivos es la búsqueda de los puntos de intersección entre dos curvas. Conocemos ahora los espacios proyectivos y sabemos que existen en ellos puntos del infinito. Este hecho será determinante en la búsqueda de las interesecciones.

Notemos que para representar los puntos de $A^2(\mathbb{R})$ en el espacio proyectivo, bastará usar la clase de equivalencia $(a : b : 1)$, relacionando cada punto (a, b) en el plano afín con $(a : b : 1)$. Esto, sin embargo, genera un conflicto al tratar de ver las curvas en el espacio proyectivo. Queremos que los puntos del espacio afín que cumplían la ecuación sigan apareciendo en el conjunto de los puntos de la curva en el espacio proyectivo.

Sea por tanto (a, b) un punto, en el espacio afín, que cumple la ecuación de la curva elíptica (1.1); entonces los puntos de la clase de equivalencia $(a : b : 1)$ también deberán cumplirla, luego debemos introducir una nueva variable, z, a la ecuación, de tal manera que, en $z = 1$, la ecuación siga siendo la misma. Puede haber varias ecuaciones que tras introducir esta variable, según como lo hagamos, cumplan las condiciones, pero teniendo en cuenta que si $(a, b, 1)$ la cumple también deben cumplirla (at, bt, t) para $t \in \mathbb{R}\backslash\{0\}$, y el caso $z = 0$, llegaremos a la necesidad del uso del método de homogeneización para encontrar la ecuación de una curva en el espacio proyectivo, una vez conocida su ecuación en el espacio afín.

1.3. Homogeneización

Es el proceso de convertir una ecuación en una ecuación homogénea, es decir, transformarla en otra en la que todos los términos de la ecuación han de tener el mismo grado. Dada una ecuación $f(x, y) = 0$ deberemos multiplicar cada término por la menor potencia de z de tal manera que todos los términos cumplan la condición de homogeneidad.

Por ejemplo, el proceso de homogeneización de una función $f(x, y)$ es

$$f(x, y) = 6x^4 - 5xy + 12y \longrightarrow F(x, y, z) = 6x^4 - 5xyz^2 + 12yz^3.$$

Y, así, usaremos $F(x, y, z)$ en lugar de $f(x, y)$.

Con esta idea, los puntos afines (a, b) se corresponden con los puntos proyectivos $(a : b : 1)$ en la curva proyectiva asociada.

Veamos qué ventajas nos ofrece este proceso. Para cualesquiera dos rectas proyectivas, podemos ahora observar que éstas se intersecan en el espacio proyectivo. Dadas dos rectas cualesquiera, al margen de sus interesecciones en el espacio afín

($z = 1$), podríamos tener intersecciones cuando $z = 0$. Debemos excluir como intersecciones los puntos $(0 : 0 : 0)$ ya que, para construir el espacio proyectivo, en la sección 1.1 consideramos $A^3 \setminus (0, 0, 0)$, es decir, al menos alguna coordenada no es cero.

La ecuación general para rectas en el espacio afín será $ax + by + c = 0$ y una paralela será $ax + by + d = 0$ con $c \neq d$ y $a \neq 0$ o $b \neq 0$ para que sea una recta. Si llevamos a cabo el proceso de homogeneización tendremos

$$ax + by + cz = 0, \quad ax + by + dz = 0. \tag{1.2}$$

Sabemos que no hay solución de (1.2) en el espacio afín, pero si consideramos $z = 0$ (es decir, miramos la recta en el infinito) tendremos que las ecuaciones se reducen a $ax + by = 0$ y separamos los casos:

- Si $a \neq 0$, despejando en la ecuación, tendremos que $x = \frac{-by}{a}$. Entonces, sustituyendo en la expresión del punto proyectivo $(x : y : 0) = (\frac{-by}{a} : y : 0)$ y dividiendo por y llegamos a la igualdad $(x : y : 0) = (\frac{-b}{a} : 1 : 0)$.

- De manera análoga al caso anterior, si $b \neq 0$ tenemos, tras despejar de la ecuación, $y = \frac{-ax}{b}$. Entonces, sustituyendo la variable y y dividiendo por x, $(x : \frac{-ax}{b} : 0) = (1 : \frac{-a}{b} : 0)$.

En cualquier caso hay un único punto de intersección en la recta en el infinito ($z = 0$) y si a, b son ambos no nulos, se tiene la igualdad $(\frac{-b}{a} : 1 : 0) = (1 : \frac{-a}{b} : 0)$, luego no hay ambigüedad al representar el punto en el plano proyectivo.

Notemos que cada recta vertical tiene la forma $ax + by + c = 0$ con $b = 0$, y ésta intersecará la recta en el infinito en $(0 : 1 : 0)$ y éste será el único punto en el infinito en una curva elíptica $y^2 = x^3 + ax^2 + bx + c$, así que toda recta vertical intersecará a la curva elíptica en este punto.

1.4. Puntos racionales en las curvas

Si $(a, b, c) \in A^3(\mathbb{Z})$ es solución no nula de la ecuación $x^n + y^n = z^n$, con $n \in \mathbb{N}$ y $n > 2$, entonces (at, bt, ct), con t un número racional no nulo, también será solución. Por tanto todo punto de $(a : b : c) \in P^2(\mathbb{Q})$ es solución. Por ello, tiene sentido considerar el punto proyectivo $(a : b : c)$ como solución de la ecuación.

Esto tiene una motivación, ya que al pensar en las soluciones de la ecuación $x^n + y^n = 1$, cuya forma es $(\frac{a}{c}, \frac{b}{c})$ con c positivo, y tras homogeneizar la ecuación, llegamos a $x^n + y^n = z^n$. Las soluciones corresponden a $(\frac{a}{c} : \frac{b}{c} : 1) = (a : b : c)$. No obstante, si $c = 0$ y n impar, las soluciones corresponden a un único punto $(a : -a : 0) = (1 : -1 : 0)$ que pertenece a la recta en el infinito en el plano

proyectivo. Luego el uso de espacios proyectivos nos ofrece un "contenedor" de todas las soluciones sin necesidad de distinguir los casos.

Echemos un vistazo a las curvas elípticas. Si (x, y) son los puntos que satisfacen la ecuación de la curva elíptica $y^2 = x^3 + ax^2 + bx + c$ donde $x^3 + ax^2 + bx + c$ es una cúbica no singular (lo que implica que las raíces sean distintas), podemos homogeneizar y buscar soluciones en el espacio proyectivo

$$y^2z = x^3 + ax^2z + bxz^2 + cz^3.$$

Las soluciones en el espacio afín, (x, y), serán $(x : y : 1)$ en el proyectivo, y tenemos ahora además la solución extra $(0 : 1 : 0)$ al considerar $z = 0$, que es la correspondiente a la de la recta en el infinito. Es decir, hay un único punto proyectivo de la curva elíptica en el infinito.

Nota 1.3. Dadas dos curvas definidas sobre alguno de los cuerpos $\mathbb{Q}$, $\mathbb{R}$ o $\mathbb{C}$ con grados m y n, podremos encontrar las mn intersecciones en el plano proyectivo $P^2(\mathbb{C})$.

1.5. La ley de grupo para las curvas elípticas

En el caso de las curvas elípticas y sus conjuntos de puntos podremos definir una operación binaria entre los puntos, de manera que se configuren grupos abelianos. Aquí es donde radica la importancia de las curvas elípticas para su uso en la criptografía. Trabajando curvas elípticas definidas sobre un cuerpo finito $\mathbb{F}_p$ podremos, por ejemplo, tratar los problemas del logaritmo discreto sobre el grupo de puntos de una curva elíptica, en lugar de sobre el grupo multiplicativo de un cuerpo finito, entre otras cosas.

Primero, supongamos que tenemos una curva elíptica $y^2 = x^3 + ax^2 + bx + c$, con coeficientes en un determinado cuerpo $\mathbb{F}$; tendrá sentido hablar del conjunto de puntos de la curva elíptica cuyas coordenadas pertenecen ambas al cuerpo $\mathbb{F}$. Denotaremos como $E(\mathbb{F})$ al conjunto de puntos de la curva. Usando diferentes cuerpos se llega a diferentes y variadas aplicaciones, tal como ya hemos dicho antes. Para la criptografía se usarán cuerpos finitos. Algunos de los conjuntos de puntos de la curva elíptica, como $E(\mathbb{Q})$, no presentan estructura geométrica, pero sí algebraica.

En primer lugar, queremos dar constancia de que la ley de grupo respeta al cuerpo subyacente, es decir, dados dos puntos $P, Q \in E(\mathbb{F})$, la ley de grupo (operación binaria) producirá un tercer punto $P \oplus Q$ que pertenecerá de igual manera a $E(\mathbb{F})$, es decir, ambas coordenadas estarán en $\mathbb{F}$.

Para obtener una primera idea sobre cómo realizar esta operación y "sumar" puntos en $E(\mathbb{F})$, podemos utilizar primero $E(\mathbb{R})$, que es representable geométri-

camente, y, una vez hallada la ley de grupo, las condiciones y los resultados algebraicos en este caso, ver qué podemos generalizar para los demás cuerpos.

Consideramos el conjunto de puntos reales en una curva elíptica que, según sea su ecuación, podremos ver que tiene una o dos componentes. Esto se debe a que hemos asumido que la cúbica es no singular, y una cúbica con coeficientes reales tiene, al menos, una raíz real. Por lo tanto, el número de componentes depende de las raíces de la cúbica, pues, de las 3 raíces que tiene que tener, serán 1 real y 2 complejas o las 3 reales.

Veamos así un par de curvas:

1. $y^2 = x^3 + 2$ con 1 raíz real.
2. $y^2 = x^3 - x$ con 3 raíces reales.

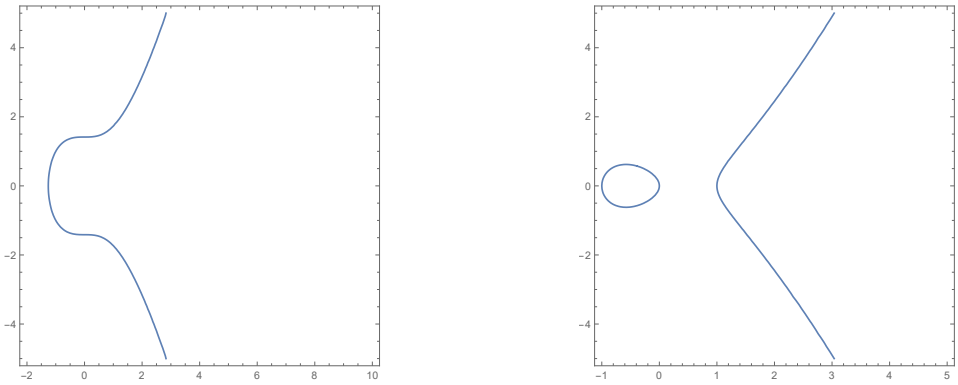

Figura 1.1: Curva con una raíz real y curva con 3 raíces reales.

Procedemos a realizar una construcción intuitiva de la ley de grupo, la cual refinaremos hasta alcanzar la deseada.

Supongamos dos puntos dados P, Q de la curva elíptica. Por la geometría euclídea, podemos trazar una única recta entre ellos dos. Además, por la identidad de Bézout, sabemos que la línea y la cúbica deben intersecarse (en el plano proyectivo $P^2(\mathbb{C})$) en tres puntos. Podría ser tan simple como hallar este tercer punto de intersección; sin embargo, vamos a ver que, de esta forma, la ley de grupo fallaría.

Dado $E(\mathbb{R})$ el conjunto de puntos de la curva, y un par de puntos $P, Q \in E(\mathbb{R})$, consideramos la recta que los une. Si $P = Q$ consideramos la recta tangente a la curva en P. Si denotamos por $P \cdot Q$ al tercer punto de la intersección, entonces "$\cdot$" es una operación binaria en $E(\mathbb{R})$. Pero hemos dicho que la ley de grupo debe configurar grupos abelianos y por tanto debe ser conmutativa, asociativa, con presencia de elemento identidad y elemento inverso, esto es:

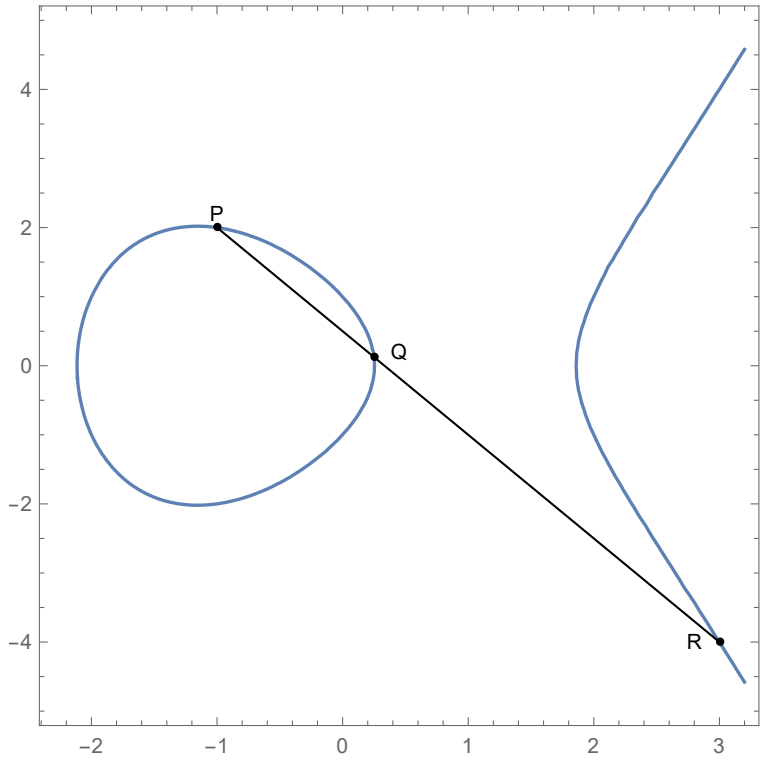

Figura 1.2: Curva elíptica donde se ejemplifica la tercera intersección entre la recta de los puntos P, Q y la curva.

- Identidad: Existe e en el conjunto $E(\mathbb{R})$ tal que $e \cdot P = P \cdot e = P$ para todo $P \in E(\mathbb{R})$.
- Inverso: Para todo $P \in E(\mathbb{R})$ existe $A \in E(\mathbb{R})$ tal que $P \cdot A = e = A \cdot P$.
- Conmutatividad: $P \cdot Q = Q \cdot P$.
- Asociatividad: $P \cdot (Q \cdot R) = (P \cdot Q) \cdot R$.

Veamos primero qué ocurre con la conmutatividad. Según la operación provisional que hemos definido, dado que la recta PQ es la misma que la recta QP, entonces $P \cdot Q = Q \cdot P$, ya que intersecarán en el mismo punto.

Podemos tratar de ver si existe un elemento identidad en la operación así definida. Encontramos aquí un primer conflicto. Llamamos O al elemento identidad, así $O \cdot P = P$ para todo $P \in E(\mathbb{R})$. Esto significa que la recta entre O y P interseca en P a la curva elíptica, luego es tangente a la curva en P. Así, toda línea tangente a la curva en P, interseca en el mismo punto O a la curva elíptica, lo que visualmente no es el caso.

Además, también es sencillo ver que la asociatividad también falla. Primero podemos notar que $P \cdot (P \cdot Q) = Q$, de manera visual. Ahora $Q = P \cdot (P \cdot Q) = (P \cdot P) \cdot Q$, luego la tangente a la curva en P cortará en un punto. Tomando la recta entre éste y Q y observando su intersección, ¿qué debería ocurrir? La recta entre $P \cdot P$ y Q debe ser tangente a la curva en Q, lo cual es obviamente imposible.

Por lo tanto, este primer intento con la operación binaria "$\cdot$" ha fallado; el primer error que hemos corroborado es la falta de elemento identidad.

1.5.1. Definiendo la operación binaria

Para ello vamos a seguir la definición que aparece, entre otros muchos textos, en [8]. El primer escollo que debemos salvar trata sobre si, realmente, $E(\mathbb{R})$ es o no vacío, pues podría serlo. Sin embargo, observaremos, mediante la representación de la curva en el plano proyectivo, como efectivamente no lo es, y además obtendremos un posible candidato a elemento identidad.

Vamos a comenzar suponiendo que tenemos un elemento identidad O ya definido. Ahora retomaremos el problema de definir la operación binaria. Sabiendo que el anterior intento falló, sean P, Q dos puntos de la curva y $P \cdot Q$ el punto de intersección de la recta PQ y la curva, y de este modo llegamos a la siguiente definición:

Definición 1.2. *Definimos la* ***operación binaria*** *del conjunto de puntos de la curva elíptica de la siguiente manera:*

$$P \oplus Q = O \cdot (P \cdot Q),$$

es decir, $P \oplus Q$ es el tercer punto de la intersección de la curva elíptica considerada y de la recta que pasa por O y por el punto $P \cdot Q$.

1.5.2. Conmutatividad

Como $P \cdot Q = Q \cdot P$, cosa que vimos en la sección 1.5 en el apartado de la conmutatividad, entonces $P \oplus Q = Q \oplus P$. Luego tenemos esta propiedad en el grupo.

1.5.3. Identidad

Veamos que O tiene sentido como elemento identidad. Esto es que, para todo $P \in E(\mathbb{R})$, se cumple

$$P \oplus O = P.$$

Consideramos $P \cdot O$; ahora sabemos que la recta entre O y $P \cdot O$ tendrá como tercera intersección P, luego, efectivamente,

$$O \cdot (P \cdot O) = P \Rightarrow P \oplus O = P.$$

1.5.4. Elemento inverso

Para entender el funcionamiento del elemento inverso, debemos primero observar la recta tangente a la curva en O, cuya existencia queda asegurada por la no singularidad de la curva elíptica. Si Q es el tercer punto de intersección entre la recta tangente a la curva en O y la curva, al elegir un punto P en la curva y trazar la recta entre P y Q, veremos que $P \cdot Q = -P$, es decir, el inverso de P. Para ello calcularemos $P \oplus (P \cdot Q)$.

Primero buscamos $P \cdot (P \cdot Q)$, es decir, la tercera intersección de la recta que une P y $P \cdot Q$ con la curva. Ésta se dará en el punto Q. Ahora, $P \oplus (P \cdot Q)$ será el tercer punto de intersección de la recta desde O a $Q = P \cdot (P \cdot Q)$. Pero, dado que hemos dicho que Q es el tercer punto de intersección de la recta tangente desde O y la curva, como los puntos de tangencia cuentan como dos puntos de intersecciones, entonces éste será O. Es decir,

$$P \oplus (P \cdot Q) = O.$$

Luego, efectivamente, $P \cdot Q$ es el inverso de P.

1.5.5. Asociatividad

Con lo que hemos visto hasta ahora, la prueba de la asociatividad sería un asunto algo farragoso, que requiere el uso de cuestiones demasiado complicadas para lo que a nosotros nos interesa, que es el uso de las curvas elípticas en la criptografía. No obstante, más adelante, tras generalizar la ley de grupo y realizar la aritmética necesaria, obtendremos unas fórmulas para calcular directamente el resultado de la operación binaria, y gracias a esas fórmulas daremos una demostración de la asociatividad de una manera general.

1.5.6. Elemento identidad

Todavía no tenemos claro cuál de los puntos es el elemento identidad, O. Veamos cuál puede ser un buen candidato, y si efectivamente nos vale.

Sabemos que si tomamos como ecuación general de las curvas elípticas

$$y^2 = x^3 + ax^2 + bx + c,$$

podremos homogeneizarla, y ya vimos que el punto $(0 : 1 : 0)$ pertenece a toda curva elíptica, y, de hecho, es el único que está en la recta en el infinito.

Nota 1.4. También vimos que el punto $(0:1:0)$ pertenece a $E(\mathbb{F})$ y éste existe sea cual sea el cuerpo $\mathbb{F}$, lo cual nos lleva a pensar en que es un buen candidato.

Consideraremos ahora las consecuencias de elegir así el elemento identidad.

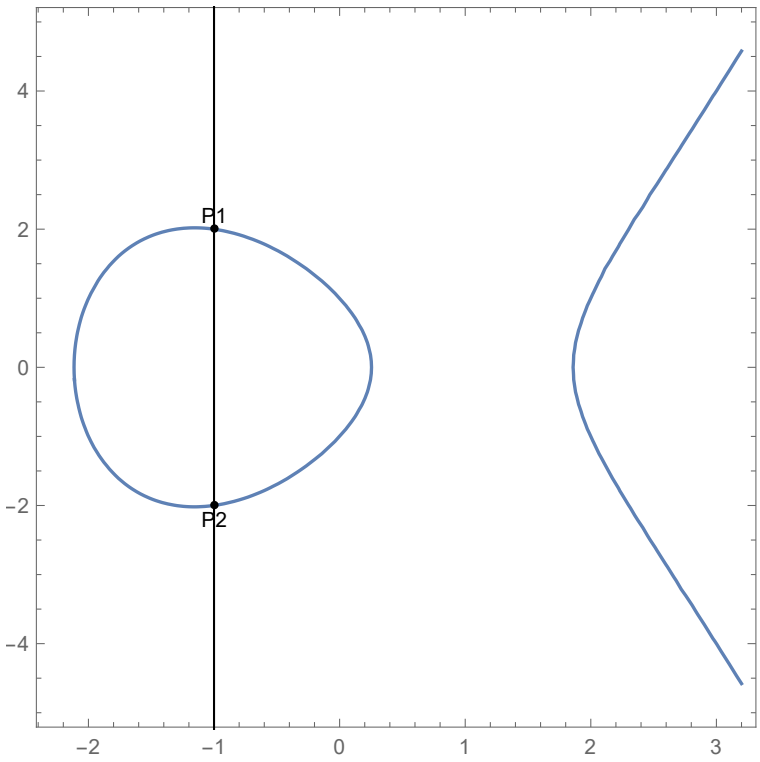

Figura 1.3: Caso de la recta vertical.

Dados P_1, P_2 puntos de la curva elíptica, cuando intersecamos la recta vertical que pasa por P_1, P_2 con la curva elíptica, automáticamente nos viene a la cabeza preguntarnos acerca de cuál será la tercera intersección, pues, como se puede ver en la figura 1.3 de la curva $y^2 = x^3 - 4x + 1$, no parece estar entre los puntos de esta curva... ¿O sí?

Ya sabemos que intersecan en un punto en la recta en el infinito. Dadas la recta $ax + by + c = 0$ y la recta en el infinito se cumple que:

- Intersecan en el punto $(0:1:0)$ bajo la condición de que $b = 0$ (es decir, que sea una recta vertical).
- Intersecan en el punto $(1:0:0)$ bajo la condición de que $a = 0$ (es decir, recta horizontal).

- Por último, interesecan en $(\frac{-b}{a} : 1 : 0) = (1 : \frac{-a}{b} : 0)$ si $a, b \neq 0$.

A nosotros nos interesa la información de que toda la recta vertical interseca a la curva elíptica en $(0 : 1 : 0)$.

Luego si denotamos a $O = (0 : 1 : 0)$, tendremos que $P_1 \cdot P_2 = O$ y así

$$P_1 \oplus P_2 = O \cdot (P_1 \cdot P_2) = O \cdot O.$$

Sin embargo, ¿de qué punto se trata $O \cdot O$? Fácil, veamos que es $(0 : 1 : 0)$.

Para todas las demás rectas consideradas, teníamos que intersecaban a la curva elíptica en 3 puntos afines o 2 afines y 1 en la recta en el infinito. Ahora $O \cdot O$ es el punto de intersección de la recta que une O y O y la curva elíptica. Entonces, 2 de los puntos que intersecan están ya en la recta en el infinito, luego esta recta no puede estar entre las ya consideradas. Por lo tanto no puede ser una recta afín y por ello debe ser la recta en el infinito, que sabemos que solo interseca a la curva elíptica en el punto O, y forzosamente $O \cdot O = O$. Esto además conlleva que $P_1 \oplus P_2 = O$; es decir, dados $P = (x : y : 1), Q = (x : -y : 1)$ que definen una recta vertical, entonces $P \oplus Q = O$.

Ahora podemos describir nuestra ley de adición. Dados P y Q puntos de la curva elíptica, hallamos la intersección de la recta PQ con la curva elíptica, $(P \cdot Q) = (x : y : 1)$, y tenemos que $P \oplus Q = (x : -y : 1)$

Así, si tenemos la curva $y^2 = x^3 - 4x + 1$ y los puntos P y Q en ella, podemos ver cómo actúa la ley de adición gráficamente en la figura 1.4.

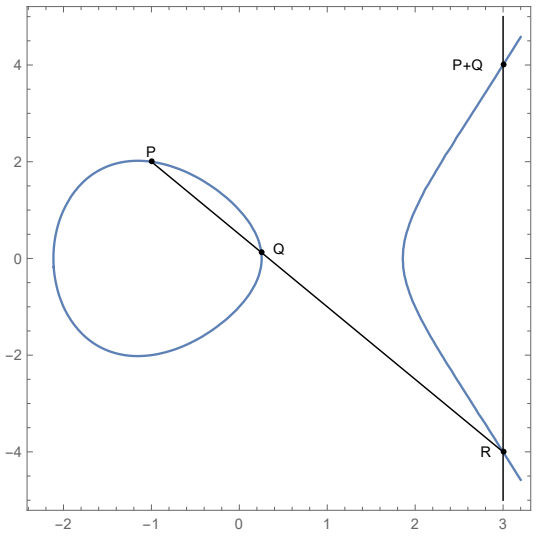

Figura 1.4: Representación visual de la ley de adición.

Para un punto $P = (x : y : 1)$ en una curva elíptica, en concreto en el plano afín, su reflejo respecto al eje x será $Q = (x : -y : 1)$, que está en la curva, y como estos puntos están en una vertical, su suma $P \oplus Q = O$, entonces tenemos $Q = -P$ según nuestra ley de grupo.

Nota 1.5. Usando asociatividad,

$$P \oplus Q \oplus (P \cdot Q) = (P \oplus Q) \oplus (P \cdot Q) = -(P \cdot Q) \oplus (P \cdot Q) = O.$$

Sin embargo, aún debemos considerar otros casos, los casos en que la recta sea tangente, es decir, que tenga un punto de intersección doble con la curva. Veamos un ejemplo:

$$P \oplus P \oplus S = O. \tag{1.3}$$

Tenemos que calcular en (1.3) el resultado de la operación $P \oplus P = 2P$. En general, para calcular $kP = P \oplus \cdots \oplus P$ con $k \in \mathbb{N}$, podemos ver fácilmente que $3P = P \oplus 2P$, y siguiendo con este racionamiento $kP = P \oplus (k-1)P = P \oplus P \oplus \cdots \oplus P$.

Surge, de este proceso de reiteración de la operación, la noción de puntos de torsión:

Definición 1.3. *Diremos que un punto P de la curva elíptica es un* ***punto de torsión*** *si existe $k \geq 2$ tal que $kP = O$.*

Veamos también la definición del orden de un punto de torsión.

Definición 1.4. *Si existe $k \geq 2$ tal que $kP = O$, k es el* ***orden del punto de torsión*** *P. Si $kP \neq O$ para todo $k \geq 1$ diremos que el orden de P es infinito.*

En el caso que nos ocupa, como consideramos una recta tangente, entonces estaremos tratando con puntos de torsión de orden 2.

Para la curva $y^2 = x^3 - 4x + 1$, obsérvese la gráfica de la figura 1.5, que la representa y muestra el caso (1.3).

1.6. Buscando una fórmula para la ley de grupo de una curva elíptica

Nuestro interés recae ahora en cómo expresar la ley de adición mediante operaciones aritméticas. A veces, no consideramos cuerpos como $\mathbb{R}$, $\mathbb{C}$ o $\mathbb{Q}$ sino $\mathbb{F} = \mathbb{F}_p$ con p un número primo. Las fórmulas y procedimientos siguientes están realizados para $p \neq 2$ y 3, por una cuestión de conveniencia. Así que sea $\mathbb{F}$ uno de estos cuerpos, y supongamos que tenemos la siguiente ecuación de la curva elíptica:

$$y^2 = z^3 + Az^2 + Bz + C.$$

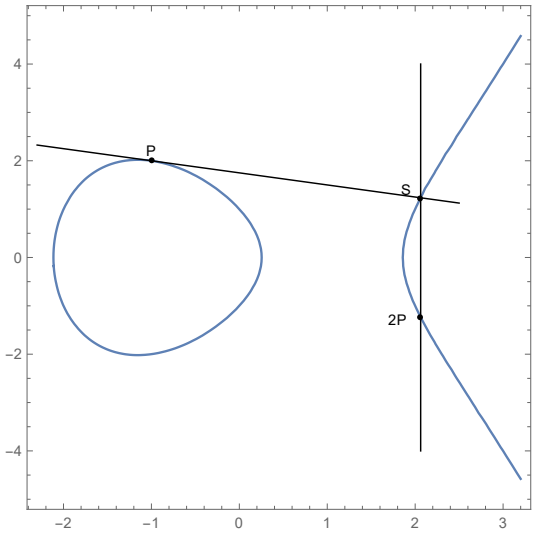

Figura 1.5: Representación gráfica de la ecuación (1.3) para P, S en el conjunto de puntos de la curva $y^2 = x^3 - 4x + 1$.

A partir de esta ecuación, podemos llegar mediante un cambio de variable, a la siguiente ecuación:

$$y^2 = x^3 + ax + b,$$

con $A, B \in \mathbb{F}$. Dicho cambio de variable es

$$z = x - \frac{A}{3}, \quad a = \frac{1}{3}(3B - A^2), \quad b = \frac{1}{27}(2A^3 - 9AB + 27C).$$

Asociado a todo polinomio $p(x) = a_n x^n + \cdots + a_1 x + a_0$ con coeficientes en el cuerpo $\mathbb{F}$, está la cantidad conocida como discriminante, Δ, que se describe en términos de las raíces del polinomio como sigue.

Dado un polinomio $p(x)$ con coeficientes en un cuerpo $\mathbb{F}$ y con n raíces, $r_1, \ldots, r_n$ (en el propio cuerpo o en una extensión algebraica adecuada), podemos expresar el discriminante Δ como

$$\Delta = a_n^{2n-2} \prod_{1 \le i < j \le n} (r_i - r_j)^2.$$

El discriminante da información sobre el polinomio; por ejemplo, si hay alguna raíz múltiple $r_i = r_j$ entonces $\Delta = 0$. Además, aunque quede definido en términos de raíces que puedan estar en una extensión algebraica del cuerpo, podrá ser expresado en términos de los coeficientes del polinomio original.

Para el caso de la cúbica $x^3 + ax + b$ entonces $\Delta = 4a^3 + 27b^2$, y la curva $y^2 = x^3 + ax + b$ define una curva elíptica cuando la cúbica es no singular (distintas raíces). Esto es, que el discriminante sea no nulo.

Ahora la curva homogeneizada es la siguiente:

$$y^2z = x^3 + axz^2 + bz^3.$$

Tiene en $O = (0 : 1 : 0)$ un punto en el infinito y todos los demás puntos son afines. El punto O es la identidad y nosotros deseamos saber cómo, dados P_1, P_2 dos puntos afines, con $P_i = (x_i : y_i : 1)$, $i = 1, 2$, hallar una expresión para la suma. Sea L la recta que une los puntos (tangente en el caso de que sean el mismo punto), y evitando que P_1 y P_2 estén en una vertical ($P_1 \oplus P_2 = O$) entonces la pendiente m es

$$m := \frac{y_2 - y_1}{x_2 - x_1} \quad \text{si} \quad x_1 \neq x_2$$

o

$$m := \frac{3x_1^2 + a}{2y_1} \quad \text{si} \quad x_1 = x_2,$$

donde la pendiente de la recta tangente se obtiene mediante derivación implícita y evaluación en (x_1, y_1) de la ecuación $y^2 = x^3 + ax + b$.

L, la recta que une los puntos P_1 y P_2, será

$$y - y_1 = m(x - x_1).$$

Despejando para y y sustituyendo en la ecuación de la curva obtenemos

$$(m(x - x_1) + y_1)^2 = x^3 + ax + b.$$

Operando,

$$x^3 - mx^2x^2 + Cx + D = 0,$$

donde las expresiones de C y D no son necesarias para lo que haremos.

Ahora las raíces de la cúbica son las tres coordenadas x de los puntos de intersección, x_1, x_2 y x_3, siendo x_3 la coordenada x de la intersección entre la recta L y la curva elíptica

$$x^3 - mx^2x^2 + Cx + D = (x - x_1)(x - x_2)(x - x_3) = x^3 - (x_1 + x_2 + x_3)x^2 + Cx + D.$$

Finalmente llegamos a la expresión de las coordenadas del punto $P_1 \oplus P_2 = P_3 = (x_3 : -y_3 : 1)$, con x_3 e y_3 expresados de la siguiente forma:

$$x_3 = m^2 - x_1 - x_2, \qquad y_3 = m(x_3 - x_1) + y_1.$$

1.6.1. Resumen de la ley de grupo

1. El elemento identidad queda definido como $O = (0 : 1 : 0)$.

2. Si $P = (x : y : 1)$ es un punto afín en la curva, entonces su inverso, $-P$, es $-P = (x : -y : 1)$.

3. Dados los puntos $P_1 = (x_1 : y_1 : 1)$, $P_2 = (x_2 : y_2 : 1)$, entonces $P_2 = -P_1$ si y solo si $P_1 \oplus P_2 = O$. Lo cual implica que la recta entre P_1 y P_2 es vertical, luego las coordenadas cumplen que $x_1 = x_2$, $y_1 = -y_2$.

4. Y, en general, la ley de adición es:

   - Si $x_1 \neq x_2$,

   $$x_3 = m^2 - x_1 - x_2, \quad y_3 = m(x_1 - x_3) - y_1, \quad m = \frac{y_2 - y_1}{x_2 - x_1}. \tag{1.4}$$

   - Si $x_1 = x_2$, pero $y_1 \neq y_2$,

   $$P_1 \oplus P_2 = O.$$

   - Si $P_1 = P_2$ e $y_1 \neq 0$,

   $$x_3 = m^2 - 2x_1, \quad y_3 = m(x_1 - x_3) - y_1, \quad m = \frac{3x_1^2 + A}{2y_1}. \tag{1.5}$$

   - Si $P_1 = P_2$ e $y_1 = 0$, entonces $P_1 \oplus P_2 = O$.
   - Además,

   $$P \oplus O = P \quad \forall P.$$

Así, la operación "$\oplus$" hace al conjunto de puntos de la curva elíptica un grupo abeliano.

Supongamos los casos en que los denominadores de las fórmulas de la pendiente son 0. En la primera fórmula estaríamos diciendo que $x_1 - x_2 = 0$ y hemos excluido este caso en la fórmula (1.4). En el segundo caso, $2y_1 = 0$, hemos excluido los casos con $\mathbb{F}_p$ y $p = 2$ o 3, luego $y_1 = 0$ pero esto es absurdo, pues en esta fórmula (1.5) imponemos que $x_1 = x_2$. Así, para $y^2 = x^3 + ax + b$ habrá como mucho dos soluciones para un valor de x dado, luego si $y_1 = 0$ y $x_1 = x_2$ entonces $y_2 = 0$, pero por ello $P_1 = (x : 0 : 1) = P_2 = -P_1$ y este caso estaba excluido.

1.7. Número de puntos de una curva elíptica

Es importante para el uso de curvas elípticas en el ámbito de la criptografía conocer el tamaño de $E(\mathbb{F}_p)$ (el conjunto de puntos de la curva elíptica) si estamos en el cuerpo $\mathbb{F}_p$ y tenemos la curva elíptica E. Para simplificar, supongamos primos $p > 3$. Para $x_0 \in \mathbb{F}_p$ si sustituimos x_0 en $y^2 = f(x)$ hay varias posibilidades:

- La primera es que si $f(x_0)$ no es un cuadrado, entonces no hay puntos de la forma $(x_0 : y : 1)$ en la curva.
- En segundo lugar, si $f(x_0) = 0$ en $\mathbb{F}_p$ obtenemos como punto $(x_0 : 0 : 1)$ de la curva E.
- Por último, si $f(x_0) = y_0^2$, es decir es un cuadrado, obtenemos dos soluciones, $(x_0 : \pm y_0 : 1)$.

Aunque nosotros estamos en el cuerpo $\mathbb{F}_p$, podemos ver qué ocurre con los cuadrados en el grupo multiplicativo $U_p = \mathbb{F}_p \setminus \{0\}$ para hacernos una primera idea.

Consideramos el grupo multiplicativo U_p, cuyo orden es $p - 1$. Veamos que exactamente la mitad de los elementos de U_p son cuadrados (módulo p), y la otra mitad no lo son.

$$U_p = \{\overline{1}, \overline{2}, \ldots, \overline{p-1}\} = \left\{\overline{1}, \overline{2}, \ldots, \overline{\frac{p-1}{2}}, \overline{\frac{-p+1}{2}}, \ldots, \overline{-1}\right\}$$

Como $K^2 \equiv (-K)^2 \pmod{p}$ para $K = 1, \ldots, \frac{p-1}{2}$, éstos son todos los posibles cuadrados, es decir, a lo más, hay $\frac{p-1}{2}$ cuadrados en U_p.

Veamos ahora que hay exactamente $\frac{p-1}{2}$.

Suponemos que $a, b \in \mathbb{Z}$ y $a^2 \equiv b^2 \pmod{p}$. Esto es, $a^2 - b^2 \equiv 0 \pmod{p}$ o $p \mid (a^2 - b^2) = (a - b)(a + b)$ Pero p es primo, luego debe dividir a uno de los términos del producto, entonces $p \mid (a - b)$ o $p \mid (a + b)$, luego $a \equiv \pm b \pmod{p}$. Para k fijo, $k \in \{1, 2, \ldots, \frac{p-1}{2}\}$, la única forma de que $a^2 \equiv k^2 \pmod{p}$ es si $a \equiv \pm k \pmod{p}$. Luego $\{\overline{k}^2 \mid k = 1, 2, \ldots, \frac{p-1}{2}\}$ son todos los cuadrados distintos de U_p. Esto quiere decir que entre los elementos de U_p, la mitad de ellos son cuadrados y la otra mitad no. Si $a \in U_p$, y a es un cuadrado en U_p, entonces a es un resto cuadrático.

Ahora, para el tamaño de $E(\mathbb{F}_p)$, usaremos el teorema de Hasse:

Teorema 1.1. *Sea $\#E(\mathbb{F}_p)$ el número puntos en el conjunto $E(\mathbb{F}_p)$. Entonces se cumplen las desigualdades*

$$-2\sqrt{p} < \#E(\mathbb{F}_p) - (p + 1) < 2\sqrt{p}.$$

La demostración de este teorema aparece en el libro [10]. Del teorema podremos obtener cierta información valiosa para el uso de las curvas elípticas en la criptografía. Dada una curva elíptica $E(\mathbb{F}_p)$, queremos conocer la probabilidad de, dado un $x \in \mathbb{F}_p$, sea éste la coordenada x de un punto en la curva. Si $N = \#E(\mathbb{F}_p)$, entonces, excepto en 4 casos, todo punto $(x : y : 1)$ está emparejado con un punto distinto $(x : -y : 1)$ en la curva, así que sin precisar, $\frac{N}{2}$ es el número de coordenadas x de puntos en la curva, luego $\frac{N}{2p}$ es aproximadamente la probabilidad de que un número elegido aleatoriamente k sea la coordenada x de un punto en la curva. Por el teorema de Hasse,

$$p + 1 - 2\sqrt{p} < N < p + 1 + 2\sqrt{p}$$

y

$$\frac{1}{2} + \frac{1}{2p} - \frac{1}{\sqrt{p}} < \frac{N}{2p} < \frac{1}{2} + \frac{1}{2p} + \frac{1}{\sqrt{p}},$$

que, para p grandes, da una probabilidad de aproximadamente $\frac{1}{2}$.

Teorema 1.2. *Dada un número primo p grande, y una curva elíptica $E(\mathbb{F}_p)$, la probabilidad de que un $n \in \mathbb{F}_p$ sea la coordenada x de un punto $P = (x : y : 1)$ de la curva es aproximadamente $\frac{1}{2}$.*

1.8. Demostración de la asociatividad

Como hemos comentado en la subsección 1.5.5 la demostración de la asociatividad de la operación "$\oplus$" es complicada de hacer sin conocer las fórmulas que hemos estudiado en la sección 1.6. Sin embargo, una vez vistas podemos utilizarlas para tratar de demostrar la asociatividad.

En primer lugar supongamos el caso en que tenemos tres puntos distintos A, B y C de una curva elíptica $E := y^2 = x^3 + ax + b$. Podemos expresar los puntos como $A = \{c, c^3 + ac + b\}$, $B = \{d, d^3 + ad + b\}$ y $C = \{e, e^3 + ae + b\}$. Queremos ver que realmente $A \oplus (B \oplus C) = (A \oplus B) \oplus C$, por lo tanto primero calculamos $A \oplus B = (x_P : y_P : 1)$.

Para ello calcularemos primero x_P e y_P. Usaremos para estos cálculos las fórmulas de la subsección 1.6.1 que hemos introducido en Mathematica mediante el siguiente código:

```
m1[x1_, y1_, x2_, y2_] := (y2 - y1)/(x2 - x1);
m2[x1_, a_, y1_] := (3*(x1^2) + a)/(2*y1);
x3[x1_, x2_, m_] := (m*m) - x1 - x2;
y3[x1_, y1_, x3_, m_] := m*(x1 - x3) - y1;
```

Como A y B son dos puntos distintos usaremos la fórmula `m1` y las fórmulas de `x3` e `y3` para calcular las coordenadas del punto $A \oplus B$. El código en Mathematica es

```
mP = m1[c, F[c, a, b], d, F[d, a, b]];
xP = x3[c, d, mP];
yP = y3[c, F[c, a, b], xP, mP];
```

con la función F definida en Mathematica previamente como

```
F[x_, a_, b_] := Sqrt[x^3 + a*x + b];
```

Así el punto $A \oplus B$ será $(x_P : y_P : 1)$ con

$$x_P = \frac{\left(\sqrt{ad+b+d^3} - \sqrt{ac+b+c^3}\right)^2}{(d-c)^2} - c - d$$

e

$$y_P = \frac{\left(\sqrt{ad+b+d^3} - \sqrt{ac+b+c^3}\right)\left(2c+d-\frac{\left(\sqrt{ad+b+d^3}-\sqrt{ac+b+c^3}\right)^2}{(d-c)^2}\right)}{d-c} - \sqrt{ac+b+c^3}.$$

Ahora debemos calcular $(A \oplus B) \oplus C$; de nuevo los puntos son distintos entre sí, así que utilizamos las fórmulas `m1`, `x3` e `y3` definidas en Mathematica previamente para obtener el punto $(A \oplus B) \oplus C = (X_1 : Y_1 : 1)$. El código escrito en Mathematica para obtener X_1 e Y_1 es

```
M1 = m1[xP, yP, e, F[e, a, b]];
X1 = x3[xP, e, M1];
Y1 = y3[xP, yP, X1, M1];
```

Procedemos ahora a calcular $A \oplus (B \oplus C)$. En primer lugar calculamos $B \oplus C = (x_Q : y_Q : 1)$, para ello utilizamos el siguiente código en Mathematica

```
mQ = m1[d, F[d, a, b], e, F[e, a, b]];
xQ = x3[d, e, mQ];
yQ = y3[d, F[d, a, b], xQ, mQ];
```

Y así obtenemos las coordenadas del punto $B \oplus C$, que son

$$x_Q = \frac{\left(\sqrt{ae+b+e^3} - \sqrt{ad+b+d^3}\right)^2}{(e-d)^2} - d - e$$

e

$$y_Q = \frac{\left(\sqrt{ae+b+e^3} - \sqrt{ad+b+d^3}\right)\left(2d+e-\frac{\left(\sqrt{ae+b+e^3}-\sqrt{ad+b+d^3}\right)^2}{(e-d)^2}\right)}{e-d} - \sqrt{ad+b+d^3}.$$

Una vez obtenido $B \oplus C$ procedemos a calcular $A \oplus (B \oplus C) = (X_2 : Y_2 : 1)$ y el código que usamos en Mathematica es

```
M2 = m1[c, F[c, a, b], xQ, yQ];
X2 = x3[c, xQ, M2];
Y2 = y3[c, F[c, a, b], X2, M2];
```

Queremos ver que $(A \oplus B) \oplus C = A \oplus (B \oplus C)$ así que nos basta con comprobar que las coordenadas de ambos puntos coinciden, es decir que se cumple que $X_1 = X_2$ y que $Y_1 = Y_2$. Para ello, en Mathematica hacemos

```
FullSimplify[X1 - X2]
FullSimplify[Y1 - Y2]
```

y en ambos casos el resultado es cero, luego en el caso de que los tres puntos sean distintos podemos afirmar que existe asociatividad.

Veamos el caso en que dos de los puntos coincidan, por ejemplo $A = B = (c : c^3 + ac + b : 1)$ y $C = (e : e^3 + ae + b : 1)$. Primero calcularemos $(A \oplus B) \oplus C = (X_3 : Y_3 : 1)$. Empecemos calculando $A \oplus B$, en este caso usaremos la fórmula `m2` ya que son el mismo punto. El código de Mathematica es

```
mK = m2[c, a, F[c, a, b]];
xK = x3[c, c, mK];
yK = y3[c, F[c, a, b], xK, mK];
```

y los resultados de las coordenadas son

$$x_K = \frac{\left(a + 3c^2\right)^2}{4\left(ac + b + c^3\right)} - 2c$$

e

$$y_K = \frac{\left(a + 3c^2\right)\left(3c - \frac{\left(a+3c^2\right)^2}{4(ac+b+c^3)}\right)}{2\sqrt{ac + b + c^3}} - \sqrt{ac + b + c^3}.$$

Calculemos ahora $(A \oplus B) \oplus C = (X_3 : Y_3 : 1)$. Para ello ejecutamos en Mathematica el siguiente código:

```
M3 = m1[xK, yK, e, F[e, a, b]];
X3 = x3[xK, e, M3];
Y3 = y3[xK, yK, X3, M3];
```

Tenemos ahora que calcular $A \oplus (B \oplus C) = (X_4 : Y_4 : 1)$. Ejecutamos el siguiente código para calcular $(B \oplus C) = (x_L : y_L : 1)$ (notar que son dos puntos distintos):

```
mL = m1[c, F[c, a, b], e, F[e, a, b]];
xL = x3[c, e, mL];
yL = y3[c, F[c, a, b], xL, mL];
```

Obtenemos los siguientes resultados:

$$x_L = \frac{\left(\sqrt{ae+b+e^3} - \sqrt{ac+b+c^3}\right)^2}{(e-c)^2} - c - e$$

e

$$y_L = \frac{\left(\sqrt{ae+b+e^3} - \sqrt{ac+b+c^3}\right)\left(2c+e-\frac{\left(\sqrt{ae+b+e^3}-\sqrt{ac+b+c^3}\right)^2}{(e-c)^2}\right)}{e-c} - \sqrt{ac+b+c^3}.$$

Por último calculamos $A \oplus (B \oplus C) = (X_4 : Y_4 : 1)$ mediante el código

```
m4 = m1[c, F[c, a, b], xL, yL];
X4 = x3[c, xL, m4];
Y4 = y3[c, F[c, a, b], X4, m4];
```

y comprobamos que $X_3 - X_4 = 0$ y $Y_3 - Y_4 = 0$.

Para el resto de posibles casos basta tener en cuenta la forma de operar con O y que, como hemos visto, si dos puntos son iguales se utiliza `m2` para hallar la suma y si son distintos se utiliza `m1`.

Capítulo 2

Criptografía con curvas elípticas

Nos disponemos ahora, tras habernos familiarizado con las curvas elípticas y haber obtenido los conceptos necesarios para trabajar con ellas, a ver sus aplicaciones en el ámbito de la criptografía.

Estudiaremos varios sistemas criptográficos basados en las curvas elípticas. Generalmente basarán su seguridad en el problema del logaritmo discreto para las curvas elípticas. También veremos la aplicación de las curvas elípticas en procesos como las firmas digitales.

Existen criptosistemas que no utilizan curvas elípticas, lo cual nos lleva a preguntarnos el porqué del uso de las curvas elípticas. La respuesta es que los métodos criptográficos que se basan en curvas elípticas ofrecen una seguridad similar a la de los métodos tradicionales con un menor tamaño de bits de las claves. Por ejemplo, una clave de 4096 bits del método RSA da la misma seguridad que 313 bits en un sistema de curvas elípticas. Además, se ha comprobado que la generación de claves mediante métodos que aplican curvas elípticas es más rápida.

A modo de introducción al capítulo, vamos primero a ver algunos conceptos de la criptografía.

2.1. Conceptos de la criptografía

Supongamos que una persona, Ana, desea mandar un mensaje a Benito. Sin embargo, Ana desea mantener en secreto la información que va a enviar y para ello utilizará un método criptográfico que le permita cifrar su mensaje y obtener un mensaje cifrado. Cuando Benito reciba el mensaje cifrado lo podrá descifrar y leer el mensaje original.

Para cifrar su mensaje Ana utilizará una clave de cifrado generada por el método. A su vez, Benito utilizará una clave de descifrado para ser capaz de leer el mensaje. Es importante que tan solo Benito tenga acceso a esta clave de descifrado,

ya que si un intruso la conociese podría leer el mensaje de Ana sin problemas.

Se distinguen dos tipos de criptografía:

- **La criptografía simétrica**, donde la clave de descifrado y cifrado son la misma, o una puede ser fácilmente deducida de la otra. Para este tipo de criptografía Ana y Benito deberán establecer comunicación para acordar la clave.

- **La criptografía de clave pública**. No es necesario un intercambio de información para elegir la clave. Benito hace pública una clave de cifrado que Ana utilizará para cifrar su mensaje. Para leerlo, Benito solo debe utilizar la clave de descifrado que tan solo él conoce.

Muchos de los métodos criptográficos que veremos basan su seguridad en el problema del logaritmo discreto.

2.1.1. Problema del logaritmo discreto

Para un número primo p y dos números enteros a y k no nulos en módulo p, podemos realizar la operación $a^k \equiv b \pmod{p}$ y calcular b.

El problema del logaritmo discreto clásico es encontrar el valor de k dado a y b.

De manera más general, para un grupo cualquiera G, pensemos que es multiplicativo por el momento, y para $a, b \in G$, supongamos que se cumple que $a^k = b$ para algún entero k. De nuevo el problema del logaritmo discreto será hallar el valor de k dados a y b. Si ahora pensamos que G es $E(\mathbb{F}_q)$ para alguna curva elíptica E y la operación de la ley de grupo (que en curvas elípticas ya no tiene notación multiplicativa sino aditiva), entonces a y b serán puntos de la curva elíptica y tendremos que encontrar el valor de k tal que $ka = b$. Así es como queda determinado el problema del logaritmo discreto para curvas.

Algunos de los métodos utilizados para tratar de resolver el problema del logaritmo discreto son los métodos de Pohlig-Hellman o los métodos ρ y λ de Pollard (pueden ser estudiados en el libro [10]). Sin embargo, con números suficientemente grandes y bien elegidos es inviable computacionalmente tratar de resolver el problema del logaritmo discreto, y esto es lo que hace seguros los métodos criptográficos basados en este problema.

2.2. Intercambio de claves Diffie-Hellman

Este método permitirá a dos personas poner en común una clave que utilizarán para cifrar y descifrar mensajes; estamos por lo tanto ante un método de clave simétrica, pero diseñado de forma que la clave común no tiene que viajar luego no

puede ser interceptada. Supongamos que Ana quiere enviar un mensaje a Benito y quiere mantenerlo en secreto, para ello tratarán de utilizar una clave común para cifrar y descifrar. Para obtenerla llevarán a cabo los siguientes pasos:

Nota 2.1. Vamos a denotar a Ana como A y a Benito como B.

1. A y B eligen una curva elíptica $E(\mathbb{F}_p)$ definida sobre un cuerpo finito $\mathbb{F}_p$, tal que el problema del logaritmo discreto (subsección 2.1.1) sea difícil de resolver en $E(\mathbb{F}_p)$. Eligen también un punto $P \in E(\mathbb{F}_p)$, que puede ser público, tal que el subgrupo generado por P tenga orden grande. De hecho, se suelen escoger el punto P y la curva $E(\mathbb{F}_p)$ tal que el orden del subgrupo generado por P sea un número primo grande.

2. A elige un número entero a que mantendrá en secreto. Calcula y envía a B el punto $P_a = aP$.

3. B elige un número entero b, que también mantendrá en secreto, y calcula y envía $P_b = bP$ a A.

4. A calcula $aP_b = abP$.

5. B calcula $bP_a = baP$.

6. Finalmente, A y B obtienen del punto común que han hallado $abP = (x : y : 1)$, una clave. Para ello pueden, por ejemplo, evaluar en una función hash (estas funciones están definidas en la sección 2.6) la coordenada x del punto.

Un intruso que desee espiar las comunicaciones, podría interceptar de los mensajes los puntos aP y bP; además, la curva elíptica $E(\mathbb{F}_p)$, el cuerpo finito $\mathbb{F}_p$ y el punto P son públicos. Con esta información para poder obtener la clave debería resolver el conocido como problema de Diffie-Hellman.

2.2.1. Problema de Diffie-Hellman

Dados P, aP y bP en la curva elíptica $E(\mathbb{F}_p)$ calcular abP.

Para resolver este problema se debe ser capaz de resolver el problema del logaritmo discreto en $E(\mathbb{F}_p)$. Utilizando P y aP se podría hallar a resolviendo el logaritmo discreto $kP = aP$ y una vez obtenido se podría calcular abP.

2.2.2. Ejemplo de uso del método

Vamos ahora a ver un ejemplo práctico de cómo funciona el método de intercambio de claves de Diffie-Hellman. Supongamos que Ana y Benito eligen como curva elíptica la curva $E := y^2 = x^3 - 4x + 1$ definida sobre el cuerpo finito $\mathbb{F}_{132247}$. Como punto de la curva eligen $P = (216 : 20967 : 1)$, cuyo orden es 132400. Seguiremos los pasos indicados en la sección 2.2.

1. Si Ana elige como número entero $a = 6$, calculará $P_a = 6P = (8453 : 42390 : 1)$ y se lo enviará a Benito.

2. Ahora si Benito elige como número entero $b = 25$, calculará $P_b = 25P = (103488 : 730 : 1)$ y se lo enviará a Ana.

3. Por último Ana calculará $6P_b = (35152 : 97267 : 1)$ y Benito calculará $25P_a = (35152 : 97267 : 1)$. Notar que son el mismo punto y por tanto Ana y Benito han hallado un punto común del que obtener su clave de cifrado.

Para obtener la clave de cifrado podrían, por ejemplo, utilizar la primera coordenada del punto común $aP_b = abP = baP = bP_a$.

Nota 2.2. Para la realización de los cálculos de este ejercicio se ha utilizado el programa Mathematica y parte del código que aparece en [4]. En los siguientes ejemplos que realicemos también se utilizarán estas herramientas.

2.3. Convertir un mensaje en un punto de una curva elíptica

Antes de continuar con los métodos de cifrado, veamos cómo dado un mensaje podemos convertirlo en un punto de una curva elíptica. Por supuesto, damos por hecho que cualquier mensaje está ya codificado en forma de un entero no negativo; y, que, además, dicho entero no es mayor que cierta cota que dependerá del tamaño de las claves o del método criptográfico que estemos usando (si, en principio, es mayor, hay que trocearlo previamente).

Primero elegimos un número entero M y un número entero k. A continuación, elegimos un número primo p que cumpla $p > Mk$. Notar que podemos escribir todo entero l tal que $1 \leq l \leq Mk$ de manera única como $l = mk + j$ con $0 \leq m < M$ y $1 \leq j \leq k$. Ya que $Mk < p$, podemos pensar en todos los enteros $mk + j$ (de hecho en los restos $\overline{mk + j}$ módulo p) con $0 \leq m < M$ y $1 \leq j \leq k$ como diferentes elementos del cuerpo finito $\mathbb{F}_p$. (Aquí y en lo que sigue, hemos usado la notación $\overline{u}$ para referirnos a la clase de equivalencia de un entero u módulo p.)

Podremos convertir un mensaje representado por un número entero m tal que $0 \leq m < M$ en un punto P_m de una curva elíptica $E_{a,b}$ definida mediante la ecuación $y^2 = f(x) = x^3 + ax + b$ sobre el cuerpo $\mathbb{F}_p$. Ahora, para cada j con $1 \leq j \leq k$ probaremos, para $t(j) = mk + j$, si $t(j)$ módulo p es la coordenada x de algún punto $Q = (x : y : 1)$ del conjunto $E_{a,b}(\mathbb{F}_p)$. Acabamos de observar en el párrafo anterior que todos los $\overline{t(j)}$ son distintos elementos del cuerpo $\mathbb{F}_p$. Calculamos $\overline{f(t(j))}$ para ver si es un cuadrado en $\mathbb{F}_p$. Vimos en el teorema 1.2 que la probabilidad de que esto ocurra es del 50 %. Si es un cuadrado podremos hallar $\overline{y} \in \mathbb{F}_p$ tal que $\overline{y}^2 = \overline{f(t(j))}$ y el mensaje se habrá convertido en el punto $P_m = (\overline{t(j)}, \overline{y})$. Si no es un cuadrado aumentamos en uno el valor de j y probamos de nuevo. Como tenemos k enteros $t(j) = mk + j$, puesto que $1 \leq j \leq k$, la probabilidad de fallar al producir el punto P_m es aproximadamente 2^{-k}.

Veamos ahora cómo podemos obtener el mensaje original del punto $P_m = (\overline{x}, \overline{y})$. Primero consideramos $\overline{x} = \overline{mk + j}$ módulo p. Notar que $x - 1$ cumple la cadena de desigualdades $mk \leq x - 1 \leq mk + (k - 1)$, y, por lo tanto,

$$m \leq \frac{x-1}{k} \leq m + \frac{k-1}{k} < m + 1,$$

lo que implica que m, el mensaje original, se puede recuperar como la parte entera de $\frac{x-1}{k}$, es decir, $m = \lfloor \frac{x-1}{k} \rfloor$.

2.4. Método criptográfico de Massey-Omura

Este método no requiere que Ana y Benito establezcan una comunicación previa para establecer una clave privada. Los pasos que A y B han de seguir son los siguientes:

1. A y B eligen una curva elíptica E definida sobre un cuerpo finito $\mathbb{F}_p$ tal que el problema del logaritmo discreto sea difícil de resolver en $E(\mathbb{F}_p)$. Denotamos al número de puntos de la curva, $\#E(\mathbb{F}_p)$, como N.

2. A convierte su mensaje en un punto $M \in E(\mathbb{F}_p)$ utilizando el procedimiento visto en la sección 2.3. Elige un número entero m_A tal que $\text{mcd}(m_A, N) = 1$ y lo mantiene en secreto. Calcula y envía $M_1 = m_A M$ a B.

3. B elige también un número entero secreto m_B tal que $\text{mcd}(m_B, N) = 1$ y calcula y envía $M_2 = m_B M_1$ a A.

4. A calcula el inverso de m_A, que es $m_A^{-1} \in \mathbb{Z}_N$. Calcula y envía $M_3 = m_A^{-1} M_2$ a B.

5. B calcula el inverso de m_B, que es $m_B^{-1} \in \mathbb{Z}_N$. Después obtiene el mensaje de Ana calculando $M_4 = m_B^{-1} M_3 = M$.

Efectivamente se cumple que el mensaje original M es igual que M_4 ya que

$$M_4 = m_B^{-1} m_A^{-1} m_B m_A M = M.$$

Basta justificar que m_A^{-1}, que es un número entero que representa el inverso de m_A módulo N, y m_A se cancelan entre sí. Sabemos que $m_A^{-1} m_A \equiv 1$ módulo N, por lo tanto $m_A^{-1} m_A = 1 + kN$ para algún número entero k. Dado que el orden de $E(\mathbb{F}_p)$ es N, tenemos que $NR = O$ para todo punto $R \in E(\mathbb{F}_p)$. Por lo tanto, si consideramos $R = m_B M$,

$$m_A^{-1} m_A R = (1 + kN)R = R + kO = R = m_B M.$$

Utilizando lo mismo para m_B^{-1} y m_B, se cancelan y

$$M_4 = m_B^{-1} M_3 = m_B^{-1} m_B M = M.$$

Un intruso podría conocer $E(\mathbb{F}_p)$ y los puntos $m_A M$, $m_B m_A M$ y $m_B M$. Pero realmente tratar de obtener M_4 es equivalente a resolver el problema de Diffie-Hellman que vimos en la subsección 2.2.1 si denotamos a $a = m_A^{-1}$, $b = m_B^{-1}$ y a $P = m_A m_B M$ y conocemos P, bP y aP y queremos hallar abP.

2.4.1. Ejemplo de uso del método de Massey-Omura

Veamos un ejemplo de aplicación del método que acabamos de estudiar. Benito (B) y Ana (A) eligen la curva elíptica $E := y^2 = x^3 - 4x + 1$ definida sobre el cuerpo finito $\mathbb{F}_{132247}$. Calculan el número de puntos de la curva que es $\#E(\mathbb{F}_{132247}) = N = 134200$. Ahora, si Ana desea enviar un mensaje m que corresponde al punto de la curva elíptica $M = (385 : 75917 : 1)$ veamos los pasos que seguirá:

1. En primer lugar Ana elige un número entero m_A que mantendrá en secreto; supongamos que elige $m_A = 3$ que cumple que $\text{mcd}(m_A, N) = 1$, calcula $M_1 = 3M = (97862 : 64239 : 1)$ y se lo envía a B.

2. Benito elige como entero $m_B = 7$, de nuevo se cumple que $\text{mcd}(m_B, N) = 1$. Calcula $M_2 = 7M_1 = (18386 : 37828 : 1)$ y se lo envía a A.

3. A calcula m_A^{-1} módulo N, el resultado es $m_A^{-1} = 88267$ y ahora calcula $M_3 = m_A^{-1} M_2 = (122257 : 128929 : 1)$ y se lo envía a B.

4. Por último, para finalizar la comunicación y que el mensaje M llegue a B, éste debe calcular m_B^{-1} módulo N, el resultado es $m_B^{-1} = 56743$ y ahora calcula $M_4 = m_B^{-1} M_3 = (385 : 75917 : 1)$.

Vemos que, efectivamente, $M_4 = M = (385 : 75917 : 1)$ y por lo tanto el mensaje de A ha llegado hasta B.

2.5. Criptografía con clave pública de ElGamal

Este método permite las comunicaciones privadas entre dos individuos. Cuando hablamos de una clave pública queremos decir que la manera de cifrar los mensajes es accesible a cualquiera que desee saberlo, sin embargo los mensajes solo podrán ser descifrados por el receptor que ha hecho pública la clave.

Supongamos que Benito (B) va a ser quien establezca la clave pública, y por tanto el receptor de los mensajes. Para ello elige, en primer lugar, una curva elíptica E en un cuerpo finito $\mathbb{F}_p$, tal que el problema del logaritmo discreto sea difícil de resolver para el conjunto $E(\mathbb{F}_p)$. Elegirá además un punto P de la curva E; se suele escoger un punto cuyo orden sea un número primo grande. Por último, elige un número entero s que mantendrá en secreto y calcula $D = sP$. La clave pública consiste de la curva E, el cuerpo finito $\mathbb{F}_p$ y los puntos D y P.

Para mandar un mensaje a B se deben seguir los siguientes pasos:

1. Se obtiene la clave pública.

2. Se expresa el mensaje como un punto $M \in E(\mathbb{F}_p)$ como vimos en la sección 2.3.

3. Se elige un número entero k que se debe mantener en secreto y se calculan $M_1 = kP$ y $M_2 = M + kD$.

4. Se envían a B los puntos M_1 y M_2.

Para obtener el mensaje original basta con calcular $M_2 - sM_1 = M$.
Esta operación funciona porque

$$M_2 - sM_1 = (M + kD) - s(kP) = M + k(sP) - skP = M.$$

Un intruso que desee espiar las comunicaciones conocería la información pública y los puntos M_1 y M_2. De nuevo el problema del logaritmo discreto impide a un intruso obtener s a partir de D y P u obtener k a partir de P y M_1, y por tanto no será capaz de espiar las comunicaciones.

En este método es crucial, para la seguridad del mismo, que el emisario no reutilice el entero k, sino que lo cambie en cada comunicación. Si lo reutilizase para encriptar dos mensajes M y H, entonces un intruso se daría cuenta de esto porque $M_1 = H_1 = kP$. Ahora podría calcular $H_2 - M_2 = H - M$. Supongamos que el primer mensaje enviado, por ejemplo M, se hace público al día siguiente. El intruso podría fácilmente utilizar M para calcular H.

2.5.1. Ejemplo de la criptografía con clave pública de ElGamal

Supongamos que Benito va a ser quien establezca la clave pública, para ello elige la curva elíptica $E := y^2 = x^3 - 4x + 1$ definida sobre el cuerpo finito $\mathbb{F}_{132247}$. Como punto de la curva elige $P = (11 : 21425 : 1)$ y como número entero $s = 5$. Debe mantener s en secreto y calculará $D = 5P = (43742 : 77301 : 1)$. La clave pública está compuesta por D, P, la curva elíptica y el cuerpo finito.

Supongamos ahora que alguien desea enviar un mensaje a Benito. El emisor, para cifrar su mensaje, elige el número entero $k = 30$ y convierte su mensaje en un punto de la curva elíptica. Supongamos que el mensaje corresponde al punto $M = (173 : 27427 : 1)$. Ahora calcula $M_1 = 30P = (116403 : 57120 : 1)$ y $M_2 = M + 30D = (76392 : 9825 : 1)$ y se los envía a Benito. Benito, para obtener el mensaje M calcula $M_2 - 5M_1$. El resultado es $(173 : 27427 : 1)$, que efectivamente corresponde al punto M.

2.6. Firma digital de ElGamal

Estudiaremos ahora un método para firmar documentos de manera digital, sin que sea posible suplantar la identidad del emisor ni copiar su firma digital.

Supongamos que Ana quiere firmar un documento de manera digital. Una primera idea sería escanear su firma y adjuntarla en el documento, sin embargo un estafador podría copiarla en otros documentos y hacerse pasar por Ana. No parece una forma adecuada de firmar el documento. Además debe ser posible verificar que la firma es realmente de Ana.

El método que veremos a continuación está basado en el problema del logaritmo discreto y ofrece una solución para firmar digitalmente los documentos.

En primer lugar, Ana creará una clave pública. Para ello elige una curva elíptica E definida sobre un cuerpo finito $\mathbb{F}_p$ tal que el problema del logaritmo discreto sea difícil de resolver en $E(\mathbb{F}_p)$. Elegirá también un punto $C \in E(\mathbb{F}_p)$, de nuevo se suele elegir C tal que su orden, al que denotamos N, sea un número primo grande. También debe elegir un número entero a, calcular $D = aC$ y una función $f : E(\mathbb{F}_p) \longrightarrow \mathbb{Z}$. Las únicas condiciones que la función debe cumplir son que su imagen sea grande y que solo un número reducido de entradas produzca una salida concreta.

La información que se hará pública es E, $\mathbb{F}_p$, C, f y D. No es necesario hacer público el orden de C. Para firmar se llevan a cabo los siguientes pasos:

1. Representar el mensaje como un número entero m. Si se cumple que $m > N$ se debe elegir otra curva o utilizar una función hash.

2. Elegir un número entero aleatorio k tal que $\mathrm{mcd}(k, N) = 1$ y calcular $R = kC$.

3. Calcular $s \equiv k^{-1}(m - af(R))$ módulo N.

El mensaje firmado será (m, R, s). En esta ocasión Ana no mantendrá en secreto el mensaje (notar que no ha sido cifrado, aparece m directamente).

Para verificar la firma, el receptor deberá realizar el siguiente proceso:

1. Obtendrá la información pública que Ana ha publicado previamente.

2. Calculará $V_1 = f(R)D + sR$ y $V_2 = mC$.

3. La firma será valida si $V_1 = V_2$.

Acabamos de ver que la firma es considerada válida si se cumple la igualdad $V_1 = V_2$, esto se debe a que

$$V_1 = f(R)D + sR = f(R)aC + skC = f(R)aC + (m - af(R))C = mC = V_2.$$

Hemos utilizado que $sk \equiv m - af(R)$ módulo N, por ello $sk = m - af(R) + zN$ para algún número entero z. De aquí se deduce que

$$skC = (m - af(R))C + zNC = (m - af(R))C + O = (m - af(R))C.$$

Esto explica por qué la congruencia que define a s la hemos tomado módulo N.

Veamos ahora cómo podría ser superada la seguridad del método. Supongamos que un intruso fuese capaz de resolver el problema del logaritmo discreto en $E(\mathbb{F}_p)$, por lo tanto sería capaz de hallar a mediante C y $D = aC$. También usando R y C podría calcular k y, como conoce $m, f(R)$ y s, podría utilizar $ks \equiv m - af(R)$ módulo N para hallar a. Si $d = \mathrm{mcd}(f(R), N) \neq 1$, entonces la ecuación $af(R) \equiv m - ks$ módulo N tendrá d soluciones para a. Mientras d sea un número pequeño, el intruso podrá probar cada posibilidad hasta obtener $D = aC$.

Una vez conocido a, el intruso podrá suplantar la firma de Ana.

Además de que a y k deben mantenerse en secreto, es importante que no se reutilice el mismo k para firmar dos mensajes distintos. Veamos por qué.

Supongamos que Ana firma con el mismo k los mensajes m y h, obteniendo los mensajes firmados (m, R, s) y (h, R, v). Como se puede ver, es fácil reconocer el uso reiterado de k dado que R aparece en ambos mensajes. De las ecuaciones de s y v llegamos a

$$ks \equiv m - af(R) \quad \text{mód}\, N,$$
$$kv \equiv h - af(R) \quad \text{mód}\, N$$

y, si los restamos, $k(s - v) \equiv m - h$ módulo N.

Para $d = \text{mcd}(s - v, N)$ habrá d posibles valores para k. Un intruso puede probar cada uno de ellos hasta que se cumpla $R = kC$, y una vez conocido k puede hallar a.

Sin embargo, un intruso que quiera falsificar la firma no necesita resolver el problema del logaritmo discreto. Si es capaz de hallar R y s tales que la igualdad $V_1 = V_2$ se cumpla, será capaz de suplantar la identidad de Ana. Si elige un punto R deberá resolver el problema del logaritmo discreto $sR = mC - f(R)D$ para hallar el entero s, pero si elige un número entero s deberá resolver la ecuación $s \equiv k^{-1}(m - af(R))$ módulo N para hallar $R = (x, y)$. Esta ecuación es, al menos, tan difícil de resolver como el problema del logaritmo discreto, por lo que no parece práctico.

Parece por lo tanto que eligiendo una curva elíptica y un cuerpo finito adecuados no será fácil falsificar la firma de Ana. Existe un método que, dado un mensaje firmado por Ana que haya sido interceptado, permite crear nuevos mensajes firmados de la misma manera que el mensaje interceptado, pero carecerán de sentido pues no es posible elegir el contenido del mensaje. Es decir, si obtuviésemos un mensaje firmado por Ana seríamos capaces de crear mensajes que pasarían el proceso de verificación de la firma pero no seríamos capaces de elegir la información que el mensaje transmite.

Una desventaja de este método de firma digital es que el mensaje firmado es aproximadamente unas 3 veces más largo que el mensaje original. Esto, obviamente, no es práctico, pues provocará problemas de almacenamiento en los dispositivos. Pensemos, por ejemplo, que el mensaje que deseamos enviar ocupa un millón de bits; si es necesario que lo firmemos, enviar un mensaje tan pesado podría ocasionar problemas.

Una solución a este problema son las *funciones hash*. Una *función hash* es una función que dada una entrada de tamaño cualquiera (por ejemplo un mensaje de millones de bits) produce una salida de un tamaño fijado (por ejemplo 124 bits). Por lo tanto, una función hash es una función

$$H : \mathbb{N} \cup \{0\} \to M$$

con M un subconjunto de los enteros no negativos. Dado que el conjunto de partida de una función hash es más grande que el de llegada, habrá ciertos valores $x, y \in \mathbb{N}$ para los que $H(x) = H(y)$. Cuando esto ocurre se habla de colisiones de hash. Estas colisiones deben ser aleatorias y no debe ser sencillo descubrir una manera de lograrlas. Para ello se imponen unos requisitos a las funciones hash, que aparecen descritos en [9]:

1. Resistentes al cálculo de preimagenes: dado un número entero y debe ser difícil hallar m tal que $H(m) = y$.

2. Resistencia fuerte a las colisiones de hash: No debe ser fácil encontrar un par $x \neq y$ tal que $H(x) = H(y)$.

3. Resistencia (débil) a las colisiones de hash: Para x y $H(x)$ debe ser difícil encontrar $y \neq x$ tal que $H(y) = H(x)$.

4. Dado un mensaje m debe ser rápido calcular $H(m)$.

La razón de que se exijan las condiciones (1) y (2) es para prevenir que un intruso sea capaz de producir mensajes con un valor hash concreto, o dos mensajes con el mismo valor hash. Esto ayuda a evitar la falsificación de la firma. Un ejemplo de función hash es la denominada MD5 (Message-Digest Algorithm 5) que produce salidas de 128 bits. Esto es que, dada MD5 la función

$$\text{MD5} : \mathbb{N} \cup \{0\} \to M,$$

se cumple que M es el conjunto $\{0, 1, 2, \ldots, 2^{128} - 1\}$.

Si Ana utiliza una función hash, la firma será $(H(m), R_H, s_H)$. Para comprobar que la firma es válida, quien reciba el mensaje deberá realizar los siguientes pasos:

1. Obtener la información pública de Ana.

2. Calcular $v_1 = f(R)D + s_H R_H$ y $V_2 = H(m)C$.

3. Si $V_1 = V_2$ la firma es válida.

2.6.1. Ejemplo de uso de la firma digital de ElGamal

Vamos ahora a firmar un mensaje mediante el método que acabamos de ver en esta sección. Supongamos que Ana debe firmar una transferencia bancaria en la que envía 6500 euros, luego el mensaje a firmar será $m = 6500$. En primer lugar Ana debe crear una clave pública, para ello elige una curva elíptica, supongamos que es $E := y^2 = x^3 - 4x + 1$ definida sobre el cuerpo finito $\mathbb{F}_{132247}$. Elige ahora un punto de la curva, supongamos que elige $C = (14 : 127091 : 1)$, cuyo orden es $N = 132400$; notar que no estamos teniendo cuidado en que el problema del logaritmo discreto sea difícil de resolver en $E(\mathbb{F}_{132247})$ ni en que el orden del punto elegido sea un número primo grande ya que estamos haciendo este ejemplo simplemente para aclarar el funcionamiento del método.

Para acabar de crear la clave pública Ana debe elegir un número entero a, por ejemplo $a = 6$, y calcular $D = 6C = (64415 : 15242 : 1)$, y por último Ana elige la función $f : E(\mathbb{F}_{132247}) \longrightarrow \mathbb{Z}$ tal que $f((x : y : 1)) = x$. Así la clave pública de Ana estará compuesta por los puntos C y D, por la curva elíptica y el cuerpo finito y la función f.

Veamos ahora el procedimiento seguido por Ana para firmar el mensaje $m = 6500$.

1. Ana elige el número entero $k = 7$, que cumple que $\text{mcd}(N, k) = 1$ y calcula $R = 7C = (28877 : 78084 : 1)$.

2. Calcula s módulo N, el resultado es $s = 51834$.

Por lo tanto el mensaje firmado es $(6500, (28877 : 78084 : 1), 51834)$. Vamos a realizar el proceso de verificación de la firma.

1. Primero calculamos $V_1 = (98154 : 37170 : 1)$.

2. Calculamos ahora $V_2 = (98154 : 37170 : 1)$.

Luego efectivamente se cumple que $V_1 = V_2$ y por lo tanto podemos estar seguros de que la transferencia bancaria ha sido realizada por Ana.

2.7. El algoritmo de firma digital

El algoritmo se basa originalmente en grupos multiplicativos de cuerpos finitos, sin embargo existe una versión basada en curvas elípticas, conocida como ECDSA (Elliptic Curves Digital Signature Algorithm), que es la que vamos a estudiar. Realmente es una versión modificada del esquema de firma de ElGamal.

Supongamos que Ana desea firmar un documento convertido en un número entero m (de hecho, se suele firmar la imagen por una función hash del documento como hemos visto en la sección anterior). Para ello elegirá una curva elíptica E definida sobre un cuerpo finito $\mathbb{F}_p$ de tal manera que el tamaño del conjunto de puntos de la curva, $\#E(\mathbb{F}_p) = fr$, con r un número primo grande y f un número entero pequeño, normalmente $1, 2$ o 4. Elige también un punto base G en $E(\mathbb{F}_p)$ de orden r. Por último, Ana elegirá un número entero a, que debe mantener en secreto, y calculará $Q = aG$. La información que se hará pública es $\mathbb{F}_p, E, r, G, Q$. Para firmar el mensaje, Ana sigue los siguientes pasos:

1. Elige un número entero k tal que cumpla que $1 \leq k < r$.

2. Calcula $R = kG = (x, y)$.

3. Calcula $s \equiv k^{-1}(m + ax)$ módulo r.

El mensaje firmado será (m, R, s) y el receptor que desee confirmar la validez de la firma realizará el siguiente procedimiento:

1. Calculará $u_1 \equiv s^{-1}m$ módulo r y $u_2 \equiv s^{-1}x$ módulo r.

2. Calculará $V = u_1G + u_2Q$.

3. La firma es válida si $V = R$.

Esto se debe a que

$$V = u_1G + u_2Q = s^{-1}mG + s^{-1}xQ = s^{-1}(mG + xaG) = kG = R.$$

La diferencia principal entre el método ECDSA y el método de ElGamal está en el proceso de verificación de la firma. En el caso de ElGamal la ecuación de verificación requiere realizar tres veces la operación mK, con m un número entero y K un punto de la curva elíptica, dado que la ecuación de verificación es $f(R)D + sR = mC$. Sin embargo, en el caso de ECDSA tan solo se lleva a cabo dos veces esa operación. Esto provoca que la verificación en el caso del método ECDSA sea mucho más rápida, ya que este tipo de operaciones son costosas a nivel computacional.

2.7.1. Ejemplo del algoritmo de firma digital

De nuevo Ana desea firmar una transferencia bancaria en la cual paga 8000 euros. Sin embargo, ahora desea utilizar el algoritmo que acabamos de estudiar. El mensaje a firmar será $m = 8000$, Ana debe crear de nuevo una clave pública, para ello elegirá la curva elíptica $G := y^2 = x^3 + 4x + 1$ definida sobre el cuerpo finito $\mathbb{F}_{143833}$. El tamaño del conjunto de puntos de la curva es $\#G(\mathbb{F}_{143833}) = 144062$; notar que si factorizamos este número obtenemos $144062 = 2 \cdot 72031$, por lo tanto $\#G(\mathbb{F}_{143833}) = fr$, con $f = 2$ y $r = 72031$ que es un número primo. Ana elige como punto base $P = (1 : 28898 : 1)$ cuyo orden es $r = 72031$. Como número entero a Ana elige $a = 8$ y calcula $Q = 8P = (21201 : 73531 : 1)$. Hará pública la información $\mathbb{F}_{143833}$, G, r, P, Q.

Para firmar el mensaje llevará a cabo el siguiente procedimiento:

1. Supongamos que elige el número entero $k = 55$.
2. Calcula $R = 55P = (114773 : 90563 : 1)$.
3. Calcula $s \equiv 57439$ en módulo 72031.

Y Ana habrá obtenido el mensaje firmado, que es

$$(8000, (114773 : 90563 : 1), 57439).$$

Ahora si queremos verificar la firma realizamos las siguientes operaciones:

1. Calculamos $u_1 \equiv 52759$ mód 72031 y $u_2 \equiv 65443$ mód 72031.
2. Calculamos $V = (114773 : 90563 : 1)$.

Como podemos observar se cumple que $V = R$, luego la firma es verdadera.

2.8. Un sistema criptográfico basado en la factorización

Los métodos criptográficos con curvas elípticas que se han estudiado hasta ahora, en este trabajo, basaban su seguridad en el problema del logaritmo discreto. Sin embargo, algunos de los métodos criptográficos se basan también en la dificultad para factorizar un número que sea producto de dos primos. Un ejemplo es el sistema RSA, que debe su nombre a las iniciales de sus desarrolladores, Rivest, Shamir y Adleman. Este sistema de clave pública, creado en 1978, permite establecer comunicaciones seguras entre dos individuos, sin necesidad de un contacto previo.

Supongamos que Ana desea enviar un mensaje privado a Benito. Para ello Benito elige una pareja de números primos grandes y distintos p y q, calcula su producto $m = pq$ y la indicatriz de Euler de éste, $\phi(m) = (p-1)(q-1)$. Es importante que los número p y q sean lo suficientemente grandes como para hacer difícil la descomposición de m. Escoge ahora otro entero positivo e tal que $\text{mcd}(\phi(m), e) = 1$. La clave pública está formada por m y e, luego deben ser accesibles.

El número entero que representa el mensaje que Ana quiere enviar a Benito es M tal que $0 < M < m$. Los dos únicos números menores que m que no son primos con él son p y q, luego la probabilidad de que M no sea primo con m es la misma que la de encontrar un factor no trivial de m dividiéndolo por un número al azar. Así que podemos dar por prácticamente seguro que M será primo con m.

El texto cifrado será el único entero C que cumple $C \equiv P^e \pmod{m}$ para $0 < C < m$. Ana enviará a Benito el número C como su mensaje cifrado. Para recuperar el mensaje original, Benito debe hallar el entero d tal que $ed \equiv 1$ módulo $\phi(m)$. Es seguro que existe un número d así debido a que $\text{mcd}(\phi(m), e) = 1$ y es fácil hallarlo mediante el algoritmo de Euclides. Ahora, si Benito calcula C^d se cumple que

$$C^d \equiv P$$

en módulo m.

Si un intruso fuese capaz de descompener m obtendría p y q y sería capaz de calcular $\phi(m)$. Podría entonces resolver $ed \equiv 1 \pmod{\phi(m)}$ para hallar d y descifrar el mensaje. Es por esto que los números primos p y q deben ser lo suficientemente grandes como para hacer imposible la descomposición de m en un tiempo razonable.

Nos preguntamos si existe una versión del método RSA que utilice las curvas elípticas. La respuesta es que sí, por ejemplo podemos estudiar el sistema desarrollado por Koyama-Maurer-Okamoto-Vanstone, aunque no sea muy utilizado en la práctica.

Como siempre, supongamos que Ana desea enviar un mensaje privado a Benito. Seguirán el procedimiento mostrado a continuación:

1. Benito elige dos primos distintos lo suficientemente grandes p, q tales que $p \equiv q \equiv 2 \pmod 3$ y calcula $n = pq$.

2. Escoge ahora dos enteros e, d que cumplan que $ed \equiv 1 \pmod{\text{mcm}(p+1, q+1)}$.

3. Benito hace accesible su clave pública que consiste de n y e. Mantiene en secreto d, p y q.

4. Ana representa su mensaje como una pareja de puntos $(m_1, m_2) \pmod n$. Considera el par (m_1, m_2) como un punto M de la curva elíptica E dada por

$$y^2 = x^3 + b$$

módulo n, donde $b = m_2^2 - m_1^3 \pmod n$, aunque no hay necesidad de calcular b.

5. Ana calcula eM en la curva E obteniendo $C = (c_1, c_2)$ y le envía a Benito el resultado C.

6. Benito para obtener el mensaje original M calcula $dC = M$.

Ciertos puntos deben ser estudiados antes de hablar de la seguridad del método:

- Las fórmulas para la ley de adición en la curva E no utilizan el valor b, por lo tanto no es necesario calcularlo, aunque no es difícil.

- Para calcular eM y dC se utilizan las fórmulas de la ley de grupo en una curva elíptica (en este caso en la curva E), con los cálculos hechos en módulo m. Nos encontraremos con expresiones como $\frac{y_2 - y_1}{x_2 - x_1}$, que se transformarán en enteros módulo n hallando el inverso multiplicativo de $(x_2 - x_1)$ módulo n. Para esto se requiere que $\text{mcd}(x_2 - x_1, n) = 1$. Notar que el mcd puede ser $1, p, q$ o n. Las posibilidades de que sea p o q son muy bajas. Si es n, entonces la pendiente es infinita y la suma (entendida como la operación de la ley de grupo) de esos puntos será O. Se seguirán las leyes para trabajar con O.

  Además, por el teorema chino de los restos, un entero módulo n puede ser considerado como una pareja de enteros, uno módulo p y el otro módulo q. Esto permite considerar un punto en $E(\mathbb{Z}_n)$ como un par de puntos, uno de E módulo p y el otro de E módulo q. Tenemos

$$E(\mathbb{Z}_n) = E(\mathbb{F}_p) \oplus E(\mathbb{F}_q). \tag{2.1}$$

- Utilizando la ecuación (2.1) se puede calcular el tamaño de $E(\mathbb{Z}_n)$, que es igual que $\#E(\mathbb{F}_p)\dot{\#}E(\mathbb{F}_q)$.

- Notar que el orden del grupo es independiente de b. Si Benito escoge una curva elíptica aleatoria $y^2 = x^3 + Ax + B$ definida sobre $\mathbb{Z}_n$, entonces tendrá que calcular el tamaño del grupo, calculándolo módulo p y módulo q. Esto es inviable si p y q son elegidos para hacer difícil la factorización de n. Además, si Benito fija la curva elíptica, Ana tendrá dificultades a la hora de encontrar puntos M en la curva. Si realiza el proceso de elegir primero la coordenada x como el mensaje, entonces al resolver $y^2 \equiv m^3 + Am + B$ (mód n) para y, se enfrenta al problema de calcular raíces cuadradas módulo n. Si Benito fija tan solo A y permite a Ana elegir B de tal manera que su punto esté en la curva, las elecciones de e y d requieren que el orden del grupo sea independiente de B. Esta es la situación en nuestro procedimiento.

Ahora podemos estudiar cómo un intruso podría tratar de romper el método y obtener el mensaje. Si el intruso es capaz de factorizar n como pq, entonces conocerá $(p+1)(q+1)$, lo que le permitirá hallar un d tal que $ed \equiv 1$ módulo $(p+1)(q+1)$. Si dispone de esta información, es capaz de descifrar el mensaje de Ana y obtener la información que se desea transmitir.

Supongamos ahora que el intruso no conoce la factorización de n, pero conoce el exponente para la desencriptación d. Es muy probable, por tanto, que sea capaz de factorizar n.

2.8.1. Ejemplo de un sistema criptográfico basado en la factorización

Ana desea enviar un mensaje a Benito, y quiere que la comunicación sea privada. Para ello seguirán el procedimiento que acabamos de ver en esta sección.

1. Benito elige como números primos $p = 659$ y $q = 809$. Notar que $p \equiv q \equiv 2$ módulo 3. Calcula ahora $n = 659 \cdot 809 = 533131$.

2. Toma $e = 49$ y $d = 16729$.

3. Benito hace públicos n y e. Mantendrá en secreto d, p y q.

Tenemos por tanto la clave pública establecida por Benito. Ahora, si Ana desea enviar a Benito el mensaje $M = (137, 2345)$ módulo 533131, realizará el siguiente proceso:

1. Considera el par $(137, 2345)$ como un punto $M = (137 : 2345 : 1)$ de la curva elíptica $E := y^2 = x^3 + 262017$.

2. Calcula $eM = C = (246301 : 523335 : 1)$, y le envía C a Benito.

En este punto Benito ya puede conocer el mensaje que Ana deseaba enviarle, para ello tan solo debe calcular $dC = (137 : 2345 : 1)$ y obtendrá M, como podemos ver.

2.9. Método de factorización con curvas elípticas

Vamos ahora a estudiar un método que, aunque no es un método criptográfico, es interesante, ya que ofrece un modo de buscar la factorización de un número y por tanto resulta útil en el ámbito de la criptografía. En particular, puede ayudarnos a prevenir ataques a la seguridad del método que hemos visto en la sección 2.8.

Estudiaremos el método de factorización con curvas elípticas desarrollado por Henrick Lenstra, cuyas siglas en inglés son ECM. Es una de las herramientas punteras en las técnicas de factorización y, por tanto, una de las formas de ataque a la seguridad de ciertos métodos criptográficos que se basan en la dificultad de factorizar un cierto número.

Podemos considerar este método una generalización del método Pollard $p-1$, por lo tanto es interesante comparar ambos. En primer lugar podemos describir cómo funciona el método Pollard $p-1$. Dado un entero n que deseamos factorizar, elegiremos un número entero a tal que $1 < a < n$, por lo tanto $\text{mcd}(a, n) < n$. Si a y n no son coprimos, entonces $\text{mcd}(a, n)$ es un factor no trivial de n y hemos tenido éxito.

Supongamos que $\text{mcd}(a, n) = 1$, luego se cumple que $a \in U_p$ para todo p que divida a n y además $a^{p-1} = 1$ en U_p. Esto último es consecuencia del teorema pequeño de Fermat. Luego p dividirá al $\text{mcd}(a^{p-1} - 1, n)$, y por lo tanto tenemos que $\text{mcd}(a^{p-1} - 1, n) > 1$. Para tratar de averiguar p realizaremos lo siguiente. Dado que se cumple que $a^{p-1} = 1$ en U_p se tiene que $a^k = 1$ en U_p para todo k divisible por $p - 1$, así que $\text{mcd}(a^{p-1} - 1, n) > 1$ implica que $\text{mcd}(a^k - 1, n) > 1$.

Esperamos, por tanto, que n sea divisible por un número primo p para el cual $p - 1$ sea el producto de pequeñas potencias de números primos pequeños. Si se cumple esto, simplemente bastará con probar pequeños valores para k, con la esperanza de encontrar un k que cumpla que $\text{mcd}(a^k - 1, n)$ sea no trivial. Si n es divisible por un número primo p para el cual $p - 1$ es el producto de pequeñas potencias de números primos pequeños, seremos capaces de hallar ese k.

Como hemos comentado antes, el método ECM es una generalización del método Pollard $p-1$. En lugar de tomar el grupo multiplicativo U_p, consideraremos una curva elíptica E y su conjunto de puntos $E(\mathbb{F}_p)$ sobre el cuerpo finito $\mathbb{F}_p$. Además no consideraremos un número $a \in U_n$ sino que consideraremos la curva E y su conjunto de puntos $E(\mathbb{Z}_n)$ y elegiremos un punto $P \in E(\mathbb{Z}_n)$. Ahora la condición $a^k = 1$ en U_p pasará a ser la condición $kP = O$ en $E(\mathbb{F}_p)$, con O el elemento identidad en el grupo. Buscamos valores de k que sean productos de potencias pequeñas

de números primos pequeños y tendremos éxito si hay un número primo p que divida a n y tal que $\#E(\mathbb{F}_p)$ sea el producto de pequeñas potencias de números primos pequeños.

Una gran diferencia entre estos métodos es que, en el caso del método Pollard $p-1$, tan solo habrá un grupo que estudiar asociado a p. Sin embargo, en el caso del método ECM, habrá un gran número de curvas elípticas E cuyos conjuntos $E(\mathbb{F}_p)$ son grupos finitos abelianos, lo que ofrece una gran variedad.

Podemos ahora, dado un número n, tratar de factorizarlo usando el algoritmo ECM de Lenstra. Para ello seguimos los siguientes pasos:

1. En primer lugar comprobamos que $\text{mcd}(n, 6) = 1$. Esto se hace para evitar tratar los casos de los factores 2 y 3, que complican bastante el tratamiento con curvas elípticas.

2. Comprobaremos que $n \neq m^k$. Para esto basta probar que las raíces $\sqrt{n}$, $\sqrt[3]{n}$, $\ldots, \sqrt[l]{n}$ no son enteros para $l = \lceil \frac{\ln n}{\ln 2} \rceil$.

3. Elegimos una cota B.

4. Elegimos una curva $E_{a,b}(\mathbb{Z}_n) : y^2 = x^3 + ax + b$ y un punto $P = (x : y : 1)$ de la siguiente manera:

   - Elegimos enteros aleatorios $x, y, a \in [0, n-1]$.
   - Calculamos $b = (y^2 - x^3 - ax)$ módulo n.
   - Calculamos $d = \text{mcd}(4a^3 + 27b^2, n)$. Si $d = n$, volvemos a empezar cambiando la elección de x, y, a. Si $1 < d < n$, d es un factor válido de n y hemos acabado. De otro modo $d = 1$, luego tenemos una pseudocurva elíptica en $\mathbb{Z}_n$ y un punto $P = (x : y : 1)$ en esa curva.

5. Calculamos las potencias de números primos más grandes que sean menores o iguales que la cota.

6. Suponemos que si $k = p_1^{a_1} p_2^{a_2} \ldots p_r^{a_r}$ para $p_1, p_2, \ldots, p_r$ números primos empezando desde 2, se cumplirá que $kP = O$. Calcularemos kP mediante el siguiente método:

   $$P \to 2P \to 4P \to \cdots \to p_r^{a_r}(p_{r-1}^{a_{r-1}} \ldots 3^{a_2} 2^{a_1} P) = kP.$$

   En cada duplicación esperamos hallar una pendiente que no pueda ser calculada por culpa de que el mcd del denominador y n sea uno. Si esto ocurre, seguimos con las operaciones. Sin embargo, si el mcd es un divisor propio de n devolvemos el factor como solución, y si el mcd es n simplemente aumentamos la cota B o probamos con otra curva.

2.10. La criptografía Post-Cuántica

Como hemos dicho, los métodos criptográficos basados en curvas elípticas ofrecen algunas ventajas respecto a los métodos criptográficos tradicionales, como puede ser el ahorro de espacio o la mayor velocidad de generación de las claves. Sin embargo siguen sin solucionar uno de los problemas que existen en la criptografía, el desarrollo de la computación cuántica.

En primer lugar, vamos a explicar muy brevemente el concepto de la computación cuántica. Los ordenadores que utilizamos a día de hoy almacenan los datos en conjuntos de bits, que no son otra cosa que cadenas de ceros y unos.

La computación cuántica, sin embargo, se basa en la física de la mecánica cuántica y, en lugar del bit, se utilizan los cubit. Supongamos que tenemos un ordenador cuya capacidad son 3 bits y por tanto habrá $2^3 = 8$ posibles estados de memoria. Si pensamos ahora en un ordenador cuya capacidad sea de 3 cubits, su memoria podrá estar en una superposición de los 8 posibles estados de memoria que los 3 bits ofrecían. Con esto queremos decir que el estado de nuestro ordenador cuántico de 3 cubits se expresa como un punto en la esfera unidad compleja 8 dimensional, es decir el estado tendrá la forma $(b_0, \ldots, b_7)$ donde cada $b_i \in \mathbb{C}$ y se cumple $\sum_{i=0}^{7} |b_i|^2 = 1$. Luego cada b_i corresponde a uno de los posibles 8 estados que los 3 bits ofrecen, y $|b_i|^2$ corresponde a la probabilidad de estar en el estado i-ésimo.

Como se puede intuir, la diferencia de capacidad entre el ordenador de 3 bits y el de 3 cubits es enorme. Esto nos lleva al porqué la computación cuántica podría suponer un problema para los métodos de cifrado actuales. La potencia de computación de un ordenador cuántico con tantos cubits como bits tiene un ordenador actual sería inmensa siempre que existan algoritmos cuánticos para el problema en cuestión. Esto es así tanto para la factorización de enteros como para el problema del logaritmo discreto, lo cual supone, por ejemplo, que se podrían descifrar mensajes e información cifrada mediante los métodos que basan su seguridad en la dificultad de resolver el problema del logaritmo discreto, o en la dificultad de factorizar un número entero, como el método RSA. Un ordenador cuántico sería capaz de resolver estas cuestiones en un tiempo razonable, debido a los algoritmos cuánticos y su potencia, y por lo tanto algunos de los métodos criptográficos conocidos quedarían inutilizados.

Sin embargo, a día de hoy, la computación cuántica es todavía una utopía y aunque se han desarrollado algunos ordenadores cuánticos, su potencia es menor que la de los ordenadores clásicos actuales. Esto no significa que no deba tenerse en cuenta este problema y ya se trabaja en el desarrollo de métodos criptográficos que lo solucionen.

Capítulo 3

Dos ejemplos de curvas elípticas utilizadas en métodos criptográficos

Vamos a estudiar ahora un par de curvas elípticas que tienen aplicaciones reales muy usadas. La primera que veremos será la **curva 25519**, más conocida como "la curva de WhatsApp", dado que en ella se basa la criptografía de esta plataforma; con un enfoque mucho más informático, en [6] se puede ver más información sobre el uso de esta curva. La segunda es la **curva Secp256k1,** que es la curva en la que se basa Bitcoin para firmar mensajes mediante el algoritmo ECDSA visto en la sección 2.7.

3.1. Curva 25519. La curva de WhatsApp

En esta sección estudiaremos brevemente uno de los casos de uso de la criptografía de curvas elípticas más conocidos. Nos referimos a la criptografía que hay detrás de las conversaciones y llamadas que se realizan a través de la aplicación WhatsApp, que a día de hoy está presente en el día a día de mucha gente.

En el año 2016 se inició en WhatsApp un proceso de cifrado de todos los mensajes enviados entre sus usuarios para que tan solo ellos fueran capaces de descifrarlos. Como podemos ver en la figura 3.1, al iniciar una nueva conversación en WhatsApp se nos indica que las llamadas y los mensajes están seguros con cifrado de extremo a extremo. Vamos a estudiar el método utilizado para lograr este cifrado.

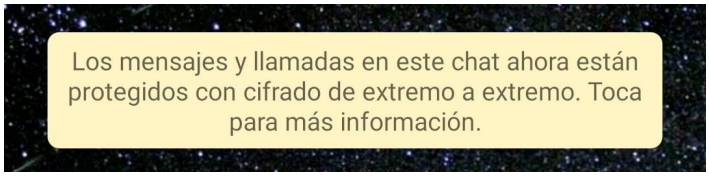

Figura 3.1: Captura del mensaje que aparece al iniciar una nueva conversación en WhatsApp.

3.1.1. La curva 25519

El método utilizado para cifrar los mensajes en WhatsApp utiliza una curva elíptica, la conocida como **curva 25519**. El nombre hace referencia al cuerpo finito sobre el que se define la curva elíptica, que es $\mathbb{F}_{2^{255}-19}$. Esta curva pertenece a la familia de curvas de Montgomery sobre $\mathbb{Z}_p$, que son curvas elípticas sobre $\mathbb{Z}_p$, con $p \geq 5$ un número primo, dadas por la conocida como ecuación de Montgomery

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

para $A \in \mathbb{Z}_p$ tal que $A^2 - 4$ no es resto cuadrático módulo p y el discriminante de la curva es $\Delta = A^2 - 4 \neq 0$.

En las curvas definidas mediante la ecuación de Montgomery se cumple el siguiente teorema:

Teorema 3.1. *Sea $p \geq 5$ un número primo, y A un entero tal que $A^2 - 4$ no es un resto cuadrático módulo p. Dada E la curva elíptica definida como $y^2 = x^3 + Ax^2 + x$ sobre el cuerpo finito $\mathbb{F}_p$, definimos la función $X_0 : E(\mathbb{F}_{p^2}) \longrightarrow \mathbb{F}_{p^2}$ de la siguiente manera: $X_0(O) = 0$ y $X_0((x : y : 1)) = x$. Sea, así mismo, n entero y q elemento del cuerpo $\mathbb{F}_p$. Entonces, existe $s \in \mathbb{F}_p$ tal que $X_0(nQ) = s$ para todo $Q \in E(\mathbb{F}_{p^2})$ tal que $X_0(Q) = q$.*

Su demostración está hecha en [2]. Ésta es una de las razones por las cuales se utiliza una curva elíptica definida mediante una ecuación de Montgomery. Veremos que, gracias al teorema 3.1, podemos definir una función llamada X25519 que usaremos para cifrar los mensajes.

Definimos la curva 25519 (a la que denotaremos como E) como el conjunto de puntos $(x, y) \in \mathbb{Z}_{2^{255}-19} \times \mathbb{Z}_{2^{255}-19}$ que satisfacen la ecuación

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

El logaritmo discreto es difícil de resolver para estos parámetros. Respecto al método utilizado para cifrar los mensajes, se trata del método de intercambio de

claves de Diffie-Hellman, visto en la sección 2.3. Por lo tanto se debe tomar un punto base $Q \in E(\mathbb{Z}_{2^{255}-19})$ que puede ser público. Se utiliza $Q = (9 : 39420360 : 1)$. Este punto base tiene orden cercano a $2^{255} - 19$, es decir genera la gran mayoría de los puntos de $E(\mathbb{Z}_{2^{255}-19})$.

3.1.2. La función X25519

Para entender la función X25519 es importante recordar que la curva 25519 quedaba definida mediante una ecuación de Montgomery. Además utilizaremos el teorema 3.1.

Ahora para el caso de la curva 25519 tomamos $p = 2^{255} - 19$. Definimos el cuerpo $\mathbb{F}_p = \mathbb{Z}_{2^{255}-19}$, notar que 2 no es un cuadrado en $\mathbb{F}_p$. Definimos $\mathbb{F}_{p^2}$ como el cuerpo $\mathbb{Z}_{2^{255}-19}[\sqrt{2}]$. Tomamos $A = 486662$ que cumple que $A^2 - 4$ no es un cuadrado en $\mathbb{F}_p$. Definimos $E := y^2 = x^3 + Ax^2 + x$ como la curva elíptica sobre F_p. Definimos la función

$$X_0 : E(\mathbb{F}_{p^2}) \longrightarrow \mathbb{F}_{p^2}$$

mediante $X_0(O) = 0$ y $X_0((x : y : 1)) = x$. Por último definimos la función

$$X : E(\mathbb{F}_{p^2}) \longrightarrow \{\infty\} \cup \mathbb{F}_{p^2}$$

mediante $X(O) = \infty$ y $X((x : y : 1)) = x$.

Dados $n \in 2^{254} + 8\{0, 1, 2, \ldots, 2^{251} - 1\}$ y $q \in \mathbb{F}_p$, según el teorema 3.1 la función X25519 produce un único $s \in \mathbb{F}_p$ tal que $X_0(nQ) = s$ para todo $Q \in E(\mathbb{F}_{p^2})$ tal que $X_0(Q) = q$. Definiremos las entradas y salidas de la función X25519 como secuencias de bytes. El conjunto de bytes será, por definición, $\{0, 1, \ldots, 255\}$. Ahora, para cada entero $s \in \{0, \ldots, 2^{256} - 1\}$ definimos

$$\underline{s} = (s \bmod 256, \lfloor s/256 \rfloor \bmod 256, \ldots, \lfloor s/(256^{31}) \rfloor \bmod 256).$$

El conjunto de las claves públicas de la curva 25519 es

$$\{\underline{q} : q \in \{0, 1, \ldots, 2^{256} - 1\}\}$$

y el conjunto de las claves privadas es

$$\{\underline{n} : n \in 2^{254} + 8\{0, 1, 2, \ldots, 2^{251} - 1\}\}.$$

Ahora definimos la función X25519 de la siguiente manera:

$$\text{X25519} : \{Cprivadas\} \times \{Cpublicas\} \longrightarrow \{Cpublicas\},$$

siendo $\{Cprivadas\}$ y $\{Cpublicas\}$ los conjuntos de claves privadas y públicas respectivamente. Fijados q y n tales que

$$q \in \{0, 1, \ldots, 2^{256} - 1\}, \qquad n \in 2^{254} + 8\{0, 1, 2, \ldots, 2^{251} - 1\},$$

por el teorema 3.1 habrá un entero $s \in \{0, 1, 2, \ldots, 2^{255} - 20\}$ que cumple que $s = X_0(nQ)$ para todo $Q \in E(\mathbb{F}_{p^2})$ tal que $X_o(Q) = q$ mód $2^{255} - 19$. Por lo tanto, X25519$(\underline{n}, \underline{q})$ se define como $\underline{s}$.

3.1.3. Protocolo simplificado de cifrado

Veamos cómo funciona el protocolo simplificado de cifrado con la curva 25519 descrito en [1].

Usaremos la siguiente notación: para $a, b \in \mathbb{Z}_{2^{255}-19}$ números de 256 bits, el punto $P \in E$ será un punto de coordenada b en el eje de las x y la función X25519(a, b) devolverá la coordenada x del punto aP.

Para generar las claves cada usuario de WhatsApp posee un procedimiento para hallar números aleatorios de 256 bits. Una vez hallado un número a de 256 bits, éste será la clave privada y la clave pública asociada a a es X25519$(a, 9)$.

Al instalar WhatsApp se generan multitud de pares de claves y se envían las claves públicas al servidor de WhatsApp. La primera de las claves públicas es la clave de identificación y el resto son las llamadas claves de un solo uso que son firmadas con la clave de identificación y se reponen cuando es necesario. El par de las claves de identificación será denotado como $I_A = (I_{Apriv}, I_{Apb})$, y los pares de las claves de un solo uso se denotan como $U_A = (U_{Apriv}, U_{Apb})$. Veamos el método de cifrado en una conversación:

1. Si A desea iniciar una conversación con B se solicita al servidor las claves públicas de B, la de identificación I_{Bpb} y una de un solo uso U_{Bpb}. La clave de un solo uso U_{Bpb} se borra entonces del servidor y A genera unas claves efímeras $E_A = (E_{Apriv}, E_{Apb})$.

2. A calcula una clave K_1 de tipo AES-256 de la siguiente manera. Primero calcula X25519(E_{Apriv}, U_{Bpb}) y después aplica la función hash SHA al valor obtenido, es decir,
$$K_1 = \text{SHA}(\text{X25519}(E_{Apriv}, U_{Bpb})).$$

3. A cifra entonces el mensaje a enviar m con el método AES-256 utilizando la clave K_1 y envía a B, en la cabecera del mensaje, E_{Apb} y I_{Apb}.

4. Ahora B puede hallar K_1 de sus claves privadas y las claves enviadas en la cabecera del mensaje. Una vez obtenga K_1 puede obtener m.

A partir de esta primera clave establecida entre A y B, para los siguientes mensajes crearán nuevas claves K_i, donde $K_i = \text{SHA}(K_{i-1})$.

Nota 3.1. Hemos utilizado en este método la función SHA, que es una de las funciones hash de las que hablamos en la sección 2.6, desarrollada por el "NIST" (National Institute of Standards and Technology).

3.1.4. Propiedades

Por último, vamos a comentar qué propiedades y ventajas ofrece el uso de la curva 25519 para cifrar los mensajes. Como se observa en [2] tenemos las siguientes propiedades:

- **Alta seguridad.** El tamaño del cuerpo primo utilizado es de 2^{255} lo cual hace que el problema del logaritmo discreto sea difícil de resolver para la curva 25519. Es decir, la seguridad del método es alta.
- **Alta velocidad de computación.** La curva 25519 permite obtener una alta velocidad de computación sin disminuir el nivel de seguridad. Esto se debe en parte a la elección de la curva en forma Montgomery y a la elección del cuerpo $\mathbb{F}_{2^{255}-19}$.
- **Claves pequeñas.** Las claves públicas y privadas son de 32 bytes, ya que la clave privada es un elemento de $\mathbb{F}_{2^{255}-19}$ y, gracias al teorema 3.1, la clave pública no es un punto de la curva sino que es la coordenada x de un punto $P = (x : y : 1)$ de la curva, luego también es un entero de 32 bytes.
- **No hay validación de las claves.** Toda cadena de 32 bytes es admitida como una clave pública en la curva 25519. En otros criptosistemas basados en el método de intercambio de claves de Diffie-Hellman se validan las claves públicas.

Como podemos observar, la curva 25519 es adecuada para cifrar los mensajes de WhatsApp, pues ofrece velocidad y seguridad además de que no ocupa demasiado espacio en memoria. Esto último es importante pues WhatsApp es una aplicación desarrollada para dispositivos móviles y éstos a veces carecen de espacio de almacenamiento.

3.2. Curva Secp256k1

Veamos ahora otra curva elíptica. Se trata de la curva Secp256k1. Esta curva es utilizada por Bitcoin para firmar mensajes como, por ejemplo, las transferencias de la moneda virtual entre los usuarios de Bitcoin. Para firmar los mensajes Bitcoin utiliza el algoritmo ECDSA visto en la sección 2.7 usando esta curva. El nombre de la curva hace referencia al "SEC" (Standards for Efficient Cryptography), documento donde se define la curva, y el tamaño del cuerpo finito sobre el que se define la curva que es $\mathbb{F}_p$ con $p = 2^{256} - 2^{32} - 2^9 - 2^8 - 2^7 - 2^6 - 2^4 - 1$.

La curva se define mediante la ecuación

$$y^2 = x^3 + 7 \tag{3.1}$$

sobre el cuerpo finito $\mathbb{F}_{2^{256}-2^{32}-997}$ (notar que el número 997 son el resto de sumandos del primo p definido antes). Como podemos observar, la ecuación de la curva (3.1) es sencilla; sin embargo, el cuerpo finito sobre el que está definida es $F_{2^{256}-2^{32}-997}$, lo cual hace que la curva ofrezca seguridad para firmar los mensajes pues el logaritmo discreto es difícil de resolver con estos parámetros.

Dado que la curva Secp256k1 se utiliza para firmar mensajes mediante el algoritmo ECDSA necesitaremos utilizar un punto base, y en este caso se utiliza como punto base $G = (x : y : 1)$. Siguiendo la expresión de las coordenadas x e y que aparece en [3] utilizaremos sus formas hexadecimales que son

$$x = 79\text{BE}667\text{EF}9\text{DCBBAC}55\text{A}06295\text{CE}870\text{B}07 \\ 029\text{BFCDB}2\text{DCE}28\text{D}959\text{F}2815\text{B}16\text{F}81798$$

e

$$y = 483\text{ADA}7726\text{A}3\text{C}4655\text{DA}4\text{FBFC}0\text{E}1108\text{A}8 \\ \text{FD}17\text{B}448\text{A}68554199\text{C}47\text{D}08\text{FFB}10\text{D}4\text{B}8.$$

El orden n del punto G es

$$n = \text{FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF} \\ \text{EBAAEDCE}6\text{AF}48\text{A}03\text{BBFD}25\text{E}8\text{CD}0364141.$$

De forma casi anecdótica, y aunque realmente la curva Secp256k1 esta definida sobre el cuerpo finito $\mathbb{F}_{2^{256}-2^{32}-997}$, en la figura 3.2 podemos observar una representación de diche curva en el cuerpo de los números reales.

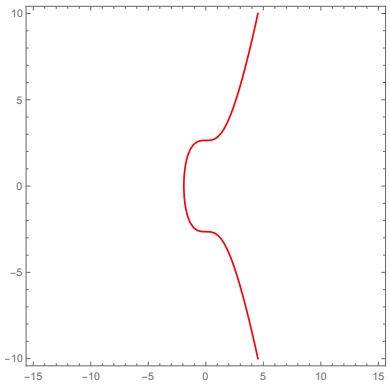

Figura 3.2: Representación de la curva Secp256k1 en el cuerpo de los reales.

3.2.1. Algunas características de la curva Secp256k1

La curva Secp256k1 es una de las curvas de Koblitz definidas por el SECG (Standards for Efficient Curves Group), como podemos ver en [7]. Bitcoin utiliza esta curva de Koblitz para verificar quiénes son los propietarios de las criptomonedas, entre otras cosas. Las curvas de Koblitz son curvas elípticas definidas mediante la ecuación

$$y^2 = x^3 + ax + b$$

sobre un cuerpo finito $\mathbb{F}_p$. En el caso de la curva Secp256k1, como ya hemos visto, $a = 0$ y $b = 7$.

Pero, ¿por qué utilizar la curva Secp256k1? Como se indica en [5], la curva se construye de una manera no aleatoria, lo que permite obtener una mayor eficiencia en la computación. Si la implementación está lo suficientemente optimizada se consigue una velocidad de computación un 30 % mayor a la obtenida mediante otras curvas.

Esto es muy importante para Bitcoin. Dado que utilizan el algoritmo ECDSA basado en la curva Secp256k1 para firmar los mensajes de las transferencias de las criptomonedas, es importante que se firmen de una manera rápida, para lograr que las transferencias sean casi instantáneas.

Otra de las ventajas de usar la curva Secp256k1 es que en el conjunto

$$\text{Secp256k1}(\mathbb{F}_{2^{256}-2^{32}-997})$$

es difícil resolver el problema del logaritmo discreto y por tanto el uso de la curva Secp256k1 ofrece seguridad en las transferencias de la criptomoneda.

Como curiosidad, la curva Secp256k1 no había sido casi nunca utilizada hasta que Bitcoin comenzó a usarla. Tras el uso de la curva por parte de Bitcoin, dicha curva ha comenzado a ser más utilizada y su popularidad ha aumentado considerablemente.

Capítulo 4

Conclusiones

Gracias a este Trabajo Fin de Grado he sido capaz de apreciar varios aspectos de las matemáticas en los que no había reparado antes. En primer lugar he observado lo importante que es relacionar las ideas y los conocimientos que se poseen. Como se puede ver, en el trabajo se relacionan dos ideas que a priori no tienen por qué ir ligadas, como lo son las curvas elípticas y la criptografía. De esta unión de conocimientos se llega al desarrollo de nuevos métodos criptográficos que ofrecen interesantes posibilidades.

También he podido estudiar cómo se pasa de la teoría matemática a su aplicación. Considero que esto es muy importante pues muchas veces conocemos la teoría y no sabemos cómo aplicarla para resolver problemas del mundo real. En este caso he estudiado las curvas elípticas y resultados referidos a ellas y después, gracias a estos resultados, he podido entender los métodos criptográficos que se basan en curvas elípticas y que se utilizan en el mundo real, como por ejemplo en la aplicación WhatsApp.

El último de los aspectos que quiero resaltar es la presencia de las matemáticas en nuestro día a día, muchas veces sin ser nosotros conscientes de ello. Al estudiar el ejemplo de la curva 25519 he sido consciente de que todos los días, al enviar un mensaje mediante la plataforma WhatsApp, utilizo la criptografía basada en curvas elípticas. Esto me ha hecho reflexionar sobre la gran cantidad de veces que las matemáticas facilitan nuestro día a día, y lo importante que es seguir investigando e invirtiendo recursos en la ciencia.

Por último, quiero agradecer a mi tutor por toda la ayuda que me ha dado durante la realización de este trabajo, y a los profesores que me han enseñado a entender y utilizar las matemáticas.

Bibliografía

[1] Irene Ayerra Balduz: *Criptografía y curvas elípticas. La curva WhatsApp*, Trabajo Fin de Grado, Universidad de Zaragoza, 2018.

[2] Daniel J. Bernstein: Curve25519: new Diffie-Hellman speed records, 207–228, *Public Key Cryptography - PKC 2006 (International Workshop on Public Key Cryptography)*, Lecture Notes in Computer Science, Springer, 2016.

[3] John D. Cook: Bitcoin key mechanism and elliptic curves over finite fields, `https://www.johndcook.com/blog/2018/08/14/bitcoin-elliptic-curves/`, 2018.

[4] José Luis Gómez Pardo: Criptografía y curvas elípticas, *La Gaceta de la RSME* **5** (2002), no. 3, 737–777.

[5] Azine Houria, Bencherif Mohamed Abdelkader y Guessoum Abderezzak: A comparison between the secp256r1 and the Koblitz secp256k1 bitcoin curves, *Indonesian Journal of Electrical Engineering and Computer Science* **13** (2019), no. 3, 910–918.

[6] Antonio Juano Ayllón: *Criptografía y Seguridad en WhatsApp*, Trabajo Fin de Máster, Universidad Nacional de Educación a Distancia (UNED), Zaragoza, 2016.

[7] Santoshi Pote y Virendra Sule y B. K. Lande: Arithmetic of Koblitz Curve Secp256k1 used in Bitcoin Cryptocurrency based on One Variable Polynomial Division, 5 pp., *2nd International Conference on Advances in Science Technology (ICAST)*, K. J. Somaiya Institute of Engineering and Information Technology, University of Mumbai, Maharashtra, India, 2019.

[8] Thomas R. Shemanske: *Modern Cryptography and Elliptic Curves. A Beginner's Guide*, American Mathematical Society, Providence (Rhode Island), 2017.

[9] Juan Luis Varona Malumbres: *Recorridos por la teoría de números*, segunda edición, Colección "Textos Universitarios", RSME-Electolibris, 2019.

[10] Lawrence C. Washington: *Elliptic curves: number theory and cryptography*, Chapman & Hall CRC, Boca Raton (Florida), 2003.