Ministry of Higher Education and Scientific Research
Salahaddin University-Erbil
College of Science - Department of Mathematics

Final Year Research Project

Elliptic Curve Cryptography

Submitted by
Runas Kawther Aziz

Supervisor
Dr. Andam Ali Mustafa

A Project submitted in partial fulfillment of the requirements for the Degree of B.Sc. in Mathematics

Abstract

Elliptic Curve Cryptography (ECC) is a public key cryptographic system that uses the properties of elliptic curves over finite fields to provide security. It is a relatively new cryptographic technique that has gained popularity due to its efficiency, security, and flexibility compared to other public key systems such as RSA.

ECC involves two keys, a public key and a private key, and uses complex mathematical operations to encrypt and decrypt messages. The security of ECC is based on the difficulty of solving the discrete logarithm problem in the elliptic curve group, which is believed to be computationally infeasible.

In summary, ECC is a powerful cryptographic technique that offers efficient and secure public key encryption. It has become a popular choice in modern applications due to its many advantages over traditional public key systems, and its continued development is expected to have a significant impact on the field of cryptography. This study focuses simply on ECC and some of its component outcomes.

Keywords: Elliptic Curve, Cryptography, Encryption, Decryption.

1. Introduction

Elliptic curve cryptography (ECC) is a public-key encryption method that uses elliptic curves to establish a secure communication channel between two parties. It is considered to be more efficient than traditional methods, such as RSA, while offering the same level of security.

In ECC, a private key is generated by multiplying a random number, known as the private key, with a point on an elliptic curve. The resulting point on the curve is used as the public key, which can be shared with anyone who needs to communicate with the owner of the private key. The strength of the security in ECC is based on the difficulty of calculating the private key from the public key.

One of the key advantages of ECC is its efficiency. Compared to RSA, ECC requires smaller key sizes to provide the same level of security. For example, a 256-bit ECC key provides the same security as a 3072-bit RSA key. This makes ECC a popular choice for mobile devices and other systems with limited resources [2].

Another advantage of ECC is its resistance to attacks based on quantum computing. While traditional encryption methods, such as RSA, can be easily broken by quantum computers, ECC has been shown to be resistant to such attacks.

Despite its advantages, there are some potential weaknesses in ECC that must be considered. One such weakness is the possibility of a side-channel attack, in which an attacker can use information about the physical properties of the system to extract the private key. Additionally, the security of ECC depends on the selection of the elliptic curve used, and some curves have been found to be vulnerable to attack.

Overall, ECC is a promising method of encryption that offers a high level of security with greater efficiency than traditional methods. However, it is important to carefully consider the potential weaknesses and vulnerabilities when implementing ECC in a system [6].

There are several encryption algorithms that can be used in elliptic curve cryptography (ECC), including:

1) Elliptic Curve Diffie-Hellman (ECDH): This algorithm is used for key exchange between two parties. It involves each party generating a private key and a public key based on an elliptic curve. The parties then exchange their public keys and use them to generate a shared secret key that can be used for encryption and decryption.

2) Elliptic Curve Integrated Encryption Scheme (ECIES): This algorithm is used for encrypting data between two parties. It involves using the ECDH algorithm to generate a shared secret key, which is then used to encrypt the data using a symmetric encryption algorithm, such as AES. The encrypted data is then sent to the recipient, who can use their private key to decrypt it.

3) Elliptic Curve Digital Signature Algorithm (ECDSA): This algorithm is used for verifying the authenticity of digital signatures. It involves the use of a private key to generate a digital signature, which can be verified using the corresponding public key. The algorithm ensures that the signature is only valid for the specific message it was created for and that it cannot be forged or tampered with.

All of these algorithms are based on the properties of elliptic curves and are designed to provide a high level of security while being efficient in terms of computation and communication overhead. They are widely used in various applications, such as secure communication protocols, digital signatures, and mobile device security [1].

1.1 Elliptic Curves

First and foremost, elliptic curves have nothing to do with ellipses. Ellipses are formed by quadratic curves. Elliptic curves are always cubic. [Note: Elliptic curves are called elliptic because of their relationship to elliptic integrals in mathematics. An elliptic integral can be used to determine the arc length of an ellipse.]. The simplest possible "curves" are, of course, straight lines. The next simplest possible curves are conics, these being quadratic forms of the following sort

$$ax^2 + bxy + cy^2 + dx + ey + f = 0$$

If $b^2 - 4ac$ is less than 0, then the curve is either an ellipse, or a circle, or a point, or the curve does not exist; if it is equal to 0, then we have either a parabola, or two parallel lines, or no curve at all; if it is greater than 0, then we either have a hyperbola or two intersecting lines. (Note that, by definition, a conic is the intersection of a plane and a cone.) - The next simplest possible curves are elliptic curves. An elliptic curve in its "standard form" is described by

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

for some fixed values for the parameters a and b. This equation is also referred to as Weierstrass Equation of characteristic 0. (The equation shown involves multiplications and additions over certain objects that are represented by x, y, a, and b. The values that these objects acquire are

meant to be drawn from a set that must at least be a ring. The characteristic of a ring is the number of times you must add the multiplicative identity element in order to get the additive identity element. If adding the multiplicative identity element to itself, no matter how many times, never gives us the additive identity element, we say the characteristic is 0. Otherwise, there must exist an integer p such that $p \times n = 0$ for all n. The value of p is then the characteristic of the ring. In a ring of characteristic 2, the elements 2,4, etc., are all equal to 0. In a ring of characteristic 3, the elements 3,6, etc., are all equal to 0.) Elliptic curves have a rich structure that can be put to use for cryptography [5].

Figure 1 shows some elliptic curves for a set of parameters (a, b). The top four curves all look smooth (they do not have cusps, for example) because they all satisfy the following condition on the discriminant of the polynomial $f(x) = x^3 + ax + b$:

$$4a^3 + 27b^2 \neq 0 \qquad (1)$$

[Note: The discriminant of a polynomial is the product of the squares of the differences of the polynomial roots. The roots of the polynomial $f(x) = x^3 + ax + b$ are obtained by solving the equation $x^3 + ax + b = 0$. Since this is a cubic polynomial, it will in general have three roots. Let's call them r_1, r_2, and r_3. Its discriminant will therefore be

$$D_3 = \prod_{i<j}^{3} (r_i - r_j)^2$$

which is the same as $(r_1 - r_2)^2(r_1 - r_3)^2(r_2 - r_3)^2$. It can be shown that when the polynomial is $x^3 + ax + b$, the discriminant reduces to

$$D_3 = -16(4a^3 + 27b^2)$$

This discriminant must not become zero for an elliptic curve polynomial $x^3 + ax + b$ to possess three distinct roots. If the discriminant is zero, that would imply that two or more roots have coalesced, giving the curve a cusp or some other form of nonsmoothness. Non-smooth curves are singular. It is not safe to use singular curves for cryptography.

The bottom two examples in Figure 1 show two elliptic curves for which the condition on the discriminant is violated. For the one on the left that corresponds to $f(x) = x^3$, all three roots of the cubic polynomial have coalesced into a single point and we get a cusp at that point. For the one on the right that corresponds to $f(x) = x^3 - 3x + 2$, two of the roots have coalesced into the point where the curve crosses itself. These two curves are singular. As mentioned earlier, it is not safe to use singular curves for cryptography. Note that since we can

write $y = \pm\sqrt{x^3 + ax + b}$ elliptic curves in their standard form will be symmetric about the x-axis. It is difficult to comprehend the structure of the curves that involve polynomials of degree greater than 3. To give the reader a taste of the parameters used in elliptic curves meant for real security, here is an example:

$$y^2 = x^3 \quad +3176890812513255034763174764138276932727469559 27x \\ +79052896607878758718120572025718535432100651934$$

This elliptic curve is used in the Microsoft Windows Media Digital Rights Management Version 2 [3].

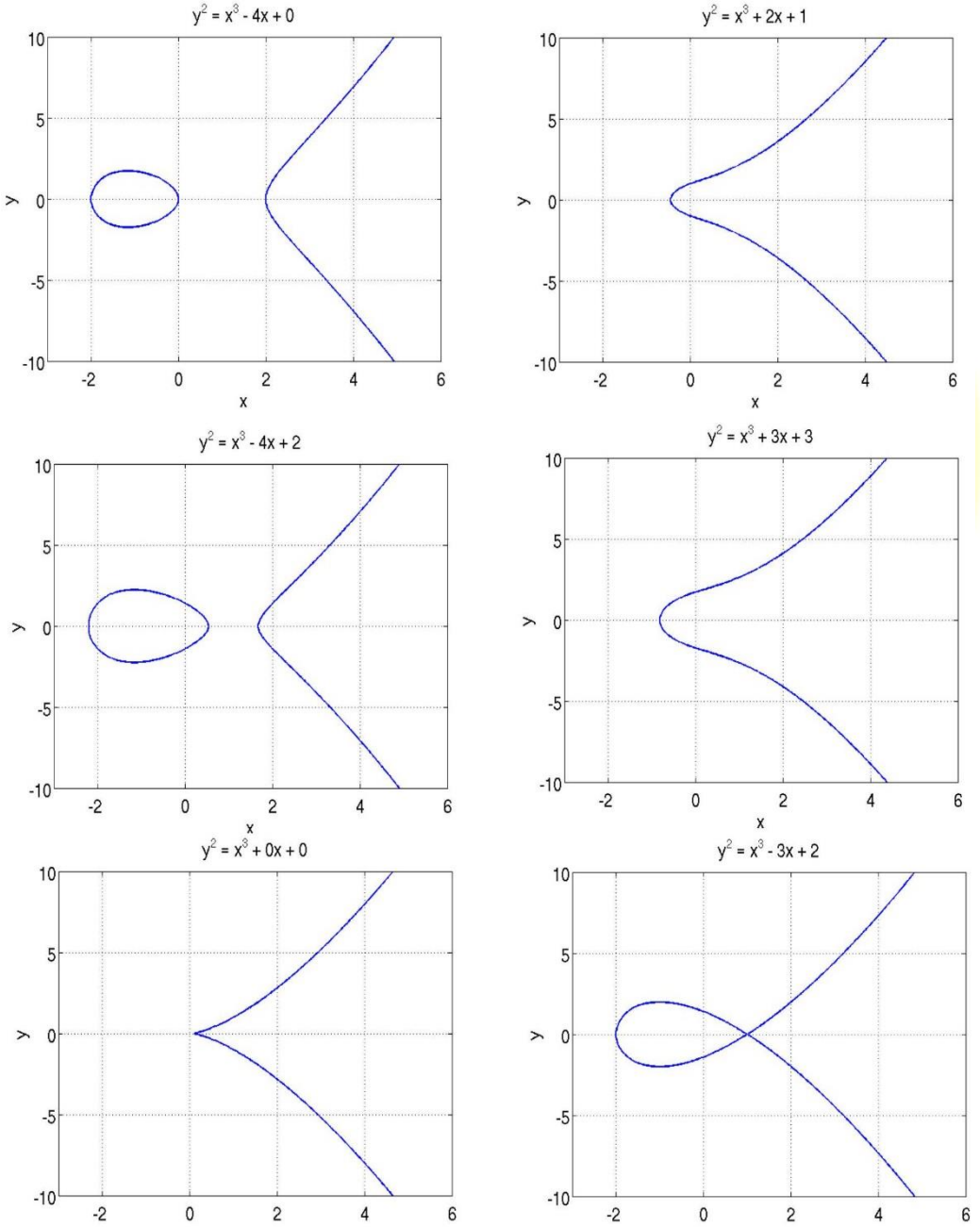

Figure 1: Elliptic curves for different values of the parameters a and b.

2. Materials and Methods

2.1 A Group Operator Defined for Points on an Elliptic Curve:

The points on an elliptic curve can be shown to constitute a group. Recall that a group needs the following:

1) a group operator;
2) an identity element with respect to the operator;
3) closure and associativity with respect to the operator;
4) the existence of inverses with respect to the operator.

The group operator for the points on an elliptic curve is, by convention, called addition. Its definition has nothing to do with the conventional arithmetic addition. To add a point P on an elliptic curve to another point Q on the same curve, we use the following rule:

We first join P with Q with a straight line. The third point of the intersection of this straight line with the curve, if such an intersection exists, is denoted R. The mirror image of this point with respect to the x-coordinate is the point $P + Q$. If the third point of intersection does not exist, we say it is at infinity [7].

The upper two curves in Figure 2 illustrate the addition operation for two different elliptic curves. The values for a and b for the upper curve at the left are -4 and 0, respectively. The values for the same two constants for the upper curve on the right are 2 and 1, respectively. But what happens when the intersection of P and Q is at infinity? We denote the point at infinity by the special symbol O and we then show that this can serve as the additive identity element for the group operator. We now stipulate that that $P + \mathrm{O} = P$ for any point on the curve.

We define the additive inverse of a point P as its mirror reflection with respect to the x coordinate. So, if Q on the curve is the mirror reflection of P on the curve, then $Q = -P$. For any such two points, it would obviously be the case that the third point of intersection will be at infinity. That is, the third point of intersection will be the distinguished point O.

We will further stipulate that that $\mathrm{O} + \mathrm{O} = \mathrm{O}$, implying that $-\mathrm{O} = \mathrm{O}$. Therefore, the mirror reflection of the point at infinity is the same point at infinity. Now we can go back to the issue of what happens to $P + Q$ when the intersection of two points P and Q is at infinity, as would be the case when P and Q are eath other's mirror reflections with regard to the x-axis. Obviously, in this case, the intersection of P and Q is at the distinguished point O, whose mirror reflection

is also at O. Therefore, for such points, $P + Q = O$ and $Q = -P$. We have already defined the additive inverse of a point P as its mirror reflection about the x-axis.

What is the additive inverse of a point where the tangent is parallel to the y-axis? The additive inverse of such a point is the point itself. That is, if the tangent at P is parallel to the y-axis, then $P + P = O$. - In general, what does it mean to add P to itself? To see what it means, let's consider two distinct points P and Q and let Q approach P. The line joining P and Q will obviously become a tangent at P in the limit. Therefore, the operation $P + P$ means that we must draw a tangent at P, find the intersection of the tangent with the curve, and then take the mirror reflection of the intersection [1].

Obviously, if the tangent at P intersects the curve at infinity (as would be the case when a line parallel to the y-axis is tangent to the curve), meaning at the distinguished point O, then $P + P = O$. Such a P would be its own inverse. For an elliptic curve

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

we define the set of all points on the curve along with the distinguished point O by $E(a, b)$. $E(a, b)$ is a group with the "addition" operator. $E(a, b)$ is obviously closed with respect to the addition operation.

We can also show geometrically that the property of associativity is satisfied. Every element in the set obviously has its additive inverse in the set. Since the operation of "addition" is commutative, $E(a, b)$ is an abelian group. Just for notational convenience, we now define multiplication on this group as repeated addition. Therefore,

$$k \times P = P + P + \cdots + P$$

with P making k appearances on the right. Therefore, we can express $P + P$ as $2P, P + P + P$ as $3P$, and so on. The two curves at the bottom in Figure 2 show us calculating $2P$ and $3P$ for a given P. The values of a and b for the lower curve on the left are -4 and 2, respectively. The values for the same two constants for the lower curve on the right are both 3 [9].

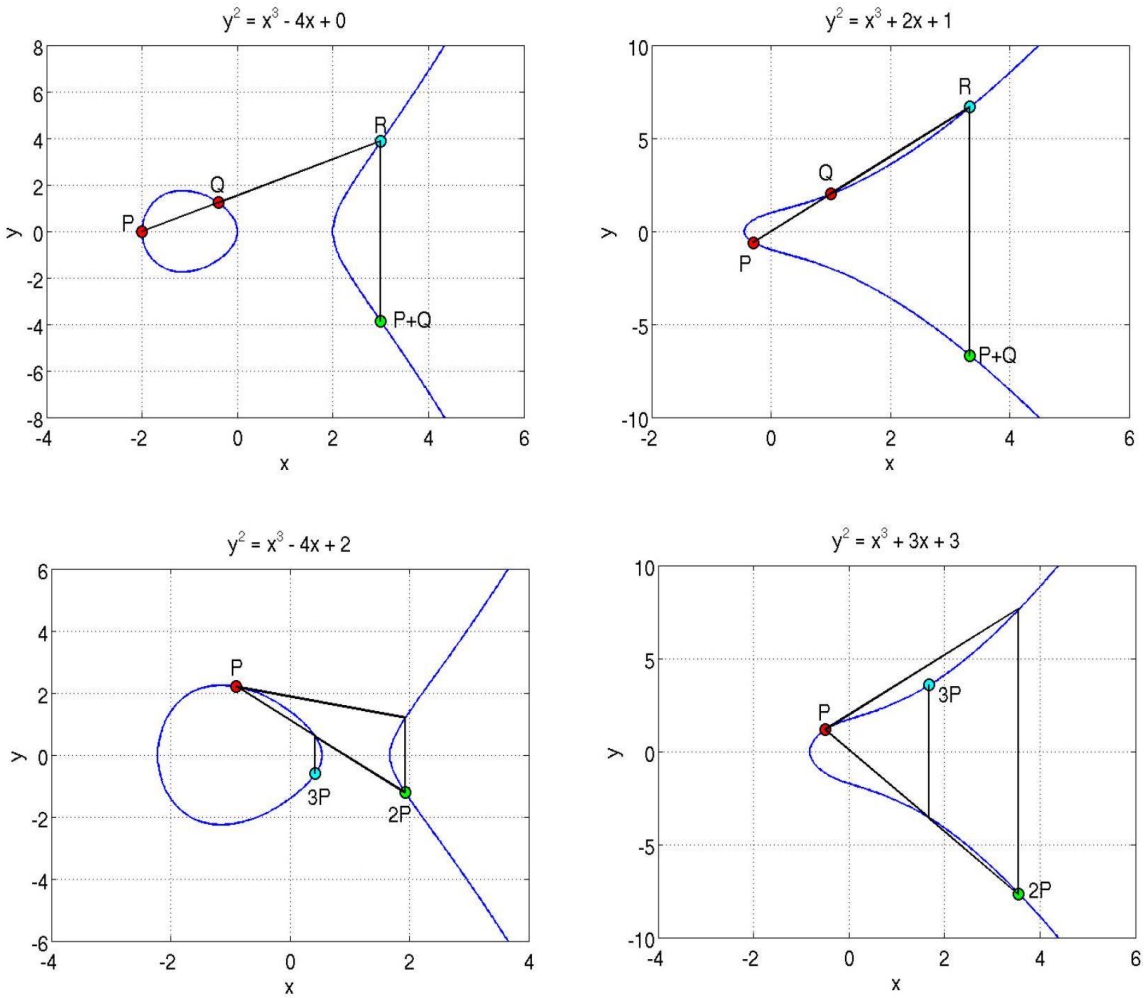

Figure 2. The Characteristic of the Underlying Field and the Singular Elliptic Curves

The examples of the elliptic curves shown so far were for the field of real numbers. These fields are of characteristic zero because no matter how many times you add the multiplicative identity element to itself, you'll never get the additive identity element.

The group law can also be defined when the underlying field is of characteristic 2 or 3. [When we consider real numbers modulo 2, we have an underlying field of characteristic 2. By the same token, when we consider real numbers modulo 3, we have an underlying field of characteristic 3.]

But now the elliptic curve $y^2 = x^3 + ax + b$ becomes singular. While singular elliptic curves do admit group laws of the sort, such groups, although defined over the points on the elliptic curve, become isomorphic to either the multiplicative or the additive group over the underlying field itself, depending on the type of singularity. That fact makes singular elliptic curves unsuitable for cryptography because they are easy to crack.

To show that the elliptic curve $y^2 = x^3 + ax + b$ becomes singular when the characteristic of the underlying field is 2, let's look at the partial derivatives of the two sides of the equation of this curve:

$$2ydy = 3x^2dx + adx$$

implying

$$\frac{dy}{dx} = \frac{3x^2+a}{2y} \qquad (2)$$

A point on the curve is singular if $\frac{dy}{dx}$ is not properly defined. This would be the point where both the numerator and the denominator are zero. [When only the denominator goes to zero, the slope is still defined even though it is ∞.] So, the elliptic curve $y^2 = x^3 + ax + b$ will become singular if it contains a point (x, y) so that

$$\begin{aligned} 3x^2 + a &= 0 \\ 2y &= 0 \end{aligned}$$

and the point (x, y) satisfying these two equations lies on the curve. Let's now consider the case when the underlying field is of characteristic 2. In this case, we go back to Equation (2) above and see that, since 2 is the same thing as 0 for such a, the derivative $\frac{dy}{dx}$ will not be defined at $x = \sqrt{\frac{-a}{3}}$. Therefore, the curve $y^2 = x^3 + ax + b$ will be singular for some values of a that can be obtained by substituting $x = \sqrt{\frac{-a}{3}}$ in the equation of the curve.

Let's now consider the case of a field of characteristic 3. In this case, since 3 is the same thing as 0, we can write for the curve slope from Equation (2):

$$\frac{dy}{dx} = \frac{a}{2y}$$

This curve becomes singular if we should choose $a = 0$.

In general, when using the elliptic curve equation $y^2 = x^3 + ax + b$, we avoid underlying fields of characteristic 2 or 3 because of the nature of the constraints they place on the parameters a and b in order for the curve to not become singular [10].

2.2 An Algebraic Expression for Adding Two Points on An Elliptic Curve

Given two points P and Q on an elliptic curve $E(a, b)$, we have already pointed out that to compute the point $P + Q$, we first draw a straight line through P and Q. We next find the third intersection of this line with the elliptic curve. We denote this point of intersection by R. Then $P + Q$ is equal to the mirror reflection of R about the x-axis.

In other words, if P, Q, and R are the three intersections of the straight line with the curve, then

$$P + Q = -R$$

This implies that the three intersections of a straight line with the elliptic curve must satisfy

$$P + Q + R = 0$$

We will next examine the algebraic implications of the above relationship between the three points of intersection. - The equation of the straight line that runs through the points P and Q is obviously of the form:

$$y = \alpha x + \beta$$

where α is the slope of the line, which is given by

$$\alpha = \frac{y_Q - y_P}{x_Q - x_P}$$

For a point (x, y) to lie at the intersection of the straight line and the elliptic curve $E(a, b)$, the following equality must obviously hold

$$(\alpha x + \beta)^2 = x^3 + ax + b \qquad (3)$$

since $y = \alpha x + \beta$ on the straight line through the points P and Q and since the equation of the elliptic curve is $y^2 = x^3 + ax + b$.

For there to be three points of intersection between the straight line and the elliptic curve, the cubic form in Equation (3) must obviously have three roots. We already know two of these roots, since they must be x_P and x_Q, correspond to the points P and Q.

Being a cubic equation, since Equation (3) has at most three roots, the remaining root must be x_R, the x-coordinate of the third point R. Equation (3) represents a monic polynomial. What that means is that the coefficient of the highest power of x is 1. A property of monic

polynomials is that the sum of their roots is equal to the negative of the coefficient of the second highest power. Expressing Equation (3) in the following form:

$$x^3 - \alpha^2 x^2 + (a - 2\alpha\beta)x + (b - \beta^2) = 0 \tag{4}$$

we notice that the coefficient of x^2 is $-\alpha^2$. Therefore, we have

$$x_P + x_Q + x_R = \alpha^2$$

We therefore have the following result for the x-coordinate of R :

$$x_R = \alpha^2 - x_P - x_Q \tag{5}$$

Since the point (x_R, y_R) must be on the straight line $y = \alpha x + \beta$, we can write for y_R :

$$\begin{aligned} y_R &= \alpha x_R + \beta \\ &= \alpha x_R + (y_P - \alpha x_P) \\ &= \alpha(x_R - x_P + y_P) \end{aligned} \tag{6}$$

To summarize, ordinarily a straight line will intersect an elliptical curve at three points. If the coordinates of the first two points are (x_P, y_P) and (x_Q, y_Q), then the coordinates of the third point are

$$x_R = \alpha^2 - x_P - x_Q \tag{7}$$

$$y_R = \alpha(x_R - x_P) + y_P \tag{8}$$

We started out with the following relationship between P, Q, and R

$$P + Q = -R$$

we can therefore write the following expressions for the x and the y coordinates of the addition of two points P and Q :

$$x_{P+Q} = \alpha^2 - x_P - x_Q \tag{9}$$

$$y_{P+Q} = -y_P + \alpha(x_P - x_R) \tag{10}$$

since the y-coordinate of the reflection $-R$ is negative of the y-coordinate of the point R on the intersecting straight line [4].

2.3 An Algebraic Expression for Calculating $2P$ from P

Given a point P on the elliptical curve $E(a, b)$, computing $2P$ (which is the same thing as computing $P + P$), requires us to draw a tangent at P and to find the intersection of this tangent with the curve. The reflection of this intersection about the x-axis is then the value of $2P$. Given

the equation of the elliptical curve $y^2 = x^3 + ax + b$, the slope of the tangent at a point (x, y) is obtained by differentiating both sides of the curve equation

$$2y\frac{dy}{dx} = 3x^2 + a$$

We can therefore write the following expression for the slope of the tangent at point P :

$$\alpha = \frac{3x_P^2 + a}{2y_P} \tag{11}$$

Since drawing the tangent at P is the limiting case of drawing a line through P and Q as Q approaches P, two of the three roots of the following equation (which is the same as Equation (3) you saw before):

$$(\alpha x + \beta)^2 = x^3 + ax + b \tag{12}$$

must coalesce into the point x_P and the third root must be x_R. As before, R is the point of intersection of the tangent with the elliptical curve.

As before, we can use the property that sum of the roots of the monic polynomial above must equal the negative of the coefficient of the second highest power. Noting two of the three roots have coalesced into x_P, we get

$$x_P + x_P + x_R = \alpha^2$$

Substituting the value of α from Equation (11) in the above equation, we get

$$x_R = \alpha^2 - 2x_P = \left(\frac{3x_P^2 + a}{2y_P}\right)^2 - 2x_P \tag{13}$$

Since the point R must also lie on the straight line $y = \alpha x + \beta$, substituting the expression for x_R in this equation yields

$$\begin{aligned} y_R &= \alpha x_R + \beta \\ &= \alpha x_R + (y_P - \alpha x_P) \\ &= \alpha(x_R - x_P) + y_P \\ &= \frac{3x_P^2 + a}{2y_P}(x_R - x_P) + y_P \end{aligned} \tag{14}$$

To summarize, if we draw a tangent at point P to an elliptical curve, the tangent will intersect the curve at a point R whose coordinates are given by

$$\begin{aligned} x_R &= \left(\frac{3x_P^2 + a}{2y_P}\right)^2 - 2x_P \\ y_R &= \frac{3x_P^2 + a}{2y_P}(x_R - x_P) + y_P \end{aligned} \tag{15}$$

Since the value of $2P$ is the reflection of the point R about the x-axis, the value of $2P$ is obtained by taking the negative of the y-coordinate:

$$\begin{aligned} x_{2P} &= \left(\frac{3x_P^2+a}{2y_P}\right)^2 - 2x_P \\ y_{2P} &= \frac{3x_P^2+a}{2y_P}(x_P - x_R) - y_P \end{aligned} \tag{16}$$

2.4 Elliptic Curves Over Z_p for Prime p:

The elliptic curve arithmetic we described so far was over real numbers. These curves cannot be used as such for cryptography because calculations with real numbers are prone to roundoff error. Cryptography requires error-free arithmetic. That is after all the main reason for why we introduced the notion of a finite field.

However, by restricting the values of the parameters a and b, the value of the independent variable x, and the value of the dependent variable y to belong to the prime finite field Z_p, we obtain elliptic curves that are more appropriate for cryptography:

$$y^2 \bmod p = (x^3 + ax + b) \bmod p \tag{17}$$

subject to the modulo p version of the same smoothness constraint on the discriminant as we had for the case of real numbers:

$$(4a^3 + 27b^2) \bmod p \neq 0 \bmod p$$

We will use the notation $E_p(a, b)$ to represent all the points (x, y) that obey the above equation. $E_p(a, b)$ will also include the dis- tinguished point O, the point at infinity. So the points in $E_p(a, b)$ are the set of coordinates (x, y), with $x, y \in Z_p$, such that the equation $y^2 = x^3 + ax + b$, with $a, b \in Z_p$ is satisfied modulo p and such that the condition $4a^3 + 27b^2 \neq 0 \pmod{p}$ is fulfilled.

Obviously, then, the set of points in $E_p(a, b)$ is no longer a curve, but a collection of discrete points in the (x, y) plane (or, even more precisely speaking, in the plane corresponding to the Cartesian product $Z_p \times Z_p$).

Since the points in $E_p(a, b)$ can no longer be connected to form a smooth curve, we cannot use the geometrical construction to illustrate the action of the group operator. That is, given a point P, now one cannot show geometrically how to compute $2P$, or given two points P and Q, one cannot show geometrically how to determine $P + Q$. However, the algebraic expressions we derived for these operations continue to hold good provided the calculations are

carried out modulo p. Note that for a prime finite field Z_p, the value of p is its characteristic. Elliptic curves over prime finite fields with $p \leq 3$, while admitting the group law, are not suitable for cryptography.

We should also mention that you can also define an elliptic curve when the coordinates are drawn from the set $(Z/pZ)^{\times}$ for any positive integer p. The notation $(Z/pZ)^{\times}$ it consists of the set of all integers that are coprime to N with the group operator betting integer multiplication modulo N.

As we will see in the next section, elliptic curves can also be defined over finite fields Z_{2^m} also commonly called binary finite fields. Binary finite fields have characteristic 2 [7].

2.5 Elliptic Curves Over Finite Field Z_{2^m}

For hardware implementations of ECC, it is common to define elliptic curves over a Finite Field Z_{2^m}. What makes the binary finite fields more convenient for hardware implementations is that the elements of Z_{2^m} can be represented by n-bit binary code words.

The addition operation in Z_{2^m} is like the XOR operation on bit patterns. That is $x + x = 0$ for all $x \in Z_{2^m}$. This implies that a finite field of the form Z_{2^m} is of characteristic 2. As mentioned before, the elliptic curve we showed earlier $(y^2 = x^3 + ax + b)$ is meant to be used only when the underlying finite field is of characteristic greater than 3. The elliptic curve equation to use when the underlying field is described by Z_{2^m} is

$$y^2 + xy = x^3 + ax^2 + b,\ b \neq 0 \tag{18}$$

The constraint $b \neq 0$ serves the same purpose here that the constraint $4a^3 + 27b^2 \neq 0$ did for the case of the elliptic curve equation $y^2 = x^3 + ax + b$. The reason for the constraint $b \neq 0$ is that the discriminant becomes 0 when $b = 0$. As mentioned earlier, when the discriminant becomes zero, we have multiple roots at the same point, causing the derivative of the curve to become ill-defined at that point. In other words, the curve has a singularity at the point where discriminant is 0.

Shown in Figure 3 are six elliptic curves described by the analytical form $y^2 + xy = x^3 + ax^2 + b$ for different values of the parameters a and b. The four upper curves are non-singular. The parameters a and b for the top-left curve are 2 and 1, respectively. The same parameters for the top-right curve are 2 and -1, respectively. For the two non-singular curves in the middle row, the one on the left has 0 and 2 for its a and b parameters, whereas the one on the right has -3 and 2.

The two curves in the bottom row are both singular, but for different reasons. The one on the left is singular because b is set to 0.

The fact that the equation of the elliptic curve is different when the underlying field is Z_{2^m} introduces the following changes in the behavior of the group operator:

- Given a point $P = (x, y)$, we now consider the negative of this point to be located at $-P = (x, -(x + y))$.
- Given two distinct points $P = (x_P, y_P)$ and $Q = (x_Q, y_Q)$, the addition of the two points, represented by (x_{P+Q}, y_{P+Q}), is now given by

$$\begin{aligned} x_{P+Q} &= \alpha^2 + \alpha - x_P - x_Q - a \\ y_{P+Q} &= -\alpha(x_{P+Q} - x_P) - x_{P+Q} - y_P \end{aligned} \tag{19}$$

with

$$\alpha = \frac{y_Q - y_P}{x_Q - x_P} \tag{20}$$

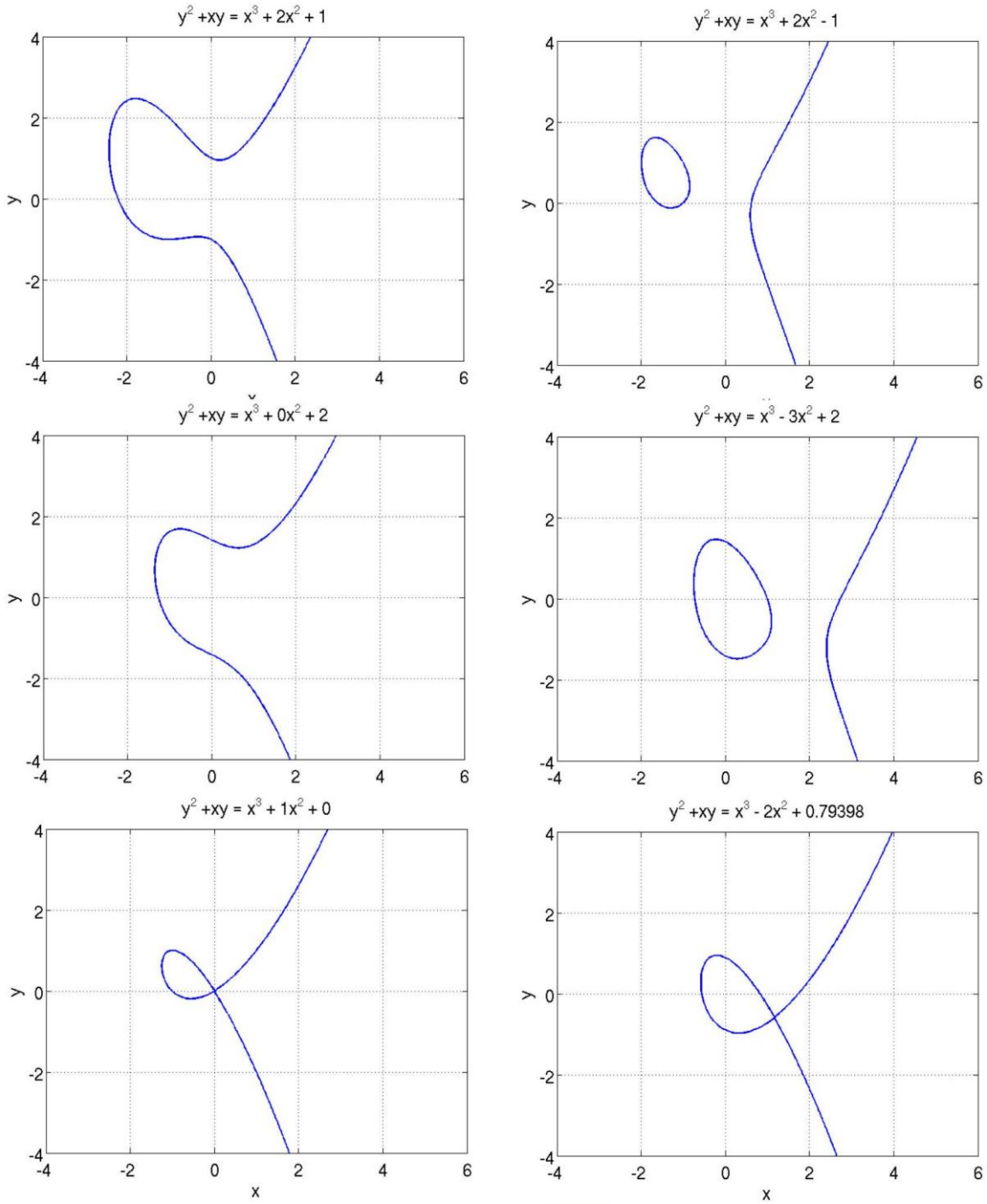

Figure 3. Elliptic curves meant to be used with finite fields.

To double a point, that is to calculate $2P$ from P, we now use the formulas

$$\begin{aligned} x_{2P} &= \alpha^2 + \alpha - a - 2x_P \\ y_{2P} &= -\alpha^2 - \alpha + a + (2+\alpha)x_P - \alpha x_{2P} - y_P \end{aligned} \tag{21}$$

with

$$\alpha = \frac{3x_P^2 + 2ax_P - y_P}{2y_P + x_P} \tag{22}$$

This value of α is obtained by differentiating both sides of $y^2 + xy = x^3 + ax^2 + b$ with respect to x and writing down an expression for $\frac{dy}{dx}$ just as we derived the expression for α in Equation (11).

Since the results for doubling shown in Equation (21) can be obtained from those in Equation (19) by letting x_Q approach x_P, which in our case can be simply accomplished by setting $x_Q = x_P$, the reader may be puzzled by the very different appearances of the expressions shown for y_{P+Q} and y_{2P}. If you set $x_Q = x_P$ in the expression for y_{P+Q}, then both the y-coordinate expressions can be shown to reduce to $-\alpha^3 - 2\alpha^2 + \alpha(3x_P + a - 1) + 2x_P + a - y_P$

The above results are derived in a manner that is completely analogous. As before, we recognize that the points on a straight line passing through two points (x_P, y_P) and (x_Q, y_Q) are given by $y = \alpha x + \beta$ with $\alpha = \frac{y_Q - y_P}{x_Q - x_P}$. To find the point of intersection of such a line with the elliptic curve $y^2 + xy = x^3 + ax^2 + b$, as before we form the equation

$$(\alpha x + \beta)^2 + x(\alpha x + \beta) = x^3 + ax^2 + b \tag{23}$$

which can be expressed in the following form as a monic polynomial:

$$x^3 + (a - \alpha^2 - \alpha)x^2 + (-2\alpha\beta - \beta)x + (b - \beta^2) = 0 \tag{24}$$

Reasoning as before, this cubic equation can have at most three roots, of which two are already known, those being the points P and Q. The remaining root, if it exists, must correspond to the point to the point R, which the point where the straight line passing through P and Q meets the curve again. Again, using the property that the sum of the the roots is equal to the negative of the coefficient of the second highest power, we can write

$$x_P + x_Q + x_R = \alpha^2 + \alpha - a$$

We therefore have the following result for the x-coordinate of R :

$$x_R = \alpha^2 + \alpha - a - x_P - x_Q \tag{25}$$

Since this point must be on the straight line $y = \alpha x + \beta$, we get for the y-coordinate at the point of intersection $y_R = \alpha x_R + \beta$. Substituting for β from the equation $y_P = \alpha x_P + \beta$, we get the following result for y_R :

$$y_R = \alpha(x_R - x_P) + y_P \tag{26}$$

The negative of a point $R = (x_R, y_R)$ is given by $-R = \left(x_R, -(x_R + y_R)\right)$. Since the point $\left(x_{P+Q}, y_{P+Q}\right)$ is located at the negative of the point R at (x_R, y_R), we can write the following result for the summation of the two points P and Q :

$$\begin{aligned} x_{P+Q} &= x_R = \alpha^2 + \alpha - x_P - x_Q - a \\ y_{P+Q} &= -(x_R + y_R) = -\alpha\left(x_{P+Q} - x_P\right) + x_{P+Q} - y_P \end{aligned} \tag{27}$$

The result for doubling of a point can be derived in a similar manner. Figure 4 shows these operations in action. The two figures in the topmost row show us calculating $P + Q$ for the two points P and Q as shown.

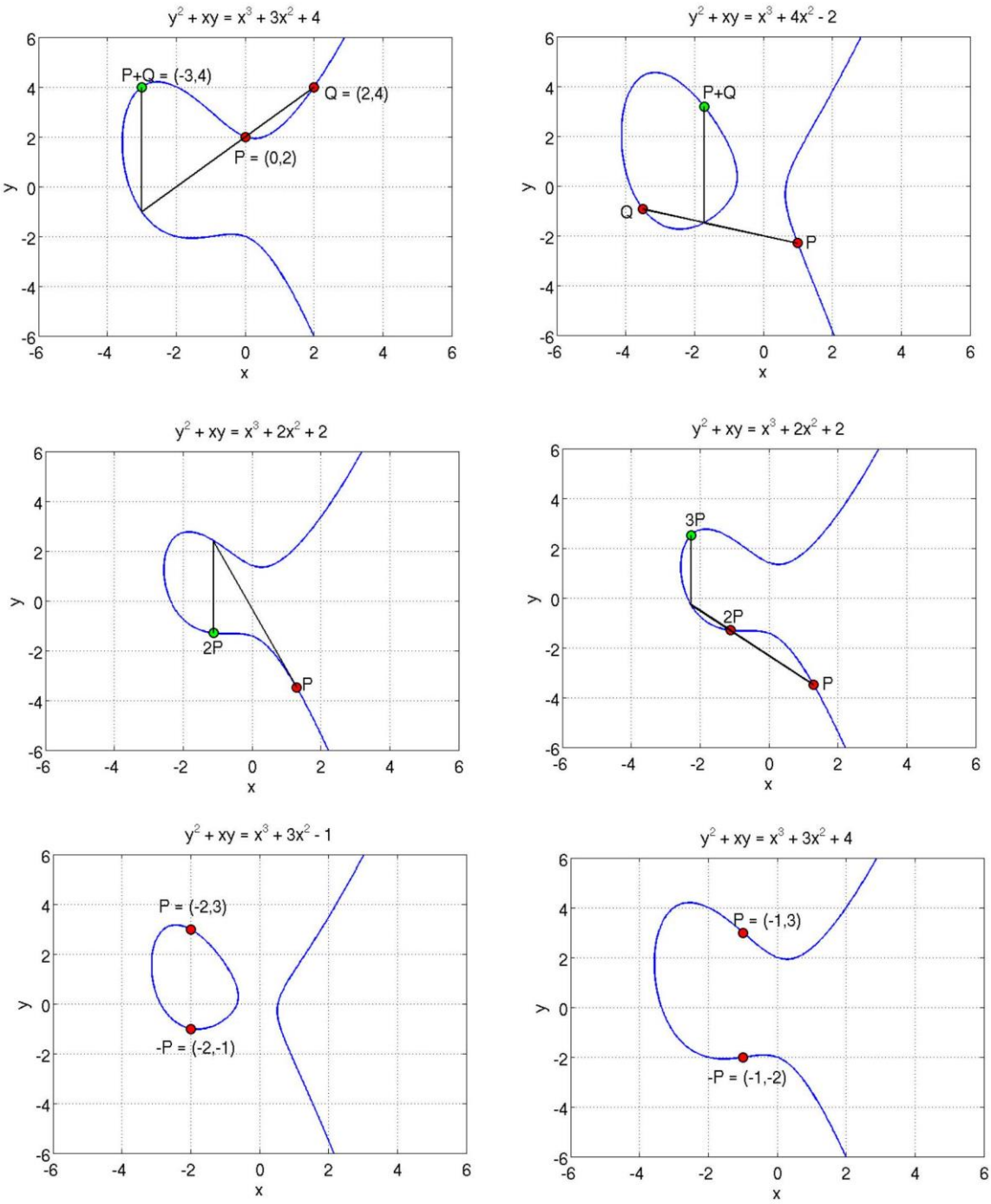

Figure 4. This figure is from middle row shows the doubling of a point and the figure on the right the tripling of a point.

3. Results and Discussion

We will use the notation $E_{2^n}(a,b)$ to denote the set of all points $(x,y) \in Z_{2^n} \times Z_{2^n}$, that satisfy the equation

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

with $a \in Z_{2^n}$ and $b \in Z_{2^n}$, along with the distinguished point O that serves as the additive identity element for the group structure formed by the points on the curve. Note that we do not allow b in the above equation to take on the value which is the additive identity element of the finite field Z_{2^n}.

If g is a generator for the field Z_{2^n}, then all the element of Z_{2^n} can be expressed in the following form

$$0,1,g,g^2,g^3,\dots\dots,g^{2^n-2}$$

This implies that the majority of the points on the elliptic curve $E_{2^n}(a,b)$ can be expressed in the form (g^i,g^j), where $i,j = 0,1,\dots,n-2$. In addition, there may be points whose coordinates can be expressed $(0,g^i)$ or $(g^i,0)$, with $i = 0,1,\dots,n-2$. And then there is, of course, the distinguished point O.

The order of an elliptic curve, that is the number of points in the group $E_{2^n}(a,b)$ is important from the standpoint of the cryptographic security of the curve. [Note: When we talk about the order of $E_{2^n}(a,b)$, we must of course include the distinguished point O.]

Hasse's Theorem addresses the question of how many points are on an elliptic curve that is defined over a finite field. This theorem says that if N is the number of points on $E_q(a,b)$ when the curve is defined on a finite field Z_q with q elements, then N is bounded by

$$|N-(q+1)| \leq 2\sqrt{q}$$

As mentioned previously, N includes the additive identity element O.

Since the Finite field Z_{2^n} contains 2^n elements, we can say that the order of $E_{2^n}(a,b)$ is equal to 2^n+1-t where t is a number such that $|t| \leq \sqrt{2^n}$. - An elliptic curve defined over a Finite field Z_{2^n} is super singular if $2 \mid t$, that is if 2 is a divisor of t. [Supersingularity is not to be confused with singularity. When an elliptic curve is defined over real numbers, singularity of the curve is related to its smoothness. More specifically, a curve is singular if its slope at a point is not defined. Super singularity, on the other hand, is related to the order of E_{2^n} and how this order relates to the number of points in the underlying finite field.]

Should it happen that $t = 0$, then the order of E_{2^n} is $2n+1$. Since this number is always odd, such a curve can never be super singular. Super singular curves defined over fields of characteristic 2 (which includes the binary finite fields Z_{2^n}) always have an odd number of points, including the distinguished point O. Super singular curves are to be avoided for

cryptography because they are vulnerable to the MOV attack. More on that later. Is $b \neq 0$ a Sufficient Condition for the Elliptic Curve $y^2 + xy = x^3 + ax^2 + b$ to Not Be Singular? In general, we want to avoid using singular elliptic curves for cryptography for reasons already indicated. We indicated that when using a curve of form $y^2 + xy = x^3 + ax^2 + b$, you want to make sure that $b \neq 0$ since otherwise the curve will be singular.

We will now consider in greater detail when exactly the curve $y^2 + xy = x^3 + ax^2 + b$ becomes singular for the case when the underlying field consists of real numbers. Toward that end we will derive an expression for the discriminant of a polynomial that is singular if and only if the curve $y^2 + xy = x^3 + ax^2 + b$ is singular. The condition which will prevent the discriminant going to zero will be the condition under which the curve $y^2 + xy = x^3 + ax^2 + b$ will stay nonsingular.

To meet the goal stated above, we will introduce the coordinate transformation

$$y = Y - \frac{x}{2}$$

in the equation

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

The purpose of the coordinate transformation is to get rid of the troublesome term xy in the equation. Note that this coordinate transformation cannot make a singularity disappear, and neither can it introduce a new singularity. With this transformation, the equation of the curve becomes

$$Y^2 - \frac{x^2}{4} = x^3 + ax^2 + b$$

which can be rewritten as

$$Y^2 = x^3 + \left(a + \frac{1}{4}\right)x^2 + b$$

The polynomial on the right-hand side of the equation shown above has a singular point wherever its discriminant goes to zero. In general, the discriminant of the polynomial

$$a_3z^3 + a_2z^2 + a_1z = 0$$

is given by

$$D_3 = a_1^2a_2^2 - 4a_0a_2^3 - 4a_1^3a_3 + 18a_0a_1a_2a_3 - 27a_0^2a_3^2$$

Substituting the coefficient values for our case, $a_3 = 1, a_2 = \left(a + \frac{1}{4}\right), a_1 = 0$, and $a_0 = b$, in the general formula for the discriminant of a cubic polynomial, we get for the discriminant

$$D_3 = -4b\left(a + \frac{1}{4}\right)^3 - 27b^2$$

This simplifies to

$$D_3 = \frac{1}{16}[-64a^3b - 48a^2b - 12ab - b - 432b^2]$$

which can be expressed as

$$D_3 = -\frac{1}{16}b[64a^3 + 48a^2 + 12a + 432b + 1]$$

Obviously, if $b = 0$, the discriminant will become 0. However, it is also obvious that even when the $b = 0$ condition is satisfied, certain values of a and b may cause the discriminant to go to 0. As with the super singular curves, elliptic curves that are singular are to be avoided for cryptography because they are vulnerable to the MOV attack [5].

3.1 Elliptic Curve Cryptography:

That elliptic curves over finite fields could be used for cryptography was suggested independently by Neal Koblitz (University of Washington) and Victor Miller (IBM) in 1985.

Just as RSA uses multiplication as its basic arithmetic operation (exponentiation is merely repeated multiplication), ECC uses the "addition" group operator as its basic arithmetic operation (multiplication is merely repeated addition).

Suppose G is a user-chosen "base point" on the curve $E_q(a, b)$, where $q = p$ for some prime p when the underlying finite field is a prime finite field and $q = 2^n$ when the underlying finite field is a Finite field.

In accordance with how the group operator works, $k \times G$ stands for $G + G + G + \cdots + G$ with G making k appearances in this expression. - Now suppose our message consists of an integer M and we encrypt it by calculating $C = M \times G$. FFor the purpose of visualization, think of $M \times G$ as the two-dimensional point G being added to itself M times through the geometric construction.

Now the question is whether an adversary with knowledge of all of the parameters of the curve $E_q(a, b)$ and of the point G can decrypt C and figure out the value of the message

integer M. [Bear in mind that whereas M is an integer, C just like G is a point on the elliptic curve. In that sense, M and C are two different types of entities.]

The core notion that ECC is based on is that, with a proper choice for G, whereas it is relatively easy to calculate $C = M \times G$, it can be extremely to recover M from C even when an adversary knows the curve $E_q(a, b)$ and the G used. Recovering M from C is referred to as having to solve the discrete logarithm problem. [To understand why finding M from C is referred to as solving the discrete logarithm problem: Note that word "addition" for the group operator for $E_q(a, b)$ is a matter of convention and convenience.

As you already know a group operator is typically referred to as addition and denoted ' + ', whereas the second operator when the group becomes a ring is typically called multiplication and denoted ' $\times$ '. So, there is nothing wrong with choosing to express $G + G + G + \cdots + G$ more generically as $G \circ G \circ G \circ \ldots \circ G$ if we do not want to get confused by mental associations with the '+' operator. Now let's see what we mean by a logarithm. As you know, if $a = b^n$ then $n = \log_b\ a$.

We are at a liberty to write b^n as $b \times b \times b \ldots \times b$, or even as $b \circ b \circ b \ldots \circ b$ if we assume that the operator $\circ$ stands for multiplication. If we want to recover the number of times b participates in $a = b \circ b \circ b \ldots \circ b$ we take the logarithm of a to the base b. By the same token, if we want to determine the number of times G participates in $C = G \circ G \circ G \circ \ldots \circ G$, we take the "logarithm" of C to the base G.

An adversary could try to recover M from $C = M \times G$ by calculating $2G, 3G, 4G, \ldots, kG$ with k spanning the size of the set $E_q(a, b)$, and then seeing which one of the results matched C. But if q is sufficiently large and if the point G on the curve $E_q(a, b)$ is chosen carefully, that would take much too long [8].

3.2 Elliptic Curve Diffie-Hellman Secret Key Exchange

A community of users wishing to engage in secure communications with ECC chooses the parameters q, a, and b for an elliptic curve-based group $E_q(a, b)$, and a base point $G \in E_q(a, b)$. A selects an integer PR_A to serve as his/her private key. A then generates $PU_A = PR_A \times G$ to serve as his/her public key. A makes publicly available the public key PU_A. B designates an integer PR_B to serve as his/her private key.

As was done by A, B also calculates his/her public key by $PU_B = PR_B \times G$. In order to create a shared secret key (that could subsequently be used for, say, a symmetric-key based

communication link), both A and B now carry out the following operations: A calculates the shared secret key by

$$K = PR_A \times PU_B \tag{28}$$

B calculates the shared secret key by

$$K = PR_B \times PU_A \tag{29}$$

The calculations in Eqs. (19) and (20) yield the same result because

$$\begin{aligned} K \text{ as calculated by } A \quad &= PR_A \times PU_B \\ &= PR_A \times (PR_B \times G) \\ &= (PR_A \times PR_B) \times G \\ &= (PR_B \times PR_A) \times G \\ &= PR_B \times (PR_A \times G) \\ &= PR_B \times PU_A \\ &= K \text{ as calculated by } B \end{aligned}$$

To discover the secret key, an attacker could try to discover PR_A from the publicly available base point G and the publicly available PU_A. Recall, $PU_A = PR_A \times G$. But this requires solving the discrete logarithm problem which, for a properly chosen set of curve parameters and G, can be extremely hard.

To increase the level of difficulty in solving the discrete logarithm problem, we select for G a base point whose order is very large. The order of a point on the elliptic curve is the least number of times G must be added to itself so that we get the identity element O of the group $E_q(a, b)$.

We can also associate the notion of order with an elliptic curve over a finite field: The order of an elliptic curve is the total number of points in the set $E_q(a, b)$. This order is denoted $\#E_q(a, b)$. Since the integers PR_A, PU_A, PR_B, and PU_B must all be less than the order of the base point G, this value must also be made publicly available.

The base point G is also known as the generator of a subgroup of $E_q(a, b)$ whose elements are all given by $G, 2G, 3G$, ..., and, of course, the identity element O. For the size of the subgroup to equal the degree of the generator G, the value of n must be a prime when the underlying field is a Finite field Z_{2^n}[3].

3.3 ECC For Digital Rights Management

ECC has been and continues to be used for Digital Rights Man- agement (DRM). DRM stands for technologies/algorithms that allow a content provider to impose limitations on the who's and how's of the usage of some media content made available by the provider.

ECC is used in the DRM associated with the Windows Media framework that is made available by Microsoft to third-party vendors interested in revenue-generating content creation and distribution. In what follows, we will refer to this DRM as WMDRM.

The three main versions of WM-DRM are Version 1 (released in 1999), Version 2 (released in 2003, also referred to as Version 7.x and Version 9), and Version 3 (released in 2003, also known as Version 10). All three versions have been cracked. As you would expect in this day and age, someone figures out how to strip away the DRM protection associated with, say, a movie and makes both the unprotected movie and the protection stripping algorithm available anonymously on the web. In the meantime, the content provider (like Apple, Sony, Microsoft, etc.) comes out with a patch to fix the exploit. Thus continues the cat and mouse game between the big content providers and the anonymous "crackers."

Again, as you would expect, the actual implementation details of most DRM algorithms are proprietary to the content providers and distributors. But, on October 20, 2001, an individual, under the pseudonym Beale Screamer, posted a detailed description of the inner workings of the WM-DRM Version 2. This information is still available at the URLs http://cryptome.org/ ms-drm.htm and http://cryptome. org. beale-sci-crypt. htm where you will find a command-line tool named FreeMe for stripping away the DRM protection of the older versions of Windows Media documents. Since Version 2 is now considered out of date, the main usefulness of the information posted at the web site lies in its educational value.

WM-DRM Version 2 used elliptic curve cryptography for exchanging a secret session key between a user's computer and the license server at the content provider's location.

The ECC used in WM-DRM V. 2 is based on the first elliptic curve $y^2 = x^3 + ax + b$. The ECC algorithm used is based on the points on the curve whose x and y coordinates are drawn from the finite field $(Z/pZ)^{\times}$, with the number p set to

$$p = 78596310237942882237669478944689739620749856895 1$$

In the WM-DRM ECC, all are represented using 20 bytes. Here is the hex representation of the modulus p shown above:

$$p = 0x89abcdef012345672718281831415926141424f7$$

We also need to specify values for the parameters a and b of the elliptic curve $y^2 = x^3 + ax + b$. As you would expect, these parameters are also drawn from $(Z/pZ)^{\times}$and their values are given by

$$\begin{aligned} a &= 31768908125132550347631747641382769327274695592 7 \\ b &= 79052896607878758718120572025718535432100651934 \end{aligned}$$

Since all numbers in the ECC implementation under consideration are stored as blocks of 20 bytes, the hex representations of the byte blocks stored for a and b are

$$\begin{aligned} a &= 0x37a5abccd277bce87632ff3d4780c009\text{ ebe 41497} \\ b &= 0x0dd8dabf725e2f3228e85f1ad78f\text{ dedf9328239e} \end{aligned}$$

Following the discussion, the ECC algorithm would also need to choose a base point G on the elliptic curve $y^2 = x^3 + ax + b$. The x and the y coordinates of this point in the ECC as implemented in WM-DRM are

$$\begin{aligned} G_x &= 77150721626264982617064826856557988990776925417 6 \\ G_y &= 39015751024655662852527945926651499556253319665 5 \end{aligned}$$

The 20-byte hex representations for these two coordinates are

$$\begin{aligned} G_x &= 0x8723947fd6a3a1e53510c07dba38daf0109fa120 \\ G_y &= 0x445744911075522d8c3c5856d4ed7acda379936f \end{aligned}$$

As mentioned, an ECC protocol must also make publicly available the order of the base point. For the present case, this order is given by

$$\#E_p(a, b) = 78596310237942882237669302488171495761268615742 9$$

With the elliptic curve and its parameters set as above, the next question is how exactly the ECC algorithm is used in WM-DRM. When you purchase media content from a Microsoft partner peddling their wares through the Window Media platform, you would need to download a "license" to be able play the content on your computer. Obtaining the license consists of your computer randomly generating a number $n \in Z_p$ for your computer's private key. Your computer then multiplies the base point G with the private key to obtain the public key. Subsequently your computer can interact with the content provider's license server in the manner described to establish a secret session key for the transfer of license related information into your computer.

In order to ensure that only your computer can use the downloaded license, WM-DRM makes sure that you cannot access the private that your computer generated for the ECC algorithm. Obviously, if you could get hold of that n, you could pass the encrypted content file and the private key to your friend and they would be able to pretend to be your vis-a-vis the license server. WM-DRM hides an RC4 encrypted version of the private key in the form of a linked list in which each nodes stores one half of the key. When DRM software is cracked, it is usually done by what is known as "hooking" the DRM libraries on a computer as they dump out either the keys or the encrypted content.

4. Conclusions

In conclusion, Elliptic Curve Cryptography (ECC) is a modern cryptographic technique that uses the properties of elliptic curves over finite fields to provide efficient and secure public key encryption. ECC offers advantages over traditional public key systems such as RSA, including smaller key sizes, resistance to attacks from quantum computers, and flexibility.

ECC has become increasingly popular in modern applications due to its efficiency and security, and its continued development is expected to have a significant impact on the field of cryptography. However, it is important to note that the security of ECC depends on the correct implementation of the system and the use of strong parameters.

Overall, ECC is a powerful cryptographic tool that has enabled secure communications, digital signatures, and other applications. Its importance in modern cryptography is expected to grow as technology advances and security concerns become increasingly important.

References

[1] Washington, L. C. (2008). *Elliptic curves: number theory and cryptography*. CRC press.

[2] Smith, C. H., & Lennon, M. J. J. (1993). Elliptic curve public key cryptography. Kluwer Academic Publishers.

[3] Blake, I. F., Seroussi, G., & Smart, N. P. (1999). Elliptic curve cryptography: theory and implementation. Springer.

[4] Hoffstein, J., Pipher, J., & Silverman, J. H. (2008). An introduction to mathematical cryptography. Springer.

[5] Corbellini, A. (2015). Elliptic curve cryptography: a gentle introduction. Retrieved from https://andrea.corbellini.name/2015/05/17/elliptic-curve-cryptography-a-gentle-introduction/

[6] Bartlett, M. (n.d.). Elliptic curve cryptography (ECC) tutorial. Certicom. Retrieved from https://www.certicom.com/content/certicom/en/developers/ecc-tutorial.html

[7] Surdhar, P. (2017). Elliptic curve cryptography explained. Retrieved from https://www.youtube.com/watch?v=NF1pwjL9-DE

[8] Entrust. (n.d.). Elliptic curve cryptography (ECC). Retrieved from https://www.entrust.com/resources/ecc/

[9] Boneh, D. (1999). Elliptic curve cryptography. In Advances in Cryptology - EUROCRYPT'99 (pp. 232-241). Springer.

[10] Certicom. (n.d.). Elliptic curve cryptography. Retrieved from https://www.certicom.com/content/certicom/en/technology/elliptic-curve-cryptography.html