Group Theory and Modern Cryptography

To what extent does group theory provide a mathematical backdrop for modern cryptography?

Group 5: Mathematics

4000 words excluding typeset expressions

Contents

1 Foreword

This paper is for research purposes only. Although the information presented is as close to accurate as possible, one should never implement a cryptographic system themselves unless they know exactly what they're doing. **I am not liable for any damages caused in testing any cryptographic concepts referenced in this paper.** Cryptography's greatest weakness is human-error, and virtually *all* breaks in cryptographic security result from poor implementation.[1]

2 Introduction and Aim

2.1 Topic and Research Question

For my Extended Essay, I will be exploring the relationship between Group Theory, a mathematical field generalizing symmetry and codifying number systems, and modern cryptography. Cryptography has seen a number of advancements in the computer-age, most notable of which is the use of 'elliptic curves.' In this paper, I intend to show how a group-theoretical perspective on cryptography helps understand elliptic curve cryptosystems. My guiding question is: *to what extent does group theory provide a mathematical backdrop for modern cryptography?*

2.2 Motivating Idea

It's undeniable that our modern-day world is reliant on cryptography. Every time a phone sends a text, a browser connects to a server, an email gets sent off, or a monetary transaction is made, our devices are performing many hundreds of math operations to ensure our data are 'encrypted.' But what does 'encryption' mean?

'Encryption' is the process of disguising a message to be, loosely speaking, hidden to all *except* the intended recipient. This is the process of converting a 'plaintext' message into a jumbled 'ciphertext.' Converting a plaintext message (a string/array of characters) into a ciphertext is known as an 'enciphering' or 'encrypting' transformation. Likewise the reverse operation of recovering the plaintext message from a ciphertext is known as the *deciphering transformation.*[2] If we denote the plain and ciphertext $\mathcal{P}$ and $\mathcal{C}$ respectively, the enciphering map f and its inverse f^{-1} we obtain the following diagram:

[1] Harold Thimbleby. "Human factors and missed solutions to Enigma design weaknesses". In: *Cryptologia* 40.2 (2016), pp. 177–202. DOI: 10.1080/01611194.2015.1028680.

[2] Neal Koblitz. *A Course in Number Theory and Cryptography.* 2nd ed. Graduate Texts in Mathematics No. 114. Springer, 1994. ISBN: 9781461264422, p. 54.

$$\mathcal{P} \xrightarrow{f} \mathcal{C} \xrightarrow{f^{-1}} \mathcal{P}$$

2.3 The Fundamentals of Symmetric Cryptography

The intuitive implementation has the two transacting parties agree upon the nature of the map f in secret, beforehand. This type of system has a special name, *"Symmetric Key Cryptography"*, reflecting that both parties have the same common secret. Historical examples include the 'Caesar cipher,' where f is a shift operation that maps each letter a number of places ahead. For instance, the parties might decide beforehand the map f: *shift each character forward 3 letters*, implying the inverse map f^{-1}: *shift each character back 3 letters*. This might look like so: [3]

$$\mathcal{P} \sim A, B, C, D, E, F, ...Y, Z$$
$$\mathcal{C} \sim D, E, F, G, H, I, ...B, C$$

For example, if one wanted to encrypt the message $\mathcal{P}$ = "HELLO", the ciphertext would be $\mathcal{C}$ = "KHOOR" which they could send off to their friend to decode with the inverse, subtract-three-letters map. Note that, even in this simple example, repeated letters, word length, and other syntactic information provide a wealth of information about the nature of the plaintext, making it vulnerable to cryptanalysis techniques.[4]

Mathematically, a Caesar cipher starts with encoding each letter as a number from 0, A, to 25, Z, depending on the alphabet of choice. We represent this map, shifting n places forward, with addition. Importantly, this operation must 'wrap around' back to zero when it exceeds 'z' in the alphabet. This process, known as *'modular arithmetic'*, is like circling a clock, where after reaching 25 we wrap back to 0. So, in our case, shifting 'z' by 3 letters looks like so: $25 + 3 \bmod 26 \equiv 2$, which reads "25 plus 3 *is congruent to* 2 *modulo* 26." With this in mind, we obtain for each letter $p \in \mathcal{P}$:

$$f(p) = p + n \bmod 26$$

This isn't a very sophisticated cryptographic scheme... For instance, comparing the most commonly occurring letters in the ciphertext to those of the English alphabet (frequency analysis) could easily break these schemes,

[3] Ceaser Ciphers are just linear transformations in disguise. We could consider groups of two letters, *digraphs*, resulting in f being matrix-vector multiplication with a 2x2 invertible matrix. Similarly, we could consider invertible *affine* transformations of the form $f : \vec{x} \to A\vec{x} + \vec{b}$, where A encodes a scaling factor.

[4] Koblitz, *A Course in Number Theory and Cryptography*, p. 56.

broadly referred as 'substitution ciphers.'[5] Modern protocols are more resistant, changing one letter of plaintext yields a completely different ciphertext, making them impervious to this technique.

In general, Symmetric-key cryptography (*viz.* predetermined, shared secret) is well understood. For instance, Claude Shannon proved that the so-called *'one-time-pad'* encryption technique was mathematically unbreakable. In full generality, he showed that if a random-generated key is at least as long as the plaintext (specifying a random[6] character for each plaintext character in the same alphabet), then performing a caesar cypher shift on each *individual* plaintext character by the value of the random character, yields a cryptosystem that is *mathematically* impenetrable. Or equivalently, performing a random modular addition on each character of the plaintext, with the shared secret being the sequence of random shifts.[7] Although modern systems seek smaller key sizes for long messages, this worst-case scenario demonstrates the strength of symmetric-key algorithms in general. The 'one-time-pad' gets its name from WWII, when the KGB would distribute palm-sized pads with one-time-keys and a table to ease in conversion. Such pads were often made of flammable materials to be burned with no trace.[8]

3 What is 'Public-Key' Cryptography

3.1 The 'Public-Key' Paradox

In the modern world, it's impractical to require that every shared secret be determined ahead of time. If a user wants to connect to a twitter server over a secure connection, how could symmetric-key encryption be employed?—there's no way to share in secret beforehand. Beyond symmetric encryption, if two parties want to transact over an eavesdropped channel, is there any way they could do so in an encrypted matter? The intuitive answer might be no; how can a shared secret be 'shared' without any man-in-the-middle being able to obtain that same key. But this defys the ubiquity of encryption on the internet that we rely on daily. How could this be?

[5] Koblitz, *A Course in Number Theory and Cryptography*, p. 54.

[6] There is a paramount distinction between *random* and *pseudorandom*. *True* randomness has to be derived from a non-computer source (for the most part). Commonly approximates include the least-significant-digit of a mouse position, the frequency of keypresses, etc.
(ibid., p. 92)

[7] Claude E. Shannon. *Communication Theory of Secrecy Systems*. Vol. 28. The Bell System Technical Journal, 1949, pp. 656–715. DOI: 10.1002/j.1538-7305.1949.tb00928.x.

[8] Robert Edward Lewand. "The Perfect Cipher". In: *The Mathematical Gazette* 94.531 (2010), pp. 401–411. ISSN: 00255572. URL: http://www.jstor.org/stable/25759724 (visited on 05/21/2022).

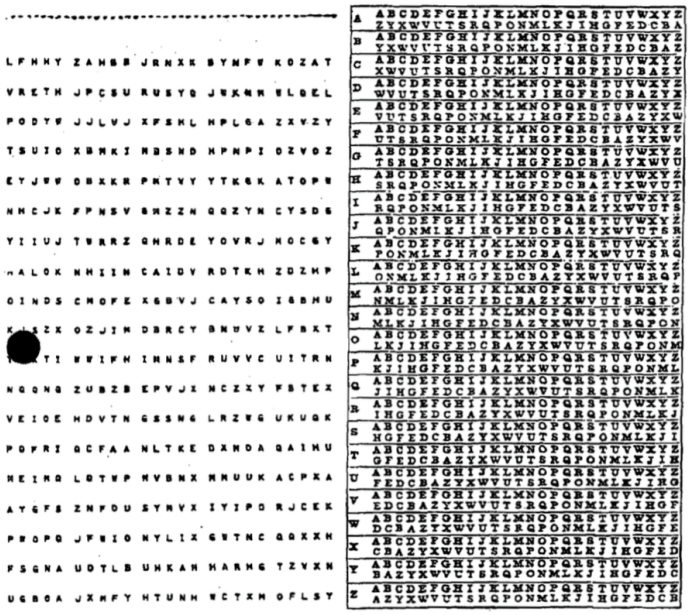


Figure 1: Format of a one-time-pad used by the NSA[9]

3.2 One-way Operations

Consider the intuitive fact that some operations are more difficult to do forwards than in reverse—certain 'one-way functions.' If I mixed two different-colored paints together and asked whether you, *a priori*, could deduce the two initial colors given the end result, could you? Although it's quite easy to check whether any two colors combine to match the end result, there isn't an *easy* operation that returns the initial colors. This happens to be true in mathematical contexts as well.

The foremost 'one-way operation,' employed to this day, is that of multiplying large numbers with minimal divisors. This makes sense from a practical standpoint; breaking up a large number into factors usually requires computationally expensive trial divisions, to use a jargon, its *asymptotic time-complexity is exponential*, meaning that the time required to reach an answer scales exponentially with the size of the number. Algorithms like the general-number-field-sieve quicken this slightly,[10] but regardless it is vastly faster to compute multiplication, the time complexity of which scales *at worst* with the square of the number.[11]

[10]Carl Pomerance. "A Tale of Two Sieves". In: *Notices of the AMS* 43.12 (1996), 1473–1485, p. 1482.

[11]Viktor Bunimov. "Area and time efficient modular multiplication of large integers". In: *Proceedings IEEE International Conference on Application-Specific Systems, Architectures, and Processors. ASAP 2003* (2003). DOI: 10.1109/ASAP.2003.1212863.

3.3 Diffie-Hellman Key Exchange

The Diffie-Hellman protocol was the first "asymmetric" protocol,[12] using the above 'multiplication' intuition as a starting point. The end result of the key exchange is a commonly shared number, which can then function as the key for parties to establish symmetric encryption.

Between two parties, Alice A and Bob B, the protocol first agrees on a common 'generator' integer g, which can be done over an eavesdropped channel. Each party chooses another *secret* integer in private: their private key. Say, Alice chooses a and Bob chooses b. Then, each party exponentiates the generator by their private key. So Alice and Bob compute g^a and g^b respectively. Each then sends this result to the other over the insecure channel. Finally, they both exponentiate the received numbers by their private key. For example, Alice would receive g^b over the network, then compute$(g^b)^a$. By the commutativity of integer multiplication, multiplying g by itself b times *then* a times is the same as multiplying g by itself a times *then* b times. Explicitly, with Alice and Bob's private-keys in amber and blue respectively:

$$(g^{a})^{b} = g^{ab} = (g^{b})^{a}$$

The one caveat to the above description, however, is that these multiplications are performed *modulo* n for some integer n. As discussed before, this means numbers wrap around 'like a clock'; if g^a exceeds n, the remainder of $g^a \div n$ is returned. This should make sense, because computers have limited memory and because exponentiation by a (large) number yields an incomprehensibly large result—too large to send over a network or perform computation with. Additionally, computers can perform modular exponentiation *very* performantly by leveraging tricks in binary.[13]

Our relation from before looks like:

$$(g^{a} \bmod p)^{b} \bmod = g^{ab} \bmod p = (g^{b} \bmod p)^{a} \bmod p.$$

Importantly, though, Alice and Bob both agree on a common result $g^{ab} \bmod p$.

In the real world, this type of Diffie-Hellman transaction is used by countless devices every second to allow secure connections over the internet. Browsers have a predetermined set of generators from NIST which they use to establish a shared secret with a web-server which is then used

[12] Precisely, it's a non-authenticated key exchange, but DH can be easily modified to send arbitrary messages in a scheme known as 'Elgamal encryption,' which is true "public-key" cryptography.
(Taher Elgamal. "A Public-Key Cryptosystem and a Signature Scheme Based on Discrete Logarithms". In: *IEEE Transactions on Information Theory* 31.4 [1985], pp. 469–472. DOI: 10.1109/TIT.1985.1057074)

[13] Bunimov, "Area and time efficient modular multiplication of large integers".

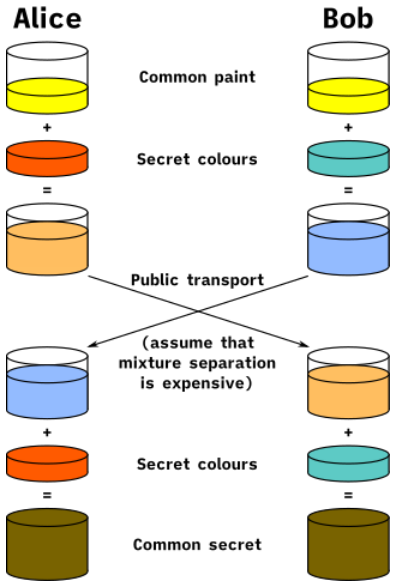


Figure 2: Diffie-Hellman protocol depicted visually, with mixing and separating paint replacing multiplication and factorization. (public domain)

to implement symmetric encryption to facilitate fast data transfer.[14] Every time you are to connect to a `https://...` website the client (you) and the server are performing this type of key exchange. Although, modern implementations might use a different 'number system' to perform the repeated multiplications...

3.4 A Proof that: $(a \bmod p)^k \bmod p = a^k \bmod p$

Here I offer a short proof verifying the ideas above. Given the linearity of modularity over multiplication:

$$a \times b \bmod n = (a \bmod n) \times (b \bmod n) \bmod n$$

I will show that $(a \bmod p)^k \bmod p = a^k \bmod p$ with induction.

Basis case, $P(1)$ or $k = 1$, holds since,

$$a^1 \bmod p = (a \bmod p)^1 \bmod p \qquad \text{(basis)}$$

Assume $P(k)$ holds for all k like so:

[14] Elaine Barker et al. "Recommendation for Pair-Wise Key Establishment Using Integer Factorization Cryptography". In: *National Institute of Standards and Technology* 800-56B (2018). DOI: 10.6028/NIST.SP.800-56Br2. URL: https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-56Br2.pdf.

$$a^k \bmod p = (a \bmod p)^k \bmod p \qquad \text{(P(k))}$$

Now, multiplying $\bmod p$ both sides by $(a \bmod p)$ and applying the given:

$$\begin{aligned} a^k \bmod p \times a \bmod p \quad \bmod p &= (a \bmod p)^k \bmod p \times a \bmod p \quad \bmod p \\ (a^k \times a) \bmod p &= ((a \bmod p)^k (a \bmod p)) \bmod p \\ a^{k+1} \bmod p &= (a \bmod p)^{k+1} \bmod p \qquad \text{(P(k+1))} \end{aligned}$$

Hence, since $P(1)$ holds and $P(k) \Rightarrow P(k+1)$, the proposition holds for all positive integers. ■

4 An Introduction to Group Theory

Transitioning from the methods described earlier, I will now explore how many of these concepts fit into a group-theoretical perspective and how these tools can illuminate a path to inventing stronger cryptosystems.

4.1 The Group Axioms

A group is a set G, equipped with a binary operation mapping two elements to another of the form $* : \ G * G \to G$ such that the following conditions hold:[15]

Associativity
For all $a, b, c \in G, \ a * (b * c) = (a * b) * c$
Identity
There exists an $e \in G$ such that $a * e = e * a = a \ \ \text{for all } a \in G$
Inverse
Each $x \in G$ has an inverse $x^{-1} \in G$ with $x * x^{-1} = x^{-1} * x = e$ [16]

Under this axiomatic definition of a group, a few canonical examples might come to mind.

For instance, the integers form a group under addition $(\mathbb{Z}, +)$. It's worth verifying for yourself that these do fit the definition of a group. Our identity element is 0, since $0 + a = a + 0 = a$ for all $a \in \mathbb{Z}$. Addition of integers is clearly associative. And, every number has a unique inverse, $a^{-1} := -a$.

Similarly, the real numbers (excluding 0) might form a group under multiplication, write $(\mathbb{R} \setminus \{0\}, \times)$, $\mathbb{R}^\times \setminus \{0\}$, or just $\mathbb{R}^\times$. We verify it has a

[15] Dan Saracino. *Abstract Algebra*. 2nd ed. 2008. ISBN: 1577665368, p. 16.

[16] Note that uniqueness of inverses is not given in the axioms for a group, since it follows from the Identity and Associativity requirement and the existence of a (not strictly unique) inverse. If b, c are left and right inverses respectively of an element a with identity 1, then considering $c = 1 * c = (b * a) * c = b * (a * c) = b * 1 = b$, gives us $b = c$ as required.

two-sided identity, 1; that its operation is associative; and that every element has an inverse $a^{-1} := \frac{1}{a}$, since $a \times \frac{1}{a} = \frac{1}{a} \times a = 1,\ \forall a \in (\mathbb{R} \setminus \{0\}, \times)$.

To help intuition,consider that the integers $(\mathbb{Z}, \times)$ *do not* form a group under multiplication, since inverses are not suitably defined, i.e., there is no integer a such that $a \times 2 = 2 \times a = 1$.

4.2 The Case for Commutativity and Abelian Groups

One key property that left out of the group axioms is *commutativity*. At first glance, this might seem like an obvious condition since many of the introductory examples do obey commutativity. In fact, non-commutative groups abound, and the *vast* majority of groups are not commutative.[17] This distinction happens to be so important that commutative groups have a special name: *abelian groups*.[18] Likewise, groups that violate this condition are sometimes called *non-abelian groups*. In general, non-abelian groups correspond to symmetries that change the backdrop for another symmetry to occur. For instance, the group of symmetries on a square, D_8, is non-abelian. If you labeled the vertices of a square, you'd find performing a rotation and than a reflection is not, in general, the same as performing a reflection then a rotation[19]. Despite this, many of the examples we have and will look at happen to be abelian. For instance, the additive and multiplicative (sans 0) groups of $\mathbb{Z}, \mathbb{Q}, \mathbb{R}$ and $\mathbb{C}$, are 'abelian.' To highlight the importance of commutativity, consider that many properties of electric charge, as described by the Standard Model of particle physics, follow directly from the commutativity of the circle group $U(1)$ or $\mathbb{T}$.[20] But this is far beyond the scope of basic group theory.

4.3 Applying Group Theory to Cryptography: Cyclic Groups

The groups that are of importance to (current) cryptography are abelian groups. Consider the Diffie-Hellman example before. Both parties combined a series of steps involving their generator and private keys. If the underlying multiplicative group were not abelian, $(g^a)^b$ doesn't necessarily equal $(g^b)^a$. This provokes the question, were the 'numbers' in these cases abelian groups?

Yes. The group of integers under addition *modulo* n (as in the Caesar cipher) is denoted $(\mathbb{Z}/n\mathbb{Z}, +)$, $(\mathbb{Z}/n\mathbb{Z})^+$ or just $\mathbb{Z}/n\mathbb{Z}$ and is called the

[17] Michael Aschbacher. *The Status of the Classification of the Finite Simple Groups*. Vol. 51. Notices of the American Mathematical Society 7. American Mathematical Society, 2004, pp. 736–740. URL: https://www.ams.org/notices/200407/fea-aschbacher.pdf.

[18] Richard M. Foote David S. Dummit. *Abstract Algebra*. 3rd ed. Wiley, 2004. ISBN: 9780471433347, p. 17.

[19] Working with the presentation of D_8, $\langle a, x \mid a^4 = x^2 = e,\ xax^{-1} = a^{-1} \rangle$, one would say that x and a don't commute—corresponding to the reflection and rotation generators respectively.

[20] Erick J. Weinberg. *Classical Solutions in Quantum Field Theory*. Cambridge University Press, 2012, pp. 81–107. ISBN: 9781139017787. DOI: 10.1017/CBO9781139017787.

'cyclic group of order n.' In an earlier example, we performed operations in $(\mathbb{Z}/26\mathbb{Z}, +)$. The notation here is meaningful. Without straying too afar, the slash corresponds to 'quotienting' out the integers by p times the integers, creating a (quotient-) group where 1 is indistinguishable from $p+1$ and indeed $2p+1, 3p+1, -p+1, -2p+1...$, called an *equivalence relation.*[21] This formalizes the imprecise notions of 'going around a clock face.'

Contrastingly, performing multiplication in a modular fashion forms the group $(\mathbb{Z}/n\mathbb{Z}, \times)$ or $(\mathbb{Z}/n\mathbb{Z})^\times$. This group behaves a little differently and consists of the elements $\{0, 1..., n-1\}$ that are relatively prime to n, i.e., share no nontrivial divisors. In this multiplicative group, it takes more care to define inverses. Inverses can be computed using the Euclidean Algorithm or by leveraging the symmetries in modular multiplication.[22] Essentially, in $(\mathbb{Z}/q\mathbb{Z})^\times$, the multiplicative inverse of a number n is a number m such that,

$$nm \equiv 1 \mod q$$

The counterpart in a more familar context, say $\mathbb{R}^\times$, is that the inverse of a number a is a number $\frac{1}{a}$, where $a \times \frac{1}{a} = 1$.

Now, I will prove that the elements of $(\mathbb{Z}/q\mathbb{Z})^\times$ are only those relatively prime to q. The proof is by contradiction:

If $\gcd(n, q) = d$ with $d \neq 1$ (i.e., they share a nontrivial divisor), assume there does exist some m for which $nm \equiv 1 \mod q$. Expanding this out,

$$\begin{aligned} nm &\equiv 1 \mod q \\ nm - 1 &\equiv 0 \mod q \\ \Rightarrow\ & p | nm - 1 \\ \Rightarrow\ & d | nm - 1 \\ \Rightarrow\ & d | nm - 1 - nm \text{ (since } d | nm) \\ \Rightarrow\ & d | 1 \end{aligned}$$

As 1 is not divisible by anything except 1, our assumption was false and n has no inverse and is not $\in (\mathbb{Z}/q\mathbb{Z})^\times$.[23] ■

Notice that for prime p, every number $< p$ is relatively prime to p, so the order (number of elements) of the group $(\mathbb{Z}/n\mathbb{Z})^\times$ is $p-1$, which you might notice is the same as the order of $(\mathbb{Z}/n\mathbb{Z})^+$. This fact will be useful in a bit.

4.4 Discrete Logarithm Problem

In this section, I want to explore the formal statement of 'why' certain operations are difficult to reverse, although this section gets quite abstract coming from a basic introduction of group theory.

[21] Saracino, *Abstract Algebra*, p. 104.
[22] Ibid., p. 36.
[23] Koblitz, *A Course in Number Theory and Cryptography*, p. 19.

In a Diffie-Hellman key exchange over $(\mathbb{Z}/q\mathbb{Z})^\times$, Alice and Bob were using the map f,

$$f(n) = g^n \bmod q$$

or, written in terms of their respective groups,

$$f : \mathbb{Z} \to (\mathbb{Z}/q\mathbb{Z})^\times$$

This f is a function between groups (a *group homomorphism*) from $\mathbb{Z}$ *onto* a ***subset*** of $(\mathbb{Z}/q\mathbb{Z})^\times$ surjectively, namely the subset generated by g^n which I will denote H. The inverse mapping here (the 'computationally difficult' part), isn't defined for the entirety of dom f, since f takes an infinite group $\mathbb{Z}$ to a subset H of the finite group $(\mathbb{Z}/q\mathbb{Z})^\times$.

Roughly outlining the proof of the one-way operation's "difficulty," it can be shown that any finite abelian group must consist of cyclic groups as building blocks, cyclic groups (the integers under addition modulo some n). [24] Hence, our subset H (which one can prove forms a group as well, a 'subgroup')[25] must be one of these cyclic groups. We can then write this sort-of inverse map f^{-1} as,

$$\log_g : H \to (\mathbb{Z}/m\mathbb{Z}, +)$$

Ascertaining this 'hidden' cyclic subgroup H is known to be a difficult problem, and this difficulty of computing the original integer n is infeasible since $\log_g$ does not map back to all the integers: just some additive group of the integers modulo some m. This is known as the *discrete logarithm problem.*[26] As discussed before with division, there is no known algorithm that can solve a discrete logarithm with time-complexity that scales polynomially (on a non-quantum computer). It remains an open question in theoretical computer-science whether such a polynomial-time algorithm exists, though all of modern cryptography rests on this principle.[27]

5 Group Law on an Elliptic Curve

5.1 Definition of an Elliptic Curve

In a complete turn of events (that will hopefully make sense in the end), I want to explain the notion of *Elliptic Curves.* Elliptic curves are a family of algebraic[28] curves defined as the solution set of a polynomial equation.

[24] By the (aptly named) Fundamental Theorem of Finitely Generated Abelian Groups. (Saracino, *Abstract Algebra*, p. 134)

[25] Koblitz, *A Course in Number Theory and Cryptography*, p. 43.

[26] Ibid., p. 97.

[27] Chris Lomont. *The Hidden Subgroup Problem - Review and Open Problems.* 2004. DOI: 10.48550/ARXIV.QUANT-PH/0411037. URL: https://arxiv.org/abs/quant-ph/0411037.

[28] Specifically a dimension one algebraic variety since it admits a group structure.

Namely, an elliptic curve is the set of solutions (x, y) to an equation of the form:[29] [30]

$$y^2 = x^3 + ax + b, \text{ for constants } a, b.$$

Importantly, the behavior of this curve is dependent on the 'field' on which one defines it. Field, in mathematics, refers to a structure where any two elements have a well-defined notion of addition, subtraction, multiplication and division (except by a 0 element). Notable examples include the rationals $\mathbb{Q}$, the reals $\mathbb{R}$, and the complex numbers $\mathbb{C}$. Not the integers since, take, for example: $1, 2 \in \mathbb{Z}$, but $\frac{1}{2} \notin \mathbb{Z}$. We will look at the case of the reals shortly, but this is an important distinction, hinting towards the possibility of a discrete context to come.

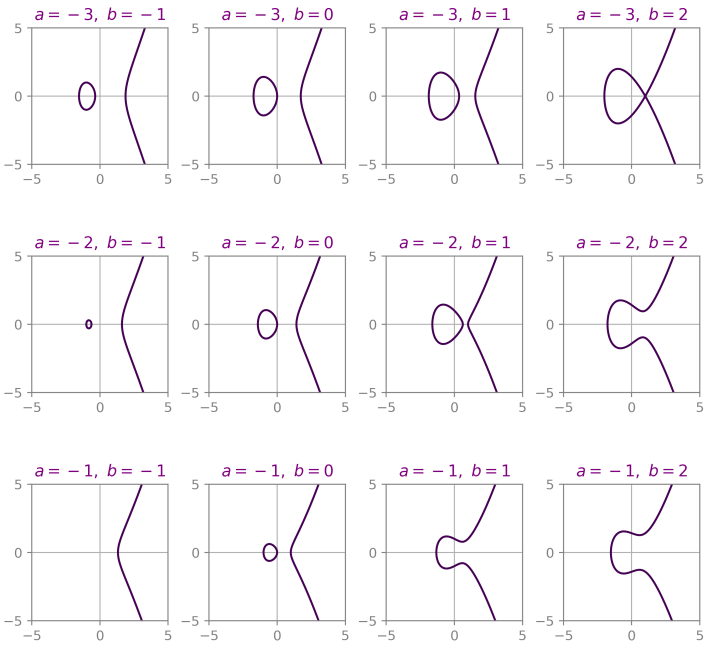


Figure 3: A few selected elliptic curves plotted over the reals $\mathbb{R}$ with matplotlib. In the form, $y^2 = x^3 + ax + b$, with a from -3 to 0 and b from -1 to 2.

[29] Koblitz, *A Course in Number Theory and Cryptography*, p. 10.

[30] The general form, $y^3 + a_1xy + a_2y = x^3 + a_3x^2 + a_4x + a_5$, can always be transformed to the reduced form preserving the group structure iff the field K is not $\mathbb{F}_2$ or $\mathbb{F}_3$.

5.2 Elliptic Curves over the Reals

In Figure 3, we see the general shape of an elliptic curve. This complicated shape arises from square rooting a general depressed cubic of the form $y = x^3 + px + q$. Without the x^2 term, this cubic is guaranteed to have only one inflection point at $x = 0$, since its second derivative $y'' = 6x$ does not depend on p or q, and since square roots preserve the inflection points. Furthermore, the square-rooting increases values from $0 < y < 1$ and decreases values when $y > 1$ with a fixed point at 1. Over $\mathbb{R}$, square roots are limited to inputs greater than 0, but since we take both branches ($\pm$) of the root, we see reflection symmetry across the y-axis (this will be important in a moment).[31] See Figure 4.

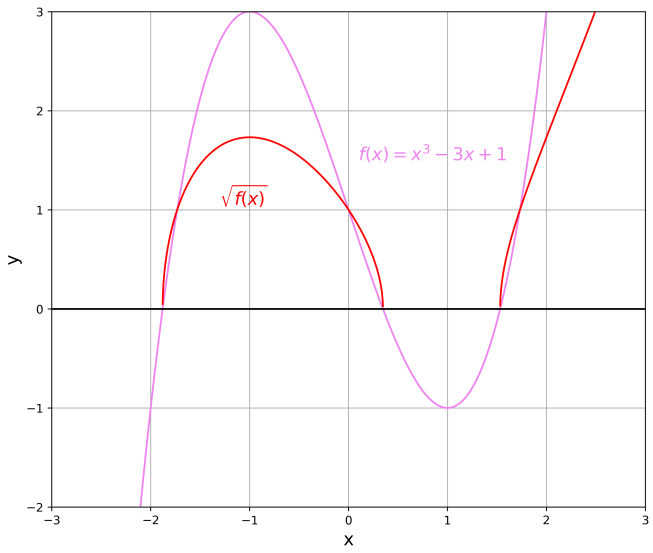


Figure 4: A depiction of how one branch of an elliptic curve arises from square-rooting a depressed cubic.

[31] Chris Sangwin Ngoc Vo Michael Haese Mark Humphries. *Mathematics: Analysis and Approaches HL*. Haese Mathematics. Haese Mathematics, 2019. ISBN: 1925489590.

5.3 Geometric Secant-Tangent Construction

It turns out, Elliptic curves are special because one can define a reasonable notion of addition with points on the curve. The one caveat is that we require a point 'at infinity' I to do so. Formally, this means were working in the projective real plane $\mathbb{RP}^2$ or $\mathbb{R}^2 \cup \{\infty\}$. But with this, we obtain the property that any secant drawn through two points on the curve must intersect another third point on the curve or this point at infinity I. Thus we will define the addition of two points P and Q to be this third point reflected across the y-axis.[32]

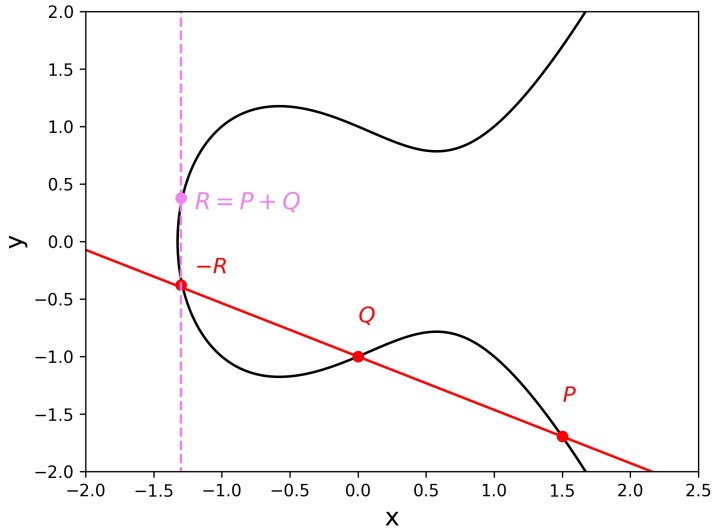


Figure 5: Point addition of two points P and Q, defined by reflecting the third point of intersection across the y-axis, on $y^2 = x^3 - x + 1$.

5.4 Addition of Two Points

We can make this idea rigorous by using some geometric tools. Our third point $-R$ is defined as the intersection of our elliptic curve $y^2 = x^3 + ax + b$ and some secant line $y = mx + l$. Given the points $P, Q = (x_p, y_p), (x_q, y_q)$ we can write down the slope:

$$m = \frac{y_p - y_q}{x_p - x_q}$$

[32]Note, without reflecting across the y-axis, this operation is not associative. It also makes sense since three colinear points sum to the point-at-infinity $A + B + C = I$ and hence, $A + B = -C$.

Proceeding with the simultaneous equations

$$y^2 = x^3 + ax + b \tag{1}$$
$$y = mx + l \tag{2}$$

Substituting (2) into (1),

$$\begin{aligned}(mx + l)^2 &= x^3 + ax + b \\ m^2x^2 + 2mlx + l^2 &= x^3 + ax + b \\ \Rightarrow 0 &= x^3 - (m^2)x - (2ml)x + (b - l^2)\end{aligned} \tag{3}$$

Now, this cubic has three real solutions, namely the roots of P, Q and R respectively. Writing these as factors:

$$(x - x_p)(x - x_q)(x - x_r) = x^3 - (m^2)x^2 - (2ml)x + (b - l^2)$$

The next step is to expand the left side out and equate coefficients.

$$\begin{aligned}LH &= x^3 - (x_p + x_q + x_r)x^2 + (x_px_q + x_px_r + x_qx_r)x - x_px_qx_r \\ RH &= x^3 \qquad\qquad - (m^2)x^2 \qquad\qquad\qquad - (2ml)x + (b - l^2).\end{aligned}$$

This means, using the purple coefficients,

$$\begin{aligned}m^2 &= x_p + x_q + x_r \\ \therefore \quad x_r &= m^2 - x_p - x_q.\end{aligned} \tag{4}$$

Substituting the x coordinate of $-R$, x_r into the point-slope form of the line with $m = \frac{y_p - y_q}{x_p - x_q}$ we obtain,

$$\begin{aligned}y_r &= y_q + m(x_r - x_q) \\ &= y_p + m(x_r - x_p).\end{aligned}$$

Flipping our result across the y-axis to obtain R, we have (finally),[33]

$$\begin{aligned}P \ + \ Q \ &:= \ R \\ (x_p, y_p) + (x_q, y_q) &:= \left(m^2 - x_p - x_q, \quad -y_q - m(x_r - x_q)\right),\end{aligned}$$

$$\text{where } m = \tfrac{y_p - y_q}{x_p - x_q}.$$

■

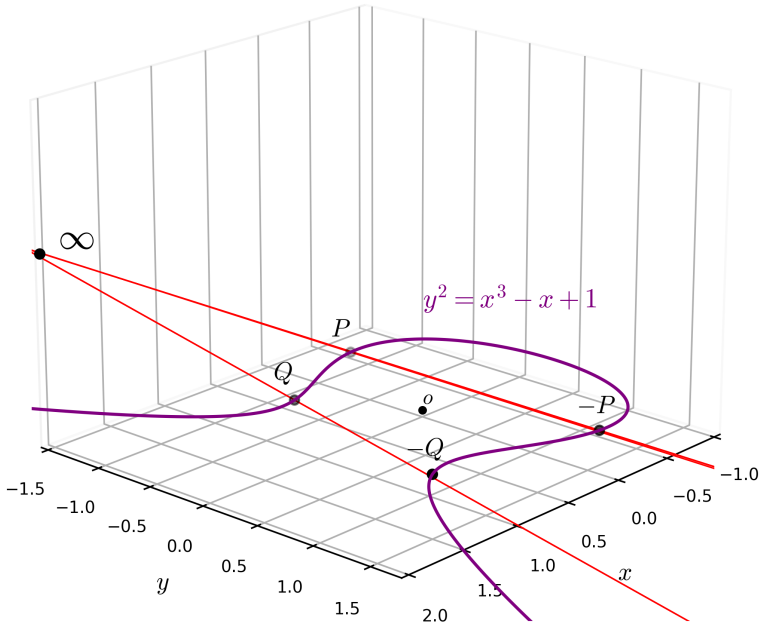


Figure 6: Parallel lines drawn through inverse points, all intersecting at point at infinity I over $y^2 = x^3 - x + 1$. As in, $P + (-P) = I = Q + (-Q)$.

5.5 Inverse Points and Infinity

In the case of adding two points opposite each other over the y-axis, one says they intersect the third point "at infinity." In general, over a projective plane, we revise the parallel postulate to say "any two parallel lines share one point of intersection, namely the point at infinity." We see this below; like train tracks meeting the horizon, the two parallel, vertical lines meet at ∞ or I.

Because of this construction, I acts as a left- and right-sided identity element. Since, for any ($\mathbb{R}$-rational) point A with inverse $-A$ on the elliptic-curve,

$$\begin{aligned} A + (-A) &= I \\ A + (-A) + A &= I + A \\ \Rightarrow \quad A + I &= I + A = A \end{aligned} \tag{5}$$

[33] As long as $P \neq Q$ and $P \neq -Q$.

5.6 Adding a Point to Itself

The case that was left out of the earlier pictorial representation of point addition is when a point is added to itself. Not to fear, however, since we can replace the secant through two points with a tangent through one. The resultant point is the resultant intersection flipped across the y-axis. The slope of this tangent can be found through implicit differentiation.

$$\begin{aligned} d(y^2) &= d(x^3 + ax + b) \\ 2ydy &= 3x^2dx + adx \\ \frac{dy}{dx} &= \frac{3x^2 + a}{2y}. \end{aligned}$$

Thus with,

$$y_r - y_p = m(x_r - x_p)$$

$$\text{where } m = \tfrac{3x_p^2+a}{2y_p},$$

we can use the same result as before with $P = Q$:

$$\begin{aligned} m^2 &= x_p + x_q + x_r \\ \Rightarrow \quad x_r &= m^2 - 2x_p \end{aligned} \tag{6}$$

and, using the point-slope form and flipping across the y-axis,

$$y_r = -y_p - m(x_r - x_p)$$

■

*A word on notation: the choice to use $+$ to denote this binary operation is entirely arbitrary, and many authors use alternate notation to represent this ($\times, \oplus, \otimes, \star$ etc.). Using $+$ hints that adding a point P to itself n times might be written nP, but for sake of consistency with the theory to come, I will write this P^n meaning $P + P + ... + P$, n-times. This will make sense in a moment.

6 Group structure of $E[K]$

6.1 Verifying the Group Axioms

By the higher powers of mathematical serendipity, this forms a group! A very special group at that. We can write this elliptic curve group as $E[\mathbb{R}]$ for the case in the reals. Or, $E[K]$ for any field K (again, a structure with $+, -, \times, \div$).

Now, we will (at least, heuristically) verify that the group axioms hold, we first see in eq (5) that we have a two-sided identity, I, given by the point at infinity. Hence,

$$P + I = I + P = I \text{ for all } P \in E[K]$$

Next, we prove the existence of an inverse for every element by noticing that our curve is symmetric about the y-axis. Similarly our identity element I is self-inverse. Hence,

$$\text{for all } P \in E[K], \text{ there exists an inverse } -P \text{ such that } P + (-P) = I$$

This operation is associative as well, but unfortunately a rigorous proof of this fact is unprecedentedly untame, especially through this geometric depiction of elliptic curves. With more powerful tools from algebraic geometry, this proof is almost trivial, following nicely from tracking values on points of a curve (namely $\text{Div}^0(E)$) and the relationship between poles and zeros, as stated by the Riemann-Roch theorem.[34] Lastly, this binary operation of adding two points is *commutative* as well, following geometrically from the secant construction or through the formulas.

As a result of meeting these conditions, $E[K]$ forms an abelian group under our binary operation.

6.2 Finite Fields $\mathbb{F}_p$

This elliptic curve group action holds for other fields K as well. Most important to cryptography is elliptic curves over "finite" fields. A finite field, denoted $\mathbb{F}_p$ for p prime,[35] is a set of ordered pairs (x, y) with addition, subtraction and multiplication taken *modulo* p.

As I alluded to earlier, the condition that p be prime is so that every element $n < p$ has a multiplicative inverse, since, to have a multiplicative inverse, $\gcd(n, p) = 1$. Really, finite fields are the elements of $\mathbb{Z}/p\mathbb{Z}$ with addition and multiplication in $(\mathbb{Z}/p\mathbb{Z})^+$ and $(\mathbb{Z}/p\mathbb{Z})^\times$. The field axioms readily follow from the group axioms of the multiplicative and additive groups.

6.3 Elliptic Curves over Finite Fields, $K = \mathbb{F}_p$

Using the same equations from before, except with operations taken *modulo* p, this elliptic curve still forms a group over a finite field, $E[\mathbb{F}_p]$. The elements of this group are the points,

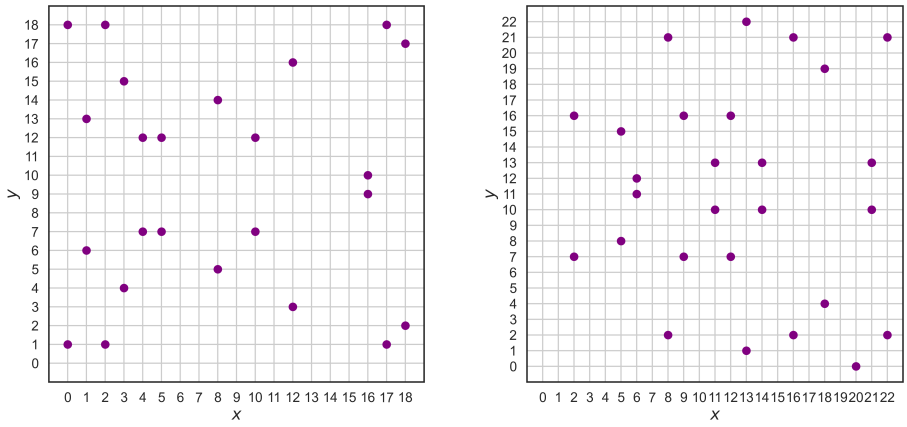
$$E[\mathbb{F}_p] = \left\{(x, y) \mid y^2 \equiv x^3 + ax + b \bmod p\right\} \cup \{\infty\}.$$

[34] Joseph H. Silverman. *The Arithmetic of Elliptic Curves*. Graduate Texts in Mathematics No. 106. Springer, 1986. ISBN: 9780387962030, p. 66.

[35] Technically, p can be a prime power. One might write $\mathbb{F}_q$ where $q = p^r$.

with the additional ∞ element serving as the identity.

Plotting examples of these groups over finite fields, we see the geometric intuition breaks down somewhat.



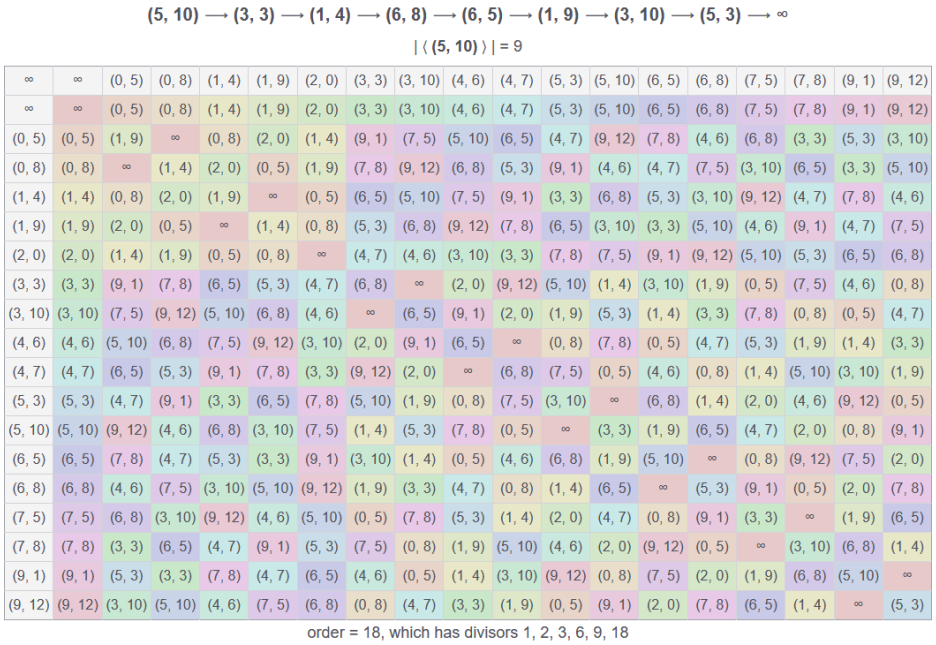
(a) $y^2 = x^3 + 15x + 1$ over $\mathbb{F}_{19}$ (b) $y^2 = x^3 + 12x + 17$ over $\mathbb{F}_{23}$

Figure 7: Select elliptic curves plotted over finite fields. Notice the symmetry about $\frac{p}{2}$.

In an attempt to better understand the group structure, I coded an interactive web-app in Typescript that computes the multiplication table of the group elements, known as a Cayley table. The symmetry about the diagonal indicates the commutativity as well.

Using this elliptic curve group, a new form of Diffie-Hellman can be formulated. Given a suitable generator point $g \in E[\mathbb{F}_p]$, all the repeated multiplications can be performed on this elliptic curve group. The benefit of doing this is that the structure of an elliptic curve group $E[\mathbb{F}_p]$ is much more complicated than that of $(\mathbb{Z}/q\mathbb{Z})^\times$. The process is similar to before: examining the *subgroup structure* of $E[\mathbb{F}_p]$ reveals the difficulty of uncovering the 'hidden subgroup' defined by the discrete logarithm. Simply put, when working in $(\mathbb{Z}/q\mathbb{Z})^\times$, there are number-theoretical properties making it easier to narrow down guesses, especially when computing a discrete logarithm. Diem, C. (2011) showed that that the discrete logarithm problem for elliptic curves has an asymptotic time complexity greater than that of $(\mathbb{Z}/q\mathbb{Z})^\times$.[36] The culmination of this research into elliptic curve cryptography is the fact that most public key protocols utilize elliptic curves in one way or another.

[36]Claus Diem. "On the discrete logarithm problem in elliptic curves". In: *Foundation Compositio Mathematica* 147 (2010), pp. 75–104. DOI: `https://doi.org/10.1112/S0010437X10005075`, p. 76.



∞	∞	(0, 5)	(0, 8)	(1, 4)	(1, 9)	(2, 0)	(3, 3)	(3, 10)	(4, 6)	(4, 7)	(5, 3)	(5, 10)	(6, 5)	(6, 8)	(7, 5)	(7, 8)	(9, 1)	(9, 12)
∞	∞	(0, 5)	(0, 8)	(1, 4)	(1, 9)	(2, 0)	(3, 3)	(3, 10)	(4, 6)	(4, 7)	(5, 3)	(5, 10)	(6, 5)	(6, 8)	(7, 5)	(7, 8)	(9, 1)	(9, 12)
(0, 5)	(0, 5)	(1, 9)	∞	(0, 8)	(2, 0)	(1, 4)	(9, 1)	(7, 5)	(5, 10)	(6, 5)	(4, 7)	(9, 12)	(7, 8)	(4, 6)	(6, 8)	(3, 3)	(5, 3)	(3, 10)
(0, 8)	(0, 8)	∞	(1, 4)	(2, 0)	(0, 5)	(1, 9)	(7, 8)	(9, 12)	(6, 8)	(5, 3)	(9, 1)	(4, 6)	(4, 7)	(7, 5)	(3, 10)	(6, 5)	(3, 3)	(5, 10)
(1, 4)	(1, 4)	(0, 8)	(2, 0)	(1, 9)	∞	(0, 5)	(6, 5)	(5, 10)	(7, 5)	(9, 1)	(3, 3)	(6, 8)	(5, 3)	(3, 10)	(9, 12)	(4, 7)	(7, 8)	(4, 6)
(1, 9)	(1, 9)	(2, 0)	(0, 5)	∞	(1, 4)	(0, 8)	(5, 3)	(6, 8)	(9, 12)	(7, 8)	(6, 5)	(3, 10)	(3, 3)	(5, 10)	(4, 6)	(9, 1)	(4, 7)	(7, 5)
(2, 0)	(2, 0)	(1, 4)	(1, 9)	(0, 5)	(0, 8)	∞	(4, 7)	(4, 6)	(3, 10)	(3, 3)	(7, 8)	(7, 5)	(9, 1)	(9, 12)	(5, 10)	(5, 3)	(6, 5)	(6, 8)
(3, 3)	(3, 3)	(9, 1)	(7, 8)	(6, 5)	(5, 3)	(4, 7)	(6, 8)	∞	(2, 0)	(9, 12)	(5, 10)	(1, 4)	(3, 10)	(1, 9)	(0, 5)	(7, 5)	(4, 6)	(0, 8)
(3, 10)	(3, 10)	(7, 5)	(9, 12)	(5, 10)	(6, 8)	(4, 6)	∞	(6, 5)	(9, 1)	(2, 0)	(1, 9)	(5, 3)	(1, 4)	(3, 3)	(7, 8)	(0, 8)	(0, 5)	(4, 7)
(4, 6)	(4, 6)	(5, 10)	(6, 8)	(7, 5)	(9, 12)	(3, 10)	(2, 0)	(9, 1)	(6, 5)	∞	(0, 8)	(7, 8)	(0, 5)	(4, 7)	(5, 3)	(1, 9)	(1, 4)	(3, 3)
(4, 7)	(4, 7)	(6, 5)	(5, 3)	(9, 1)	(7, 8)	(3, 3)	(9, 12)	(2, 0)	∞	(6, 8)	(7, 5)	(0, 5)	(4, 6)	(0, 8)	(1, 4)	(5, 10)	(3, 10)	(1, 9)
(5, 3)	(5, 3)	(4, 7)	(9, 1)	(3, 3)	(6, 5)	(7, 8)	(5, 10)	(1, 9)	(0, 8)	(7, 5)	(3, 10)	∞	(6, 8)	(1, 4)	(2, 0)	(4, 6)	(9, 12)	(0, 5)
(5, 10)	(5, 10)	(9, 12)	(4, 6)	(6, 8)	(3, 10)	(7, 5)	(1, 4)	(5, 3)	(7, 8)	(0, 5)	∞	(3, 3)	(1, 9)	(6, 5)	(4, 7)	(2, 0)	(0, 8)	(9, 1)
(6, 5)	(6, 5)	(7, 8)	(4, 7)	(5, 3)	(3, 3)	(9, 1)	(3, 10)	(1, 4)	(0, 5)	(4, 6)	(6, 8)	(1, 9)	(5, 10)	∞	(0, 8)	(9, 12)	(7, 5)	(2, 0)
(6, 8)	(6, 8)	(4, 6)	(7, 5)	(3, 10)	(5, 10)	(9, 12)	(1, 9)	(3, 3)	(4, 7)	(0, 8)	(1, 4)	(6, 5)	∞	(5, 3)	(9, 1)	(0, 5)	(2, 0)	(7, 8)
(7, 5)	(7, 5)	(6, 8)	(3, 10)	(9, 12)	(4, 6)	(5, 10)	(0, 5)	(7, 8)	(5, 3)	(1, 4)	(2, 0)	(4, 7)	(0, 8)	(9, 1)	(3, 3)	∞	(1, 9)	(6, 5)
(7, 8)	(7, 8)	(3, 3)	(6, 5)	(4, 7)	(9, 1)	(5, 3)	(7, 5)	(0, 8)	(1, 9)	(5, 10)	(4, 6)	(2, 0)	(9, 12)	(0, 5)	∞	(3, 10)	(6, 8)	(1, 4)
(9, 1)	(9, 1)	(5, 3)	(3, 3)	(7, 8)	(4, 7)	(6, 5)	(4, 6)	(0, 5)	(1, 4)	(3, 10)	(9, 12)	(0, 8)	(7, 5)	(2, 0)	(1, 9)	(6, 8)	(5, 10)	∞
(9, 12)	(9, 12)	(3, 10)	(5, 10)	(4, 6)	(7, 5)	(6, 8)	(0, 8)	(4, 7)	(3, 3)	(1, 9)	(0, 5)	(9, 1)	(2, 0)	(7, 8)	(6, 5)	(1, 4)	∞	(5, 3)

Figure 8: App on my personal website computing the Cayley table of $y^2 = x^3 + 3x + 12$ over $\mathbb{F}_{17}$. Clicking on an element computes the subgroup it generates, the focused point being $(5, 10)$ here.

7 Conclusion

In this paper, I sought to explore **to what extent does group theory provide a mathematical backdrop for modern public-key cryptography?** From my findings, it's clear that group theory *does* indeed provide a rich backdrop for discussing cryptographic concepts. Especially since, as computing power grows cheaper, the use of group-theoretical techniques to strengthen existing principles is paramount. For instance, the Diffie-Hellman protocol works on any abelian group. Further research is underway at examining other candidates for 'secure' groups, such as the study of hyperelliptic curves and other higher-genus abelian varieties (a solution set to a polynomial forming an abelian group).

In the process of researching, it's clear that there are two different mathematical perspectives introducing elliptic curves. The decision was made to continue with the simpler route, defining the elliptic curve as the solution to a polynomial equation. However, this is not strictly necessary, since elliptic curves arise in many other contexts, like through the intersection of two quadric surfaces (like a sphere and an hyperboloid for instance). The other route is a more theoretical perspective, using tools from algebraic geometry

and topology to reach the same conclusions. The benefit of this advanced approach is that tools can be *generalized* to reveal new perspectives. Doing so also would decrease the 'arbitrary-ness' of some of the topics explored in this paper. Coming at a topic from the top-down can make certain difficult proofs and perspectives trivial, as I mention a couple times in the paper.

Another limitation of this paper is the imprecise comparison between elliptic curve cryptography and traditional cryptography. Going into it, I intended to explore the nitty-gritty details of the elliptic curve group structure, detailing exactly why the hidden-subgroup problem and discrete logarithm problem were more difficult. This proved to be much more challenging than anticipated, and, although I satisfied some of my own curiosity, I realized it would be an impractical undertaking to convey the wealth of prerequisites to reach the conclusions I wanted. As a result, my research lacks the 'point' I wanted it to have, rather opting for a more expository lens on this complicated subject. Regardless, I find this subject fruitful to learn because of its approachability and its underappreciated influence on modern-day life.

Bibliography

A History of U.S. Communication Security. Vol. 1. The David G. Boak Lectures. National Security Agency, 1973. URL: https://www.governmentattic.org/18docs/Hist_US_COMSEC_Boak_NSA_1973u.pdf.

al., Elaine Barker et. "Recommendation for Pair-Wise Key Establishment Using Integer Factorization Cryptography". In: *National Institute of Standards and Technology* 800-56B (2018). DOI: 10.6028/NIST.SP.800-56Br2. URL: https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-56Br2.pdf.

Aschbacher, Michael. *The Status of the Classification of the Finite Simple Groups*. Vol. 51. Notices of the American Mathematical Society 7. American Mathematical Society, 2004, pp. 736–740. URL: https://www.ams.org/notices/200407/fea-aschbacher.pdf.

Bunimov, Viktor. "Area and time efficient modular multiplication of large integers". In: *Proceedings IEEE International Conference on Application-Specific Systems, Architectures, and Processors. ASAP 2003* (2003). DOI: 10.1109/ASAP.2003.1212863.

David S. Dummit, Richard M. Foote. *Abstract Algebra*. 3rd ed. Wiley, 2004. ISBN: 9780471433347.

Diem, Claus. "On the discrete logarithm problem in elliptic curves". In: *Foundation Compositio Mathematica* 147 (2010), pp. 75–104. DOI: https://doi.org/10.1112/S0010437X10005075.

Elgamal, Taher. "A Public-Key Cryptosystem and a Signature Scheme Based on Discrete Logarithms". In: *IEEE Transactions on Information Theory* 31.4 (1985), pp. 469–472. DOI: 10.1109/TIT.1985.1057074.

Koblitz, Neal. *A Course in Number Theory and Cryptography*. 2nd ed. Graduate Texts in Mathematics No. 114. Springer, 1994. ISBN: 9781461264422.

Lewand, Robert Edward. "The Perfect Cipher". In: *The Mathematical Gazette* 94.531 (2010), pp. 401–411. ISSN: 00255572. URL: http://www.jstor.org/stable/25759724 (visited on 05/21/2022).

Lomont, Chris. *The Hidden Subgroup Problem - Review and Open Problems*. 2004. DOI: 10.48550/ARXIV.QUANT-PH/0411037. URL: https://arxiv.org/abs/quant-ph/0411037.

Michael Haese Mark Humphries, Chris Sangwin Ngoc Vo. *Mathematics: Analysis and Approaches HL*. Haese Mathematics. Haese Mathematics, 2019. ISBN: 1925489590.

Pomerance, Carl. "A Tale of Two Sieves". In: *Notices of the AMS* 43.12 (1996), 1473–1485.

Saracino, Dan. *Abstract Algebra*. 2nd ed. 2008. ISBN: 1577665368.

Shannon, Claude E. *Communication Theory of Secrecy Systems*. Vol. 28. The Bell System Technical Journal, 1949, pp. 656–715. DOI: 10.1002/j.1538-7305.1949.tb00928.x.

Silverman, Joseph H. *The Arithmetic of Elliptic Curves*. Graduate Texts in Mathematics No. 106. Springer, 1986. ISBN: 9780387962030.

Thimbleby, Harold. "Human factors and missed solutions to Enigma design weaknesses". In: *Cryptologia* 40.2 (2016), pp. 177–202. DOI: 10.1080/01611194.2015.1028680.

Weinberg, Erick J. *Classical Solutions in Quantum Field Theory*. Cambridge University Press, 2012, pp. 81–107. ISBN: 9781139017787. DOI: 10.1017/CBO9781139017787.