Why Do We Care About Primes?

Alex Kolar

May 2024

Abstract

This paper looks into the multifaceted world of prime numbers and their significance in cryptography, mathematics education, and beyond. Beginning with a historical overview, it explores the foundational role of primes in number theory, highlighting key theorems and conjectures. The paper then delves into the practical applications of primes, focusing on their crucial role in modern cryptographic systems such as RSA and Diffie-Hellman key exchange. Through a examination of these cryptographic techniques, it illustrates how the unique properties of prime numbers underpin the security of digital communication. It also explores the connections between prime numbers and diverse fields such as nature, mathematics education, and fractals. To conclude, the paper's purpose is to emphasize the pervasive influence of prime numbers, emphasizing their fundamental importance in society.

1 Introduction

The prime numbers are everywhere, or so we are told. Prime numbers are a staple of every elementary mathematics education, yet I feel as if our education system has failed us with providing any "real world" applications of prime numbers in mathematics courses. They tend to pop up as counterexamples or display a weird property when plugged into an equation. While this is interesting in a purely mathematical sense, I wanted to investigate places where prime numbers show up in our daily lives. Primes are introduced as a fundamental math topic, but do they share a similar importance to concepts such as addition and multiplication? In this paper, I want to expose the reader to cryptography, the main way by which humans have adapted the properties of prime numbers into a useful tool for randomization and secrecy.

1.1 What are Primes?

It would be remiss of this author to assume the reader already possesses the necessary information to follow along with the remainder of this paper, so I will take the time to lay out a foundation of the topic here.

Simply put, prime numbers are positive integers greater than 1 whose only divisors are 1 and itself. In set notation we have $P = \{x \in N | x \bmod y = 0 \implies y = 1 \text{ or } y = x, x > 1, y \in N\}$. There are infinitely many prime numbers, and every integer has a unique prime factorization, that is, given $x \in Z$, we can write $x = p_1 \cdot p_2 \cdots p_n$ where $n \in N$ and $p_1, \cdots p_n \in P$. This fact is known as the Fundamental Theorem of Arithmetic. This unique prime factorization is very difficult to determine given a sufficiently large number, a fact that we will use later on in the application section.

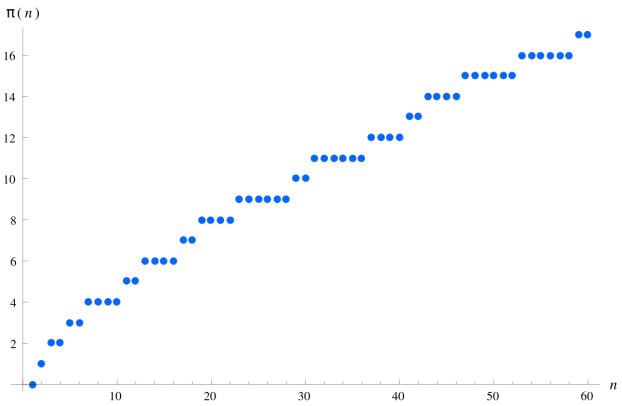

Figure 1: Prime Counting Function

1.2 History of Primes

The first civilization to extensively research the prime numbers was the Greeks. In Euclid's famous book *Elements*, he formulated the Fundamental Theorem of Arithmetic. It is noteworthy that Euclid lived in a time before the prevalence of algebra, meaning he defined prime numbers geometrically: "A number measured by a unit alone". Roughly 100 years after Euclid published elements, another Greek mathematician Eratosthenes devised an algorithm to calculate all of the primes less that an integer n which is now know as the Sieve of Eratosthenes.

After these rather elementary results, there was a gap of nearly 2000 years before a significant result in prime number theory appeared, this gap is commonly know as the Dark Ages.

Pierre de Fermat is now considered one of the great mathematicians of the 1600s, and his most famous result involving prime numbers is essential for the applications discussed today. Fermat's Little Theorem states for any integer a and any prime p, we have the following:

$$a^{p-1} \equiv 1 \bmod p$$

We will use this result specifically in RSA cryptography.

After Fermat's Theorem, a young Carl Friedrich Gauss stumbled across prime number theory. He was particularly interested in the distribution of the prime numbers, so he studied the Prime Counting Function, as shown in Figure 1. The Prime Counting Function is positive function which the inputs are positive integers and the outputs are the number of primes less than that integer. Gauss wanted to model this function and fortunately, he was concurrently working on a problem involving the logarithmic integral. Thus, he derived the Prime number Theorem, which says the distribution of the prime numbers is relatively regular. That is, given a positive integer n, the number of primes less than n is roughly $\frac{1}{log(n)}$.

Continuing the theme of hunting for a way to represent the distribution of primes, in 1859, German mathematician Bernhard Riemann introduced a new way to model the primes, this time based on the zeros of a complex function and the logarithmic integral. This model provides a much better representation of the Prime Counting Function. The model's true nature however can not be fully verified without the solution to the Riemann Hypothesis.

1.3 Famous Prime Problems

The Riemann Hypothesis isn't the only famous problem closely related to the properties of prime numbers. Three of the most notable open questions in mathematics related to patterns within the set of prime numbers

1.3.1 Goldbach Conjecture

The Goldbach Conjecture is one of the oldest unsolved problems in mathematics. It was a simple observation conjectured by Christian Goldbach in 1742. It states that every even number is the sum of two prime numbers. That is, every even number has a unique decomposition into a sum of primes. No counterexample has been found, and no proof has been discovered.

1.3.2 Twin Prime Conjecture

It is known that generally the gaps between the primes get larger the further one looks down the number line. In a way, the Twin Prime Conjecture contradicts that statement. The Twin Prime Conjecture states that there are infinitely many primes that differ by 2 from another prime. Primes that differ from another prime by 2 are known as twin primes. For example, 11 and 13 are a twin prime pair.

1.3.3 Beal Conjecture

The Beal Conjecture is slightly lesser know and a modification of the equation popularlized by Fermat and Pythagoras.

$$A^x + B^y = C^z$$

Beal's Conjecture is if the above statement holds true for positive integers A, B, C and $x, y, z \geq 3$, then A, B, C must have a common prime factor.

The introduction of these conjectures serves to emphasize the magnitude of the work being done by prominent mathematicians on the properties of the primes. In a more general sense, the properties of the primes are important because within every integer are hidden primes, that is, it has a prime factorization. So in a round about way, all integers are primes in disguise, proving things about the prime numbers can give us deeper insights into the integers as a whole, which in itself is important. However, that is not the complete purpose of this paper, as we still want to examine the more direct impact primes have on our lives.

2 Cryptography

2.1 Cybersecurity

Before cryptography can be introduced, we need to first ensure the basics of cybersecurity are understood. Cybersecurity is the practice by which we protect devices and data from unauthorized access.

It is very easy for an entity to monitor your activity on the internet. It is necessary for someone or thing see the data that you send and receive to make sure your data is sent to and received from the proper recipient and sender. So that raises the question: How do we

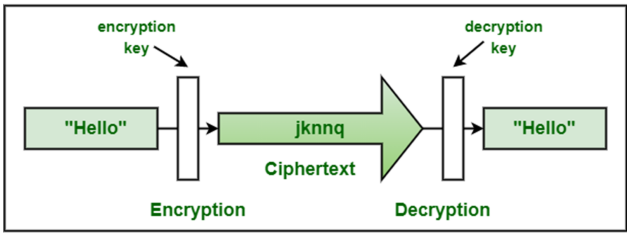

Figure 2: Basic Asymmetric Encryption Sequence

ensure that these "men in the middle" can't unlawfully see our data? The answer lies in cryptography.

Cryptography is the art of writing or solving codes, that is, turning plain text into some sort of incomprehensible sequence of characters that for our purposes we will refer to as cipher text, then reverting the cipher text to plain text when the correct entity receives it. Encryption is absolutely vital for the internet to function as it does today, as it protects all messages, passwords, queries etc. from unwanted attention. Figure 2 shows the basic sequence by which an encryption/decryption algorithm would work. A message is encrypted by one user using an encryption key, and then that cipher text is sent to another user. When that user receives the text, they use a decryption key to decrypt that cipher text, turning it back to plain text. This scheme is know as *asymmetric* encryption because the encryption and decryption keys are distinct. So now the question becomes how can we generate a key pair? When thinking about the components of encryption algorithms like plain text, cipher text, and keys, it is useful think of text as numbers and keys as operations on those numbers. In the world of computers, text is converted to a sequence of numbers before it can be sent.

2.2 RSA

The most popular key generation algorithm is RSA. In 1977 three mathematicians from MIT, Ron Rivest, Adi Shamir, and Leonard Adleman, formulated a way to generate an infinite number of key pairs using the properties prime numbers. RSA is prevalent anywhere the internet is used, and the key generation algorithm is as follows.

Begin by choosing two primes p and q. In practice, these primes are between 1024 and 2048 bits long, meaning between 309 and 617 digits. For the examples, we will use smaller primes for simplicity. After choosing the primes, we compute $n = pq$ and $(p-1)(q-1)$. Then we choose a number that is coprime to $(p-1)(q-1)$, we will call this number e. Next, we will find d such that $de \equiv 1 \mod (p-1)(q-1)$. Now we have our public and private keys, with the public key being (n, e) and private key being (d, p, q).

Now we can encrypt a message m into cipher text c in the following way:

$$m^e \mod n = c \tag{1}$$

And decrypt:

$$c^d \mod n = m \tag{2}$$

To show an example of this algorithm, we can begin with two (small) primes $p = 7$, $q = 11$, their product $pq = n = 77$, and $(p-1)(q-1) = 60$. Next, a number coprime to 60

is 7, so $e = 7$. Finally, we can calculate that $7 \cdot 43 = 301 \bmod 60 = 1$. Thus, $d = 43$. So now we have a public private key pair, public: $(60, 7)$ private: (43,7,11).

Let's say we want to send a simple private message using this key pair. The message we want to send is 8. The sender should use the formula in Equation 1 to compute the cipher text. $8^7 \bmod 77 = 57$. It should be noted here that modular exponentiation is not a bottleneck in this process as the result can be calculated without calculating m^e using a algorithm known as binary exponentiation.

Now the recipient can decrypt the message using Equation 2. $57^43 \bmod 77 = 8$.

A very important observation here that we always use the recipient's public key to encrypt the message, so the recipient can use their private key to decrypt the message.

2.3 RSA Proof

To exemplify the importance of primes in this algorithm, we'll go through a brief proof. First, however, we should isolate the claim that is being made. In RSA, we are using that for integers m and c with $c = m^e \bmod n$, where n is the product of primes p and q and e is a natural number coprime to $(p-1)(q-1)$, that we have $c^d \bmod n = m$, where d is a number such that $de \equiv 1 \bmod n$. This is equivalent to the claim $(m^e)^d \equiv m \bmod n$, so this is what we will prove.

Claim: $(m^e)^d \equiv c^d \equiv m \bmod n$

Proof. First we should note that e and d always exists. e exists because every number has at least one number less than it but also coprime to it. d exists because if we consider the unit group $U((p-1)(q-1))$, we know that $e \in U((p-1)(q-1))$ by definition. Thus as e is an element in a group, it must also have a multiplicative inverse in that group. That multiplicative inverse is d.

Using the definition of the modulus, we get the following result:

$$de = 1 + (p-1)(q-1)$$

Thus,

$$\begin{aligned} m^{ed} &= m^{1+(p-1)(q-1)} \\ &= m \cdot m^{(p-1)(q-1)} \\ &= m \cdot (m^{(p-1)})^{(q-1)} \\ &\equiv m(1)^{q-1} \bmod p && \text{(Fermat's Little Theorem)} \\ &\equiv m \bmod p \end{aligned}$$

Similarly,

$$\begin{aligned} m^{ed} &= m^{1+(p-1)(q-1)} \\ &= m \cdot m^{(p-1)(q-1)} \\ &= m \cdot (m^{(q-1)})^{(p-1)} \\ &\equiv m(1)^{p-1} \bmod q && \text{(Fermat's Little Theorem)} \\ &\equiv m \bmod q \end{aligned}$$

Thus, we have $m^{ed} \equiv m \bmod p$ and $m^{ed} \equiv m \bmod q$. Now consider the system of equations:

$$x \equiv m \bmod p$$
$$x \equiv m \bmod q$$

The Chinese Remainder Theorem tells us that this system of equations has a unique solution for x in modulo pq. We know due to the reflexive property of equivalence relations that:

$$m \equiv m \bmod p$$
$$m \equiv m \bmod q$$

and we just showed that:

$$m^{ed} \equiv m \bmod p$$
$$m^{ed} \equiv m \bmod q$$

Thus by the Chinese Remainder Theorem we have $m^{ed} \equiv m \bmod pq$, which equates to $m^{ed} \equiv m \bmod n$, as desired.

□

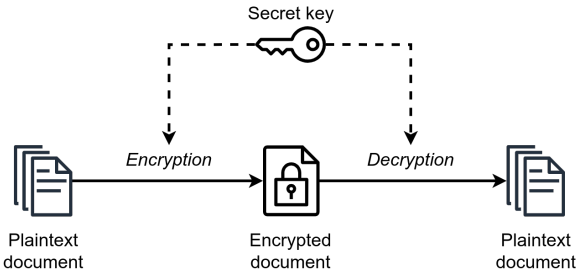

Figure 3: Basic Symmetric Encryption Sequence

2.4 Symmetric Cryptography

RSA is an example of asymmetric cryptography, meaning that two keys are required to complete the encryption/decryption sequence. Another popular form of encryption is known as symmetric cryptography. Symmetric cryptography uses the same key for encryption and decryption, which allows it to work faster. Thus, it is typically used on large blocks of stored data that need to be able to be retrieved quickly. However, say for example Bob and Alice have data that they need to encrypt using the same symmetric encryption key. Since this key is used for encryption and decryption, it can't be public like in RSA, otherwise an attacker would be able to both read and write ciphertext. So, Bob and Alice need a secure way to generate a key over an open communication channel. This simplifies to both parties semi-independently generating the same pseudorandom number (as again we should remember that keys in cryptography are just very large numbers). Once again, we can resort to the properties of the prime numbers to solve our problem.

2.5 Diffie-Hellmann

In 1976, Whitfield Diffie and Martin Hellman detailed an algorithm to secretly exchange a symmetric key. The procedure is as follows:

To begin, Alice and Bob mutually decide upon two numbers, g and n. g is a relatively small integer called a generator, and n is a large prime number. Then Alice decides upon a number a, such that $1 < a < n$, and Bob decides upon a number b, such that $1 < b < n$. a is private to Alice and b is private to Bob. Alice can then compute $g^a \bmod n$ and send the result over the open communication channel, and Bob can compute $g^b \bmod n$ and send it to Alice. Now the numbers in the public domain are $g, n, g^a \bmod n$, and $g^b \bmod n$. a and b are private to Alice and Bob respectively. Now Alice can take $g^b \bmod n$ and raise it to the a, yielding $(g^b)^a \bmod n$ or $g^{ab} \bmod n$. Likewise, Bob can take $g^a \bmod n$ and raise it to the b, yielding $(g^a)^b \bmod n$ or $g^{ab} \bmod n$. Thus, both Alice and Bob have arrive at the number $g^{ab} \bmod n$ without ever directly sending it over the channel. Now they can use it as their key for symmetric encryption/decryption.

While it is apparent that primes play a role in the Diffie-Hellmann Key Exchange, it is not obvious why they are used, and why this algorithm provides security.

2.6 Primes and the Discrete Logarithm Problem

When we perform the Diffie-Hellmann Key Exchange, we rely on one key component for security. That is, we hope that an attacker given $g^a \bmod n$ cannot compute what a is. This computation is known as the discrete log problem. Of course, we are familiar with the continuous log problem, that is, given $g^x = b$, we have efficient methods to compute $x = log_g(b)$, but with the discrete log problem, given $g^x \bmod n = b$, we need to solve $x = dlog_g(b)$. It is unable to obtain this d efficiently using any of the algorithms developed throughout the history of computer science. The discrete log problem is know as an exponential problem, meaning the time it takes for a computer to solve the problem increases exponentially as the size of n increases. There is no exact proof that this problem can only be solved by exponential algorithms, but a significant amount of effort has been put in trying to find an efficient solution, but we've come up empty handed.

Looking into the math behind Diffie-Hellmann, we first notice that we begin by choosing a large prime number n, where n is of the form $2p + 1$ with p being prime, and a generator g. Primes of the form $2p + 1$ are known as safe primes and are the ideal candidates for the primes used in the Diffie-Hellmann Key Exchange. We choose n to be prime to create a group. We know from group theory that Z_n (all of the positive integers less than n) form a group under the operation multiplication mod n. g is chosen to be a number that "generates" Z_n, that is every element in Z_n can be written as some power of g. For example, given prime $n = 7$ ($n = 2(3) + 1$). We can find a generator of Z_7 because we know that the order of a generator is equal to the size of the group. The size of Z_7 is 6 ($2p$ or $2{\cdot}3$). We also know that the order of every element divides size the group. Since we chose n to be $2p + 1$, we know that the elements must have order 2, p, or $2p$ or 2,3, or 6 in our example (we will disregard the element of order 1). Thus we just need to find an element that is not of order 2 or p. Going back to our example we are looking for elements that are not of order 2 or 3, consider $2 \in Z_7$. $2^2 \equiv 4 \bmod 7$, $2^3 \equiv 1 \bmod 7$, thus 2 has order 3. Next consider $3 \in Z_7$. $3^2 \equiv 2 \bmod 7$, $3^3 \equiv 6 \bmod 7$, so 3 does not have order 2 or 3, thus it must have order 6 and is therefore our generator. Now that we know that 3 is a generator, we know that $3^x \bmod 7$ has the potential to be any element in Z_7 for $1 < x < 8$.

To exemplify the importance of primes in this scheme, if we try the same process with a nonprime n, say 8, we know from group theory that Z_8 is not a group under the operation of multiplication, that is because the elements 2,4, and 6 do not have an inverse. We could try with just the unit group of 8 under multiplication, $U(8) = \{1, 3, 5, 7\}$, but we can easily see that this group is not cyclic as each element is it's own inverse meaning there are no generators. To generalize, we know that for every nonprime n, Z_n is not a group. Additionally, the Primitive Root Theorem tells us that $U(n)$ is not a cyclic group unless $n = 1, 2, 4, p^k$, or $2p^k$ for some odd prime p and $k \geq 1$. Of course, if $n = p$, $U(n) = Z_n$. Thus, the security of the discrete log problem falls apart when we consider nonprime n's.

3 Connections and Reflections

For the last part of this paper, I'd like to look into the connections between prime numbers and some of the topics we covered in the MATH400 class this semester. As primes are somewhat of an omnipresent mathematical topic, they pop up in various scenarios.

3.1 Nature

It is famously known that prime numbers show up in the life cycles of cicadas. Cicadas only come out every 7,13, or 17 years. It has been conjectured that this is because they evolved to avoid the life cycles of their predators. For example, if they came out every 8 years, they would align with predators who have 1 year life cycles, 2 year life cycles, 4 year life cycles, and 8 year life cycles. Instead, they only direct interact with predators with 1 and 7 year life cycles.

3.2 Learning Math

Since prime numbers are such a foundational part of math, it makes sense that there is some sort of correlation between an individual's understanding of prime numbers and their mathematical proficiency. Studies have show that certain students with Autism Spectrum Disorder (ASD) have an innate ability to classify numbers as prime. It has been theorized that this is because subjects with ASD have some sort of system that allows them to visual numbers as image and group them in a way that could be seen as similar to factoring (5). Another story was told by a famous British neurologist named Oliver Sacks who challenged two young men with ASD to game of naming large prime numbers. According to Sacks, the participants were able to name prime numbers with 10 digits. This research signifies that there is some sort of fundamental ability associated with prime numbers.

3.3 Fractals

There's also a connection between the distribution of prime numbers and fractals. The Ulam Spiral, for example, visually represents the distribution of prime numbers in a spiral pattern, which exhibits fractal like characteristics. Fractal theory is interestingly interlinked with number theory, and it likely that more discoveries will lead to further knowledge on the distributions of the primes, as an exhorbitant amount of research is being done on the subject.

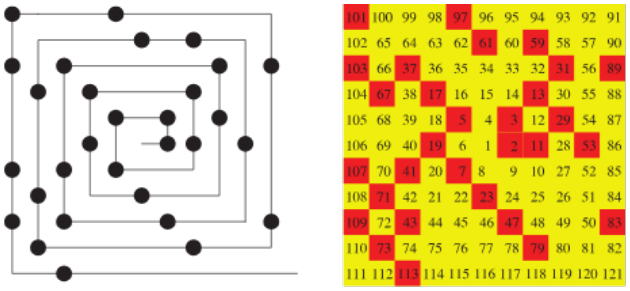

Figure 4: Ulam Spiral

3.4 Reflections

To answer the question I posed at the beginning of the paper, I think primes are of utmost importance in a mathematical education. With the research for this paper I really wanted to look for "real world" applications of the primes, but I wasn't coming up with a lot that I would have considered practical. Of course, there was the cryptography aspect of the primes which was very interesting, but other than that, I wasn't really coming up with much. However, I think somewhere along the way I realized that my idea of "practicality" was a bit unreasonable. At this point, I was too far in to rework the entire premise, but I did want to mention something here. My standard for practicality was a bit naive. I think I now realize that there is practicality in the fact that primes simply exist. Without primes, there simply is no number system. I think in the in class presentation, it would have been nice to incorporate this arc into lecture.

References

[1] B. Lynn, "Generators," Number Theory - Generators, https://crypto.stanford.edu/pbc/notes/numbertheory/gen.html (accessed May 11, 2024).

[2] C. Postal, "How diffie-hellman key exchange provides encrypted communications: Upguard," UpGuard, https://www.upguard.com/blog/diffie-hellman. (accessed May 11, 2024).

[3] D. Rountree, Security for Microsoft Windows System Administrators: Introduction to Key Information Security Concepts. Amsterdam: Syngress, 2011.

[4] D. R. Guichard, "When is U(n) Cyclic An Algebraic Approach," Mathematics Magazine, vol. 72, no. 2, pp. 139–142, Apr. 1999.

[5] Loconsole, M., Regolin, L. Are prime numbers special? Insights from the life sciences. Biol Direct 17, 11 (2022). https://doi.org/10.1186/s13062-022-00326-w

[6] "Symmetric and asymmetric encryption: Which is better?," Encryption Consulting, https://www.encryptionconsulting.com/education-center/symmetric-vs-asymmetric-encryption/. (accessed May 11, 2024).