Math 116 Pre-Studying

Michael Pham

Spring 2025

Contents

CHAPTER 1

AN INTRODUCTION TO CRYPTOGRAPHY

1.1 Simple Substitution Ciphers

1.1.1 Caesar (Shift) Ciphers

To begin with, we will explore one of the most basic ciphers: the *Caesar Cipher*. In this cipher, each letter in our unencrypted message (the *plaintext*) gets shifted by some amount to encrypt our message (the *ciphertext*). A basic example of a Caesar Cipher is seen in Example 1.1.

Example 1.1 (Basic Caesar Cipher Walkthrough). Suppose we had the following ciphertext:

```
j s j r d k f q q n s l g f h p g w j f p y m w t z l m n r r n s j s y q z h n z x
```

Then, if we knew that it had been shifted up by five letters, we can apply the reverse, yielding us:

```
j s j r d k f q q n s l g f h p g w j f p y m w t z l m n r r n s j s y q z h n z x
-----------------------------------------------------------------------------------
e n e m y f a l l i n g b a c k b r e a k t h r o u g h i m m i n e n t l u c i u s
```

Then, breaking up the decrypted plaintext and breaking it into words appropriately, along with supplying the right punctuation, we get:

```
Enemy falling back.  Breakthrough imminent.  Lucius.
```

Now, the question may arise: what happens if the message had a letter such as `d`? There are no letters that are five before `d`. Well, in that case, we instead wrap away to the end of the alphabet.

More concretely, for `d`, it becomes `y` in our ciphertext.

A nice way to visualize the Caesar cipher then is to place the letters in our alphabet in a circle rather than a line in order to encapsulating this "wrapping" property. This creates what we call a *Cipher Wheel*, as is illustrated in Figure 1.1.

Due to how the method works, Caesar Ciphers are also referred to as *Shift Ciphers*. Now, an issue with the cipher arises in the following scenario:

Suppose that Alice wants to send a secret message to Bob. Using the Caesar Cipher, she encrypts her message and sends it to Bob.

However, Eve intercepts the message! The issue with this encryption method is immediately obvious: it's

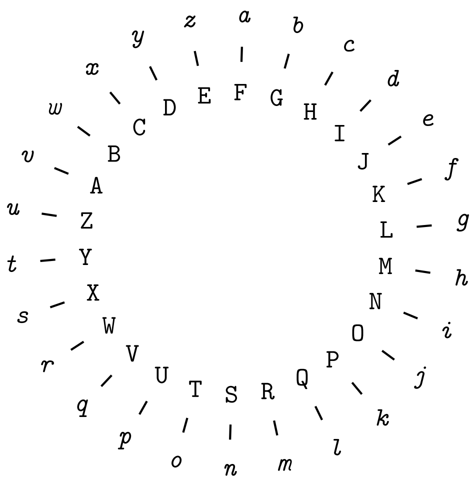

Figure 1.1: Cipher Wheel with offset of 5

incredibly easy to decrypt, with Eve only needing to try out 26 shifts at most to crack the message.

The question becomes then: is there a way to make the message harder to be decrypted? One way in which Alice can do so is to use a more complicated replacement scheme.

1.1.2 Simple Substitution Ciphers

> **Example 1.2** (An Alternate Scheme). A scheme that Alice can employ is to replace every occurrence of a letter with another one: for example, consider replacing every occurrence of $\mathtt{z}$ with $\mathtt{a}$, and vice-versa. And then you continue with $\mathtt{y} \leftrightarrow \mathtt{b}, \mathtt{x} \leftrightarrow \mathtt{c}, \ldots, \mathtt{m} \leftrightarrow \mathtt{n}$.

This is what we call a *Simple Substitution Cipher*, where each letter is replaced by another letter. In fact, the Caesar Cipher is an example of a simple substitution cipher, although many other types exist as well.

One way to view substitution ciphers is to think of them as a function $f : \{\mathtt{a}, \ldots, \mathtt{z}\} \to \{\mathtt{A}, \ldots, \mathtt{Z}\}$.

> **Remark 1.3.** For the remainder of these notes, we will be using lowercase for the plaintext, and uppercase for the ciphertext to make it more distinguishable.

Now, for an encryption function to be valid, it must be *injective* – that is, no two plaintext letters get mapped to the same ciphertext letter.

1.1.3 Cryptanalysis of Simple Substitution Ciphers

A natural question to ask is: how many simple substitution ciphers are there?

Well, we proceed as follows to figure this out:

1. First, $\mathtt{a}$ has 26 letters it can be mapped to.
2. Next, $\mathtt{b}$ has 25 letters it can be mapped to.
3. We continue until $\mathtt{z}$ which has 1 letter it can be mapped to.

Putting this all together, we observe then that we have:

$$26 \cdot 25 \cdots 2 \cdot 1 = 26! = 403291461126605635584000000$$

In other words, there are over 10^{26} different simple substitution ciphers. We define each associated encryption table to be a *key*.

Now, suppose that Alice tries to send a message to Bob. However, Eve intercepts it and tries to decrypt it by trying every possible substitution cipher. We define *Cryptanalysis* to be the process of decrypting a message without knowing the underlying key.

Now, since there are over 10^{26} different simple substitution ciphers, even if Eve can check one million cipher alphabets per second, it'd still take over 10^{13} years to crack.

However, this is where one of the most important ideas of this course comes into play:

> **Theorem 1.4** (Enemy Knows Best)**.** The security of an encryption system depends on the best known method to break it.

Going back to simple substitution ciphers, we note that a fatal flaw exists: the English language is *not* random. For example, `q` is almost always followed by `u`. Certain letters also appear more frequently than others, as seen in Figure 1.2.

With that in mind, Eve can simply count the frequency of the letters in the secret message and make an educated substitution for them. Thus, the actual number of tries needed is much lower than the massive 10^{26}.

By decreasing frequency		In alphabetical order	
E	13.11%	A	8.15%
T	10.47%	B	1.44%
A	8.15%	C	2.76%
O	8.00%	D	3.79%
N	7.10%	E	13.11%
R	6.83%	F	2.92%
I	6.35%	G	1.99%
S	6.10%	H	5.26%
H	5.26%	I	6.35%
D	3.79%	J	0.13%
L	3.39%	K	0.42%
F	2.92%	L	3.39%
C	2.76%	M	2.54%
M	2.54%	N	7.10%
U	2.46%	O	8.00%
G	1.99%	P	1.98%
Y	1.98%	Q	0.12%
P	1.98%	R	6.83%
W	1.54%	S	6.10%
B	1.44%	T	10.47%
V	0.92%	U	2.46%
K	0.42%	V	0.92%
X	0.17%	W	1.54%
J	0.13%	X	0.17%
Q	0.12%	Y	1.98%
Z	0.08%	Z	0.08%

Figure 1.2: Frequency of letters in English text

1.2 Divisibility and GCD

In this section, we will delve into the foundations of algebra and number theory. A lot of cryptography is built on top of this.

With that in mind, we will first discuss some basic number theory – that is, the study of integers.

To begin our discussion of the integers, $\mathbb{Z}$, we recall that they form a ring – that is, an abelian group under $+$ and $\times$ obeys associativity, distributivity, and has a multiplicative identity 1.

Now, with that in mind, we note that multiplication doesn't always have an inverse: that is, we can't always divide an integer by another integer. This leads to the concept of "divisibility":

Definition 1.5 (Divisibility). Let a and b be integers such that $b \neq 0$. Then, we say that b divides a (or a is divisible by b) if there exists an integer c such that:

$$a = bc$$

We denote this by $b \mid a$. If b doesn't divide a, we write $b \nmid a$.

Example 1.6. We note that $847 \mid 485331$ as $485331 = 847 \cdot 573$.

On the other hand, $355 \nmid 259943$, as when we try to divide 259943 by 355, we get a remainder of 83. That is, $259943 = 355 \cdot 732 + 83$.

There are various divisibility properties, some of which are listed here:

Proposition 1.7. Let $a, b, c \in \mathbb{Z}$ be integers. Then,

1. If $a \mid b$ and $b \mid c$, then $a \mid c$.
2. If $a \mid b$ and $b \mid a$, then $a = \pm b$.
3. If $a \mid b$ and $a \mid c$ then $a \mid (b+c)$ and $a \mid (b-c)$.

Proof. Suppose that $a, b, c \in \mathbb{Z}$. Now, we will prove the first statement:

Proof. Suppose that $a \mid b$ and $b \mid c$. Then, this means that there exists x and y such that:

$$\begin{aligned} b &= ax \\ c &= by \\ &= (ax)y \\ &= a(xy) \end{aligned}$$

And since $xy \in \mathbb{Z}$, we note then that exists a $z = xy \in \mathbb{Z}$ such that $c = az$; this is precisely what it means for $a \mid c$. □

Next, we prove the second statement:

Proof. Suppose that $a \mid b$ and $b \mid a$. Then, there exists x, y such that:

$$\begin{aligned} b &= ax \\ a &= by \end{aligned}$$

Then, this means that:

$$\begin{aligned} b &= ax \\ &= (by)x \\ 1 &= yx \end{aligned}$$

Then, this means that $x = y = \pm 1$. With that in mind, we see then that $a = \pm b$ as desired. □

Finally, we prove the last statement:

Proof. Suppose that $a \mid b$ and $a \mid c$. Then, there exists $x, y \in \mathbb{Z}$ such that:

$$\begin{aligned} b &= ax \\ c &= ay \end{aligned}$$

Then, we observe that:

$$\begin{aligned} b + c &= ax + ay \\ &= a(x + y) \\ b - c &= ax - ay \\ &= a(x - y) \end{aligned}$$

Then, there exists $z_1, z_2 \in \mathbb{Z}$ such that $z_1 = x + y$ and $z_2 = x - y$ such that $az_1 = b + c$ and $az_2 = b - c$. This is precisely what it means for $a \mid (b + c)$ and $a \mid (b - c)$. □

Thus, we have proven all three properties as desired. ■

Definition 1.8 ((Greatest) Common Divisor)**.** Let $a, b \in \mathbb{Z}$. Then, we define a common divisor d to be a positive integer that divides both a and b. That is, $d \mid a$ and $d \mid b$.

As its name suggests then, the Greatest Common Divisor (GCD) is the greatest d such that $d \mid a$ and $d \mid b$. This is denoted by $\gcd(a, b)$.