Deciphering the Secrets of Ciphers

Mengxiang Jiang

Topic Number 14: Ciphers or cryptography (secret codes)
Course: Math 645 A Survey of Mathematical Problems

May 2, 2025

Abstract

This paper presents an expository survey of classical cryptography, emphasizing the evolution of cipher techniques alongside the underlying mathematical principles. Beginning with a historical overview, from the rudimentary Caesar cipher to more complex systems like the polyalphabetic and polygraphic substitution ciphers, the discussion illustrates how early encryption methods paved the way for modern cryptographic practices. Key mathematical tools such as modular arithmetic, permutations, and matrix operations are examined in relation to cipher construction and cryptanalysis. Moreover, the paper describes the implementation of encryption and decryption processes using computer simulations and interactive web-based tools, demonstrating the practical application of statistical techniques such as frequency analysis and the Kasiski examination. By synthesizing historical context, theoretical underpinnings, and practical implementations, this work underscores the enduring relationship between mathematics and the art of secure communication.

Contents

1 Introduction 3

2 Historical Development of Classical Ciphers 3
- 2.1 Caesar Cipher . . . 3
- 2.2 Substitution Cipher . . . 4
- 2.3 Caesar's Letter . . . 4
- 2.4 Polyalphabetic Cipher . . . 5
- 2.5 The Confederacy's Failed Vigenère . . . 5
- 2.6 Polygraphic Substitution Cipher . . . 6
- 2.7 One-Time Pads . . . 7

3 Mathematical Foundations 8
- 3.1 Modular Arithmetic . . . 8
- 3.2 Frequency Analysis . . . 8
- 3.3 Kasiski Examination . . . 9
- 3.4 Matrix Manipulation . . . 10
- 3.5 One-Time Pads: Mathematics and Attacks . . . 12

4 Playing With Ciphers 12

5 Conclusion 13

6 References 14

1 Introduction

In the hidden corridors of history, secret messages have long played a pivotal role, whispered across dim rooms and exchanged under cover of uncertainty. The decoding of these covert communications shifted the balance of power and altered the fate of not only the individuals using them but also of their nations. As such, classical cryptography emerged as a vital means of securing communications, and according to historian David Kahn, probably "appears spontaneously - as its parents, language and writing" [4]. Grounded in mathematics, these early methods not only secured important communications during challenging times but also laid the groundwork for the sophisticated encryption systems of today. The following discussion examines the historical evolution, mathematical foundations, and practical implementations of classical cryptography, providing insights into how these time-tested techniques continue to shape the art of secure communication.

2 Historical Development of Classical Ciphers

The progression of cryptographic techniques reflects a long history of evolving ideas, many of which reappear under different names and forms. However, there is a common way of presenting plaintext, the message one wants to send, and the corresponding ciphertext, the encrypted message (the example here is using a Caesar cipher with a shift of 7):

Plaintext: `helloworld`
Ciphertext: `OLSSVDVYSK`

Notice that the standard practice is to use lowercase for the plaintext and uppercase for the ciphertext, and the font should be monospaced to align the two. Now, let us take a look at some of the key ciphers in history.

2.1 Caesar Cipher

The Caesar cipher is one of the simplest encryption schemes. Attributed to Julius Caesar by the Roman historian Gaius Suetonius Tranquillus, this method replaces each letter with another letter a fixed number of places later in the alphabet, typically a shift of three [4]. Thus we have the following cipher alphabet:

Plaintext: `abcdefghijklmnopqrstuvwxyz`
Ciphertext: `DEFGHIJKLMNOPQRSTUVWXYZABC`

Despite its simplicity, the Caesar cipher established the fundamental concept of secret keys and substitution.

2.2 Substitution Cipher

A natural generalization of the Caesar cipher is the substitution cipher, in which each letter of the plaintext is replaced by another letter (usually of the same alphabet, but not necessarily). Here is an example of a cipher alphabet:

Plaintext: `abcdefghijklmnopqrstuvwxyz`
Ciphertext: `MNRUYWZDOQSLPTXJGKBVACHIEF`

In the math section, we will see that this scheme has significantly more variation compared to the Caesar cipher, yet it still has a vulnerability to pattern analysis. Despite this, the substitution cipher remains a fundamental building block for many more advanced ciphers. Indeed, Julius Caesar himself actually utilized one during his campaign in Gaul (modern day France) [2]. The use of a substitution cipher in this context is particularly interesting, as it was critical for a strategic military operation that could have changed the course of Caesar's life and Roman history. The letter was sent to Cicero, who was in charge of the siege of the city of Alesia, and it was crucial for him to receive the message without it being discovered by the enemy. Below is a translation of the Caesar's own account.

2.3 Caesar's Letter

Then with great rewards he (Caesar) induces a certain man of the Gallic horse to convey a letter to Cicero. This he sends written in Greek characters, lest the letter being intercepted, our measures should be discovered by the enemy. He directs him, if he should be unable to enter, to throw his spear with the letter fastened to the thong, inside the fortifications of the camp. He writes in the letter, that he having set out with his legions, will quickly be there: he entreats him to maintain his ancient valor. The Gaul apprehending danger, throws his spear as he has been directed. Is by chance stuck in a tower, and, not being observed by our men for two days, was seen by a certain soldier on the third day: when taken down, it was carried to Cicero. He, after perusing it, reads it out in an assembly of the soldiers, and fills all with the greatest joy. – *The Gallic Wars Book 5 Chapter 48* [2]

2.4 Polyalphabetic Cipher

The next great innovation is an evolution of the substitution cipher that only appeared in history around 1500 years after Caesar [4]. It is the polyalphabetic cipher where the substitution mapping changes throughout the message according to a repeating key. One of the most renowned examples of this approach is by Blaise de Vigenère in his 1585 book *Traicté des Chiffres* [4]. The Vigenère cipher uses multiple alphabets determined by a keyword that is repeated to the same length as the message, and applying a different Caesar cipher based on the current letter of the keyword. Here is an example using the keyword `FAST`:

Plaintext: `helloworld`
Keyword: `FASTFASTFA`
Ciphertext: `MEDETWGKQD`

This dynamic approach helps obscure the frequency patterns inherent in the original text, making the cipher more resistant to simple pattern analysis while still relying on the secrecy of the key. However, we will see that it is also still vulnerable to a technique called Kasiski examination, which we will cover in the math section. Still, the Vigenère cipher was a significant advancement in cryptography and remained the best form of cryptography for centuries. But that is only the case when applied correctly, as our next story illustrates.

2.5 The Confederacy's Failed Vigenère

During the American Civil War, the Union had a relatively simple cipher that involved writing the plaintext words in row order on a rectangular grid and then creating the ciphertext by reading them off in column order [4]. For instance a 4 × 4 plaintext is shown below:

four	score	and	seven
years	ago	our	fathers
brought	forth	on	this
continent	a	new	nation

Figure 1: Union's rectangular grid for a message

and the corresponding ciphertext by reading in column order is:

```
FOURYEARSBROUGHTCONTINENTSCOREAGOFORTHAANDOURONNEWSEVENFATHERSTHISNATION
```

There were a few more complications included as well, but despite its simplicity, it remained relatively secure.

The Confederacy, on the otherhand, chose to use the Vigenère cipher as one of its systems, and on paper it should have been much more secure than that of the Union [4]. However, practical missteps undermined the cipher's strength: word divisions were maintained in the ciphertext, key parts of the messages were left unencrypted, and only three fixed fifteen-letter keys were repeatedly used. These operational oversights transformed a theoretically robust system into one that Union cryptanalysts could easily break, exposing the vulnerabilities in Confederate secret communications.

2.6 Polygraphic Substitution Cipher

Polygraphic substitution ciphers, a fairly recent invention, encrypt groups of letters rather than individual characters. A well-known example in this category is by Lester Hill in his 1929 paper *Cryptography in an Algebraic Alphabet*, applying a transformation to blocks of text[4]. For instance, the plaintext `helloworld` can be divided into blocks of length 2, and then each block is transformed into a corresponding ciphertext block according to a predetermined scheme. The following table provides a simplified illustration of such a transformation (which is determined by matrix multiplication, which we will cover in the math section):

Plaintext Block	**Ciphertext Block**
`he`	`XM`
`ll`	`CD`
`ow`	`QP`
`or`	`RI`
`ld`	`SN`

By encrypting multiple letters at once, polygraphic substitution ciphers obscure the statistical properties of the plaintext, thereby enhancing the security of the encoded message. However, if the attacker obtains a pair of plaintext and ciphertext, they can break the cipher.

2.7 One-Time Pads

The one-time pad is essentially a Vigenère cipher taken to its ultimate extreme [5]. In a standard Vigenère cipher, a short key is repeated over the course of the plaintext, introducing vulnerabilities through recurring patterns. However, when the key is as long as the message and drawn from a random, non-repeating source, such as a passage from a book, the process becomes a one-time pad. For example, consider encrypting `helloworld` using the first ten letters from a book passage, say `ALICEWASBE` from Lewis Carrol's *Alice's Adventures in Wonderland* [3].

Plaintext: `helloworld`
Keyword: `ALICEWASBE`
Ciphertext: `HPTNSSOJMH`

In this setup, each letter of is paired with a unique letter from the key, exactly as in the Vigenère method, but the key is as long or longer than the message. This non-repeating, unpredictable key removes the statistical clues that could otherwise be exploited, ensuring perfect secrecy and making the one-time pad unbreakable as long as the pad is not used again. Genevieve Grotjan Feinstein, a prolific American mathematician and cryptanalyst during World War II, was able to crack reused Russian one-time pad systems [4].

3 Mathematical Foundations

Classical cryptography rests on a foundation of mathematical concepts that ensure both the functionality and security of cipher systems. Below are some of the key concepts that relate to each of the ciphers described before.

3.1 Modular Arithmetic

The Caesar cipher relies on modular arithmetic by shifting each letter of the plaintext by a fixed number of positions. Its encryption function can be expressed as

$$E(x) = (x + k) \mod 26,$$

where x represents a letter's numerical equivalent and k is the amount of shifts [5]. Due to the limited number of possible shifts (only 25), the Caesar cipher is extremely vulnerable to brute-force attacks. Additionally, once the substitution pattern is recognized, simple frequency analysis quickly reveals the underlying plaintext.

3.2 Frequency Analysis

The substitution cipher replaces each letter of the plaintext with another letter according to a fixed permutation. To illustrate its vulnerability, consider the example passage:

```
in the shadowed halls of encoded whispers secret messages
flow like silent rivers though cloaked in cipher the familiar
pulse of language endures betraying hints to those who
know its rhythm
```

When this passage is encrypted using a substitution cipher, every letter is swapped with another according to a predetermined key. Although the resulting ciphertext appears scrambled, the overall frequency distribution of letters remains unchanged. In English, certain letters such as `E` appear with predictable regularity, and cryptanalysts can exploit these frequency patterns to deduce the substitution mapping. The tables below include one from Simon Rubinstein-Salzedo's *Cryptography* which presents the typical frequency distribution of English letters, which is preserved even after encryption [5].

This preserved distribution makes the substitution cipher particularly sus-

Letter	Count	Frequency (%)	Rank
E	21	13.1%	1
S	15	9.4%	2
H	13	8.1%	3
I	12	7.5%	4
O	11	6.9%	5
T	11	6.9%	6
N	9	5.6%	7
A	9	5.6%	8
R	9	5.6%	9
L	9	5.6%	10
D	6	3.8%	11
W	5	3.1%	12
G	5	3.1%	13
F	4	2.5%	14
C	4	2.5%	15
U	4	2.5%	16
P	3	1.9%	17
M	3	1.9%	18
K	3	1.9%	19
Y	2	1.3%	20
V	1	0.6%	21
B	1	0.6%	22

(a) Passage Frequencies

Letter	Frequency (%)	Rank
E	12.702	1
T	9.056	2
A	8.167	3
O	7.507	4
I	6.966	5
N	6.749	6
S	6.327	7
H	6.094	8
R	5.987	9
D	4.253	10
L	4.025	11
C	2.782	12
U	2.758	13
M	2.406	14
W	2.360	15
F	2.228	16
G	2.015	17
Y	1.974	18
P	1.929	19
B	1.492	20
V	0.978	21
K	0.772	22
J	0.153	23
X	0.150	24
Q	0.095	25
Z	0.074	26

(b) Rubinstein-Salzedo's Global Frequencies

ceptible to frequency analysis, as the statistical properties of the plaintext remain visible despite the superficial scrambling.

3.3 Kasiski Examination

The Kasiski examination is a powerful method for estimating the key length of a polyalphabetic cipher like the Vigenère [5]. Consider the following with the original plaintext

```
attackatdawnnowattackatdawnnow
```

which is 30 letters long. The chosen key is `LEMON`, repeated to match the plaintext length, yielding:

```
LEMONLEMONLEMONLEMONLEMONLEMON
```

Using the Vigenère encryption process, the plaintext is transformed into the ciphertext:

```
LXFOPVEFRNHRZCJLXFOPVEFRNHRZCJ
```

In this ciphertext, observe that the sequence `LXFOP` appears twice, once starting at position 1 and again starting at position 16. The distance between these two occurrences is $16 - 1 = 15$ characters. Because the key is applied cyclically, the key length must be a divisor of the distance between repeated sequences. The factors of 15 are 1, 3, 5, and 15. Disregarding the trivial case of 1 and considering the context of a polyalphabetic cipher, a key length of 5 is the most plausible candidate, which indeed is the actual length of the key. Once the key length is determined through the Kasiski examination, the ciphertext can be divided into 5 groups (one for each character of the key). Frequency analysis can then be applied separately to each group, treating each as if it were produced by a simple Caesar cipher, to deduce the individual shifts. This process ultimately reveals the entire key and allows the original plaintext to be recovered.

3.4 Matrix Manipulation

The mathematics of the Hill cipher uses concepts from linear algebra to encrypt blocks of text. In this system, plaintext letters are first converted into numerical equivalents and arranged into vectors. The key matrix, an invertible square matrix over the integers modulo 26, is then used to transform these plaintext vectors into ciphertext vectors by means of matrix multiplication. Mathematically, if p is the plaintext vector and K is the key matrix, the ciphertext vector c is computed as

$$c \equiv Kp \pmod{26}.$$

For example, consider a 2 × 2 key matrix

$$K = \begin{bmatrix} 3 & 3 \\ 2 & 5 \end{bmatrix},$$

and let the plaintext block be `hi`, where `h = 7` and `i = 8`. The plaintext vector is

$$p = \begin{bmatrix} 7 \\ 8 \end{bmatrix}.$$

Multiplying the key matrix by the plaintext vector, we obtain

$$Kp = \begin{bmatrix} 3 \times 7 + 3 \times 8 \\ 2 \times 7 + 5 \times 8 \end{bmatrix} = \begin{bmatrix} 21 + 24 \\ 14 + 40 \end{bmatrix} = \begin{bmatrix} 45 \\ 54 \end{bmatrix}.$$

Taking the result modulo 26 gives

$$c \equiv \begin{bmatrix} 45 & \mod 26 \\ 54 & \mod 26 \end{bmatrix} = \begin{bmatrix} 19 \\ 2 \end{bmatrix}.$$

The numerical values 19 and 2 correspond to the letters `T` and `C`, so the ciphertext for it is `TC`.

Despite its elegant design, the Hill cipher is vulnerable to a known-plaintext attack. If an attacker obtains enough plaintext-ciphertext pairs, they can set up a system of linear equations modulo 26 to solve for the key matrix. This is due to the key fact that a square matrix is invertible if and only if its rows are linearly independent. For an $n \times n$ key matrix, if the attacker knows n linearly independent plaintext blocks and their corresponding ciphertext blocks, these can be arranged into matrices p and c so that

$$c \equiv Kp \pmod{26}.$$

Provided that p is invertible modulo 26, the key matrix can be recovered by computing

$$K \equiv cp^{-1} \pmod{26}.$$

This approach effectively undermines the cipher's security if sufficient plaintext is available, demonstrating the importance of key management and the inherent vulnerabilities of linear encryption schemes when exposed to known-plaintext attacks.

3.5 One-Time Pads: Mathematics and Attacks

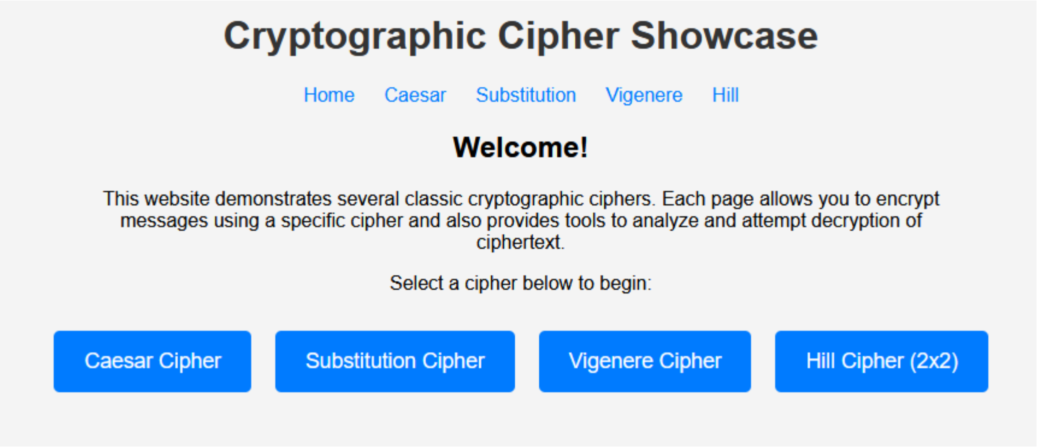
The one-time pad represents the ideal of perfect secrecy by employing a key that is as long as the message and composed of truly random characters. This method is mathematically equivalent to a Vigenère cipher with a non-repeating, random key. When used correctly, every possible plaintext is equally likely, rendering the ciphertext theoretically unbreakable. In practice, however, the security of a one-time pad hinges on the strict non-reuse of the key; any key reuse or compromise introduces vulnerabilities, as it allows the determination of the key length through Kasiski.

4 Playing With Ciphers

Just as students cannot fully learn mathematics just by attending lectures and must practice what they learn through homework, it is also not enough to just read about these ciphers without at least trying them out yourself. And unfortunately, if one were to go about this using only pen and paper, it would be a very tedious process. Fortunately, by following a nice blog post by Jonas Bromso about creating a javascript website on github pages, I was able to create a simple website that allows you to play around with the ciphers that we have discussed in this paper [1]. The website is hosted on GitHub Pages and can be found at the following link:

`https://kumj2028.github.io/`

The website is designed to be user-friendly, allowing users to easily select the cipher they wish to use as well as showcasing the methods of breaking them.



Cryptographic Cipher Showcase

Welcome!

This website demonstrates several classic cryptographic ciphers. Each page allows you to encrypt messages using a specific cipher and also provides tools to analyze and attempt decryption of ciphertext.

Select a cipher below to begin:

Figure 2: Screenshot of the website

The website is still a work in progress, but it is fully functional, and I plan to add more features in the future. It is fully open source, and the code can be found at the following link:

`https://github.com/kumj2028/kumj2028.github.io`

Please feel free to check it out and contribute to the project.

5 Conclusion

Ultimately, classical cryptography serves as a compelling case study in the interplay between mathematical theory and real-world application. Its evolution highlights how historical contexts and practical needs drive innovation, while the underlying mathematics ensures that even the simplest cipher carries a depth of structure worthy of rigorous study. The enduring relevance of these early cryptographic methods is evident in the way they inform modern encryption techniques and continue to inspire new research directions. As we look to the future, the lessons learned from classical cryptography will undoubtedly contribute to the development of even more secure methods of communication in an increasingly digital world.

6 References

References

[1] Jonas Bromso. Experimenting with github pages and javascript, 2019.

[2] Julius Caesar et al. *The Gallic Wars*. The Internet Classics Archive, 1869. Translated by W. A. McDevitte and W. S. Bohn.

[3] Lewis Carroll. *Alice in Wonderland*. Macmillan, 1865. Project Gutenberg.

[4] John F Dooley. *History of cryptography and cryptanalysis*. Springer, 2018.

[5] Simon Rubinstein-Salzedo. *Cryptography*, volume 260. Springer, 2018.