

Unit 2- Introduction to Classical Cryptography

This unit gives an overview of classical cryptography.

At the end of the unit, the learner should be able to:

- Describe the main cryptographic categories used in classical cryptography: substitution and transposition ciphers
- Explain the principle of substitution encipherment using Caesar cipher and mono-substitution ciphers
- Identify the characteristics of poly-substitution using Playfair and Vigenère ciphers as examples
- Explain transposition ciphers such as rail-fence and row transposition ciphers
- Define the characteristics of product ciphers that serve as the basis for modern cryptography

Target Audience: Third Year Students - BSc Software Engineering, BSc Computer Science

1. Classical Cryptography

Cryptography may be split into two epochs: the pre-computer and computer eras. The cryptographic schemes in the pre-computer era were predominantly pencil-and-paper language-based ciphers called ***classical (or historical) ciphers (aka cryptograms)***. The computer era began at the end of the Second World War (WWII) with the invention of the first computer and coincided with the emergence of ***modern cryptography***.

Cryptography took birth around 2000 BC in Egypt, where hieroglyphs were carved on tombs to tell the story of deceased rulers and kings in a concealed manner to impress the reader. Cryptography evolved through time with alphabetic and writing systems. The first recorded encryption method, known as the “skytale”, was devised and used by the Greek Spartans for military purposes in the fifth century BC [1]. The origins of classical cryptography can be traced back to ancient civilizations such as Egypt and Greece and were primarily done using symbols and the alphabet by hand. The knowledge of the strengths and weaknesses of classical ciphers have immensely contributed to the development of modern ciphers.

There are two broad categories of classical cryptographic techniques: **Substitution** and **Transposition**.

2. Substitution Ciphers

A **substitution cipher** is a method of encryption whereby a character (or group of symbols) of the plaintext message is replaced by another character (or group of symbols) in some "secret" manner to produce the ciphertext. For decryption, the reverse procedure is performed on the ciphertext to return the original plaintext.

The following substitution ciphers shall be studied in this unit:

1. Caesar Cipher
2. Monoalphabetic Cipher
3. Vigenère Cipher
4. Playfair Cipher
5. Hill Cipher

2.1. Caesar Cipher

The first-ever recorded substitution cipher in history is that used by the Roman Emperor Julius Caesar and was named Caesar Cipher[2]. It involved the replacement of each letter of the plaintext with another letter three steps ahead in the alphabet. If the end of the alphabet set is reached, the next character from the beginning of the alphabet was chosen. Hence **A** is substituted by **D**, **B**→**E**, **C**→**F**, ..., **X**→**A**, **Y**→**B**, and **Z**→**C**. Here is an example given below.

Caesar Cipher

Key : 3

Plain text : The quick brown fox jumps over a lazy dog

Cipher text: WKH TXLFN EURZQ IRA MXPSV RYHU D ODCB GRJ

Deciphered : THE QUICK BROWN FOX JUMPS OVER A LAZY DOG

Originally, the key used was 3, but any whole number between 1 and 25 can be used. If the letters are coded, such that 'a' is 0, 'b' = 1, 'c' = 2, ..., 'z' = 25, then the result can be calculated using the formulae below.

Encryption: $c = (p + k) \bmod n$

Decryption: $p = (c - k) \bmod n$

where p is the plaintext code, c – the ciphertext code, k is the key, and n is the size of the alphabet. *mod* is the remainder after the integer division operator (sometimes denoted by %).

2.2. Monoalphabetic Cipher

An improvement over Caesar Cipher is to use an arbitrary permutation of the letters of the alphabet as the key such that any plaintext alphabet is replaced by another arbitrary character.

Plain : abcdefghijklmnopqrstuvwxyz

Key : ZLGOTJQXKIUDNFHCYVAEWRBMP

For example, using the above key, the following plaintext shall yield the ciphertext below. This is called the monoalphabetic cipher.

Plaintext: The big brown fox jumps over the lazy dog

Ciphertext: EXT LKQ LVHSF JHB IWNCA HRTV EXT DZPM OHQ

Hence it is more difficult to "guess" the key as there are $26!$ ($= 4 \times 10^{26}$) combinations of keys.

However, using a technique known as *frequency analysis* of characters, the mapping of plaintext on the ciphertext can be deduced or guessed.[2]

2.3. Playfair Cipher

In order to counteract letter frequency analysis, double-letter substitution is used instead of single-letter. One such cipher is the **Playfair Cipher** which was invented by British scientist Sir Charles Wheatstone in 1854, but named after his friend Baron Playfair. It was based on the construction of a **5 x 5 matrix** of letters constructed using a keyword.

Given the keyword, e.g., **CYBERSECURITY**, the Playfair Matrix is as follows: First, the unique letters of the keyword are inserted in the matrix, for I and J the same cell is used. The remaining letters of the alphabet are used to complete the filling of the matrix.

C	Y	B	E	R
S	U	I/J	T	A
D	F	G	H	K
L	M	N	O	P
Q	V	W	X	Z

Key : CYBERSECURITY

Plain text : to be or not to be

Plain text : TO BE OR NO TX TO BE

Cipher text: HX ER PE OP HE HX ER

Deciphered : TO BE OR NO TX TO BE

The plaintext is encrypted in pairs of distinct letters.

For example, **T** and **O** fall on the same column, so we take the next character downwards, i.e. **T**→**H** and **O**→**X**

B and **E** fall on the same row, so we take the next character to the right, i.e. **B**→**E** and **E**→**R**

OR forms a rectangle and maps onto **PE**.

2.4. Hill Cipher

The Hill Cipher is another multi-letter cipher developed by the mathematician **Lester S. Hill** in 1929 based on matrix multiplication.

An invertible $n \times n$ matrix, K , represents the key, where the entries are $0..m - 1$, and m is the size of the alphabet. Then the text block is a column matrix, P of size n representing the code of the characters in the block.

So encryption is $C = (K \times P) \bmod 26$

and decryption is $P = (K^{-1} \times C) \bmod 26$

where K^{-1} is the inverse of matrix K .

As an example:

$P = \text{hello}$

The key matrix $K = \begin{bmatrix} 12 & 19 \\ 9 & 23 \end{bmatrix}$

'h' = 7, 'e' = 4

$$\begin{bmatrix} 12 & 19 \\ 9 & 23 \end{bmatrix} \times \begin{bmatrix} 7 \\ 4 \end{bmatrix} \bmod 26 = \begin{bmatrix} 4 \\ 25 \end{bmatrix} = \begin{bmatrix} E \\ Z \end{bmatrix}$$

Reminder for finding matrix inverse:

Inverse of 2×2 matrix

If $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, then

$$A^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

Note: A^{-1} exists only when $ad - bc \neq 0$

2.5. Vigenère Cipher

The **Vigenère Cipher** is an example of a **polyalphabetic substitution cipher**, whereby each character of the plaintext is replaced by different characters (at different positions) in the ciphertext, based on a secret key. It was developed by an Italian cryptologist Giovan Battista Bellaso in 1553. In 1586 **Blaise de Vigenère**, a French cryptologist published a variant of this cipher (the autokey cipher) and was misattributed its invention.

The key must be of the same length as the plaintext. If the key is shorter, it is repeated in order to make up for the remaining length of the plaintext, as shown in the example below.

Plain text : thebigbrownfoxjumpsoverthelazydog

Key : penguin

Attack Key : penguinpenguinpenguinpenguinpengu

Cipher text: ILRHCOOGSJ TZWKYYZVMWITVGNYTNOCQUA

Deciphered : THEBIGBROWNFOXJUMPSOVERTHELAZYDOG

The ciphertext is obtained by "adding" the plaintext to the key, each letter of the alphabet being encoded with a number as shown in the table below:

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25

3. Transposition Ciphers

Transposition ciphers are a type of cryptographic technique that involves rearranging the characters or elements of a message to create the ciphertext. Unlike substitution ciphers, where individual characters are replaced with other characters, transposition ciphers do not change the characters themselves but alter their positions within the message. The security of transposition ciphers relies on the difficulty of reconstructing the original message without knowledge of the specific rearrangement rule.

There are several **types of transposition ciphers**, each with its own method of rearrangement:

1. Rail Fence Cipher
2. Columnar Transposition
3. Route Cipher
4. Permutation Ciphers

Transposition ciphers can provide some level of security, especially when combined with other cryptographic methods. However, they are generally considered weaker than modern encryption techniques, as they are susceptible to attacks based on pattern recognition and frequency analysis. The frequency distribution of plaintext characters is NOT obfuscated in the ciphertext.

Nevertheless, historical examples of transposition ciphers, such as the scytale used by the ancient Greeks and the route cipher employed during the American Civil War, demonstrate their relevance in the evolution of cryptography. Today, transposition ciphers are primarily used for educational purposes and for creating puzzles or challenges rather than for securing sensitive information in modern cryptographic applications.

3.1. Rail Fence Cipher

Rail Fence Cipher: In this technique, the plaintext is written diagonally in a zigzag pattern across a set number of "rails" or lines. The ciphertext is obtained by reading the characters row by row from the top to the bottom. The key is the number of rails used for the cipher, $2 \leq \text{key} \leq n/2$, where n is the size (number of characters) of the message.

Example:

Rail Fence Cipher

Plain text : thebigbrownfoxjumpsoverthelazydog

Key : 3

```
T . . . I . . . O . . . O . . . M . . . V . . . H . . . Z . . . G
. H . B . G . R . W . F . X . U . P . O . E . T . E . A . Y . O .
. . E . . . B . . . N . . . J . . . S . . . R . . . L . . . D . .
```

Cipher text: TIOOMVHZGHBGRWFXUPOETEAYOEBNJSRLD

Deciphered : THEBIGBROWNFOXJUMPSOVERTHELAZYDOG

3.2. Columnar Transposition Cipher

Here, the message is written into the rows of a grid, with the width determined by a specific keyword or key phrase e.g. a number sequence or a character sequence. The incomplete last row is padded with a filler letter (e.g. 'X'). The ciphertext is formed by reading the columns of the grid in the (alphabetical or numerical) order specified by the keyword.

Columnar Transposition Cipher

Plain text : thebigbrownfoxjumpsoverthelazydog

Key : 2 6 1 5 3 4

```
2 6 1 5 3 4
T H E B I G
B R O W N F
O X J U M P
S O V E R T
H E L A Z Y
D O G X X X
```

Cipher text: EOJVLGTBOSH DINMRZXGFPTYXBWUEAXHRXOEO

Note that if the ciphertext of columnar transposition is re-encrypted using the same algorithm, the resulting ciphertext is more difficult to break.

3.3. Route Cipher

This cipher involves writing the plaintext into a geometric shape, such as a grid or a table, and then reading it out in a specific pattern to generate the ciphertext. The plaintext can fill an $n \times n$ grid, and the ciphertext is obtained by reading the data in a zigzag, up-down, left-right, or spiral pattern.

Example of Route Cipher using a 5x5 grid

Plain text : universitedesmascareignes

Cipher text: SETISMCAREDEUNIVERASSENGI

Input Pattern (Down-Up)

U↓	E↗	D↘	E↗	I↘
N↓	T↑	E↓	R↑	G↓
I↓	I↑	S↓	A↑	N↓
V↓	S↑	M↓	C↑	E↓
E↖	R↗	A↖	S↗	S↘

Output Pattern (Spiral)

U↖	E←	D←	E↖	I↖
N↓	T↖	E↖	R↑	G↑
I↓	I↓	S↑	A↑	N↑
V↓	S↖	M→	C↗	E↑
E↖	R→	A→	S→	S↗

4. Product Ciphers

In order to create a more complex encryption scheme, one may be tempted to perform multiple encryptions. For example, a substitution followed by another substitution would produce a more complex substitution. However, this new complex substitution shall have the same drawbacks as the simple substitution.

Similarly, one transposition followed by another produces a more complex transposition, which still suffers from the drawbacks of single transpositions.

However, it has been observed that a combination of substitution followed by transposition operations yields more complex encryption schemes void of the disadvantages of each of them

individually. These encryption schemes are known as **product ciphers** and are found at the heart of modern block ciphers.

5. References:

- [1] D. Kahn, *The Codebreakers: The Story of Secret Writing*, Revised. New York, 1996.
- [2] W. Stallings, *Cryptography and Network Security - Principles and Practice*. 2017.

6. Useful Websites:

<https://www.dcode.fr/en>