Classic Substitution Ciphers

Melisa Azizović
University of Novi Pazar
Department of Computer Science
Novi Pazar, Serbia
melisa.azizovic@uninp.edu.rs

Emruš Azizović
University of Novi Pazar
Department of Computer Science
Novi Pazar, Serbia
azizovic.emrus@gmail.com

Abstract—**Classic substitution ciphers were analysed in the paper. First, we listed the definitions and the basic terms which was necessary for successful monitoring of the content of the paper. We defined the substitution ciphers and demonstrated text encryption and decryption on examples. We analysed the classic substitution ciphers, Caesar's and Vigenère's cipher. The proposed methods are explained in detail and supported by concrete examples.**

Keywords- Classical cryptography, Substitution ciphers, Caesar's cipher, Vigenère's cipher

I. INTRODUCTION (HEADING 1)

Cryptography is a science that uses mathematics and linguistics to encrypt data. Thus, a secret communication between two parties is enabled so that the third party cannot comprehend the meaning of the message without the key or if it does not apply cryptanalysis. Within classical cryptography, there are two types of algorithms. The first type consists of substitution ciphers where each letter of a plaintext is replaced by another letter. While the second type consists of transposition ciphers where the letters in the plaintext change their positions, resulting in an anagram (Dujella and Maretić, 2007).

Today, it is a generally accepted fact that information is the most valuable. Breaking and creating ciphers dates back to ancient times, in the history of cryptography we distinguish two major periods. Contemporary cryptography period is known as modern cryptography, but the period before the advent of the Internet is called classic cryptography.

The aim of the paper is to analyse classic substitution ciphers and to indicate that contemporary cryptosystems play an important role in modern technology. Furthermore, the aim of the paper is to describe in detail the most prominent algorithms for each of the analysed ciphers. A specific application is also presented, that is, encryption and decryption were performed on some examples.

The paper is organized as follows. In the first chapter, the basic terminology to be used in the text is defined. While the second chapter gives a brief description of substitution ciphers, the third chapter provides a detailed overview of classic substitution ciphers. Caesar's ciphers and Vigenère's cipher are described.

Cryptography deals with data in digital form, the encryption and decryption procedure is mathematical in nature, and is carried out automatically with the aid of a computer. For these reasons, modern cryptology mainly relies on computer science, but is greatly aided by number theory.

II. BASIC TERMINOLOGY

A. Maintaining the Integrity of the Specifications

In this part, we will list some definitions and terms that are necessary for successful following the content of this paper - cryptography and cryptanalysis of substitution ciphers according to (Dujella and Maretić, 2007).

Cryptology is a science that deals with studying and defining methods for protecting information (encryption) and studying and finding methods for revealing encrypted information (decryption). For this purpose, it uses its knowledge of mathematics, statistics and linguistics. Cryptology includes cryptography and cryptanalysis.

Cryptography is a scientific discipline that deals with the study of methods that enable communication security between two individuals - senders and receivers of messages. The term cryptography is derived from the Greek language and means secret writing.

Cryptanalysis (Codebreaking) or decryption is a scientific discipline that studies procedures for reading hidden messages without knowing the rules of encryption and the key, using knowledge from mathematics, statistics and linguistics.

Substitution ciphers: each element of the plaintext is replaced with some other element, according to a predetermined transformation. Depending on the number of transformations, they can be monoalphabetic and polyalphabetic.

Any directed action by a cryptanalyst is referred to as an attack.

Plaintext is the text that the sender wants to send to the recipient. He encrypts the message in order to protect it from a

potential adversary, which involves transformations according to a pre-agreed rule or key. This process is called encryption.

Ciphertext - cryptographically coded (protected) message.

Key - a parameter (most often secret) that is used in the message transformation process.

Encryption (Enciphering) – the process of transforming plain text into a cipher using a cryptographic key.

Decryption (Decrypting) is the process of transforming ciphers into plain text using a cryptographic key.

III. SUBSTITUTION CHIPERS

Substitution ciphers replace each letter of the plaintext with a letter from the cipher alphabet. With this procedure, each letter that we get in the encrypted text will be in the same position in the text as in the plaintext, only it will change its identity. There may be any number of ciphered alphabets. That's why we call the systems that use only one ciphered alphabet monoalphabetic systems, and those that use more ciphered alphabets we call the polyalphabetic systems. The most prominent substitution ciphers are the Caesar's cipher, the Vigenère's cipher, the Hebrew cipher and the Playfair's cipher.

In substitution ciphers, any element of the plain text (a letter, a bit, a group of bits or letters) is replaced by some other character, letter, bit or some other sign according to a predetermined transformation. The term substitution cipher comes from the fact that something changes, that is, every letter of our message is substituted (Veinović and Adamović, 2013).

In this paper, we will present classic substitution ciphers. The substitution ciphers are characterized by the fact that the substitution procedure is very easy to remember. If the decryption key is simple, the adversary can easily break the code and read the message (text). Among the oldest and simplest substitution ciphers is the one characterized by the fact that in the upper row the letters are written in alphabetical order, while in the lower row the letters are arranged in reverse alphabetical order.

IV. STANDARD (CLASSIC) SUBSTITUTION CHIPERS

Cryptography is the study of methods for sending messages in a form that can only be understood by the intended recipient. Classic cryptography includes ciphers similar to Caesar's and Vigenère's, which represent the process of encrypting data using a secret private key that only the individuals within the communication know. This is why these systems are also known as secret-key cryptosystems.

In this part, we will describe simple symmetric cryptosystems. For the analysis we will use the English alphabet of 26 letters. Additionally, a correspondence is introduced between the letters of the English alphabet (A-Z), considered as the international alphabet, and whole numbers (0-25) as follows:

TABLE I. LETTERS OF THE ALPHABET AND THEIR ORDINAL NUMBERS

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

The reason for using the numerical equivalent of a letter lies in the manner in which some cryptosystems encrypt and decrypt the plaintext - where the numerical representation of the letter itself is used. If one were to work with an open text in the Serbian language, then the Serbian letters with dialectal signs are replaced with those without them.

Instead of letters in the message, we will often use their numerical equivalents given by the above table, and then the set is marked with

$$\mathbb{Z}_{26} = \{0,1,\ldots,25\}$$

For the operations of addition and multiplication the modulo 26 set Z26 is a commutative ring with unity. We will write down the sum and the product modulo 26 of elements a, b ∈ Z26 as

$$(a+b) \bmod 26, \quad a \cdot b \bmod 26$$

The result of addition and multiplication modulo m is a unique remainder when divided by number m from the set {0,1,...,m}. Modular arithmetic can be represented as a clock arithmetic. In the case of substitution ciphers, that module is generally equal to 26, which means that it is the same as the number of letters of the alphabet that we are encrypting (Ibrahimpašić, 2011).

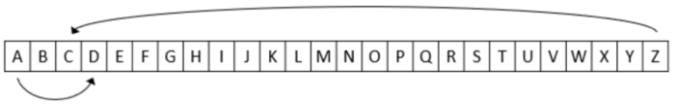

Figure 1. A sample of Caesar's cipher

In addition to the original Caesar's cipher, there are other versions such as the arbitrary displacement, Caesar's cipher with permuted alphabet and key.

Example 1. Plain text

P R I Z R E N

with a shift of 3 spots to the right, we encode it into

S U L C U H Q

where after the last letter Z, the letters A, B, C follow, which means that the alphabet continues cyclically.

If we move on to the numerical equivalents of the alphabet, then we describe Caesar's cipher as a cryptosystem where the plaintext domain, the cipher domain and the key domain are equal to the set Z26, and we can replace the described system with a mathematical model. According to (Dujella and Maretić, 2007), the mathematical model of the Caesar's cipher is represented by the following definition.

Definition 1. According to (Dujella and Maretić, 2007). Let $\mathbb{P} = \mathbb{C} = \mathbb{K} = \mathbb{Z}_{26}$

The encryption and decryption functions are

$e_n: Z_{26} \rightarrow Z_{26}, e_n(x)=(x+n) \bmod 26,$

$d_n: Z_{26} \rightarrow Z_{26}, d_n(y)=(y-n) \bmod 26,$

where $0 \leq n \leq 25$ is the key of this cipher.

One of the methods of decryption is to examine all possible keys successively until some understandable meaningful text is obtained, which is possible since the number of keys is small. That is, the number of keys is exactly the same as the number of the letters of the alphabet - 26.

Example 2. For the cipher HXIGXGX, examining the keys successively, we get:

HXIGXGX for n=0,

JZKIZIZ for n=1,

IYJHYHYfor n=2,

KALJAJA for n=3, which is the name of the fortress in Prizren.

For the encryption procedure itself, for practical reasons is used the so-called Caesar's circle shown in the following figure (Figure 2.).

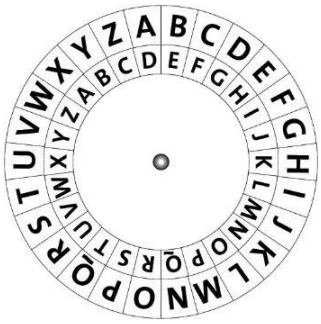

Figure 2. Caesar's circle - codebook

Although the Caesar's cipher is easy to break with the so-called brute force, that is, by examining all the keys successively, it is quite suitable to illustrate the decryption method, which usually uses frequency analysis of letters. Every language has some letters that appear more often than others and this data is well known. Of course, in order to be able to apply this method, it is desirable to know in advance what language the text is written in.

A. Ciphers with date shift

In order to make the encrypted message as difficult as possible to decrypt without a key, it is possible to change the number of shifts for each letter. One way is to use the date when the message was sent. For example, if you want to send a message on October 21, 2022. October is the 10th month of the year, and we can write the date then according to the American system as 10-21-22. If we remove the hyphens, the number is 102122. That number is then repeated above each letter of the message:

Example 3. Writing a simple substitution cipher with a date shift

1 0 2 1 2 2 1 0 2 1 2 2 1 0 2 1 2 2 1

U N I V E R S I T Y O F P R I Z R E N

To encode a message based on image 1., the letter U needs to be moved by one spot, and it then becomes V. N is not moved (because of the number 0), so it remains N. I is moved by 2 spots, and becomes K, and so on for all the letters that are in the message record. If the scroll goes beyond the last letter Z, then it continues from the beginning of the alphabet. The final message, using the key 102122 and grouping the letters into groups, gives the following result:

VNKWGTTIVZ QH QRKATGO

To decode the message, it is necessary to write the key above the message and the movements within the alphabet are to be done backwards. If the scroll goes beyond the first letter A, then it is necessary to return to the very end of the alphabet and then continue backwards. This code type is not monoalphabetic. The last word PRIZREN becomes QRKATGO in the encrypted message, which makes date-shifted ciphers extremely difficult to crack.

B. Ciphers with a keyword

Headings, or heads, are organizational devices that guide the reader through your paper. There are two types: component heads and text heads.

With the mentioned form of the Caesar's cipher, it is evident that both parties in the communication must remember the entire ciphered alphabet as a key that is created by randomly selecting letters, which is sometimes very inefficient. In order to solve this, it is possible to build an encrypted alphabet so that the sender and receiver agree on a single keyword or phrase that will represent the key. An alphabet with a cipher is created by placing a key at its beginning without spaces in-between and without letters that are repeated several times, while the rest is filled with letters that are not in the key, successively from the beginning of the alphabet. The biggest advantage of this type of arrangement of the encrypted alphabet is that the key is easy to remember (Matić, 2015).

One of the simpler ways to create an encryption alphabet is to use a keyword or phrase.

Example 3. Simple substitution ciphers with the word MARS. If we take the keyword "MARS" as an example, the procedure is as follows: Under each letter of the alphabet, write the default keyword, in this case "MARS", and then write the remaining letters of the alphabet successively.

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

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

The keywords are easy to remember and each keyword creates a different substitution cipher table and the substitution cipher itself. The encryption and decryption procedure are carried out identically to the previously described methods.

A Caesar's cipher using a keyword is a well-known monoalphabetic substitution cipher. This cipher is distinctive in that the encryption function is a permutation with a keyword.

Key ciphers consist of a keyword and a number that indicates the position of the spot in the plaintext alphabet from where the keyword starts, and the remaining letters are written in alphabetical order.

Example 4. Caesar's cipher with keyword and specific position

The following example shows the procedure for encrypting with a Caesar's cipher with a keyword where the key is K=(MARS, 5). The ciphered alphabet is created by entering the key word from the specified position. We don't write the letters that are in the key word anymore, and the rest is filled with letters that are not written in the key, starting from the first letter of the alphabet.

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

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

We can see that from the specified number in the key, in our case 5, we start writing the key word after which we write down the remaining letters of the alphabet successively if they have not been used until then.

Example 5. Encrypt UNIVERSITY OF PRIZREN

Based on the substitution in example 4, with the key K= (MARS, 5) we encrypt the specified text into:

TJSUEOPSQY KM LOSZOEJ

In our example, the letters Y and Z are not changed, because the word MARS does not contain letters beyond the position of the letter S. If a word containing the letter Y is used, it will change the entire alphabet except for the letter Z. It is desirable that the word applied for encryption does not contain repeating letters.

If the cipher word changes every week, it is not convenient to agree on which word will be used on each occasion. It is recommended that one of the modes for an easier agreement is to use newspapers or magazines that are available and accessible to everyone to find keywords. It is necessary to choose a word that can be used for substitution, write down the page on which the word is found, the line and the ordinal number of the word on the page. This information can then be written at the end of the encrypted text. For example, the number 12-10-7 means that the keyword is on the 12th page, in the 10th row and is the 7th in line in that row (Ibrahimpašić, 2011)..

C. Vigenère's cipher

According to (Singh and Šifre, 2003), this cipher was first recorded by Giovan Battista Bellaso in the mid-16th century, but in the 19th century the credit was wrongly attributed to Blaise de Vigenère, a French diplomat and cryptographer. For three centuries it was believed that this cipher was impossible to break, but in the mid-19th century Friedrich Kasiski published a general method for deciphering the Vigenère's cipher.

Because of its simplicity, Caesar's cipher was very easily quickly deciphered. A step forward is the Vigenère's cipher, where each letter in the text is represented by one of n possible letters (where n is the length of the key). Vigenère's cipher was so effective at the time that it was called the uncrackable cipher. The reason was that instead of one ciphered alphabet, as many as 26 were used (Singh and Šifre, 2003).

The encryption method is very similar to the Caesar's. The difference is that in this case the key is made up of a block of letters, i.e., shorter words. That is why the Vigenère's cipher is an example of a block cipher.

Example 6. If we take as the key the word RIS (this example refers to the letters of the Serbian Latin alphabet), which consists of the 17th, 8th, and 18th letters of the alphabet (Table 1.), then we will move the first letter of the plaintext forward by 17 spots (that is, cyclically to the right), the second for 8 spots, the third for 18 spots, the fourth for 17 spots, etc. If we were to encrypt INFORMATIKA with the default key, we would get ZVXFZERBABI.

I N F O R M A T I K A

R I S R I S R I S R I

Z V X F Z E R B A B I

Decryption is analogous to the encryption, only we move backwards (i.e., cyclically to the left).

According to (Dujella and Maretić, 2007), the Vigenère's cipher can also be described algebraically by the following definition:

Definition 3. "Let $m \in N$ be the length of the keyword. Then

$$\mathbb{P} = \mathbb{K} = \mathbb{C} = (\mathbb{Z}26)^m,$$

i.e., the plaintext space, the key space and the cipher space is equal to the set of all ordered m-tuples of elements from Z26. Encryption and decryption functions for the key

$$\mathbb{K} = (k_1,\ldots,k_m) \in (\mathbb{Z}26)^3 \text{ are given with}$$

$$e_K(x)=(x_1+k_1,\ldots,x_m+k_m) \bmod 26,$$

$$d_K(y)=(y_1-k_1,\ldots,y_m-k_m) \bmod 26,$$

$$\text{where } x = (x_1,\ldots,x_m), y = (y_1,\ldots,y_m) \in (\mathbb{Z}26)^m".$$

Using the numerical equivalents in the previous example we have:

Encryption is much simpler with the use of Vigenère's square (Figure 3.). The encrypted letter is located in the intersection of the column that begins with the letter contained in the plaintext and the line that begins with the letter of the keyword. For the decryption a completely analogous procedure is applied.

	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
1	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	A
2	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B
3	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C
4	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D
5	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E
6	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F
7	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G
8	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H
9	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I
10	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J
11	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K
12	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L
13	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M
14	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N
15	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
16	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
17	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
18	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
19	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
20	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
21	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
22	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
23	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
24	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
25	Z	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
26	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

Figure 3. Vigenère's square

In Vigenère's square, the top row represents the 26 letters of the English alphabet. While under each letter there is a differently ciphered alphabet, i.e., each is shifted by one letter from the previous alphabet (Veinović and Adamovuć, 2013). The complexity of this cipher is that a person can, for example, encrypt the first letter using the 8th order alphabet, the second letter using the 14th order alphabet, the third letter using the 9th order, etc. In order for the person to know which line to use, a keyword (i.e., a key) is introduced that is known only to the recipient and the sender.

If we do not know the key, statistical methods are used to find the length of the key, and then frequency analysis allows us to find the key. In order for encryption to be as secure as possible, the keyword should be as long as possible. A completely secure system is the one with an "infinitely" long key. Also, the system is more secure if the entire key consists of random characters, but this is not very practical.

CONCLUSION

The history of classic cryptography is a very extensive topic. Historically, the most interesting part of the classic cryptography is certainly the period of the First and Second World Wars, in which ciphers and cryptanalysts played a major role. Substitution ciphers are very important for cryptology as a science. Due to its long development period, various cryptosystems have been developed, each of which tried to correct some flaw of the previous one. These flaws were usually revealed by new aspirations and efforts for successful attacks on specific ciphers.

The results of both this and related research indicate that there is a great interest and need for greater expansion and improvement of cryptographic models, because today cryptosystems play a significant role in modern technology, i.e., technologies involving communication rely on ciphers to ensure security and privacy.

The goals of the work have been proven, the cryptosystems listed in this paper, as well as some other important systems, were a necessary basis for the development of modern security computer systems that today drive all aspects of human life. With the increasing development of information technologies and the Internet as a transmitter of messages and information, there is a need to protect them.

In terms of further research, after this initial overview of the state of the field, a more detailed analysis of Hebrew and Playfair ciphers is planned, as well as research into cryptosystems with a secret key – the symmetric cryptosystems. As well as improving the implementation of proposed solutions in the field of substitution ciphers in those models.

REFERENCES

[1] A. Dujella, M. Maretić, Kriptografija, Element, Zagreb, 2007.

[2] A. Dujella, Teorija brojeva, Školska knjiga, Zagreb, 2019.

[3] D. Kahn, The Code-Breakers: The Story of Secret Writing, The Macmillan Company, New York, 1967.

[4] I. Matić, Uvod u teoriju brojeva, Odjel za matematiku, Sveučilište u Osijeku, 2015.

[5] M.Veinović, S.Adamović, Kriptologija 1, Univerzitet Singidunum, Beograd, 2013.

[6] S. Singh, Šifre : Kratka povijest kriptografije, Mozaik knjiga, Zagreb, 2003.

[7] G. Baumslag, B. Fine, M. Kreuzer, G. Rosenberger, A Course in Mathematical Cryptography, De Gruyter, Boston, 2015.

[8] B. Ibrahimpašić, Kriptografija kroz primjere, Pedagoški fakultet Bihač, 2011.

[9] Azizović, E., Maznikar, B. (2019). Digital competences of public notaries, Knowledge in practice -International Journal, Vol. 35 No. 5.

[10] A.Dujella, M. Maretić, Kriptografija: https://web.math.pmf.unizg.hr/~duje/kript/kriptografija.html, 24.5.2022.

[11] https://www.boxentriq.com/code-breaking, 26.5.2022.