kriativ.tech

Edição Nº 9, 11-10-2021
ISSN Print: 1646-9976 | ISSN Online: 2184-223X |
http://www.kriativ-tech.com
DOI: 10.31112/kriativ-tech-2021-10-54

Improving Caesar Cipher for greater security

António Santos
Assistant Professor at ISTEC – asisanto@my.istec.pt

Renato Vasconcelos Júnior
BSc Multimedia Engineer graduated at ISTEC – xrenaport@gmail.com

Abstract: *Before the invention of computers all cryptographic methods were calculated manually, and as such the cryptographic methods developed during that period took into account this limitation. The Caesar Cipher method was one of the first to be used and disseminated in several countries. This method is very simple, which implies that with current means you can break your security quickly and easily. However, it has a characteristic that, given its nature, any change in the method increases its safety, and like other authors in this article, it will be shown that a small change will imply some improvement in the method's safety.*

Keywords: *Encryption, replacement, Caesar's Cipher, Enhancement.*

I. Introduction

Since the beginning of time, human beings have sought to socialize within their communities as well as the most distant ones. In order for this integration/socialization to be complete, they invented sounds which allowed the elements of this group to communicate with each other, and then language appeared. But over time humans began to disperse and there was a need to communicate with humans farther and farther away and with that need they evolved into writing and sending messages. Messages could be transmitted in various ways, such as: signals, text, sounds, etc.

Based on text messages, they can be transmitted in terms of their form: clear text, steganography and encryption. The plain text as being the text in natural language, or according to [1] states that the clear text is the text of the message in ordinary language, on the other hand, [2] defines this as the text that can be understood by anyone who knows the language, as long as the message is not encoded in any way.

Taking into account the need for a message to reach the recipient, as far as possible, without unauthorized third parties having access to it, techniques have emerged that allow hiding the message. This need allied to human curiosity promoted scientific advances in this area, and, due to this gap, techniques were developed. The two best known techniques for hiding messages/data are shorthand and encryption. Steganography can be defined as the concealment of the message; [3] describes this as writing hidden from view to all. Finally, [1] writes that steganography consists in hiding the very existence of the message.

With regard to cryptography, it is understood as the basic text encoding, so that unwanted people do not have access to the content. According to [2], cryptography is the art of obtaining security by encoding messages to make them unreadable, while [4] goes further, stating that nowadays cryptography is the science of secret writing with the objective of hiding the meaning of a message. And last but not least, [5] define cryptography is the science of keeping secrets secret. Although encryption and steganography are independent, it is possible to combine the two in order to scramble and hide a message to maximize security [6].

Cryptography, according to [7], is divided into two categories depending on the type of security keys used to encrypt/decrypt

data, these techniques are: asymmetric and symmetric encryption. Symmetric or single-key encryption uses the same key to encrypt and decrypt. With this type of encryption, both the sender and the receiver know the same secret code, called a key. Messages are encrypted by the sender using the key and decrypted by the recipient using the same key [8]. [9] also writes that symmetric cryptography is a form of cryptosystem in which encryption and decryption are performed using the same key. While symmetric cryptography, as noted, uses one public key, asymmetric cryptography uses two, one public and one private. [8] define asymmetric cryptography, also called public-key cryptography, uses a pair of keys to encrypt and decrypt. With public-key cryptography, keys work in combined public and private key pairs. The public key can be freely distributed without compromising the private key, which must be kept secret by its owner. As these keys function only as a pair, encryption initiated with the public key can only be decrypted with the corresponding private key.

In addition to being of the two types mentioned above, cryptography uses two techniques: transposition and substitution [6]. According to [6] in transposition, the letters of the message are simply reorganized, effectively generating an anagram. Whereas [2] in the substitution technique, the characters of a plain text message are replaced by other characters, numbers or symbols.

According to [10], the term cipher can be defined as another word for the definition of algorithm. Generally speaking, ciphers are simpler forms of algorithms than those used today. Many of the initial ciphers were very easy to decipher. Nowadays, the principles that were developed in the old ciphers are used, however, with evolution; a lot of complexity has been implemented in order to make the message safer and more difficult to break. In other words, [2] writes that cipher text is the result when plain text is encoded using any suitable scheme

After some definitions, the article will focus on the Caesar cipher that is part of symmetric cryptography.

II. Caesar's cipher

The art of cryptography is considered to be born together with the art of writing. As civilizations evolved, human beings organized themselves into tribes, groups and kingdoms. This led to the emergence of ideas such as power, battles, supremacy and politics. These ideas further fueled people's natural need to secretly communicate with the selective recipient, which in turn ensured the continued evolution of cryptography as well. The roots of cryptography are found in civilizations: Roman and Egyptian [11].

The "Caesar Cipher" is one of the earliest known ciphers, although according to [1], Caesar was probably not the original inventor of what I now call the Caesar cipher, but he certainly made it popular and is described in "De Vita Caesarum , Divus Julius" ("The Lives of the Caesars, The Deified Julius"), written in approximately 110 BC [12] It was by Julius Caesar, Roman military and political leader, to send secret messages to his generals and allies. For Julius Caesar, the security of his information was essential in order to guarantee his success. Caesar was responsible for developing a system in which it was intended to guarantee the security of his messages, which if intercepted by his enemies, could not be decrypted without the use of a key. In case one of your messages was intercepted, your opponent could not read it. What at the time was a great advantage over the opponents of the emperor and the empire.

Despite the apparent simplicity of this cipher, even at the time, the messages intercepted by potential enemies were incomprehensible due to the high level of illiteracy that existed, often being incomprehensible because they were considered to be of foreign script.

2.1 Introduction

The Caesar cipher, according to [9], uses the substitution cipher, which involves replacing each letter of the alphabet by the letter that is three places to the right in the alphabet, that is, for each letter or character there is a letter/character corresponding encoded. In Caesar's original cipher, there is a replacement for the corresponding “encryptable” letter with a certain rotation. This rotation corresponds to the

amount of letters or characters that are traversed in the alphabet, starting from the letter to be encrypted. To decode a particular message, the recipient of the message would also have privileged access to the information to proceed with its decryption. One needs to know said rotation, to be applied to the flat message. This rotation number indicates the change or displacement of the letters – thus being the key to decrypt the message. This cipher is probably one of the best known in the world [10].

The base Caesar cipher is only defined for the 26 characters of the alphabet, hence and according to [13], spaces and punctuation are omitted, and these can be replaced by other characters, or even left as space characters, which makes easier your decryption.

As it is 26 characters, in the encryption and in the Caesar cipher there is a 3-position advance, the following is observed:

A →D, B→E, …W→Z and the rest are projected into the initials, such as: X→A, Y→B e Z→C.

If we represent each character by a decimal value according to the order of the character within the alphabet: A →0, B→1, …, X→23, Y→24, Z→25. Applying the Caesar cipher offset to these (add 3): A →0+3, B→1+3, …, X→23+3, Y→24+3, Z→25+3, would result: A →3, B→4, …, X→26, Y→27, Z→28; as position 3 represents the D, so A→D, 4 represents the E, B→E, and so on until the W→Z. From 26 onwards, the remainder of the division by 26 can be taken as the value, that is: 26=26*1+0, 27=26*1+1 e 28=26*1+2, taking the remainder of the division as the value, it follows that: X→0, Y→1, Z→2, which in turn: X→A, Y→B, Z→C. Based on this reasoning, E[x]=x+3 (mod 26) can be used, where mod 26 is the remainder of the division of x+3 by 26, x the decimal value to be transformed and E[x] represents the encryption of the character whose value is x, that is, x=0,1,2,3,4, … , 25.

The reverse encryption process, in order to recover the original text of the message from its encrypted version, is called decryption or decryption [14]. According to [13], in César's cipher, decryption is performed by shifting three spaces backwards (left).

To represent decryption, inverse to encryption, the same methodology will be used:

Taking: A →0, B→1, …, X→23, Y→24, Z→25. Applying the inverse shift of the Caesar cipher to these (subtract 3): A →0-3, B→1-3, …, X→23-3, Y→24-3, Z→25-3, would result: A →-3, B→-2, C→-1, D→0…, X→20, Y→21, Z→22; as position 0 represents A, then D→A, 1 represents the B, so E→B, and so on until the Z→W, missing the first three, then the A→X, B→Y and C→Z. Deducing the equation D[x]=x-3 (mod 26), where D[x] represents the decryption of the character with the decimal value x.

According to [15], the Caesar cipher is a cipher whose cipher text is obtained from the plain text by moving each letter a fixed value of positions, that is, it is not limited to three positions. On the other hand, [16] writes that although Caesar's cipher is reputed to have used a shift of three to the right/left, any shift pattern will work with this method, shifting to the right or left. for any number of spaces. [17] also writes that it is not mandatory to apply a change of 3 characters, any value can be used, although he emphasizes that only values strictly comprised between 0 and 26 offer different encryptions. This type of cipher is called a displacement cipher. According to [12], the displacement cipher is similar to the Caesar cipher, but the displacement (k) is introduced and this k is a decimal between 0 and 25. From here, the following formulas can be drawn: encryption: E[x]=x+k (mod 26) and decryption: D[x]=x-k (mod 26).

According to [5], in classical cryptography schemes, more specifically the symmetric ones, the encryption depends on the same secret key k. This k key is used to encrypt and decrypt. On the other hand [18] states that the Caesar Cipher is a direct standard alphabet with the specific key 3. From these two quotes it can be seen that the k used in the encryption/decryption formulas of the Shift Cipher is the secret key.

2.2 Caesar's Algorithm

From the introduction to the Caesar cipher, it can be seen that the character read

advances three letters in the alphabet, and from the X it starts at A. To demonstrate the methodology, Bowne's algorithm in Python programming language [19] was used as a basis. By observing the code during its execution, it only accepts the characters of the alphabet and in capital letters, since the first line of the code presented, which is the base of the alphabet: alpha = "ABCDEFGHIJKLMNOPQRSTUVWXYZ", indicates the alphabet used in the encryption by the Caesar cipher The following is the representative code of the Caesar cipher, according to [19]:

```
alpha = "ABCDEFGHIJKLMNOPQRSTUVWXYZ"
str_in = input("Enter message, like HELLO: ")
n = len(str_in)
str_out = ""

for i in range(n):
    c = str_in[i]
    loc = alpha.find(c)
    nloc = (loc + 3)%26
    str_out += alpha[nloc]
    print (i, c, loc,nloc, str_out)
print("Obfuscated version:", str_out)
```

For the implementation of the Caesar cipher the author, as mentioned in the paragraph above, starts with the introduction of the alphabet. This is followed by the reading of the plain text (input), which will be stored in the variable: str_in, whose size will be stored in the variable n. Then the author implements a "for" loop where he takes character by character of the word inscribed inside the string str_in and with this character he searches it in the alphabet with alpha.find(c) which returns the position of the character in the alphabet, to this value he adds 3 and find the remainder of the division by 26 (nloc = (loc + 3)%26) and introduce the character represented in the output variable, followed by printing character by character. At the end the output word is printed.

Although the decryption algorithm is not found in the [19], [9], writes that the decryption algorithm is essentially the encryption algorithm executed in the reverse direction, taking the cipher text and the secret key and transforms this into the plaintext. Taking this into account, to the previous code, where the value three is added, in this case the value three is removed: nloc = (loc - 3)%26.

When analyzing the code, it can be seen that it is very limited. First it only recognizes capital letters, giving errors with numbers, punctuations and small letters, in addition to the limitations of Cesar's cipher. According to [20], simple substitution ciphers, such as this one, can be easily broken because the cipher does not hide the underlying frequencies of the different characters of the plain text, on the other hand, [11] states that as the secret key is a value between 0 and 25, in this case three, a brute force attack can break the cipher scheme in a short period of time.

2.3 Results

The code was executed, in order to verify its real functioning, for that the word "ISTEC" was used, and the result was "LVWHF", as shown in Picture 1.

Picture 1: Caesar cipher encryption according to the algorithm of [19]

By observing Picture 1, in position 0 there is the character "I" which is in position 8 of the alphabet array, adding to this 8 the displacement value 3, we obtain 11 whose position in the alphabet corresponds to "L". This is followed by the "S" which has position 18 in the alphabet, added 3 gives 21 which corresponds to the "V", and so on, until the last character entered.

In the case of decryption, instead of adding 3 to the base value, remove 3. Taking as the base word the encrypted text: "LVWHF", it is intended that its decryption result in the word "ISTEC".

Picture 2: Caesar cipher decryption according to the algorithm of [19]

Observing Picture 2, in position 0 there is the character “L” which is in position 11 of the alphabet, subtracting from this 11 the value 3 of the displacement, we obtain 8 whose position in the alphabet corresponds to “I”. This is followed by the “V” which has position 21 in the alphabet, subtracting 3 gives 21 which corresponds to the “S”; and so on, until the last character entered.

This code has some limitations, in addition to only having the possibility of displacement equal to 3, also any other character that is not contained in the alphabet from "A" to "Z" and capitalized, is encrypted to "C" and decrypted to "Z”, that is, it takes the maximum value (25).

III. Caesar Algorithm Modified

The Caesar cipher is quite easy to break security as referred to by [20] and [21], mentioned earlier. According to [22], a modification of the Caesar algorithm serves to overcome some of the weaknesses and limitations of the Caesar cipher, which can be improved using one or more different encryption algorithms. In this article we will make some modifications to Caesar's algorithm in order to bring some security.

3.1 Caesar’s Algorithm Modified

As mentioned before, the Caesar, Bowne algorithm [19], when encrypting, advances three positions to the right in the alphabet, and moves back three places in the case of decryption. If the displacement can be chosen, we are faced with the displacement cipher, in which it advances or retreats as many places in the alphabet as the chosen value (k), with $0 \leq k \leq 25$. With regard to the two previous cases, he is only talking, as in the case of Bowne's algorithm [19], about capital characters of the alphabet. This for the present is not very viable, because if one intends to write a text with uppercase and lowercase characters, spaces and numbers, it would be limited.

In this algorithm, it is proposed to use ASCII code characters as a basis or alphabet. That is, first the read character will be converted to ASCII code and then the sum or subtraction of the displacement value modulo 256 is done, in which the displacement value varies between 0 and 255. Analytically representing: encryption: $E[x]=x+k \ (mod\ 256)$ and decryption: $D[x]=x-k \ (mod\ 256)$. Finally, to improve security, the cipher text is inverted.

To optimize the code, the encryption function will be used for the encryption and decryption, using the offset complement of k for the decryption case, that is: $D[x]=xk \ (mod\ 256)=x+(256-k) \ (mod\ 256)$.

The Python programming language was used to create code to demonstrate the method's efficiency. We started by creating a function that encrypts the plaintext given a shift of k values.

```
def encripta(frase, k):
    textoc=''
    j=0
    for i in frase:
        c=(ord(i)+k)%256
        textoc=textoc+chr(c)
        print(" ",j , i,ord(i), c, textoc )
        j=j+1
    print("Encrypted message before the
inversion: ",textoc)
return textoc[::-1]
```

By analyzing the encrypt function, it receives the sentence and the displacement, initializes two variables (textoc and j), a cycle follows that goes through the entire sentence and for each character it gets the value of the ASCII code (ord(i)) sum - if ke calculates the remainder of the division by 256. After this calculation, the chr(c) command is used, which transforms the previously calculated value into the representative character according to the ASCII code. After going through the entire sentence and transforming it with the displacements, the sentence is inverted.

To run the above function, you first ask for the offset value and the phrase, and call the encryption function, as follows:

```
k = int(input("Enter the shift value: "))
textopl = input("Enter the message to encrypt: ")
textocf=encripta(textopl, k)
print("Encrypted message:",textocf)
```

To decrypt it works the same as encrypting, it asks for the offset value and the ciphertext and calls the encryption function with the phrase and the 256-k offset, as follows:

```
print("Enter the shift value:",k)
print("Enter encrypted message:",textocf)
plano=encripta(textocf,256-k)
print("Plain text:", plano)
```

After explaining the code, testing and analysis of results follows.

2.3 Results

In order to test the implemented algorithm to verify its real operation, the text "ISTEC 2021" was used with the displacement value 25, and the result was "JKIK9\^mlb", as shown Picture 3.

```
Enter the shift value: 25
Enter the message to encrypt: ISTEC 2021
0 I 73 98 b
1 S 83 108 bl
2 T 84 109 blm
3 E 69 94 blm^
4 C 67 92 blm^\
5   32 57 blm^\9
6 2 50 75 blm^\9K
7 0 48 73 blm^\9KI
8 2 50 75 blm^\9KIK
9 1 49 74 blm^\9KIKJ
Encrypted message before the inversion:  blm^\9KIKJ
Encrypted message: JKIK9\^mlb
```

Picture 3: Caesar Cipher Encryption Using Improved Algorithm

By observing figure Picture 3, in position 0 there is the character "I" which is represented by the value 73 of the ASCII code, adding to this value 25 of the displacement, 98 is obtained, which represents the "b" in the ASCII code. The "S" follows, which is represented by the value 83 of the ASCII code, adding to this displacement value 25, we obtain 108 which corresponds to the "l" and so on, until the last character entered. Then follows the inversion of the ciphertext.

For the case of decryption, instead of adding 25 to the base value, 256-25=231 is added. Taking as the base word the encrypted text: "JKIK9\^mlb", it is intended that the decryption of this result in the word "ISTEC 2021".

```
Enter the shift value: 25
Enter encrypted message: JKIK9\^mlb
0 J 74 49 1
1 K 75 50 12
2 I 73 48 120
3 K 75 50 1202
4 9 57 32 1202
5 \ 92 67 1202 C
6 ^ 94 69 1202 CE
7 m 109 84 1202 CET
8 l 108 83 1202 CETS
9 b 98 73 1202 CETSI
Encrypted message before the inversion:  1202 CETSI
Plain text: ISTEC 2021
```

Picture 4: Caesar Cipher Decryption Using Improved Algorithm

By observing Picture 4, in position 0 there is the character "J" which is in position 74 of the ASCII code, adding to this 231 (mod 256) we obtain the value 49 whose position in the ASCII code corresponds to " 1". It follows the "K" which has position 75 in the ASCII table, adding 231 (mod 256) to it, we obtain the value we obtain 50 which corresponds to "2"; and so on, until the last character entered. And finally, the inversion takes place.

This code has some limitations, the most obvious being that when a character appears repeated in plaintext it appears repeated in ciphertext.

IV. Conclusion

The Caesar cipher, although considered one of the simplest of the implemented ciphers, was also one of the first and is undoubtedly the most versatile and as such serves as the basis for many of the other ciphers developed over time. On the other hand, given its versatility, with small changes it also changes, more or less proportionally, its security against attacks. In this case the limitation, two equal characters in plaintext remain equal in ciphertext and this may imply a break using character frequency as the basis for the break attempt.

So these days, programmers can use the Caesar cipher, modifying it according to the security needs of the data in question. For

internal data of a company, classified as unimportant, it should always be encrypted, and a simple encryption with some limitations is better than none.

V. References

[1] Holden, Joshua (2017). The Mathematics of Secrets: Cryptography from Caesar Ciphers to Digital Encryption, Princeton University Press, New Jersey

[2] Kahate, Atul (2003). Cryptography and network security, Tata McGraw-Hill, New Delhi,India.

[3] Kipper, G. (2004). Investigator´s Guide to Steganography. Auerbach Publications. USA.

[4] Paar, Christof and Pelzl, Jan (2010). Understanding Cryptography: A Textbook for Students and Practitioners. Springer-Verlag, Berlin, Germany.

[5] Delfs, Hans and Knebl, Helmut (2007). Introduction to Cryptography: Principles and Applications, Second Edition, Springer-Verlag, Berlin, Germany.

[6] Singh, Simon (1999).The Code Book, Anchor Books: The Science of Secrecy from Ancient Egypt to Quantum Cryptography. Anchor Boks. New York, USA

[7] Aggarwal, Surabhi (2016). A Review on Enhancing Caesar Cipher, International Journal of Research Science & Management, 3(6).

[8] Shrivastava, Manish, Jain, Shubham and Singh, Pushkar (2016). Content Based Symmetric Key Algorithm, International Conference on Computational Modeling and Security, Procedia Computer Science 85

[9] Stallings, William (2011). Cryptography and network security: Principles and Practice, Fifth Edition, Prentice Hall, New York

[10] Cobb, C. (2004). Cryptography for Dummies. Hoboken - New Jersey - United States of America: Wiley Publishing.

[11] Kumari, Sarita (2017). A research Paper on Cryptography Encryption and Compression Techniques, International Journal Of Engineering And Computer Science. Volume 6 Issue 4Page No. 20915-20919. DOI: 10.18535/ijecs/v6i4.20

[12] Katz, Jonathan and Lindell, Yehuda (2008). Introduction to Modern Cryptography, Chapman & Hall/CRC, Taylor & Francis Group, Florida, USA.

[13] Trappe, Wade and Washington, Lawrence (2006). Introduction to Cryptography with Coding Theory, Second Edition, Pearson Education Inc., Pearson-Prentice Hall. New Jersey, USA.

[14] Churchhouse. Robert (2004). Codes and ciphers: Julius Caesar, the Enigma and the Internet, Cambridge University Press. Cambridge, UK.

[15] Baldoni, M. W.; Ciliberto, Ciro and Cattaneo, G. M. P. (2009). Elementary Number Theory, Cryptography and Codes. Springer-Verlag, Rome, Italy.

[16] Easttom, William (2021). Modern Cryptography Applied Mathematics for Encryption and Information Security, Springer Nature Switzerland AG, Springer. Cham, Switzerland.

[17] Bauer, Craig (2013). Secret History: The Story of Cryptology, Chapman and Hall/CRC, Filadelfia, USA.

[18] Sinkov, Abraham (1966), Elementary Cryptanalysis - A Mathematical Approach, Fifth Printing, The Mathematical Association of America. Washington, USA.

[19] Bowne, Samuel (2018). Hands-On Cryptography with Python, Packt Publishing, Birmingham, UK.

[20] Schneier, Bruce (1996). Applied Cryptography, Second Edition. John Wiley & Sons

[21] Musa, Sarhan M. (2018). Network Security and Cryptography: A Self-teaching Introduction. Mercury Learning & Information. Virgínia, USA

[22] Jain, Atish; Dedhia, Ronak and Patil, Abhijit (2015). Enhancing the Security of Caesar Cipher Substitution Method using a Randomized Approach for more Secure Communication. International Journal of Computer Applications. 129(13)