Data Security and Cryptography Substitution Techniques Algorithms Implementation

Abdulmajid M. Afat
Misurata University / Department of Computer Science
Misurata, Libya
majid.afat@it.misuratau.edu.ly

***Abstract*— Cryptography and data security have been used to make data more secure for many years, military communications are having a major impact on growing cryptography field. Although the internet have been invited in the ends 1960's, but it is not appear to public until invent the World Wide Web (www) in 1989. The World Wide Web allows people to e-mail each other and transfer data. This appeared as one reason in growing use of computer. In addition needs to make e-commerce more secure. Cryptography understanding helps to find better ways or protocols to make transfer data faster and more efficient.**

This paper starts with a brief background about security and cryptography. It provides central concepts of cryptography. Also provides an account of substitution techniques algorithms and their implementation. The implementation will be as an application in c# language that reads a text, encodes it by a specified algorithm and decodes the encrypted text. In addition explaining the mathematical functions used in the algorithms.

Index Terms: **key, cryptography, cipher, algorithm, plaintext.**

I. INTRODUCTION

Personal data or other types of data transmitted around the world can be the target of hacker software that break security rules, this requires the use of certain techniques to make data secure and confident. Therefore, it is important to invest some time studying security and cryptography.

Cryptography definitions:

Cryptography change data to be understood by limited persons which have authorization.

"Cryptography is the science and study of secret writing. A cipher is a secret method of writing, whereby plaintext (or clear-text) is transformed into cipher-text, sometimes called a cryptogram" [1].

"Data security is the science and study of methods of protecting data in computer and communications systems" [1].

Received 26 Sep, 2019; revised 2 Sep, 2019; accepted 25 Sep, 2019.

Available online Sep 28, 2019.

Symmetric cipher model:

There are five components of the symmetric cipher model as follows:

Plaintext: is the input of the encryption algorithm which is taken form of the source text or message.

• Encryption algorithm: The encryption algorithm processes the plain text.

• Decryption algorithm: It is algorithm that works as an encryption algorithm in a reversal manner.

• Secret key: The secret key also appears as input to the algorithm and it must be secure.

• Ciphertext: It is garbage message as output of encryption algorithm.

For ciphertext and plaintext they can be as input or output regarding to the situation figure(1).

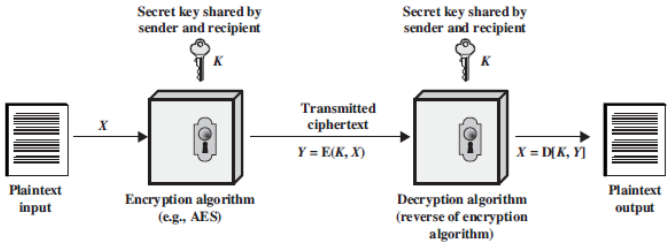

Figure 1. Model of Symmetric Cipher Model.

II. SUBSTITUTION TECHNIQUES

Cryptography algorithm:

"This is the study of techniques for ensuring the secrecy and/or authenticity of information" [2].

A substitution technique is the process of changing the plaintext by replacing letters, numbers or symbols.

2.1 Caesar cipher:

In cryptography, a Caesar cipher is the oldest known algorithm to encrypt data. It was named after Julius Caesar; it works by changing each letter in plaintext by the third letter after that letter.

Example:

Cryptography of: fubswrjudskb

Plain: 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

cipher: 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

For each letter in alphabet has a number starting by 0 as the following:

0	1	2	3	4	5	6	7	8	9	10	11	12
a	b	c	d	e	f	g	h	i	j	k	l	m
13	14	15	16	17	18	19	20	21	22	23	24	25
n	o	p	q	r	s	t	u	v	w	x	y	z

Figure 2. Alphabet Indexes.

Expressed algorithm: for each letter in plain text p there is cipher letter c.

$$c=E(3,p)=(p+3) \bmod 26$$

The decryption algorithm as follows:

$$p=D(3,c)=(c-3) \bmod 26$$

The number 3 called Key and it can be replaced by another number, where it's known by data sender and data receiver. Similarly, other characters can be included in previous algorithm.

2.2 Monoalphabetic Cipher:

In this algorithm there is no specific encryption pattern. The key is chosen by sorting alphabet and matching each letter randomly with another as follows:

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

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

Using the random key strategy makes Monoalphabetic better than Caesar.

Example:

cryptography zgdbevyglbod

The decryption algorithm is by reversing the encryption algorithm.

2.3 Playfair Cipher:

"The best-known multiple-letter encryption cipher is the Playfair" [2]. In this algorithm, the key is a two dimensional matrix in which the elements of the key word has unique letters, the remaining letters appear in their alphabetical order.

Example:

If the keyword is (LIBYA) then the matrix will be as the follow:

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

Figure 3. Alphabet Indexes.

The English alphabet has 26 letters. For this reason, the letters I and J are placed in the same cell.

After specifying the matrix, the playfair algorithm classifies the plaintext into blocks, each of which has two letters. The block must consist of two deferent letters. If they are the same the second one is replaced by the letter x.

Finally the encryption of the pair of letters in each block is done according to the following:

1. If the letters position are the same row in the matrix, they are replaced by the next letter on the right side.
2. If the letters position are at the same column in the matrix, they are replaced by the letter in the cell beneath.
3. If none of the conditions in 1 and 2 are met, then the letters are replaced by the letter in the cell where they intersect.

Example:

CR YP TO GR AP HY

EP LS ZT ET LT NL

Notes: in the plain text the same letters will encoded by deferent letters.

2.4 Vigenere cipher:

"The best known, and one of the simplest, polyalphabetic ciphers is the Vigenere cipher" [2]. It employs techniques from both Playfair and Caesar algorithms. A random key word is determined and the letters are given indexes as in figure (2). The keyword is repeated according to the plain text. The cipher text results from summation of index of each letter in the plaintext with the index of the corresponding letter from the repeated key word.

In other words assume that the plain text is:

P = p0, p1, p2,... , pn-1

K = k0, k1, k2,…, kn-1

C = C0, C1, C2, … , Cn-1

Where C calculated as the following

C = C0, C1, C2, …, Cn-1 = E(K, P) = E[(k0, k1, k2,…, km-1), (p0, p1, p2,… , pn-1)]

= (p0 + k0) mod 26, (p1 + k1) mod 26,…, (pm-1 + kn-1) mod 26,

(pm + k0) mod 26, (pm+1 + k1) mod 26,…, (p2m-1 + km-1) mod 26, … etc

Example:

Plain text= NOT IMPOSIPLE

Key = COM PUTERCOMP

15 28 31 23 32 34 18 35 10 29 23 19

2 5 6 8 9 3

Cipher text= PCFXGISJKDXT

In decryption algorithm reverse the algorithm above:

Pi=(Ci-Ki) mode 26

Decryption problem: in the example above, the second letter in the cipher text 'C' results from (14+14) mod 26=2. After implementing the decryption algorithm (absolute value (2-14) 12 mod 26=12) the resulting letter is 'M'. The problem can be solved in the following way:

if ((Ci-Ki) < 0) then Pi = (Ci-Ki) +26;

else Pi = (Ci-Ki);

Now, (2-14) +26=14 the resulted letter is O.

Note: the length of the key word should be as close as possible to the length of the plaintext and have no statistical relationship between them.

2.5 Hill Cipher:

"Another interesting multiletter cipher is the Hill cipher, developed by the mathematician Lester Hill in 1929" [2].

The key in this algorithm is specified square matrix, which has an inverse.

Definition of Matrix invers:

If A is a matrix then A^{-1} is an inverse to A, where $A \times A^{-1}=I$.

To calculate inverse of matrix A3×3.

$A^{-1} = (1/|A|) \times adj(A)$

Example:

Encryption algorithm:

if the key is $\begin{pmatrix} k_{11} & k_{21} \\ k_{12} & k_{22} \end{pmatrix}$

C = PK mod 26

$(c1\ c2) = (p1\ p2) \begin{pmatrix} k_{11} & k_{21} \\ k_{12} & k_{22} \end{pmatrix} \bmod 26$

Where pi is an index of a letter in plaintext in the position i. ci is the same with pi but in cipher text.

Decryption algorithm:

To decrypt the cipher text, the first step is to find the invers of the key as bellow:

Calculating the determinant D:

k11*k22-k21*k12=D

D inverse mod 26=D inverse

Creating a new matrix(key inverse) by performing a swap between the two elements in the main diagonal, and multiplying -1 to the secondary diagonal .

Finally by finding key inverse mod 26× (D inverse)

Example:

Key= $\begin{pmatrix} 3 & 3 \\ 2 & 5 \end{pmatrix}$

determinant=(3×5)-(3×6)=-121

-121 mod 26=9

9 inverse mod 26=3

To calculate b mod m using the following algorithm:

If b>=0 then Return reminder Else Return b-(integer(b/m)-1)*m

by swapping the elements of the main diagonal, and multiplying the elements of secondary diagonal by -1. The resulted matrix is key inverse as follow

:

Key-1= $\begin{pmatrix} 5 & -3 \\ -2 & 3 \end{pmatrix}$ ×mod 26

$\begin{pmatrix} 5 & 23 \\ 24 & 3 \end{pmatrix}$ ×3= $\begin{pmatrix} 15 & 69 \\ 72 & 9 \end{pmatrix}$

Multiplying the matrix again by mode 26, because 69,72 are more than 26.

Key inverse mod 26×3= $\begin{pmatrix} 15 & 17 \\ 20 & 9 \end{pmatrix}$

For example, to encrypt (comp) by using the key above:

Starting with the first 2 letters (c o)

$(c1\ c2) = (2\ 14) \begin{pmatrix} 3 & 3 \\ 2 & 5 \end{pmatrix}$

mod 26=(34 76) mod 26=(8 24)

The same manner to the next group(m p)

$(c3\ c4) = (12\ 15) \begin{pmatrix} 3 & 3 \\ 2 & 5 \end{pmatrix}$

mod 26=(66 111) mod26

=(14 7)

The positions of cipher text letters for the given plaintext are (8 24 14 7), which are respectively represented by iyoh.

To decrypt iyoh by using the decryption key $\begin{pmatrix} 15 & 17 \\ 20 & 9 \end{pmatrix}$

$(p1\ p2) = (8\ 24) \begin{pmatrix} 15 & 17 \\ 20 & 9 \end{pmatrix}$

mod 26= (600 352) mod 26=(2 14)

$(p1\ p2) = (14\ 7) \begin{pmatrix} 15 & 17 \\ 20 & 9 \end{pmatrix}$

mod 26=(350 301) mod 26=(12 15)

The results of decryption are (2 14 12 15), which represent letters in the fig.2 respectively c, o, m and p, which is the original plaintext.

2.6 VERNAM CIPHER:

This algorithm is works on the binary representation of the plaintext or ciphertext. As in Vigener cipher, there is a binary key of appropriate length. The cipher text comes from XOR between the binary representation of the plaintext and the binary key.

Encryption algorithm: ci = pi XOR ki.

Decryption algorithm: pi = ci XOR ki

Example:

Plaintext as binary=11111000001010

Binary key=10110

11111000001010

10110101101011

Ciphertext as binary= 01001101100001

If the plaintext is longer than the binary key, the binary key is repeated to fit the plain text.

"The one-time pad is immune to a ciphertext only attack, since the ciphertext yields no information about the plaintext, other than its length" [3].

III. PROBLEM DEFINITION

For any natural language, there is a behavior and nature to each letter. Some individual letters are use more than others. such behaviors can be used to decrypt the cipher text, and figuring the key can be easy. It is called by cryptanalysis (code breaking) - study of the ways of decrypting ciphertext without knowing the key.

Cryptanalysis can be used in attacks, or as a way to evaluate how strong the algorithm is.

IV. IDEAS

To make the cipher text more secure, multiple algorithms should be used. The plaintext is divided into different parts each of which is encrypted by deferent algorithm. This multi-encryption technique becomes an additional key that cannot be estimated and the ciphertext is not easy to analyze.

V. ALGORITHMS IMPLEMENTATION

The implementation of algorithms above is an application in c# language which was designed that receives a plaintext or a ciphertext as input and produces a ciphertext or the original plaintext as output.

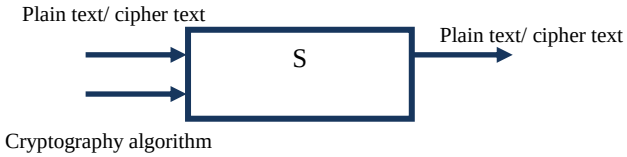

Figure 4.1.Application Diagram.

The application interface as follows:

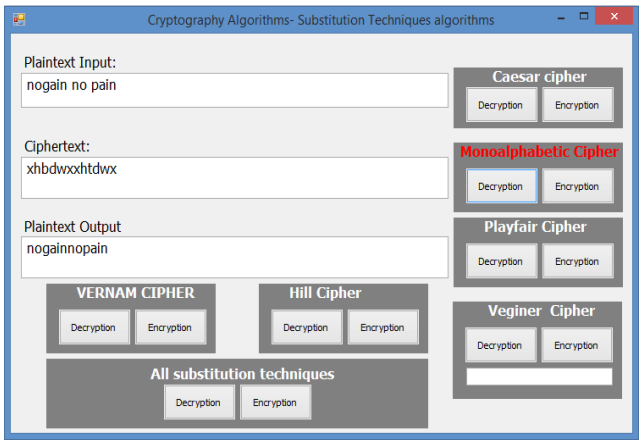

Figure 4.2. Application Interface.

To make data more secure, the application provides an option that implements all algorithms on the target plaintext.

VI. CONCLUSION

The increasing of transferring data in real life such as social media, end-to-end encryption communication and military applications leads the research to be focus on studying and inventing more powerful algorithms which can be hard to break.

VII. REFERENCES

[1] Cornelius Lowell Robling. "*Cryptography and data security*". By Addison-Wesley Longman Publishing Co., Inc., 1982.

[2] William Stallings. "*Cryptography and network security. By Pearson Education India*", 2011.

[3] Mark Stamp and Richard M. Low. "*APPLI ED CRYPTANALYSIS Breaking Ciphers in the Real World*", John Wiley & Sons 2007.