Comparison between the types of Substitute Ciphers

Jemima Abraham[1], Siddharth Nanda[2], Rajeshwari Gundla[3]
[1]U.G. Student, [2]Faculty, [3]Senior Faculty
SOE, ADYPU, Lohegaon, Pune, Maharashtra, India[1]
IT, iNurture, Bengaluru, India[2,3]

Abstract - Cryptography is an art of converting critical information into un-understandable data so that any unauthenticated person may not get hold of the sensitive data [2].Substitute cipher is one such technique which replaces one symbol with another. In substitute cipher each letter has a corresponding fixed alphabet that is substituted to get the plain text [1]. There is a variety of methods used in substituting. In this paper the comparison between the different types of substitute cipher is studied. This comparison will help in understanding which cipher is highly used, secured and difficult to crack.

Keywords - Cipher, cryptanalyst, plain text, encryption, decryption, techniques, keys, substitute, Monoalphabetic, Polyalphabetic.

I. INTRODUCTION

Today's most challenging issue is data security that affects all aspects of our day-to-day lives especially communication. However, by introducing cryptography techniques, many problems related to integrity, confidentiality, availability, authentication and authorization are solved. The security needed for applications are also provided by the cryptography techniques. Methods for securing confidential information through encryption techniques are being discovered on day to day basis [3].

Data encryption is considered to be the most efficient means to counterbalance the hackers [4]. Encryption is a process of converting messages or converting information into a form that only authorized person or group can access it. The purpose of a cryptosystem to is store data safely in a file and to ensure that the channel is secure while transferring data. However, in both scenarios mentioned above, the encrypted file does not prevent the hackers from accessing the data rather it ensures the hackers should not understand the data that is being transmitted. This process is used in various cryptographic concepts like authentication, digital cash and digital signatures. In this paper, the primary focus will be on the different techniques existing in the substitute symmetric key cipher. Further study on these techniques help in analyzing the basic approaches to symmetric encryption used these days.

II. CLASSIFICATIONS

Basic Terminologies -

Plain Text: The actual message is known as plain text.

Cipher Text: The random stream of data which is un-understandable.

Encryption Algorithm: Process of converting from plain text to cipher text.

Decryption Algorithm: Process of converting cipher text to plain text.

Key: It is the algorithm used to encrypt and decrypt the data.

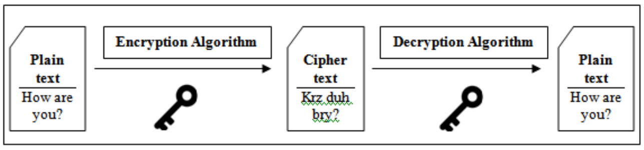


Figure 1: Simple Cryptosystem

Symmetric & Asymmetric Encryption Algorithm: There are two main methods by which encryption algorithm can be classified:

- Symmetric Key Cryptography
- Asymmetric Key Cryptography

Symmetric Key is also known as a private or secret key [7]. In this, the same key is used for both encrypting and decrypting the message. Traditional Symmetric Systems provide various security goals; they are known to provide high data confidentiality. They are entirely based on alphabets as a preliminary factor [5]. Symmetric key ciphers use substitution cipher, transposition cipher and block cipher.

Asymmetric Key uses two keys. One key, called public key is used to encrypt the message and the other key, called private key is used to decrypt the message at the receivers end. If any middle man gets hold of the public key, still it is impossible for him to assume the private key to decrypt the message [8]. This system is also called Public Key Cryptosystem.

Traditional Symmetric Key Ciphers: In "Communication Theory of Secrecy of Systems" published in the year 1949; C. Shannon explains that in classical cryptography the plaintext and the key length were the same to support secrecy through encryption [6]. Substitution and transposition are the building blocks of all encryption techniques [4]. When these two approaches come together it is called Product Cipher [2].

Substitution Cipher: As the name suggests, it substitutes the letter characters in the plaintext with a set of corresponding letters, numbers or symbols. Substitute cipher can be categorized into Monoalphabetic or Polyalphabetic ciphers.

Transposition Cipher: In this technique, the position of the alphabet is changed instead of being substituted with another alphabet. Example: Columnar transposition. In columnar transposition, the plaintext is placed horizontally with a specific alphabet, and then the cipher text is read vertically. Example of columnar transposition:

W	A	I	T	T
I	L	L	N	I
G	H	T		

Cipher Text: “WIGA LHIL TTNTI”

Substitution Cipher: The substitution cipher replaces the actual text or message with other characters thus making it impossible for the hackers to get hold of the actual message. This technique is extensively practiced as it is uncomplicated to implement.

Caesar Cipher: Caesar cipher is the best example of substitute cipher introduced initially. It is a mono-alphabetic cipher wherein each letter of the plaintext is substituted by another letter to form the cipher text. It is the easiest method of substitution cipher scheme. In this type, the cipher text is created by shifting the plaintext characters 3 steps forward, where ‘A’ become ‘D’ and ‘C’ becomes ‘F’.
Plain Text: “HIDDEN INSIDE A BLUE BOX”
Cipher Text: “KLGGHQ LQVLGH D EOXH ERA”
Note: The alphabets are wrapped around, after Z, A continues.
We can also substitute the letters in the alphabets with its corresponding integers.

A	B	C	D	E	F	G	H	I	J	K	L	M
0	1	2	3	4	5	6	7	8	9	10	11	12

N	O	P	Q	R	S	T	U	V	W	X	Y	Z
13	14	15	16	17	18	19	20	21	22	23	24	25

For this the algorithm will be as follows:
Plain text is denoted a ‘p’ along with which its cipher text is denoted with the letter ‘C’. ‘E’ denotes encryption, ‘D’ denotes decryption and ‘k’ denoted the secret key.
C = E (3, p) = (p + 3) mod 26
Any number of shifts can be applied. In that case, the formula will be:
C = E (k, p) = (p + k) mod 26
Here ‘k’ can be within 1-25. The decryption algorithm is:
P = D (k, C) = (C – k) mod 26

If the hacker finds out that Caesar cipher algorithm is used, he can use the brute force attack to easily get the plain text. Brute force attack is where the attacker uses all possible keys to determine the right key for decrypting the message.
The 3 characteristics that make Caesar cipher easy to crack are:
i). The algorithm for encryption and decryption are known.
ii). There are only 25 keys to try.
iii). The plaintext language is easily recognizable and known.
From among all the types of substitute cipher, simply shifting the plain text alphabets cyclically to get cipher text is the easiest technique to implement and is also said to be simple for the hacker to crack the code to get the plaintext message.

Monoalphabetic Cipher: Caesar cipher that was mentioned above is a type of Monoalphabetic cipher. It uses the same substitution method to get the cipher text characters for each plain text character. In Caesar cipher we see that it is easy for a hacker to crack the key as Caesar cipher provides only 25 keys in all. This pit is covered by using Monoalphabetic cipher. According to this Monoalphabetic cipher, the substitute characters symbols provide a random permutation of 26 letters of the alphabet. 26! Permutations of the alphabet go up to 4*10^26. This makes it difficult for the hacker to use brute force attack to gain the key.
However, Mono-alphabetic cipher is a kind of substitution where the relationship between a symbol in the plaintext and a symbol in the cipher text is always one-to-one and it remains fixed throughout the encryption process. These ciphers are considered highly susceptible to cryptanalysis. For example: if ‘T’ is encrypted by ‘J’ for any number of occurrences in the plain text message, then ‘T’ will always be encrypted to ‘J’.
If the plaintext is “TREE”, then the cipher text would be “ADOO” and this showcases that the cipher is probably mono-alphabetic as both the “O”s in the plaintext are encrypted with “E”s in the cipher text.
Although the hacker will not be able to use brute force attack, it is possible for him to assume the key by using the All- Fearsome Statistical Attack. If the hacker knows the nature of plaintext of any substitution cipher, then regardless of the size of the key space, he can easily break the cipher using statistical attack. Statistical attack consists of measuring the frequency distribution for characters, comparing those with similar statistics for English.
Fig.2 shows the frequencies of the letters in the English alphabet. By comparing this distribution with a histogram for the letters occurring in a piece of cipher text, you may be able to establish the true identities of the cipher text letters.

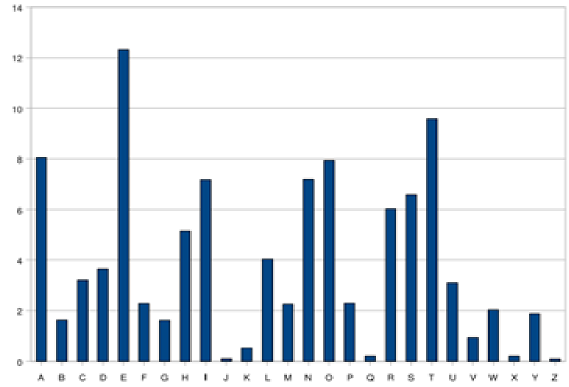


Figure 2: Relative frequencies of occurrence for the letters of the alphabet in a sample of English text.

(This figure is from Lecture 2 of "Computer and Network Security" by Avi Kak)

Polyalphabetic Cipher: The first Polyalphabetic cipher was the Alberti Cipher which was invented by Leon Battista Alberti in the year 1467. He used a random alphabet to encrypt the plaintext, but at different points he would change to a different mixed alphabet, indicating the change with an uppercase letter in the cipher text. In order to utilize this cipher, Alberti used a cipher disc to show how plaintext letters are related to cipher text letters.

In other words, each occurrence of a same letter will have a different substitute and the relationship between characters in the plaintext to a character in the cipher text is one-to-many. Playfair cipher and Vigenere Cipher are the best examples of Polyalphabetic ciphers.

On creating a polymorphic cipher, the effort is put on creating a cipher text character that is dependent on the two main factors:

i). The corresponding character

ii). The position of the plaintext character in the message.

This creates a note that the secret key must be a stream of sub-keys. These sub-keys must be depended somehow on the position of the character in the plaintext so that by using the sub-key the message can be converted into cipher text.

In plain English, we need a stream of keys k = (k1, k2, k3...) within which k1 will be used to encipher character in the plaintext and thus form a sequence on cipher text [9].

Play Fair Cipher: One character at a time substitution leaves a lot of information about the plaintext in the cipher text [4]. So play fair a Polyalphabetic cipher destroys some of the clues that map multiple characters at a time to cipher text characters. It one of the best approaches in classical encryptions, that carries out multiple-character substitution [9]. It encrypts pairs of letters unlike, Monoalphabetic cipher. In 1854, a British Scientist named Sir Charles Wheatstone invented this cipher, but the names it under the name of his friend Baron Playfair who championed the cipher at British foreign office [5].

In play fair cipher, before creating a key table an encryption key has to be chosen. While choosing the key it is necessary to make sure that there is no duplicate character in the key. Then the keys are entered in the cells of a 5 × 5 matrix key table, in a left-to-right and top-to-down pattern starting with the first cell at the top-left corner [10]. From among the 26 letters in the alphabet, only 25 are to be entered in the 5×5 key table. Usually, the letter 'J' is omitted from the table. If the plain text consists of 'J' in that case, the letter 'J' is replaced by 'I'.

Example: A sender and a receiver decide on a common key say, "SURVEY FORMAT".

So the key is placed on the key table, the first characters (going left to right) in the table is the phrase, excluding the duplicate letters. The remaining place will be filled by the leftover letters in the alphabet.

The key table will look like:

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

If there is an odd number of letters in a message, a Z is added to the last letter. If the following message, "hide money" has to be encrypted, it will be written as

HI DE MO NE YZ

Plain text can be scanned by consecutively occurring characters. The following rules are used for any given pair of plain text characters. This is done to determine the corresponding pair of cipher text characters.

Rule 1: If the pair of letters are in the same column i.e. one below the other. Then, in that case take the letter that is below these two letters.

HI DE MO **NE** YZ

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

Rule 2: If a pair of letters is in the same row i.e. one beside another, then in that case take the letter that is in the right of these pair of letters.

HI DE MO NE YZ

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

Rule 3: If the above-mentioned proceedings are false then, in that case, form a rectangle with the two letters and take the letters on the horizontal opposite corner of the rectangle.

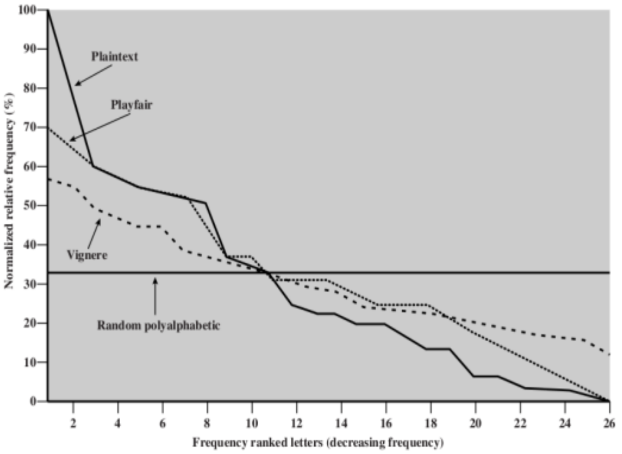
HI **DE** MO NE YZ

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

[Bold letters are the plain text and the underlined red color letters are the cipher text]

After using these above mentioned rules, the result of the encryption of ‘hide money’ with the key of ‘survey format’ would be – IK GV MA ZA AP

In World War 1, it was used as the encryption system by the British Army. It was also used by the U.S. Army and many other Allied forces in World War 2. It was thought for decades that play fair cipher is unbreakable [10]. But it turned out that is extremely easy to break. Play fair cipher does alter the frequency of individual letters but not that efficiently. According to the cryptanalysis of the Playfair cipher, it is aided by the fact that a pair of letter and its reverse will encrypt in a similar pattern. That is, if AB encrypts to XY, then BA will encrypt to YX. So by looking for words that begin and end in reversed letters one can try to compare them with plaintext words that are similar. Example of words that begin and end in reversed letters are receiver, departed, repairer, redder, denuded, etc [4]. The Playfair cipher was used mainly to protect important, yet non-critical secrets, as it is quick to use and requires no special equipment [11]. Fig. 3 shows the single-letter relative frequencies in descending order.



Relative Frequency of Occurrence of Letters

Figure 3: Single-letter relative frequencies in descending order for a class of ciphers. (This figure is taken from Chapter 2 of William Stallings: “Cryptography and Network Security”, Fourth Edition, Prentice-Hall.)

Vigenere cipher: Using different Monoalphabetic substitutions as one proceeds with the plain text is called Polyalphabetic substitution cipher. The features that are common from all these techniques are:

i). Related set of Monoalphabetic rules are followed.

ii). For a given transformation which rule is applicable is determined by the key

Vigenere cipher is one of the best examples of a Polyalphabetic cipher. Vigenere cipher uses a text string or a word as a key unlike other methods and behaves similar to Polyalphabetic Cipher but differs in using different length of key [11].

For example, let’s assume the key is ‘clue’. Each letter in the key is transformed into its corresponding numeric value: In this example, c → 3, l → 12, u → 21, and e → 5.
Thus, the key is: 3 12 21 5.

The sender and the receiver decide on a key that is ‘clue’, for which the numerical representation is ‘3 12 21 5’. Now if the sender wants to send a message saying “watch at night”, he/she will arrange the plain text and the numeric representation as follows:

W	A	T	C	H	A	T	N	I	G	H	T
3	1	2	1	5	3	1	2	1	5	3	1

He/she will now shift each plaintext alphabet by the number written below it to create cipher text as shown below:

W	A	T	C	H	A	T	N	I	G	H	T
3	1	2	1	5	3	1	2	1	5	3	1
Z	B	V	D	M	D	U	P	J	L	K	U

Each plaintext character here has been shifted by a different amount and that amount is determined by the key. The length if the key must be less than or equal to the size of the message.

The receiver uses the same key and shifts received cipher text in reverse order to obtain the plaintext for decrypting the message.

Z	B	V	D	M	D	U	P	J	L	K	U
3	1	2	1	5	3	1	2	1	5	3	1
W	A	T	C	H	A	T	N	I	G	H	T

Vigenere cipher was regularly used in the past for protecting sensitive, political and military information. It was known as the unbreakable cipher because it was difficult for cryptanalysis. It was created by twisting the standard Caesar cipher to reduce the effectiveness of cryptanalysis on the cipher text and make a cryptosystem more robust. It is significantly known as more secure than a regular Caesar Cipher.

Vigenere is divided into two cases:

i). Vernam Cipher: The keyword length is similar to the length of plain text. This is more secure than the typical Vigenere cipher.

ii). One-Time Pad: The length of the keyword is similar to the length if the plain text. The keyword is derived from a random string of alphabets. The key work can be used only once [11].

III. COMPARISON TABLE

POINTS	CAESAR CIPHER	MONO-APHABETIC CIPHER	PLAY FAIR CIPHER	VIGENERE CIPHER
Creating	Simplest technique to encrypt data.	Simple method	Simple but effort put in creating cipher text	Uses a text string or a word as a key
Attack	Brute force attack	All- Fearsome Statistical Attack	Letter frequency	-
Security	Not much secure as the cipher can be easily broken.	Secure from brute force attack.	Uses one-many relation between plain & cipher texts	Secure from brute force, statistical attack.
Can Be Used For	Insensitive data	Any official but insensitive data	Protecting important but not critical data	Protecting sensitive, political and military information

IV. CONCLUTION

This paper gives an overview of the different classical encryption techniques. The techniques are explained in detail with examples. Further, comparisons Caesar cipher, Monoalphabetic cipher, play fair cipher and Vigenere cipher are given for better understanding and usage of each cipher. This paper is written to understand different the substitution cipher methods in depth and to do a comparative study on the above mentioned substitution techniques.

V. REFERENCES

[1]. Peleg, S., & Rosenfeld, A. (1979). Breaking substitution ciphers using a relaxation algorithm. Communications of the ACM, 22(11), 598–605. doi:10.1145/359168.359174

[2]. Som, S., Kundu, M., & Ghosh, S. (2012). A Simple Algebraic Model based Polyalphabetic Substitution Cipher. International Journal of Computer Applications, 975, 8887.

[3]. Tripathi, R., & Agrawal, S. (2014). Comparative study of symmetric and asymmetric cryptography techniques. International Journal of Advance Foundation and Research in Computer (IJAFRC), 1(6), 68-76.

[4]. Kumar, M., Mishra, R., Pandey, R. K., & Singh, P. (2010). Comparing Classical Encryption With Modern Techniques. proceedings of S-JPSET, 1(1).

[5]. Stallings, W. (2006). Cryptography and Network Security, 4/E. Pearson Education India.

[6]. Ijaz Ali Shoukat , Kamalrulnizam Abu Bakar and Mohsin Iftikhar, “A Survey about the Latest Trends and Research Issues of Cryptographic Elements”, p 141, International Journal of Computer Science Issues, Vol. 8, Issue 3, No. 2, May 2011, ISSN 1694 0814.

[7]. Fontaine. C. and Galand. F. (2007), A Survey of Homomorphic Encryption for Nonspecialists, EURASIP Journal on Information Security Volume 2007, Article ID 13801, 10 pages, doi:10.1155/2007/13801, Hindawi Publishing Corporation.

[8]. Jain, N., Pal, S. K., & Upadhyay, D. K. IMPLEMENTATION AND ANALYSIS OF HOMOMORPHIC ENCRYPTION SCHEMES.

[9]. Sukalyan Som, Saikat Ghosh, “A Survey of Traditional or Character Oriented Symmetric Key Cryptography”, International Journal of Advanced Research in Computer Science, Vol. 2, No. 4, July-August 2011

[10]. Lecture Notes on "Computer and Network Security" by Avi Kak.Pdf h t t p: / / jun i ch ol l . or g/ Cr ypt a na l ysi s/ Da t a / EnglishData.php

[11]. https://www.tutorialspoint.com/cryptography/traditional_ciphers.htm