ROBUST CAESAR CIPHER AGAINST FREQUENCY CRYPTANALYSIS USING BI-DIRECTIONAL SHIFTING.

ABDULKADIR HASSAN DISINA

A thesis submitted in partial

fulfillment of the requirement for the award of the

Degree of Master of Computer Science

Faculty of Computer Science and Information Technology

Universiti Tun Hussein Onn Malaysia

March 2014

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

C	-	Ciphertext
M	-	Message
OSB	-	Odd Position Shift Backward
ESF	-	Even Position Shift Forward
ESW	-	Encryption Switch

LIST OF APPENDICES

ABSTRACT

Cryptography is the art of encoding a message such that only the sender and the intended recipient can decode. But with advent of Internet, information confidentiality faces more treats, which lead to the development of cryptographic algorithms. This thesis presented a cryptographic algorithm titled Robust Caesar cipher against frequency analysis using bidirectional shift. It encrypts message bit by bit (stream cipher) and uses one key ideology (symmetric key cipher). The sender encrypts the message before transmitting and the receiver decrypts upon receiving using the same key as used in the encryption process. It works by shifting the plaintext characters to different direction which eliminates repetition of characters in the ciphertext. Nevertheless, previous versions of Caesar cipher are prone to frequency analysis attack. In this research, Caesar cipher is enhanced to 95 characters in a tabular arrangement as digital messages are not only 26 characters. Based on this method, the sender will transpose the bits in the message according to their sequence arrangement (odd and even position) by shifting the characters in the odd position to the left and characters in the even position to the right side. The cryptographic key given by the sender will determine the shifting position of all the characters. Shifting the plaintext to different directions mitigates the problem of repetition. The experiment shows that the proposed method has an efficiency of 99.9% resistance than the earlier version of the Caesar cipher. However, the proposed method can save as an option to be integrated with other algorithms to strengthen the security.

ABSTRAK

Kriptografi adalah seni pengkodan mesej yang hanya membolehkan penghantar dan penerima membaca mesej tersebut. Tetapi dengan kemudahan Internet, kerahsiaan maklumat menghadapi pelbagai serangan dan hal ini membawakan kepada pembangunan algoritma kriptografi. Tesis ini dikemukakan berkaitan algoritma kriptografi yang bertajuk Robust Caesar cipher against frequency analysis using bidirectional shift. Ia menyulitkan mesej sikit demi sedikit (aliran cipher) dengan menggunakan satu ideologi utama (simetri cipher utama). Penghantar menyulitkan mesej sebelum menghantar kepada penerima dan penerima membaca mesej tersebut dengan menggunakan kekunci yang sama semasa proses penyulitan. Caesar cipher ini berfungsi dengan mengalihkan huruf plaintext ke arah yang berbeza untuk menghapuskan pengulangan huruf dalam mesej. Walau bagaimanapun, versi awal Caesar cipher telah terdedah kepada serangan frekuensi analisis. Dalam kajian ini, Caesar cipher mempertingkatkan huruf dalam susunan jadual mesej digital dari 26 ke 95 huruf. Berdasarkan kaedah ini, penghantar akan mengubah nada bit dalam bentuk mesej mengikut susunan urutan yang ditentukan (kedudukan ganjil dan genap) dengan mengalihkan huruf-huruf ganjil ke kiri dan huruf-huruf genap ke sebelah kanan. Kekunci kriptografi yang diberikan oleh penghantar akan menentukan kedudukan semua huruf. Peralihan plaintext ke arah yang berbeza menyelesaikan masalah pengulangan huruf. Eksperimen menunjukkan bahawa kaedah yang dicadangkan mempunyai kecekapan rintangan 99.9 % daripada versi awal cipher Caesar. Walau bagaimanapun, kaedah yang dicadangkan adalah bagi membolehkan dan disimpan sebagai pilihan untuk disepadukan dengan algoritma lain supaya dapat mengukuhkan lagi keselamatan mesej yang dihantar secara sulit.

CHAPTER 1

INTRODUCTION

1.1 Overview

Networking has become a very important aspect of every one's life if the rate of Internet usage is considered. Information is usually being sent and sometimes shares data. Some of those classified information contains account numbers, meeting venue, addresses, and other crucial information that need confidentiality. Likewise the information may contain some banks and academic related documents that require some privacy. In most cases, senders of those messages are not really concern about the security of the communication channel. Malicious individuals could tap into the unsecured channel of the communication illegitimately, to make illegal use of the resources (compromises confidentiality) or to temper (compromises availability and integrity) with the data. To secure the reliability of the communication channel, restriction methods can be applied on the channels so that only the original sender and intended receiver can unlock and decode the message. However, user might like to make sure that the message at the destination is the same as the original from the source, which means to ensure the integrity of the message. In this case, some sort of pad lock has to be applied to lock the data. This could not only be done on a physical surface, can also be applied on logical or digital data as well, but only with the help of some mathematical algorithms called

cryptography. Cryptography is the art and science of using mathematics and logics to prepare a coded or protected communication that can only be understood by the sender and intended recipient. The sender encrypts the message using a key such that the receiver needs to posses the key in order to decrypt the message into its original readable form. There are two criteria to logically use the cryptographic key, symmetric (secret key) and Asymmetric (two keys, public and private) key cryptography [1].

Cryptography is all about finding an unpredictable way to manipulate information into unreadable form, it could be simple or difficult old or modern, and it is believed that "Old is gold" which gives room for the manipulation of the old techniques in order to come up with more powerful ones [2]. Caesar cipher is one of the earliest symmetric ciphers which suffer from weakness and unreliability.

In this research, a proposed modified version of Caesar cipher is introduced which is expected to achieve exclusion of character repetition in the message when it is encrypted. It also randomizes the characters in the process of encryption to make it very difficult for the cryptanalyst to decipher it using frequency analysis. In modern world, hackers try to break a cryptographic algorithm or try to retrieve the key which is needed to encrypt a message, by analyzing the insertion or presence of repetitive bits/characters (bytes) in the message and encrypted message to find out the encryption algorithm or the encryption key. Hence, there is a need to develop a strong encryption method that will exclude the repetitive characters such that there will be no trace of character repetition. It is done to compete with other ciphers of its kind by bringing simple and reliable security to the messages that does not require advanced or extremely powerful encryption techniques. It uses more characters than the 26 English alphabets because the nowadays messages are far more than that. All the characters in the ASCII printable code table will be included so as to have all the numbers (0,1,2….9) and all the special characters. This will enable the generation of more permutations (all possible combinations) as all could be found in a single message. This research proposes an enhanced Caesar cipher that would resist frequency analysis by providing a randomized ciphertext.

1.2 Problem Statement

There are some problems attach to the Caesar cipher which leads to the proposal of this encryption algorithm. Its weakness is proven by applying frequency analysis of the English alphabets, which describe that the cipher could be decrypted easily once captured by the attacker because English alphabets are only 26 characters, and two or more alphabets of the same type in the plaintext, have same representation in the cipher text. The following problems are out lined in sequence to highlight some of the weakness that needs to be addressed in this research work. Absence of an encryption algorithm compromises the confidentiality due to unauthorized access, in the same process integrity could also be bridged [2].

i. The earlier version of Caesar cipher is very weak, and thus makes it easier to decrypt as the English letters are limited to 26 characters which means the attacker has only 26 choices to determine the letter. The frequency analysis table makes it easier, attacker could just guess the next alphabet with the help of that frequency analysis table [3].

ii. Repetition or redundancy of characters further increases the weakness of the algorithm. If an encrypted sentence that has two or more same characters or alphabets, repetition is easily identifiable as same alphabets would have same representation after the encryption [4].

1.3 Motivation

There are many supportive ideas and issues that motivate this research which is why this idea is brought and proposed, it is believed that different cryptographic methods are used by different organizations for different reason. For example; methods of data protection used in government institutions differs from methods used in banks. In this case we should use something that the attacker would never expect and make it more difficult as it seems [3] [5].

Day by day number of hackers is being increased. Existing old methods of security often fails to overcome it. Hence, as an ancient quote says it's true that "Old is Gold" those methods helps in developing new ideas to sort out problems, thus leads to discovery of a new horizon [3]. Most of the powerful encryption techniques are being discovered from the old ones. Little modification or combination of two or more algorithms makes a lot of progress which leads to achievement and satisfaction.

Hackers nowadays are always trying to break the cryptographic methods or retrieve keys by different means and one of such methods include the process of inclusion of repetitive texts or characters in a message and then encrypt it to study the behavior of the method and retrieve the key which is needed for decryption [6].

The main reason why Caesar cipher is considered very weak is because it has only 26 characters and the permutation of 26 characters is very small to try all the possibilities in a very short time [7]. It consists of the alphabet written out 26 times in different rows each alphabet shifted cyclically to the left compared to the previous alphabet, corresponding to the 26 possible Caesar ciphers. At different points in the encryption process, the cipher uses a different alphabet from one of the rows. The alphabet used at each point depends on a repeating keyword [3]. What if the characters are being increased to even double of the existing ones? If 26! is not long enough to be secured, then a cipher that could be without any repetition or redundancy in representing same characters and has 95 characters would be secured more than the previous versions [7].

1.4 Objectives

The objectives of this research are to ensure that the proposed version of Caesar cipher is strengthen to the extent that it could be stronger than all the previous versions. It surfers from problems ranging from redundancy of the characters, frequency analysis and crypto analysis. This research work is to overcome the above mentioned criticisms, the followings are the objectives in items.

1. To propose a modified version of Caesar cipher based on all the characters in the ASCII table not only the 26 English alphabets. Improve the Caesar cipher by twisting and randomizing the cipher which will increase the difficulty level of attack thus increase the reliability and security.
2. To validate the proposed algorithm on an encryption system using frequency analysis test.
3. To compare the proposed algorithm with other existing versions of Caesar cipher.

1.5 Scope

This research concentrated on the stream cipher particularly Caesar cipher which will be modified and add more characters as in the ASCII table to increase the permutation of the characters.

1.6 Contribution

This research focuses on the contribution to the message confidentiality by solving the following four issues. It is expected to come up with a Robust Caesar cipher that should have more characters than the previous versions, which will provide more permutation (possible combination) of the characters in the message. The algorithm is strengthened by randomizing (switch position) the ciphertext in the encryption process. This research will ensure that the proposed version of Caesar cipher will eliminate the redundancy of the characters after the encryption. It will also provide an option for the researchers to choose between the available versions of Caesar cipher for their encryption algorithms or for implementation. The followings are the outline of the contributions.

i. Increase the security by modifying a cryptographic technique to more reliable and secured.
ii. Give options to the users of cryptographic algorithms or stream ciphers.
iii. Give researchers another option to explore more techniques and could use this to integrate with other techniques for more security.

1.7 Summary

In substitution and transposition ciphers, there is repetition or redundancy problem that usually happens when plaintext is encrypted to ciphertext. It can be broken using frequency analysis which makes the two techniques prone or vulnerable to attack(s). Therefore a new method called Robust Caesar cipher against frequency Analysis is proposed based on the problems mentioned to overcome the situation. The proposed cipher is expected to eliminate any sort of repetition in the ciphertext. The result of this research will serve as an alternative technique for encrypting and decrypting secret messages.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

In this part of the research, all the related items are described in details to prove and disclosed how the previous development are achieved and how to proceed and develop an advanced system based on the facts and related options.

2.2 Overview of the Cipher

Thousands of years ago, the mail was based on paper and pen and it was the only means of non physical contact discussion. Messages those days are being sent through the receiver by human protocol, which means slaves are sent with a written message to the intended recipient. The slaves are not always reliable as an intruder could intercept and read or destroy the message before reaching its destination; therefore the message needs to be locked [3].

Encryption started with simple pen-and-paper methods based on letter substitutions. Then it further evolved into special machines built to encrypt messages. Today we have moved away from the more physical methods, and the focus is on digital encryption that can only be done using computers [2].

Encryption techniques can be divided into two classes: traditional encryption techniques and modern encryption techniques. Traditional encryption techniques are

pen-and-paper based techniques developed when computers did not exist, although some of these ideas can be, and have been, transformed into computer-based algorithms.

With the beginning of the Computer Era, which can be marked with the appearance of the first computer encryption techniques underwent a major change. Encryption techniques were being specifically designed for computer usage and used 'bits' instead of alphabets. These encryption techniques are called modern encryption techniques [7].

However, the idea of locking of messages such that any individual with the key can access the message is called cryptography. The word cryptography came from Greek word "kryptos" and "graphein" which means hidden and writing respectively, this means converting a written message (plain text) into unreadable form (cipher text) to prevent its confidentiality [7].

With the help of secured cryptographic algorithms, two people can communicate with each other securely. Eavesdropper could find it difficult to intercept or eavesdrop the message [8]. It further moves to ensure message integrity, authentication and digital signature. Nowadays recipient of a message can check if the message sent was not modified by the interceptor or eavesdropper and the message he received was the actual message from the original source [5][9].

Caesar cipher is considered as the most famous traditional encryption method developed by Julius Caesar between 50 and 60 BC [10] [11]. The Caesar Cipher worked on the principle of substitution, where each letter in the alphabet is substituted for another letter. In this case each letter was transposed with another three places after the original letter in the alphabet.

2.3 Types Cryptographic Algorithm

In cryptography, the message to be transmitted is usually locked (encrypted into cipher text) that both the sender and receiver needs the key by which only the message could be retrieved back (decrypted) to its original form (plain text) as at the sender side. There are two types of keys symmetric and asymmetric key encryption [8].

2.3.1 Symmetric Key Cryptography

In symmetric, only one key is used for both encryption at the sender station and decryption at the receiver station, the symmetric key is also called secret key. In the other hand asymmetric key is the opposite of symmetric in which two keys are used for different purposes (encryption and decryption), the sender encrypts the message or plaintext using the recipients' public key and at the receiver side the recipient uses his private key to decrypt the message. Moreover, asymmetric key encryption is more secured than the symmetric [2] [5].

2.3.2 Asymmetric Key Cryptography

Asymmetric cryptography is a cryptographic algorithm which requires two separate keys, one of which is secret (or private) and the other is public. Although, the two parts of this key pair are mathematically linked [8]. The public key is used to encrypt plaintext whereas the private key is used to decrypt ciphertext. In 1976, the first idea of public key cipher or encryption was introduced. It was published on a paper titled "New Direction in Cryptography" until the idea of public key cipher came, they are called symmetric or private key cipher [7].

The world first public key cipher was published in 1978 in a paper titled “A Method for Obtaining Digital Signature and Public-key Cryptosystem”, because they accurately predicted the arrival of electronic mail soon. As it was said earlier in this report, in 1978 the only means to send mail was paper mail with a signature of the sender on it. The researchers created the cipher that preserves the two important characteristics of the paper mail system [2] [3].

2.3.3 Chosen Key Type

This research focuses on private key cryptography which is widely known as symmetric key cryptography, to be used for both encryption and decryption process as already mentioned above.

2.4 Types of Cipher

Encryption of the plaintext or message into cipher text is generally categorized into two categories; stream and block cipher. This simply means that the form by which the message is transformed into, stream ciphers are usually encrypted character-by-character or in other words bit-by-bit unlike block cipher which is divided into fixed length group of bits before encryption [12].

2.4.1 Stream Cipher

Stream cipher is the encryption of the plain text bit by bit or character by character one at a time. Stream ciphers represent a different approach to symmetric encryption from block ciphers. Caesar cipher which is one of the stream ciphers was transformed into a "modern" cipher with a key by agreeing upon a key that each letter is substituted by another and that the key should be the permutation of all the alphabets or possible combinations in the characters [12].

Moreover, rail fence cipher was introduced also known as zigzag cipher, the name was driven from the way it was encoded which is a zigzag-like form. It is also a form of transposition cipher which also suffers from small key problem, the number of practical keys is very small such that the crypto analyst can try all possibilities at a given time. This form of cipher is being written down diagonally [6]. There are several ways to examine the letter frequency in the mono alphabetic stream ciphers but this research covers only two out of the three, which are the homo alphabetic and poly alphabetic cipher [12].

2.4.1.1 Mono or Homo Alphabetic

In homo alphabetic cipher, common letters are assigned to many different characters in the cipher text and rare letters like "z" which gets only few. When letters are encoded, associate cipher symbols are chosen at random and this is prone to attack as plaintext characters are usually mapped to specific ciphertext characters [6].

2.4.1.2 Poly Alphabetic

This cipher also, uses simple substitution. Each letter is shifted using either normal cyclic shifts or involutions. The substitution is changed for every letter in a previous agreed way. Changing the agreed way in the subsequent letters mask the redundancy of the plain test which means, many ciphertext characters can be mapped to one plaintext character. Hence, this method makes it difficult to be attacked by frequency analysis [6].

2.4.2 Block Cipher

A block cipher is a deterministic algorithm operating on fixed-length groups of bits, called blocks, with an unvarying transformation. Block ciphers are widely used to implement encryption of bulk data with a fixed unvarying transformation. Most block cipher algorithms are classified as iterated block ciphers, which means that they transform fixed-size blocks of plaintext into identical size blocks of ciphertext via the repeated application of an invertible transformation known as the round function, with each iteration referred to as a round [12]. This research focuses on only the stream cipher.

2.4.3 Chosen Cipher Type

One key cryptographic scheme is usually associated with stream cipher more often and the poly alphabetic cipher is the mode of operation that is chosen for this research work. Thus, to eliminates character repetition in the ciphertext.

2.5 Permutation

In mathematics, a permutation of a set of objects is an arrangement of those objects into a particular order. Since there are 26 alphabets or characters in the Caesar cipher algorithm, then the key space should be 26! Therefore, with the help of some statistical methods, the mono alphabetic cipher can be broken using frequency analysis. If the technique permits the cipher alphabet to be any rearrangement of the plaintext alphabet, then an enormous number of distinct modes of encryption can be generated. There are over Four (4) million such rearrangements, which gives rise to an equivalent number of distinct cipher alphabets. Each cipher alphabet is known as a key, if the message is intercepted by enemy who correctly assumes that a mono alphabetic substitution cipher have been used, they are still faced with the impossible challenge of checking all possible keys. If an enemy agent could check one of these possible keys every second, it would take roughly one billion times the lifetime of the universe to check all of them and find the correct one [6].

2.6 Related Work

Recently in the year 2013, another version of modified shift cipher is released titled "DEDD" means Double Encryption and Double Decryption [3]. In this cryptosystem, Alice encrypts the message twice with the public key and Bob decrypts that encrypted message twice, the procedure is as follows.

Consider Bob and Alice as a sender and receiver respectively, Bob generates a key and assigns it to Alice. Alice enciphers the message by applying "shift Cipher" and encrypts it by its length and gets cipher 1 (for 1st time). Second encryption will be done by applying shifting technique on the cipher 1 which finally encrypts the message into ciphertext. The cipher will be sent to Bob. Then Bob will decrypt this encoded message

by applying shifting technique on the method [3]. Improved shift cipher was driven from the previous cipher as it becomes stronger than all the previous versions.

Similarly, another cipher was created to improve and overcome the earlier version of the Caesar cipher by excluding repetitive terms in a message when it is said to be encrypted which will strengthen it to be more difficult or almost impossible for a crypto analyst to predict the original message (plain text) from the encrypted.

Nowadays, hackers try to break a code or cryptographic algorithm or try to retrieve the key, which is the main purpose of any encryption algorithm. Therefore, they applied SD-AREE cryptographic method to exclude repetitive terms from a message that is to be encrypted. In SD-Aree algorithm the repetitive bits or characters are minimized and there is little or no trace of any repetition in the message. For example, if a message has two characters of the same type, they will totally become different characters in the cipher text. This cipher is called "A New Modified Caesar Cipher Cryptographic Method Along with Bit-Manipulation to Exclude Repetition from a Message to be encrypted". The ASCII value of each character will be extracted from the text, which is produced after bit level encryption, and then the code is added with the ASCII value of each character [10]. Figure 2.1 shows the graphical representation of the encryption process in Modified Caesar Cipher (SD-Aree) algorithm.

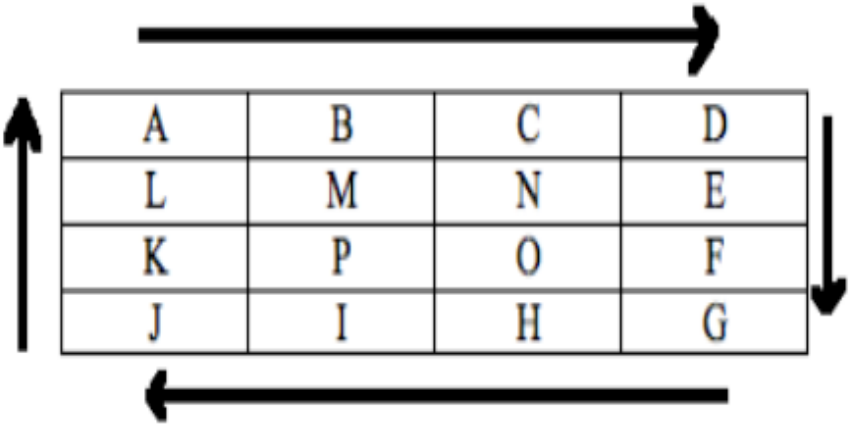

Figure 2.1. Modified Caesar Cipher [10]

2.7 Components of Symmetric Cipher

Symmetric cipher or secret key cryptography has components ranging from the original message, the algorithm, secret key and cipher text. Plain text is always referred to as the original message before the encryption process to produce the cipher text, which is referred to as unreadable form of the message. All the processes occur with the help of a key without which the encryption could not be possible. It is used to encrypt the message at the sender's end and to decrypt it at the receiver's end to get the original message or plain text [8]. Figure 2.2 shows how the components are connected to each other.

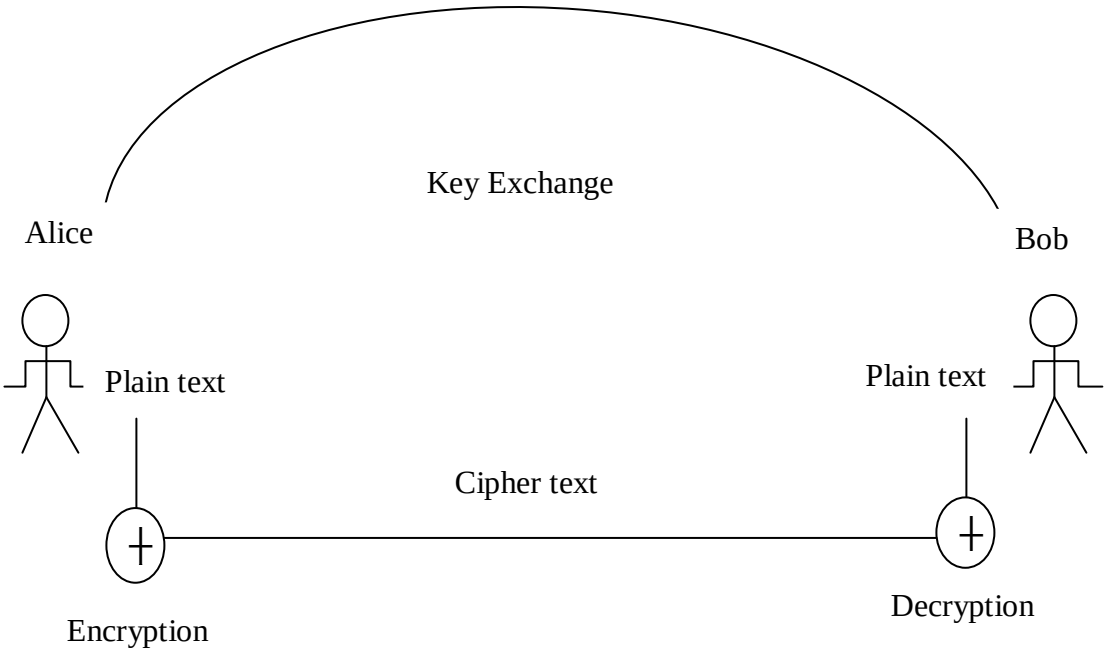

Figure 2.2 Components of Stream Cipher [8]

2.7.1 Plaintext (Message)

In Figure 2.2, plaintext is the original message in its readable form before encryption into cipher or after decryption from cipher [7]. This is any data that need to be encrypted (coded, so that only sender and receiver understand) before transmission or for the purpose of storing it for future use. In stream cipher, message is usually encrypted character by character or bit by bit starting from the first to the last alphabet of the message [8]. A good cryptographic scheme has to produce the same message as the original after decryption, any changes to the original message could render the message and the algorithm invalid [9].

2.7.2 Encryption

This is the process of transforming and converting the message by the sender into unreadable form (cipher text) using an encryption algorithm and a key [8]. To increase the security, an encryption algorithm could be a hybrid of two or more algorithms [9]. It is also defined as the process of encoding messages (or information) in such a way that eavesdroppers or hackers cannot read it [7].

2.7.3 Key

In symmetric cipher where both sender and receiver share the same key for encryption and decryption respectively, the key determines the reliability of the algorithm. It plays an important role to make sure that the message is scramble enough such that the secrecy of the message will be maintained. Symmetric key algorithms are a class of algorithms that uses the one cryptographic key (private key) for both encryption of plain text and

decryption of cipher text [8]. The key in practice represent a shared secret between two or more parties that can be used to maintain a private information link. This requires that both parties should have access to the secret key which is one of the main drawbacks of symmetric key encryption, in comparison to public-key encryption [7].

2.7.4 Decryption

Decryption is the process of reversing the cipher text into the original message (plain text) as it was at the sender station before applying it to the encryption algorithm [8]. Any adversary that can capture the cipher text should not be able to determine the content of the original message. An authorized party however, is able to decode the cipher text using a decryption algorithm that usually requires a secret decryption key that adversaries do not have access to. Any algorithm that enciphers a message and cannot decrypt the message into its original form when an appropriate key is used is considered unaccepted [7].

2.8 Types of Operations for Stream Cipher

Stream cipher like any other, has some modes by which it operates on, algorithms manipulate either the arrangement or representation of the plaintext characters to encrypt messages to ciphertext. The modes of operations are not limited to substitution, transposition, and mixed cipher [10].

2.8.1 Substitution

A substitution cipher is a method of encoding by which units of plaintext are replaced with ciphertext. The units may be single letters (the most common), pairs of letters, triplets of letters, mixture of the above and so forth. The receiver deciphers the text by performing an inverse substitution [10]. Moreover in this cipher, the units of the plaintext are retained in the same sequence in the ciphertext, but the units themselves are altered [11].

2.8.2 Transposition

A transposition cipher is a method of encryption by which the positions held by units of plaintext (which are commonly characters or groups of characters) are shifted according to a regular system, so that the ciphertext constitutes a permutation of the plaintext. That is the order of the units is changed. Mathematically a bijective function is used on the character positions to encrypt and an inverse function to decrypt [12].

2.8.3 Mixed Cipher

In this type, both methods are combined together to strengthen one another. The cipher has both substitution and shift operation integrated together. For example, a simple substitution cipher combined with a columnar transposition avoids the weakness of both. Replacing high frequency ciphertext symbols with high frequency plaintext letters does not reveal chunks of plaintext because of the transposition [6] [12]. The combined operation is the chosen mode of operation for this research.

2.9 ASCII Characters

The above acronym means American Standard Code for Information Interchange (ASCII) which includes all the printable and non printable codes and characters. In this research only the printable characters will be included as the messages of nowadays includes numbers, smalls and capital letters, multiplication signs, and other important characters [7]. When this is added to the Caesar cipher, there would be more possible permutations as all the characters could represent at least one other character [13]. Figure 2.3 shows the ASCII table which contains the printable characters to be used in this research.

Dec	Hx	Oct	Html	Chr	Dec	Hx	Oct	Html	Chr	Dec	Hx	Oct	Html	Chr
32	20	040		Space	64	40	100	@	@	96	60	140	`	`
33	21	041	!	!	65	41	101	A	A	97	61	141	a	a
34	22	042	"	"	66	42	102	B	B	98	62	142	b	b
35	23	043	#	#	67	43	103	C	C	99	63	143	c	c
36	24	044	$	$	68	44	104	D	D	100	64	144	d	d
37	25	045	%	%	69	45	105	E	E	101	65	145	e	e
38	26	046	&	&	70	46	106	F	F	102	66	146	f	f
39	27	047	'	'	71	47	107	G	G	103	67	147	g	g
40	28	050	(	(	72	48	110	H	H	104	68	150	h	h
41	29	051	)	)	73	49	111	I	I	105	69	151	i	i
42	2A	052	*	*	74	4A	112	J	J	106	6A	152	j	j
43	2B	053	+	+	75	4B	113	K	K	107	6B	153	k	k
44	2C	054	,	,	76	4C	114	L	L	108	6C	154	l	l
45	2D	055	-	-	77	4D	115	M	M	109	6D	155	m	m
46	2E	056	.	.	78	4E	116	N	N	110	6E	156	n	n
47	2F	057	/	/	79	4F	117	O	O	111	6F	157	o	o
48	30	060	0	0	80	50	120	P	P	112	70	160	p	p
49	31	061	1	1	81	51	121	Q	Q	113	71	161	q	q
50	32	062	2	2	82	52	122	R	R	114	72	162	r	r
51	33	063	3	3	83	53	123	S	S	115	73	163	s	s
52	34	064	4	4	84	54	124	T	T	116	74	164	t	t
53	35	065	5	5	85	55	125	U	U	117	75	165	u	u
54	36	066	6	6	86	56	126	V	V	118	76	166	v	v
55	37	067	7	7	87	57	127	W	W	119	77	167	w	w
56	38	070	8	8	88	58	130	X	X	120	78	170	x	x
57	39	071	9	9	89	59	131	Y	Y	121	79	171	y	y
58	3A	072	:	:	90	5A	132	Z	Z	122	7A	172	z	z
59	3B	073	;	;	91	5B	133	[	[	123	7B	173	{	{
60	3C	074	<	<	92	5C	134	\	\	124	7C	174			|
61	3D	075	=	=	93	5D	135	]	]	125	7D	175	}	}
62	3E	076	>	>	94	5E	136	^	^	126	7E	176	~	~
63	3F	077	?	?	95	5F	137	_	_	127	7F	177	&#127;	DEL

Figure 2.3 Printable ASCII Characters [13]

2.10 Attacks on Symmetric Cipher

Symmetric cipher is prone to cryptanalysis and other sort of attacks as other ciphers does. Historically, it has been susceptible to known-plaintext attack, chosen plaintext attack, differential cryptanalysis and linear cryptanalysis. Careful construction of the functions for each round can greatly reduce the chances of a successful attack [14].

2.10.1 Known Plaintext Attack

The known plaintext attack is a cryptanalysis or a model of attack where the attacker has already known or has sample of both the original message (plaintext) called a “crib”, and its corresponding cipher text (encrypted message). It is used to unveil or reveal further secret messages or information such as secret keys and code books [14] [15].

2.10.2 Chosen-Plaintext Attack

In this form of attack, the cryptanalyst or hacker chooses an arbitrary part of the original message (plaintext) and then encrypts the message with his/her key and further compares with the captured ciphertext. If the result appears to be the same, then the remaining ciphertext will be known. In the worst case, a chosen-plaintext attack could reveal the scheme's secret key. For some chosen-plaintext attacks, only a small part of the plaintext needs to be chosen by the attacker. Such attacks are known as plaintext injection attack [14].

2.10.3 Differential Cryptanalysis

In this method, the cryptanalyst observes the behavior of the cipher when some changes in the input (plaintext) are made and how the changes affects the output (cipher text). It is also like chosen plaintext attack where the attacker will try many possible keys to understand the differences [15].

2.10.4 Frequency Analysis

This technique is the study of the frequency of the letters or characters in the cipher text. Usually in any language, the frequency of the letters differs. For example if the English alphabets are considered, some letters appears many times in one sentence while others come occasionally. In English language “E”, “T”, “A” and “O” are the most common letters in messages, while “Z”, “Q” and “X” are rare. Likewise, “TH”, “ER”, “ON”, and “AN” are the most common pairs of letters (termed bigrams or digraphs), and “SS”, “EE”, “TT”, and “FF” are the most common repeats. The “nonsense phrase” "ETAOIN SHRDLU" represents the 12 most frequent letters in typical English language text. Cryptanalyst usually uses this technique to uncover the encrypted message in stream ciphers [15]. If the results shows that “E” followed by “T” are the most common letters then the ciphertext may be a transposition cipher rather than a substitution. If one of the characters has a 20% frequency, then the language may be considered as German since it has high percentage of E. Italian has 3 letters with a frequency greater than 10% and 9 characters are less than 1%.

Example

If “Happy” is considered as the message and “KDSSB” is the cipher obtained after encryption with mod 3 key. “K”, “D”, “B” has 20% frequency respectively and S has 40% which simply means it appears twice in the message. Figure 2.4 shows the percentage of frequency of English letters.

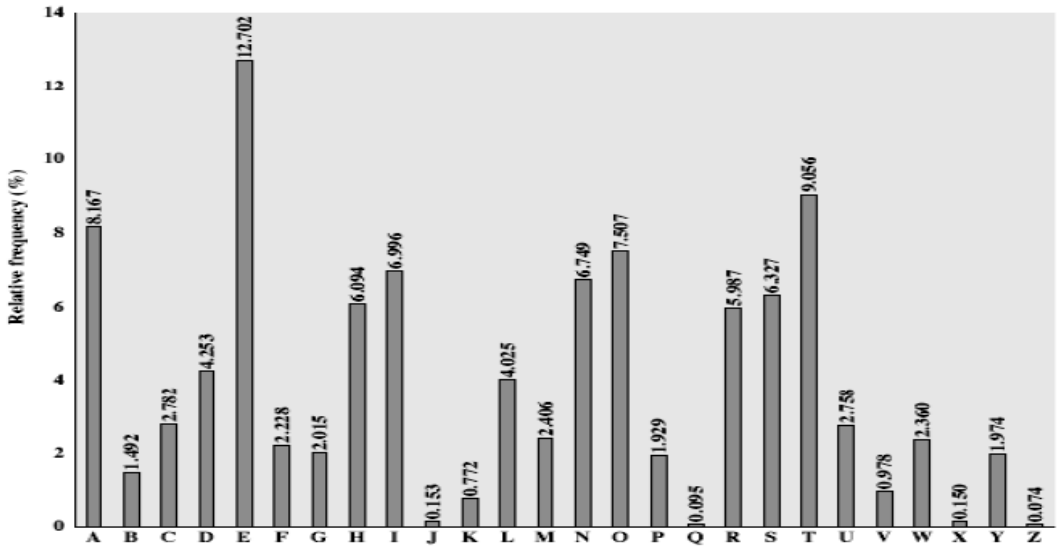

Figure 2.4 Frequency Analysis Table [15]

2.10.5 Cryptanalysis

Cryptanalysis, on the other hand is the art of “breaking” or “cracking” these encryption methods; i.e. it is the process of deducing the meaning of specially encoded messages without actually being the legitimate sender or receiver [14]. The battle of code makers versus code breakers has been going on for quite some time. More than once a new “unbreakable” cipher has been developed by code makers only to be “broken” by some code breaker [14] [16].

2.11 Stages of Cryptanalysis

In the cryptanalysis of the Caesar cipher, there exist three stages, these are:

i. Identification,
ii. Breaking and
iii. Settings.

The following section will describe each of the stages in details.

2.11.1 Identification

The first problem when a cryptanalyst or hacker captured a ciphertext is to identify or discover what sort of algorithm or cipher system was used to encrypt the message. It may have been already known cipher or new, either way the problem of identification or discovery still exist. The cryptanalyst should take into account the available collateral information such as the type of system by which the sender sends the cipher through, if the previously used system is known to the cryptanalyst or any new system which has recently appeared, cryptanalyst should then examine the message thoroughly. Short messages are difficult to decode thus the cryptanalyst should wait for more messages, if the message is long enough then the cryptanalyst should apply some mathematical test to identify which algorithm was used and the difficulty level [15].

REFERENCES

1 Kohnfelder, L. M. *Towards a practical public-key cryptosystem*. P.h.D. thesis. Massachusetts Institute of Technology; 1998.

2 Fukuda, K. & Akio T. Key-sensitivity improvement of block cipher systems based on nonlinear feedback shift registers. *Asia Pacific Conference on Circuits and Systems (APCCAS)*. IEEE, 2012.

3 Praloy, S. D. & Prasenjit, M. DEDD Symmetric-Key Cryptosystem, *International Journal of Advanced Computer Research.* 2013. Vol.3. No.1.

4 Dey, S. Nath, J. & Nath, A. An Integrated Symmetric Key Cryptographic Method-Amalgamation of TTJSA Algorithm, Advanced Caesar Cipher Algorithm, Bit Rotation and Reversal Method: SJA Algorithm. *International Journal of Modern Education & Computer Science*, 2012. vol. 4, No.5.

5 Haque, M. Secure text message transmission in a 4G compatible MIMO MCCDMA system with combined implementation of Vigenere Cipher and RSA cryptographic algorithm. *International Journal of Information Technology Convergence & Services*. 2012. Vol.2, No.5.

6 Kumar A.V. & Kumar, P. PA Substitution Cipher. *Anuj Kumar International Journal of Engineering Research & Technology (IJERT).* 2012. Vol.1.

7 Sharma, A. Bhatnagar, A. Tak, N. Sharma, A. Avasthi, J. & Sharma, P. An Approach Of Substitution Method Based On Ascii Codes In Encryption Technique, *IJASCSE*. 2012. Vol.1 No.1.

8 Babu, R. Abraham, G. & Borasia K. A Review On Securing Distributed Systems Using Symmetric Key Cryptography. *International Journal of Advances in Science and Technology*. 2012. Vol. 4, No.4.

9 Goyal, D. & Srivastava, V. RDA Algortihm: Symmetric Key Algorithm. *International Journal of Information and Communication Technology Research*. 2012. Vol. 2, No. 4.

10 Dey, S. (2012). SD-AREE: A New Modified Caesar Cipher Cryptographic Method Along with Bit-Manipulation to Exclude Repetition from a Message to be Encrypted. *arXiv preprint arXiv:1205.4279*.

11 Suriram, R. & Marimuthu, K. Designing an algorithm with high Avalanche Effect. *International Journal of Computer Science and Network Security*. 2011. Vol.11 No.1.

12 Klein, A. (2013). *Stream Cipher*. Retrieved August 20, 2013, from http://www.amazon.com/gp/search?index=books&linkCode=qs&keywords=1447150791.

13 ASCIItable (2013). *AsciiTable*. retrieved August 21, 2013, from http://www.asciitable.com/.

14 Courtois, N. Nohl, K. & O'Neil, S. (2008). Algebraic Attacks on the Crypto-1 Stream Cipher in MiFare Classic and Oyster Cards. *IACR Cryptology ePrint Archive*, *2008*, pp. 166.

15 John, K. Bruce, S. David, W. & Chris, Hall. (2005) Side Channel Cryptanalysis of Product Ciphers. retrieved August 28, 2013, from http://www.springer.com/?SGWID=5-102-0-0-0.

16 Sharif, S. O. & Mansoor, S. P. Performance analysis of stream and block cipher algorithms. *3rd International Conference on Advanced Computer Theory and Engineering (ICACTE)*. IEEE. 2010. Vol. 1, pp. V1-522.

17 Stallings, W. *Cryptography and Network Security*. 4th Ed, Pearson Education, India. 2006. pp. 162.