Global Proceedings Repository
American Research Foundation
ISSN 2476-017X
Available online at http://proceedings.sriweb.org
شبكة المؤتمرات العربية
http://arab.kmshare.net/
The 1st International Conference on Sciences and Arts (ICMSA 2017)
المؤتمر الدولي الاول للعلوم والاداب
مايو 2017 - اربيل – العراق3
http://sriweb.org/erbil/

Cryptography Using Modified Vernam - Homophonic Method Implemented by Matlab

S.S.Kadhim
College of Engineering
Al-Nahrain University.

Abstract. Cryptography is an efficient way to protect electronic information from intruders. The demand for more secure and unbreakable techniques has been increased specially after the giant leap in the information technology during the end of the twentieth century.
This paper has overviewed the types of the classical methods used in cryptography which all of them were brooked in different times during the last century. Most of the recent encryption techniques are derived or based on these traditional methods. Therefore, this paper suggested a new algorithm called modified Vernam - Homophonic method, which is a combination between Vernam, homophonic transposition techniques to overcome the weaknesses in individual one. The suggested method created a stronger algorithm to resist interception by an authorized people. This method is implemented using Matlab program because it has more features over other programming applications. Matlab has many built-in functions that makes programming and implementation of the new algorithm seems to eat a piece of cake.

General Terms
Security, Encryption, Ciphers, Cryptography.

Global Proceedings Repository
American Research Foundation
ISSN 2476-017X
Available online at http://proceedings.sriweb.org
شبكة المؤتمرات العربية
http://arab.kmshare.net/

Key Words
Caesar Cipher, Substitution Cipher, Transposition Cipher, Affine Cipher, Vernam Cipher, Homophonic Cipher, Encryption, Decryption, Cryptography, Shift Cipher, Plain Text, Cipher Text, Cryptanalysis, ASCII Code .

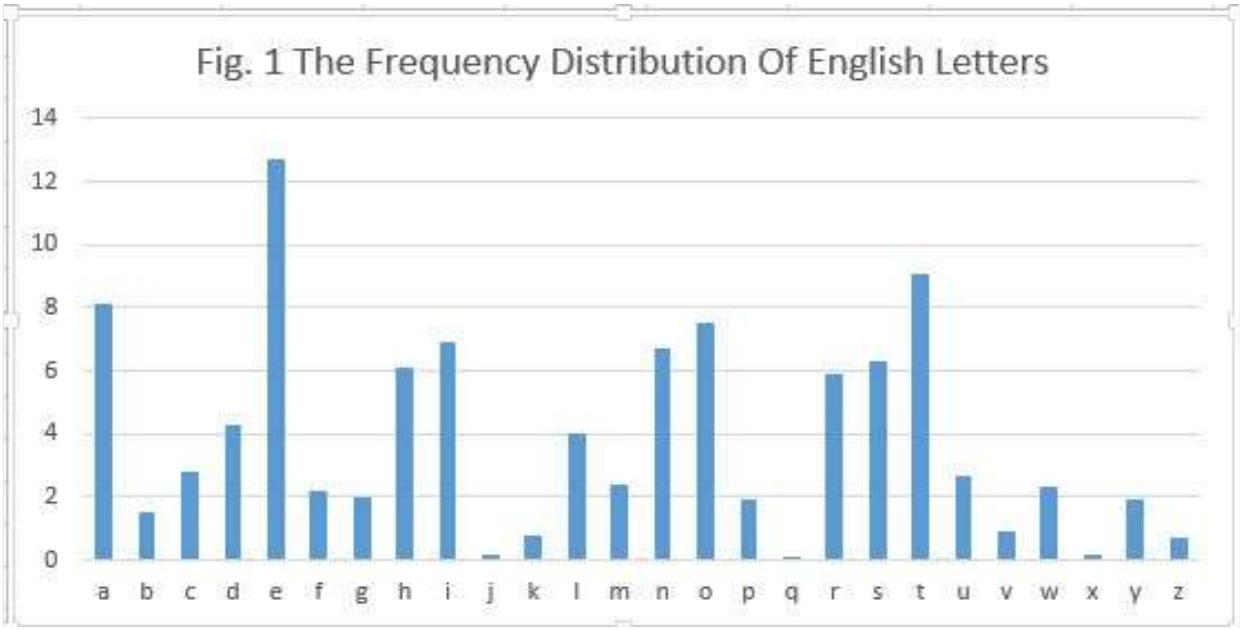
Introduction

In these days the internet has become an essential demand in our life and made the world as small village. It vanished the borders, long distances and natural terrain between countries. Email messages, cash and commercial transactions are being implemented via internet. Therefore, huge amount of data is transferred over the internet for personal or professional purpose. The channels of communication used by computers are threatened by interception via unauthorized people. Thus. Providing security to such data has become substantial. Cryptography, includes techniques that provide virtual security for data or text that is stored or transmitted across insecure networks (like internet). Virtual security is much better than physical one because it is more efficient with lower cost. In other word, cryptography uses process called encryption to scramble plaintext (ordinary text, sometimes called plaintext) into hidden or secret text (cipher text). Then, back again by a process called decryption. The root of the word encryption comes from the Greek word krypto, means hidden or secret. Cryptography should provide four objectives:

confidentiality : unauthorized people should not be able to understand the information.
Integrity: any change in the information during transmission should be detected by the receiver.
Non-repudiation: the sender should agree sending the information.
Authentication: the sender and receiver are able to confirm identity [1].

Cryptanalysis is the science that studies the encrypted message and try to find a weakness to reveal the hidden meaning of that message without necessarily knowing the algorithm or the key of encryption and this is called breaking the encryption. In fact, cryptanalyst has tools to try breaking the cipher text [2]. The English language has regularity feature, which means that some letters have more frequency in a text than others. Fig. 1 represents the frequency distribution of English letters, which shows that letter E is most frequent, letter T in the second rank, then A,O …etc. Therefore, the cryptanalytic tries to recognize the frequency of each symbol in the cipher text and the period of repetition to find matching with the frequency distribution in Fig. 1. Moreover, In general English text has common digraphs and trigraphs, which are not distributed randomly. For example, the most common digraphs are TH, IN, ER and trigraphs are THE, ING, AND ……..

Global Proceedings Repository
American Research Foundation
ISSN 2476-017X
Available online at http://proceedings.sriweb.org
شبكة المؤتمرات العربية
http://arab.kmshare.net/



Classical Cryptography can be classified into three types[3]:

1. By type of encryption process used
a. Substitution
- Monoalphabetic like:

I. Caesar method
II. .Affine method

- Polyalphabetic like:

I. Vigenere method
II. Vernam method
III. One-time pad

- Homophonic
- Polygram

b. Transposition
2. By type of keys used
a. Secret-key or symmetric
b. Public-key or asymmetric
3. By the method in which clear text is processed
a. Stream
b. Block

1.By type of encryption process used :Substitution cipher means substituting one piece of the plaintext by another piece called cipher text. In monoalphabetic cipher a plain letter is

Global Proceedings Repository
American Research Foundation
ISSN 2476-017X
Available online at http://proceedings.sriweb.org
http://arab.kmshare.net/

replaced by a cipher symbol. For example, in Caesar method a shift by 3 is used for encryption according to the following equation:

$C_i = (P_i+3) \text{ Mod}(26)$ ……(1)

C_i the cipher letter,

P_i is the plain letter

Mod26 is the module arithmetic operation of 26 English letters.

To implement the above equation, the alphabetical letters are numbered starting from 0 to 25 .By replacing each plain letter by it's equivalent number in equation 1 results the cipher letter number. Hence the Plain letters and the equivalent cipher letters are shown in Table.1

Table.1 Caesar Cipher Method Equivalent Symbols

Plain letter	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
Letter Number	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
Cipher Number	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	0	1	2
Cipher letter	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

Affine monoalphabetic cipher is a little bit more complicated than Caesar. The cipher letters are generated from the plain letters using the following mathematical equation:

$C_i = (m* P_i + k) \text{ Mod } (n)$ ……..(2)

m constant

P_i a plain letter

k the key

n=26

With very important condition, the Greatest Common Divisor (GCD) of [m,n]=1. This condition is essential to preform decryption. For example, if m=7, k=10 and the plain text wanted to be encrypedt : WAR LOST

GCD(7,26) =1

Plain Text	W	A	R	L	O	S	T
Plain numbers	22	0	17	11	14	18	19
Cipher numbers	8	10	25	9	4	6	13
Cipher Text	I	K	Z	J	E	G	N

In general, monoalphabetic encryption methods are simple to apply but very easy to break by cryptanalyst using English letters frequency distribution.

On the other hand, polyalphabetic cipher is a set of monoalphabetic methods used as well as a key to transform from plaintext to cipher one [4]. Vigenere method is one of the well

Global Proceedings Repository
American Research Foundation
ISSN 2476-017X
Available online at http://proceedings.sriweb.org
http://arab.kmshare.net/

known and simplest polyalphabetic cipher . It consists of a table of 26 Caesar cipher rows. The first row is shifted by 0, the second by 1…..till the last by 25.The Plain letters represents the upper row and the key letters are organized as the first column to the left. In this method the intersection of the plain and the key letter in the table represents the cipher letter. For Vernam method which uses binary XORing between the plaintext and the key of any length to generate cipher text.

$C_i = P_i \oplus K_i$ ……(3)

$\oplus$ Logical XOR

K_i key letter.

Finally, in one-time pad a set of non repeating keys was used . The keys were written on papers glued together into one pad. For encryption, the sender used a table like vigenere table to encrypt the plaintext with about 16 papers of keys. In decryption, the receiver used the same pad of the sender. This is quite secure and effective because there is no repetition in key but still there are significant problems. firstly in generating unlimited number of keys and secondly synchronization between sender and receiver [5].

Homophonic substitution method used more than one letter or symbol to allocate the high frequency letters. For example the letter ‘e’ and ‘t’ might be represented by six different symbols because both have the highest frequency in English. Three symbols for ‘m’ and the rest by one.Polygram substitution method compromised substituting a block of plaintext by another block of cipher text. For instance, AA might map to GO, SD map to DZ . This is certainly is done according to certain key [6] .Transposition: it is the encryption in which the places of letters in the message are changed (letters are rearranged).This provide diffusion for the cipher text because the Cryptanalytic will face difficulty to know the right order [7].

2. By type of keys: it is another way of encryption[8]. Secret-key or symmetric, in this method one key is used for encryption and decryption. Therefore this key should be secret, known only by the sender and receiver. Public-key or Asymmetric : two key are used a public key for encryption, everyone know it and a secret key for decryption known by the authorized receiver. It is clear that Asymmetric method is more secure than symmetric, but the former can become more secure by increasing the key length to at least 128 bits. In this case, the brute force needs long years to decrypt [9].

3. By the way in which clear text is processed [3]
   Stream : means the symbols of a plaintext are encrypted individually, one after the other.
   Block: in this method a block of plaintext is encrypted and so on.

THE SUGGESTED ALGORITHM

In the previous section most of classical and popular methods have been stated with brief explanation. Each of these methods has a kind of weakness in structure which made them intercepted and broken in different times in the last century. The suggested work is based on

Global Proceedings Repository
American Research Foundation
ISSN 2476-017X
Available online at http://proceedings.sriweb.org
http://arab.kmshare.net/

building a new algorithm which is a combination of some of these methods to overcome the existing weaknesses . The new encryption method which is named modified Vernam-Homophonic method contains substitution, transposition, homophonic , binary Vernam method and finally some matrix mathematical operations.

As stated before, in homophonic substitution method the more frequent letters have multi-two substitution letters option to get around frequency distribution cryptanalysis. This principle is slightly changed in the new method by replacing each letter with two letters of equal probability without depending on the number of repetition for each letter as shown in Figuer 2.

Fig. 2 The English Letter Mapping in the Modified Vernam Method

Then, replace each symbol by the ASCII equivalent number to facilitate the next operations. After that, transforming the 1 x 2N vector into an array of 2x2N dimensions followed by dual transposition. Later, convert the numbers to there 8-bit binary equivalent. Finally use Vernam method to perform XORing between the ciphered array and the key array. In modified Vernam method a symmetric key is used for encryption. It was mentioned before that the symmetric key should be not repeated or the length is long enough to deter the brute-force from cracking the cipher text. The Key of the proposed algorithm is generated randomly with long period. Matlab program is used to implement this algorithm because it has many built – in functions that make programing much easier.
The ASCII numbers for both the printed and non-printed (invisible) letters and symbols are used. The ASCII of the printed letters and symbols starts from 32 to126, for the hidden from 0 to 31 and from 127 to 255. Table. 3 shows the ASCII numbers for the printed characters.

Table.3 the ASCII of the printed characters

Char-acter		!	“	#	$	%	&	‘	(	)	*	+	,	-	.	/
ASCII	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47
Char-acter	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
ASCII	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63
Char-acter	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
ASCII	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79
Char-acter	P	Q	R	S	T	U	V	W	X	Y	Z	[	\	]	^	_
ASCII	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95

Global Proceedings Repository
American Research Foundation
ISSN 2476-017X
Available online at http://proceedings.sriweb.org
شبكة المؤتمرات العربية
http://arab.kmshare.net/

Char-acter	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
ASCII	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111
Char-acter	p	q	r	s	t	u	v	w	x	y	z	{	|	}	~	
ASCII	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	

Algorithm

A. Encryption

1. Generate a matrix 1 x N
2. Input the plain text.
3. Convert the plain text to ASCII according to table. 3
4. Create a symmetric key and convert to ASCII too.
5. Convert the ASCII of both the plain text in step 3 and key in step 4 to 8-bit binary equivalent.
6. XOR the binary bits of plain text and key in step 5.
7. Convert the results of XOR in step 6 to decimal.
8. Convert ASCII in step 7 to characters .
9. Create a map table as in Figure 2.
10. Substitute each character in step 8 by two letters according to map table in step 9.
11. From step 10 a vector of cipher text with 1*2N dimensions is obtained.
12. Convert the cipher text in step 11 to a matrix of size 2x N
13. Perform columnar transposition to matrix in step 12.

B. Decryption

1. Reverse columnar transposition is performed.
2. Convert the 2 x 2N matrix to 1x 2N matrix.
3. Use the map table in reverse direction, hence the 1x2N matrix is converted to 1xN.
4. Convert the 1 x N matrix to decimal.
5. Convert the vector in step 4 to binary.
6. Use the same key in encryption for decryption.
7. XOR the result of step 5 with the key in step 6.
8. Transform the result of step 7 to decimal.
9. Transform the decimal in step 8 to letters of the plain text.

IMPLEMENTATION OF MODIFIED HOMOPHONIC-VERNAM METHOD

1. Input the plain text. pt=ENCRYPTION.
2. Calculate the length of the plain text p including spaces between words using the Matlab built-in function (length)
   Example; N=length(pt)

Global Proceedings Repository
American Research Foundation
ISSN 2476-017X
Available online at http://proceedings.sriweb.org
شبكة المؤتمرات العربية
http://arab.kmshare.net/

3. Create a vector 1 x N. The MATLAB language does not have a dimension statement, instead Matlab allocation method is used to create a matrix. For example p=zeros (1,N), means p is an array of 1xN dimensions with all elements equal to zero .
4. Create another array (1,2N) for the cipher text.
5. Convert the plain text letters to ASCII numbers using (double) Built-in function. Example: p= double ('ENCRYPTION') as shown in Table.4.

Table. 4 The Plain Text (ENCRYPTION)and The Equivalent Decimal Numbers

Plain text	E	N	C	R	Y	P	T	I	O	N
ASCII	69	78	67	82	89	80	84	73	79	78

6. Convert the decimal numbers in step 5 to 8-bit binary using the built –in function (de2bi). For example: b=de2bi(p, 8). It should be mentioned that matlab displays the 8-bit binary in revers direction. i.e the Least Significant bit(LSB) is to the left and Most Significant bit(MSB) is to the right. Hence the result is shown in Table. 5.

Table. 5 the plain text : ENCRYPTION and the equivalent decimal and binary bits.

Plain text	E	N	C	R	Y	P	T	I	O	N
ASCII	69	78	67	82	89	80	84	73	79	78
8- bit binary Equivalent	10100010	01110010	11000010	01001010	10011010	10001010	00101010	10010010	11110010	01110010

7. To generate the Symmetric – key. Take the module of the last letter in the plain text which represents the beginning of the secret key.

$K_1 = 90 \bmod p(N)$ …….(4)
p(N)= N as shown in table. 5 with decimal number equal 78.
The number 90 is used for normalization of the ASCII numbers representing the alphabetical letters

$K_1 = 90 \bmod 78 = 12$

$K = [k1\ k2 \ldots . . kN]$ ……(5)

Equation (4) is repeated for all the key K elements.
In equation. 5, the length of the key is N equal to the length of plain text i.e key length = 10 .
K=[12 13 14 15 16 17 18 19 20 21]

8. Convert the ASCII of the key to 8-bit binary as in step 6. The result is shown in Table.6

Global Proceedings Repository
American Research Foundation
ISSN 2476-017X
Available online at http://proceedings.sriweb.org
شبكة المؤتمرات العربية
http://arab.kmshare.net/

Table.6 The decimal & the binary equivalent of the secret key

Key ASCII	12	13	14	15	16	17	18	19	20	21
Key 8-bit binary Equivalent	00110000	10110000	01110000	11110000	00001000	10001000	01001000	11001000	00101000	10101000

9. XOR the binary bits of the text in step 6 with that of the key in step 8. Table. 7 demonstrates that.

Table. 7 XOR operation between the text and Key

	Binary Equivalent	Equivalent	Result of Encryption by XOR	ASCII Equivalent	Character Equivalent
1st character	The Key Binary The Plain Text Binary	00110000 10100010	10010010	73	I
2nd character	The Key Binary The Plain Text Binary	10110000 01110010	11000010	67	C
3rd character	The Key Binary The Plain Text Binary	01110000 11000010	10110010	77	M
4th character	The Key Binary The Plain Text Binary	11110000 01001010	10111010	93	]
5th character	The Key Binary The Plain Text Binary	00001000 10011010	10010010	73	I
6th character	The Key Binary The Plain Text Binary	10001000 00001010	10000010	65	A
7th character	The Key Binary The Plain Text Binary	01001000 00101010	01100010	70	F
8th character	The Key Binary The Plain Text Binary	11001000 10010010	01011010	90	Z
9th character	The Key Binary The Plain Text Binary	00101000 11110010	11011010	91	[
10th character	The Key Binary The Plain Text Binary	10101000 01110010	11011010	91	[

10. Now each character in step 9 is replaced by two characters according to Fig.2. Since the dealing is with ASCII, therefore, the ASCII of the characters is replaced by two numbers, the first is the ASCII number of the character itself and the second is the same character number + 13 if the ASCII is <78 and character number -13 if ASCII>78.

Text from	I	C	M	]	I	A	F	Z	[	[

Global Proceedings Repository
American Research Foundation
ISSN 2476-017X
Available online at http://proceedings.sriweb.org
شبكة المؤتمرات العربية
http://arab.kmshare.net/

Cipher text	I	V	C	P	M	Z	]	P	I	V	A	N	F	S	Z	M	[	N	[	N
ASCI	73	86	67	80	77	90	93	80	73	86	65	78	70	83	90	77	91	78	91	78

11. Do double transposition for cipher text in step 10. In first transposition convert cipher text vector to an array of 2 x N dimensions.

I	V	C	P	M	Z	]	P	I	V
A	N	F	S	Z	M	[	N	[	N

Columnar transposition is performed in second one.

I	A	V	N	C	F	P	S	M	Z
Z	M	]	[	P	N	I	[	V	N

Discussing the method performance against cryptanalysis

The main target of encryption is to convert the plain text to unreadable text, In a way that makes cryptanalysis a big dilemma. Suppose a cryptanalyst intercepted a cipher text encrypted by modified Vernam method. Example.1 shows the plain and the cipher text.

Example.1 the plain text and the cipher text using Vernam-Homophonic method

The plain text	DATA THAT CAN BE READ AND UNDERSTOOD WITHOUT ANY SPECIAL MEASURES IS CALLED PLAIN TEXT.THE METHOD OF DISGUISING PLAINTEXT IN SUCH A WAY AS TO HIDE ITS SUBSTANCE IS CALLED ENCRYPTION
The cipher text	HSU*L7Y-Z:M=NJA"0/=$E1R(Z5MMRZE!@.M)56BPUCH5VBI;VHI 9-F3X@K ^-Q?<LI$O1B1[>N7^DQ=dJW[N ,c9V1m>`?`LS6 C s f I \ l _ l _ x k¥x k x k b¼U¨a T k ^ f»Y {·n-g Z | o¯z¢m c V c V ¾%³x x k t g¹b¶U©)¶n×a n¹a¬z m ¾ ±d Wô ç u ( ) ç% è- , mé` ç é) pîc × , % w j " (} pé è' 8âE 5 B LéY 7÷Dê, 9 - F

The key is unknown. Therefore, to regenerate the plain text, he tried different cryptanalysis techniques. Firstly, he applied frequency analysis to find matching between the cipher text in example.1 and English distribution in Fig.1. It is clear that , there is a large diversion between them, specially the cipher text contains many unprinted and printed symbols. Secondly, its

Global Proceedings Repository
American Research Foundation
ISSN 2476-017X
Available online at http://proceedings.sriweb.org
http://arab.kmshare.net/

overly difficult to find common digraph and trigraph words like(is, she,the, etc). Thus, using long key plus double transposition made the algorithm more stronger against attackers.The key itself and length is different for each message. The first charectar in the key depends on the last character in the plain text. For instance, if the last character in a message is Y(ASCII=89), then the first chaecter in the key is

K1=90mod 89=1

And the length of the key in this case is 255-1=254 characters.

Hence, its perferable that the message length should not be longer than 254 characters to avoid repetition of the key. In general, for a key length ≥230 characters, the brute force needs 230^{255} tries to find the right key combinations which takes million of years.

Features of the Modified Vernam Method

1. The key is generated automatically from the message itself.
2. Similar letters in a message are mapped to different symbols because each letter has a different key to encode.
3. The characteristics of the language are hidden.
4. It is too hard to be broken by brute force.
5. It is recommended that the length of the message with spaces ≤ 254 characters, to avoid key repetition.
6. It is rather difficult to implement but using MATLAB with built-in functions made the application of the algorithm much quite easier.

Conclusion

In this paper, most of the traditional encryption methods are stated with brief description and their points of weaknesses and limitations. The Vernam – Homophonic method is proposed to overcome these impairments by the combination between Vernam, Homophonic and transposition methods. The key in this method is generated automatically from the plain text and the length is made long enough to prevent the brute force from finding the right combination of the key. As a result, the frequency distribution has no intersection with that for normal English language .In addition, the similar letters in a text are not mapped to the same character, which makes the cryptanalyst confused. Furthermore, double transpositions break the adjacencies between double and triple words. Finally, Matlab is used for programming which in return facilitate the code sequences.

Refrences

Global Proceedings Repository
American Research Foundation
ISSN 2476-017X
Available online at http://proceedings.sriweb.org
http://arab.kmshare.net/

[1] Jain A., dedhia R. & Patil A., " Enhancing The Security of Ceasar Cipher Substitution method Using a Randomized Approach for More Secure Communication", International Journal of Computer Applications Vol. 129 – No.13, 2015.
[2] Schafe E., "An Inroduction to Cryptography & Cryptanalysis ", SantaClara University, eschaefer@scu.edu.
[3] Stalling W., "Cryptograpgy & Security Principles & Practice",fifth edition, 2011.
[4] Dhull S, Beniwal S., Kalva P., " Polyalphabetic Cipher Techniques Used For Encryption Purpose", International Journal of Advanced Research in Communication Science 7 Software Engineering , Vol.3, Issue-2, 2013.
[5] Venkateswaran R., Sundaram V., "Information Security : Text Encryption & Decryption with Polysubstitution Method & Combining The features of Cryptography", International Journal of Computer Applications Vol. 3 – No.7, 2010.
[6] Ravi S.& Knight K., " Bayasian Infrence for Zodiac & Other Homophonic Ciphers",
[7] Arya G., Nautiyal A., Singh S. & Handa T., "A Cipher Design With Automatic Key Generation sing The Combination of Substitution & Transposition Technique & Basic Arithmetic & Logic Operations", The SIJ Transactions on Computer Science engineering & Its Applications, Vol.1, No.1, 2013.
[8] Ebrahim M., Khan S. & Khalid U.," Symmetric Algorithm Survey : A Comparitive Analysis ", International Journal of Computer Applications Vol. 61 – No.20, 2013.
[9] Tripathi R., Agrawal S., " Comparative Study of Symmetric & Asymmetric Cryptography Techniques ", International journal of advance foundation & Research in Computers, Vol.1, Issue-6, 2014.