IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 17, Issue 2, Ver. VI (Mar – Apr. 2015), PP 18-23
www.iosrjournals.org

The Effect Of Varying Key Length On A Vigenère Cipher

Ayman Al-ahwal[1], Sameh Farid[2]
[1](Communication and electronics/ pyramid-institute for Engineering and Technology, Egypt)
[2](mechanical engineering, College/ pyramid-institute for Engineering and Technology, Egypt)

Abstract: *Vigenère cipher is one of the polyalphabetic substitution ciphers. Its weakness is the key repetition. To overcome this weakness there are many researchesgoing on to modify the key generation. In this paper a key generator function isimplemented with C++. Itgenerates a key with length depends on the message security level.This level is determined by the sender of the message, according to the security importance of the message. The security level determines the key length, which is a ratio from the plain-text message. This ratio varies from 10% to 100% from the plain-text length. This paper studies the effect of varying the key length on the performance of Vigenère cipher andits frequency analysis attack. When the key length increases the encryption and decryption time increases, the frequency analysis attack becomes more difficult and also the confusion is increased.*

Keywords *–Vigenère cipher, poly-alphabetic cipher, substitution cipher, Index of Coincidence, Kasiski-test, varying key-length.*

I. Introduction

We live in a computer-based society, this widespread use of computer networks and information in electronic form increases illegal reaches of information and causes insecurity [1]. Security is always a people problem, it is not a product, but it is the process of protecting computer networks against penetration by unauthorized persons.

The confidentiality is the major security service which is used to prevent attackers to know about the transmitted message [2]. To achieve message confidentiality, cryptography is used. It is the science or art of keeping message secure, by using difficult mathematical problems [3]to transform readable text (plain-text) into a form that is impossible to read (cipher-text) [1]. The cryptographic algorithm plus all possible keys is called a cryptosystem [3].When both the transmitter and the receiver uses the same key the cryptosystem is called symmetric or single-key, but if they use different keys the cryptosystem is called asymmetric, two-key or public-key.

Vigenère cipher is a symmetric key cryptosystem proposed by Blaise de Vigenère in the late 1500s[4]. It based on the substitution cipher. The substitution cipher is a method of encoding by which units of plain-text are replaced with cipher-text, according to a regular system. The receiver deciphers the text by performing an inverse substitution. Substitution ciphers achieve Confusion.In Shannon's original definitions, confusion refers to making the relationship between thecipher-text and the symmetric key as complex as possible.Confusion means that the key does not relate in a simple way to the cipher-text[5].

Vigenèrecipher uses Vigenère tableau (table I) of alphabets (26 ×26) [4]. It uses a key of distinct letters its length is called the key period. To encrypt a plain-text message write it without spaces, bolts and punctuation and repeat the key letters to be the same length of the plain-text. The intersection of the Vigenère tableau column of plain-text letter and the row of the key letter is the cipher-text letter [4].

The encryption algorithm is implemented in algebra by replacing the key letters and the plain-text letters by numbers (a = 0; b =1 …z = 25) and using the following equation (1) [6]:

$$C_i = E_k (X_i) = (X_i + k_i) \bmod 26 \quad (1)$$

Where $C_i = C_oC_1C_2\ldots\ldots C_n$ the cipher-text with n letters length, (E_k) is the Vigenère encryption, Xi = XoX1X2…Xn is the plain-text with n letters length, the key with m letters length is $k_i = k_ok_1k_2\ldots k_m$, and mod26 is the modulo (modulus) operation which finds the remainder of division of one number by 26.

To decrypt the cipher-text the receiver needs to know the key before decryption process. It uses the Vigenère tableau to search in the row of the key letter to get the cipher letter its column heading is the plain-text letter. The decryption algorithm is implemented in algebra by replacing the key letters and the cipher-text letters by numbers and using the following equation (2)[6]:

$$X_i = D_k (X_i) = (C_i - k_i) \bmod 26 \quad (2)$$

Where (D_k) is called the decryption of the cipher-text.

Table: Vigenère Tableau[6]

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

The science or art of retrieving the plain-text from the cipher-text without knowing the secret key is called Cryptanalysis [1].Cryptanalysis techniques for Vigenère cipher, Kasiski test and index of coincidence, are used to decrypt the cipher-text without prior knowledge of the key. These techniques are based on the repeated key weakness of the Vigenère cipher.

A. Kasiski Test:

It estimates the key length (m) that used in the Vigenère cipher. It finds the distance between the repeated groups of at least three cipher-text letters. The greatest common divisor (GCD) of all distances should be the multiple of key length [7].Once the key length is discovered, the cipher-text lines in m columns. Then, each column can be treated as mono-alphabetic substitution cipher[8]to get the matching plain-text lettersby performing the frequency analysis.

B. Index of Coincidence (IC):

It is a statistical measure gives the probability of two randomly selected letters are identical. It is calculated by equation (3) [8]:

$$\mathrm{IC}(x) = \frac{\sum_{i=0}^{25} f_i(f_i - 1)}{n(n-1)} \tag{3}$$

Where IC(x) is index of coincidence of the plain-text X, $X = X_o X_1 X_2 \ldots X_n$ is a string of n alphabetic English characters, and the frequencies of A…Z in X is $f_0 \ldots f_{25}$ respectively.
The approximate of IC(x) the probability to select two English lettersat randomcan get by equation (4) [9]:

$$\mathrm{IC}(x) \approx \frac{\sum_{i=0}^{25} f_i^2}{n^2} = \sum_{i=0}^{25} P_i^2 = 0.065 \tag{4}$$

Where P_i is the English letter frequency probability in table (2).
The IC of mono-alphabetic cipher is similar to English letter frequency:

$$\mathrm{IC}_{\text{mono-alphabetic}} \approx \mathrm{IC}_{\text{English}} = 0.065 \tag{5}$$

Due to the poly-alphabetic cipher tends to randomize the occurrences of the letters [8]its IC is decreased to:

$$\mathrm{IC}_{\text{poly-alphabetic}} \approx P_i = \frac{1}{26} = 0.038 \text{ (but no lower than 0.038)} \tag{6}$$

IC can give an estimate of the Vigenère cipher keyword length (m) by using the following formula [8]:

$$m \approx \frac{0.0265n}{(0.065 - \mathrm{IC}) + n(\mathrm{IC} - 0.0385)} \tag{7}$$

Wherem is the keyword length and n is the cipher-text length.

Table: Standard English letters Frequency (P_i) [6]

Character	Probability	Character	Probability	Character	Probability
A	**0.08167**	J	**0.00153**	S	**0.06327**
B	0.01492	K	0.00772	T	0.09056
C	0.02782	L	0.04025	U	0.02758
D	0.04253	M	0.02406	V	0.00978
E	0.12702	N	0.06749	W	0.0236
F	0.02228	O	0.07507	X	0.0015
G	0.02015	P	0.01929	Y	0.01974
H	0.06094	Q	0.00095	Z	0.00074
I	0.06966	R	0.05987		

To verify the keyword length m, divide the cipher-text into m length sub-strings and compute IC for each sub-string. If all IC values are around 0.065 then m is the correct key length, otherwise (IC=0.038) m is not the correct key length. "IC is a good estimator of the cipher period only for small m, but its predictions are less accurate for larger m values" [8].

The cryptanalysis of poly-alphabetic cipher is based on the primary weakness of the Vigenère cipher, the repeated key, to increase the security of Vigenère cipher the key length needs to be enlarged. There are several researches to increase the key length, in paper [10] generates large key from short keyword using Linear-Feedback-Shift-register (LFSR) which flatten the letter frequencies of cipher text. It leads to decrease the effectiveness of Kasiski and Index of Coincidence (IC) attacks.

Paper [11] changes the key length by using initial key value to generate the second step key will be as a result of a function that operated on the first step (initial key) and so on. After each successive encryption step the successive keys are dependent on the initial key value.

Paper [12] extends Vigenère cipher tableau to Extended ASCII characters,and increases the length of the key to 256 characters without repetition of any charactersto decrease the effectiveness of cryptanalysis attacks.

This research is aimed at contributing the developing a key generator function that generates a key with length depends on the message security level (high, medium, low or of any percentage). This level is determined by the sender of the message. The sender determines the importance of the messageaccording to its sensitivity.It also studies the effect of varying the key length on the performance of Vigenère cipher, the frequency analysis attack.The paper has the following structure: section II the proposed key generator function, section III experimental results, andsection IV conclusions and future work.

II. The Proposed Key Generator Function

The proposed key generator function uses RAND function of C++. It is a pseudo-random integer number produces unpredictable random number that it is computationally infeasible to predict the next random number and all the previous numbers in the key stream. It returns a sequence of non-related numbers each time it is called. The numbers are converted to characters based on (modules 26). The characters are stored in a file with a length based on the security level.

Thekey generator function algorithm is the following pseudo-code:

1. Select the security level.
2. Calculate the plain-text filelength (FS).
3. Put counter equals zero, and create an array of letters, key as key string.

a. If the security level is low:
- Put new file size (NFS) equals 33% of the plain-text file length (FS).

b. If the security level is Medium:
- Put NFS equals 67% of the plain-text file length.

c. If the security level is high:
- Put new file size (NFS) equals the plain-text file length (FS).

d. If the security level optional inputa percentage (p %) from the plain-text filelength (FS).
- Compute NFSequals (p *plain-text file length/100).

4. Generate random number Rand by C++ RAND ().
5. Get character Rand % 26 and use it as an index to get letter from array of letters.
6. Concatenate the character with the key.
7. Increment counter by one.
8. Compare counter with (NFS): If counter less than or equals(NFS) go-to step 4 otherwise store the key generated in a key file

At step 8 when the counter equals to the plain-text file length (FS) the key is generated is a ratio from the plain-text message or it equals its length. When the key length equals the message length the "Vigenère cipher is called one-time pad or Vernam cipher which is a perfect secure symmetric encryption algorithm. This case is not practical for real use because it is difficult to distribute the key pad securely" [2].

III. Results AND Analysis

The results of this paper are based on two assumptions:

i. The key file storage and distribution are secured.
ii. The key length is based on the sender security level.

The Vigenère cipher program runs on 32-bit operating system machine, a windows 7 enterprise, with processor Intel Core i7, 3.40 GHz with 4.00 GB Ram with Visual studio C++ 2012. The program runs to get from the user the name of plain-text file, the blank key file name, to write the generated key into it and the security level. The program calculates the encryption and the decryption time, number of flipped bits. The plain-text and cipher-text are written in text files.

The program runs to get the plain-text input message (in this paper is the first two sections of it without spaces, punctuation, bolts and numbers and its is size of 4056 characters). The user selects a different security level according to the following table (3):

Table: Vigenère Cipher performance for input

Security Level	Plain/Cipher Text Length (Character)	Key length (character)	Encryption time (ms)	Decryption time (ms)
Low	4056	1352	516	347
Medium		2704	525	353
high		4056	540	409

The relationship between the key length and the encryption and decryption time are implemented in Fig.1:

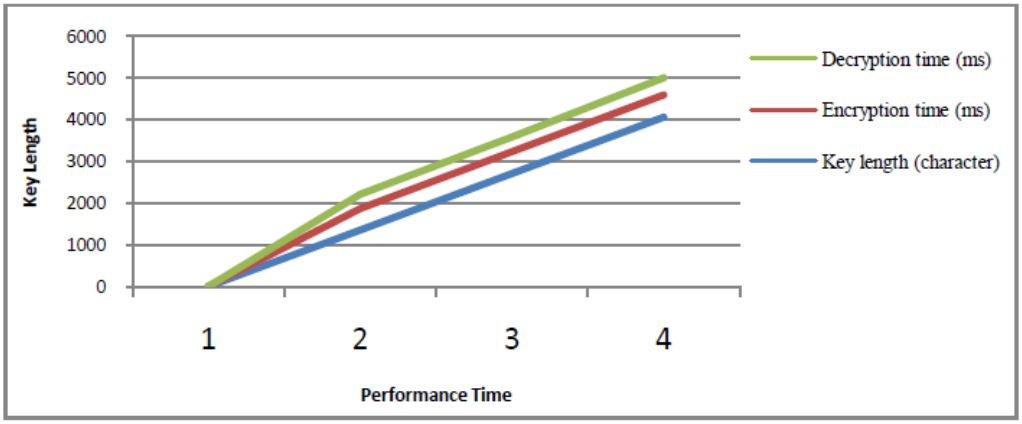

Fig: Relationship between the performance and the key length

Fig. 1 represents the performance (the encryption and decryption time) of the Vigenère cipher for the same input plain-text message. The encryption time and the decryption time increases when the key length increases.

The program calculates also the relative frequency of the cipher-text, which calculated by equation (8) as follows:

$$Relative\ Frequency\ of\ a\ \ letter = \frac{counts\ of\ \ the\ letter\ in\ message}{total\ number\ of\ \ letters\ in\ message} X100 \qquad (8)$$

The relative frequency of the cipher-text at different security level (high, medium and low) is drawn in Fig.2and fig.3 for the input plain-text file (4056 characters). The key lengths are (1338, 3042 and 4056 characters) at different security levels.

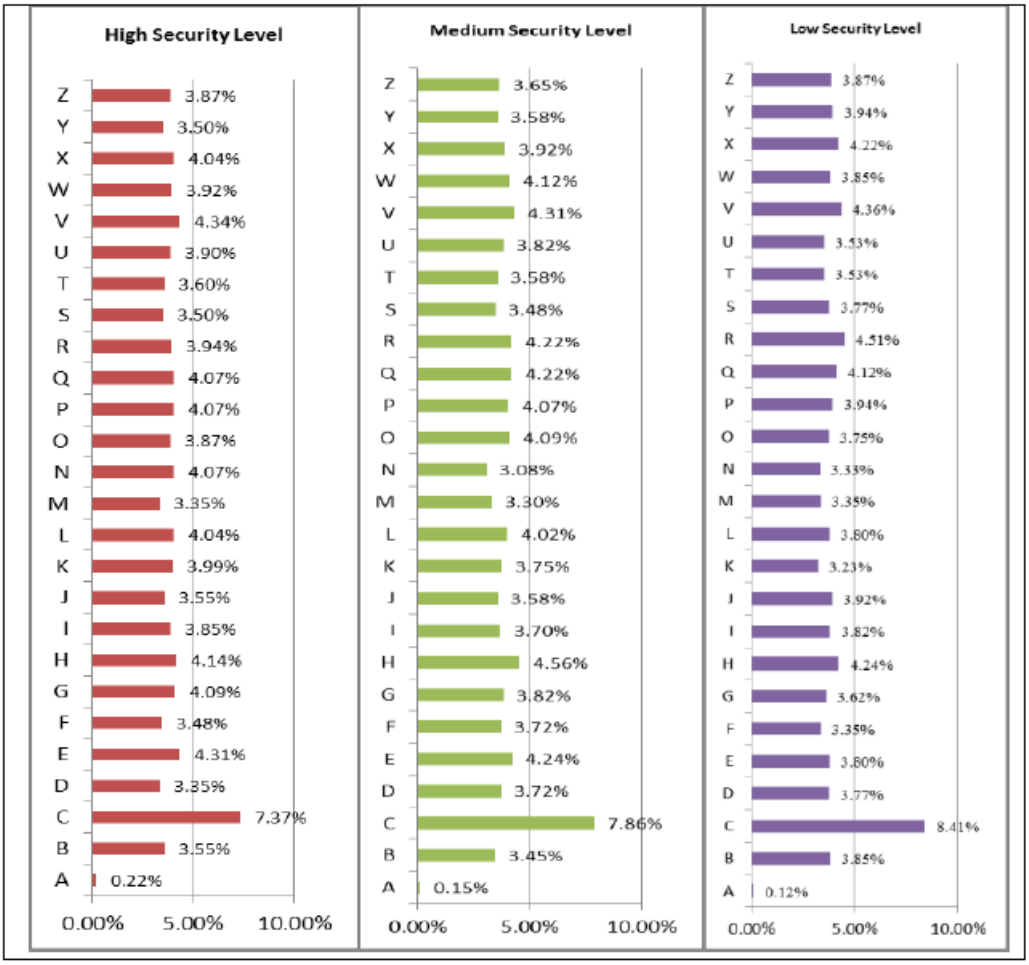

Fig. 2. Relative frequency of cipher-text at different security level

The result of Fig.2 is re-represented in Fig.3 as follows:

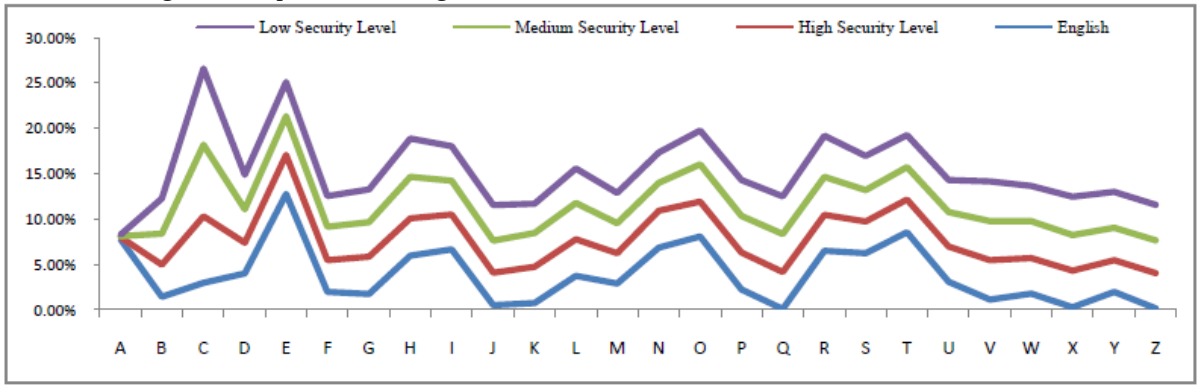

Fig. 3. Relative frequency of cipher-text at different security level (low, medium and high)

Fig.2 and Fig.3 represent the relative frequency of English letters and the relative frequency of cipher-text at different security levels (low, medium and high). The shape of curve at high security level is more flat than that of English letters and also medium and low security levels. So that when the key length increases the curve becomes more flat this reduces the effect of cryptanalysis techniques. So that the cryptanalysis process with the longer keys is harder than that with shorter keys for long plain-text input message. To proof that he program calculates the number of ones cipher-text, after convert to binary bits,by usingequation (9) as follows:

$$flipped\ bits\ Ratio = \frac{Number\ of\ ones\ in\ ciphertext - Number\ of\ ones\ in\ plaintext}{Number\ of\ ones\ in\ ciphertext} \qquad (9)$$

The program calculates the flipped bits ratio at different security level and the representation in Fig.4 as follows:

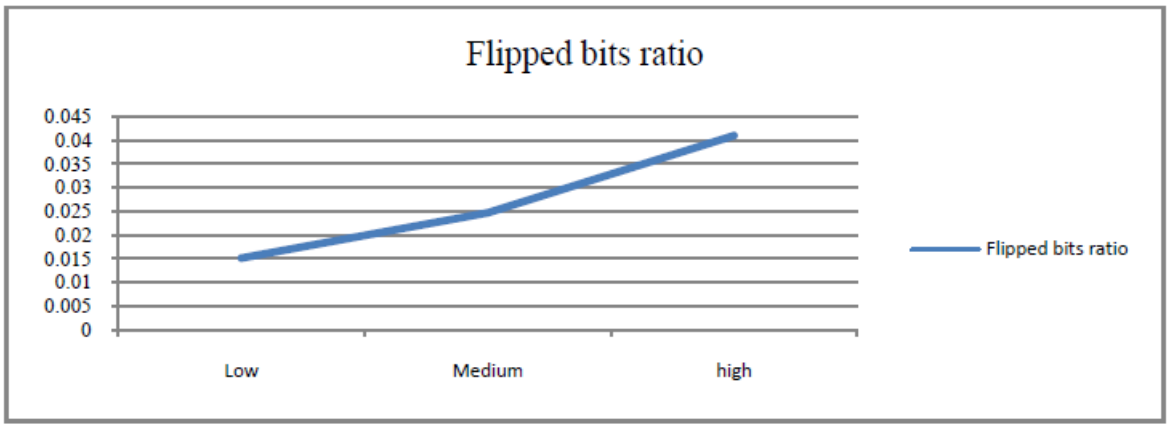

Fig.4: The relationship between the security level and flipped bits ratio

Fig. 4 showsx-axis is the security level and y-axis is the flipped bits ratio forthe same plain-text input.When the security level increasesflipped bits ratioincreases. This measures the confusion concept which makes the relationship between cipher-text and key as complex as possible (each character of the cipher-text depends on many parts of the key).

IV. Conclusion

From the analysis the Vigenère cipher performance is better with smaller key length than that of larger key length. At higher security level (i.e. at higher key length) thefrequency analysis more difficult and also confusion is increased. For more security the key length should be larger but it should be compromised with performance in the application that requires better in performance.

The future work will include the cryptanalysis effect of varying key length of the Vigenère cipher and limitations of cryptanalysis algorithms in guessing the key length and the key itself.

References

[1]. Alfred J. Menezes, Paul C. van Oorschot and Scott A. Vanstone, Handbook of Applied Cryptography(CRC Press Inc.1996).
[2]. Bruce Schneier,Applied Cryptography: Protocols, Algorithms, and Source Code in C(2nd Edition, John Wiley & Sons, 1996).
[3]. Miles E. Smid and Dennis K. Branstad, the data encryption standardPast and future(IEEE, 1992).
[4]. http://user.it.uu.se/~olgag/Cryptology/Vigenère.html
[5]. http://en.wikipedia.org/wiki/Substitution_cipher
[6]. William Stallings, cryptography and network security(5th edition, Prentice Hall, USA, 2010).
[7]. http://en.wikipedia.org/wiki/Kasiski_examination
[8]. Ranju S Kartha, Varghese Paul, Survey: Recent Modifications in Vigenère Cipher IOSR Journal of Computer Engineering (IOSR-JCE), Volume 16, 2014.
[9]. http://en.wikipedia.org/wiki/Index_of_coincidence
[10]. Abdul Razaq ,Yasir Mahmod, Faroq Ahmed, Ali Hur,Strong Key Mechanism Generated by LFSR based Vigenère CipherThe International Arab Conference on Information Technology (ACIT'2013),Sudan ,2013
[11]. Quist-Aphasia Kester, a cryptosystem based on Vigenère cipher with varying key International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), Volume 1, Issue 10, ISSN: 2278 – 1323, December 2012.
[12]. Ravindra Babu Kallam, S. Udaya Kumar, Md Abdul Rasool, A. Vinaya Babu and Puskur Pavan, An Enhanced Polyalphabetic Cipher using Extended Vigenere Table International Journal of Advanced Research in Computer Science, Mar-Apr 2011.