International Journal of Information & Network Security (IJINS)
Vol.1, No.4, October 2012, pp. 313~320
ISSN: 2089-3299

Novel Cipher Technique Using Substitution Method

Shobha Vatsa, Tanmeya Mohan, A. K. Vatsa
Shobhit University, Meerut, U.P., INDIA

Article Info

Article history:

Received Jun 05th, 2012
Revised Jul 30th, 2012
Accepted Aug 28th, 2012

Keyword:

Vowels,
Consonants,
Substitution cipher,
Cryptography,
Security.

ABSTRACT

Cryptography plays a very vital role in the field of Network Security. It is a complex and mathematically challenging field of study. Always there exists a need for confidentiality and integrity of all information being transmitted on the network that is very important for certain communicating parties. The data being transmitted needs to be kept secure and out of reach for unauthorized access. In this paper, we propose a new symmetric substitution cipher technique using a mechanism on vowel, consonants of English language and numeric digits, which exploits the disadvantages of the symmetric substitution techniques previously known to us. Also certain new aspects of encryption and decryption have been explored.

Corresponding Author:

First Author,
Shobhit University, Meerut, U.P., INDIA.
Email: shobha778@gmail.com

1. INTRODUCTION

In this age of universal electronic connectivity, of viruses and hackers, of electronic eavesdropping and electronic fraud, there is indeed no time at which security does not matter. The explosive growth in computer systems and their interconnection via network has increased the dependence of both organizations and individuals on the information stored and communicated using these systems which intern has led to a heightened awareness of the need to protect data and resources from disclosure, to guarantee the authenticity of data and messages, and to protect systems from network-based attacks.

Also the disciplines of cryptography and network security have matured, leading to the development of practical, readily available applications to enforce network security. Cryptography is the design of certain techniques for ensuring the secrecy and/or authenticity of information. Earlier the requirement of information security within an organization was primarily provided by physical and administrative means [1]. But the concept of network security became quite evident with the introduction of computers and later with introduction of distributed systems. The need of cryptographic algorithm is to avoid threat to integrity, confidentiality and availability.

Symmetric Cipher Technique is also known as Conventional, Single key, Secret Key, One – key and classical encryption techniques. This technique is based on the encryption of plain-text to cipher-text which is safe to transmit and from unauthorized access, by using a secret key in a specific encryption algorithm.
It uses the following ingredients:

a. **Plain-text:** This in an intelligible piece of information i.e. original text that needs to transferred safely to the receiver. It is the main input to the encryption algorithm.
b. **Secret Key:** This is another input to the encryption and decryption algorithm, which is the main component used for converting the plain-text to cipher-text i.e. an unintelligible form which has the useful content hidden in a way.

c. **Encryption Algorithm:** This is the actual process by which we are converting the plain-text into cipher-text.
d. **Cipher-text:** This is the output of encryption process in which we are taking plain-text and secret key as input and processed by encryption algorithm. The cipher-text can be understood as a scrambled piece of text which has useful information in secret form.
e. **Decryption Algorithm:** This algorithm is the reverse of the encryption algorithm which takes in cipher-text and secret key as inputs and produces plain-text as the output.

The receiver on receipt of this unintelligible text, called cipher-text uses the secret key and with the decryption algorithm extracts the useful hidden information sent by the source. the process is depicted in figure 1.1. The secret key can either be shared between the source and destination by a secure channel or by a trusted third party. The requirements for secure use of conventional encryption are that the opponent should be unable to decrypt cipher-text or discover the key even if he/she is in possession of a number of cipher-texts together with the plain-text that produced each cipher-text. Also the sender and receiver must have obtained copies of the secret key in a secure fashion and must keep the key secure. With the use of symmetric encryption, the principal security problem is maintaining the secrecy of the key.

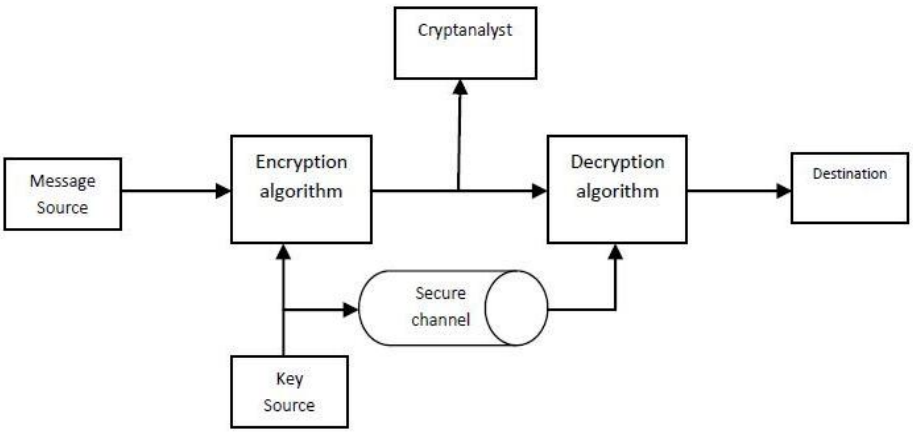

Figure 1: Model of Conventional Cryptosystem

A substitution technique is one in which the letters of plain-text are replaced by letters or numbers or symbols. If plain-text is viewed as a sequence of bits, then substitution involves replacing plain-text bit patterns with cipher-text bit patterns. The substitution technique either involves single alphabet replacement or multiple alphabet replacements, respectively known as mono-alphabetic substitution technique and poly-alphabetic substitution technique.

Mono-alphabetic substitution is a system of encryption where every occurrence of a particular plain-text letter is replaced by a cipher-text letter. the definition of a monoalphabetic substitution allows for the possibility that two distinct plain-text letters are replaced by the same cipher-text letter. However, to break this system using a known plain-text attack, we will require that any two distinct plain-text letters are replaced by two distinct cipher-text letters [2]. The Monoalphabetic Cipher (often referred to as a cryptogram) uses a KEY which is the rearrangement of the letters of the alphabet. These different letters are then substituted for the letters in the message to create a secret message. That KEY is needed to decipher the secret message [3]. A monoalphabetic substitution is one where a letter of plain-text always produces the same letter of cipher-text. The simplest examples of monoalphabetic substitutions are probably the Caesar Cipher and Atbash [4]. The problem with messages encrypted with a monoalphabetic substitution cipher is that it is very easy to break using frequency analysis [5].

In polyalphabetic substitution the cleartext letters are enciphered differently depending upon their placement in the text. As the name polyalphabetic suggests this is achieved by using several cryptoalphabets instead of just one, as is the case in most of the simpler crypto systems. Which cryptoalphabet to use at a given time is usually guided by a key of some kind, or the agreement can be to swich alphabet after each word encrypted (which, of course, presumes that the word boundaries are kept intact or indicated in some way), but the latter is seldom practiced in real life [6]. Polyalphabetic substitution ciphers are useful because they are less easily broken by frequency analysis, however if an attacker knows for instance that the message has a

period n, then he simply can individually frequency analyze each cipher alphabet. The number of letters encrypted before a polyalphabetic substitution cipher returns to its first cipher alphabet is called its period. The larger is the period, the stronger is the cipher. Of course, this method of encryption is certainly not secure by any definition and should not be applied to any real-life scenarios [7].

On basis of literature review the following problem identified with all other available substitution techniques codes can be broken either by means of brute force or character frequencies or by another way of cryptanalysis. Therefore, the purposed cipher technique is efficient and secure symmetric substitution cipher technique using a mechanism on vowel and consonants of alphabets, which exploits the disadvantages of the symmetric substitution techniques previously known to us. Also certain new aspects of encryption and decryption have been explored. It has been analyzed that there are two types of cryptanalysis attacks namely Brute-Force cryptanalysis, and Relative frequency analysis. In cryptography, a brute force attack or exhaustive key search is a strategy that can in theory be used against any encrypted data by an attacker who is unable to take advantage of any weakness in an encryption system that would otherwise make his task easier. It involves systematically checking all possible keys until the correct key is found. In the worst case, this would involve traversing the entire search space [21]. Frequency analysis is a methodology for "breaking" simple substitution ciphers, like the Caesar cipher. These ciphers replace one letter of the plain-text with another to produce the cipher-text, and any particular letter in the plain-text will always, in the simplest and most easily breakable of these cyphers, turn into the same letter in the cipher. Frequency analysis as shown in figure 1.2, is based on the fact that certain letters, and combinations of letters, appear with characteristic frequency in essentially all texts in a particular language [22].

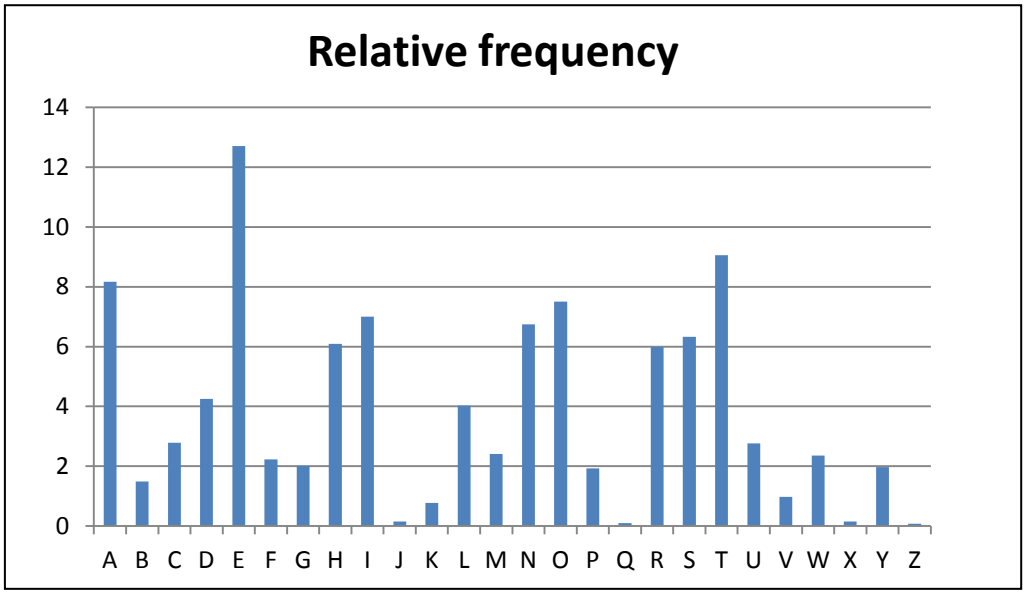

Figure 2. Relative frequency in English alphabets

This paper is organized into sections, introduction section familiarizes us to the basics of network security, and previously known techniques. It also emphasizes on the problem identification, due to the disadvantages of earlier known cipher techniques. The two cryptanalysis attacks have been discussed. The next section describes the related work that has been discovered by many scholars in the field of cryptography like Caesar cipher, Play Fair cipher, Hill cipher, Poly-alphabetic cipher, One time padding, Homophonic cipher and Aryabhata's Mathematics. Next is a detailed description of the algorithm that is being proposed in this paper with an example of the entire encryption and decryption process followed by conclusion of this paper and future scope. The related references have also been mentioned.

2. RESEARCH METHOD

The algorithm proposed in this paper uses some earlier proposed techniques and is based on some cryptographic facts. The following section describes the similar work proposed earlier and the facts on which it is based. Caesar Cipher is One of the simplest examples of a substitution cipher is the *Caesar cipher*, which is said to have been used by Julius Caesar to communicate with his army. Caesar is considered to be one of the first persons to have ever employed encryption for the sake of securing messages. Caesar decided that shifting each letter in the message would be his standard algorithm, and so he informed all of his generals of his

decision, and was then able to send them secured messages. The Caesar cipher is a *shift cipher* since the cipher-text alphabet is derived from the plain-text alphabet by shifting each letter a certain number of spaces. To encipher a message, we perform a simple substitution by looking up each of the message's letters in the top row and writing down the corresponding letter from the bottom row [9].

Play Fair Cipher, Despite its invention by Wheatstone, became known as the Play fair cipher after Lord Playfair, who heavily promoted its use. The Playfair cipher uses a 5 by 5 table containing a key word or phrase. Memorization of the keyword and 4 simple rules was all that was required to create the 5 by 5 table and use the cipher. To generate the key table, one would first fill in the spaces in the table with the letters of the keyword (dropping any duplicate letters), then fill the remaining spaces with the rest of the letters of the alphabet in order (usually omitting "Q" to reduce the alphabet to fit, other versions put both "I" and "J" in the same space). The key can be written in the top rows of the table, from left to right, or in some other pattern, such as a spiral beginning in the upper-left-hand corner and ending in the centre. The keyword together with the conventions for filling in the 5 by 5 table constitutes the cipher key [10].

Hill Cipher, In classical cryptography, is a polygraphic substitution cipher based on linear algebra. Invented by Lester S. Hill in 1929, it was the first polygraphic cipher in which it was practical (though barely) to operate on more than three symbols at once. Each letter is first encoded as a number. Often the simplest scheme is used: A = 0, B =1, ..., Z=25, but this is not an essential feature of the cipher. A block of *n* letters is then considered as a vector of n dimensions, and multiplied by an n × n matrix, modulo 26. (If one uses a larger number than 26 for the modular base, then a different number scheme can be used to encode the letters, and spaces or punctuation can also be used.) The whole matrix is considered the cipher key, and should be random provided that the matrix is invertible in $\mathbb{Z}_{26}^{n}$ (to ensure decryption is possible) [12]. A Hill cipher is another way of working out the equation of a matrix [11].

A polyalphabetic cipher is any cipher based on substitution, using multiple substitution alphabets. The Vigenère cipher is probably the best-known example of a polyalphabetic cipher, though it is a simplified special case. The Enigma machine is more complex but still fundamentally a polyalphabetic substitution cipher [13]. Polyalphabetic substitution cipher designers seem to have concentrated on obscuring the choice of a few such alphabets (repeating as needed), not on the increased security possible by using many and never repeating any [14]. The Alberti cipher by Leon Battista Alberti around 1467 was believed to be the first polyalphabetic cipher. Alberti used a mixed alphabet to encrypt a message, but whenever he wanted to, he would switch to a different alphabet, indicating that he had done so by including an uppercase letter or a number in the cryptogram [15].

The one-time pad (OTP) is a type of encryption, which has been proven to be impossible to crack if used correctly. Each bit or character from the plain-text is encrypted by a modular addition with a bit or character from a secret random key (or *pad*) of the same length as the plain-text, resulting in a cipher-text. If the key is truly random, as large as or greater than the plain-text, never reused in whole or part, and kept secret, the cipher-text will be impossible to decrypt or break without knowing the key [16]. The two main problems in one time padding are key distribution and true randomness. Distributing very long one-time pad keys is inconvenient and usually poses a significant security risk. Also high-quality random numbers are difficult to generate. The random number generation functions in most programming language libraries are not suitable for cryptographic use [17].

Figure 3: A simple homophonic substitution cipher

A homophonic cipher is a substitution cipher in which a character may have any of a number of different representations. Figure 1.3 gives one such cipher and a sample message using it [18]. The Homophonic Substitution Cipher involves replacing each letter with a variety of substitutes, the number of potential substitutes being proportional to the frequency of the letter. The point of offering several substitution options for popular letters is to balance out the frequencies of symbols in the cipher-text [19]. It can also be defined as a venerable technique for converting a clear-text sequence into a random sequence [20].

Aryabhata's Mathematics, A small resemblance of this new cipher technique has been found in research paper presented as Aryabhata's Mathematics. Vatsyayana's *Kama-sutra* describes cryptographic writing as *mlecchita-viklapa*. In his commentary on the *Kama-sutra*, Yasodhara describes two kinds of this writing:

- ***Kautiliyam.*** In this letter substitutions are based on phonetic relations, such as vowels becoming consonants. A simplification of this form is called *durbodha*.
- ***Muladeviya*.** Its cipher alphabet consists of pairing letters and using the reciprocal ones:

a kh gh c t ñ n r l y
k g ṅ – p ṇ m ṣ s s

with all the other letters remaining unchanged. These codes can serve as sub-units in cryptographic transformations [8].

3. PROPOSED WORK

It is discussed in following steps.

[***Step I***] This algorithm uses a key which comprises of two parts:

a. A 5-digit key- This part of the key contains numeric digits {1,2,3,4,5} arranged in some random order. This part of the key is used for further arrangements if substitution matrices.
b. A keyword- This part of the key either contains the keyword "**UP**" or "**DOWN**". "**UP**" refers previous and "**DOWN**" refers next.

[***Step II***] The arrangement of the numeric digits, intern forms two more keys, one is the arrangement of vowels {a,e,i,o,u} and the numbers {6,7,8,9,0} in the same sequence as that of the key digits. Ex. If the key digit is "53142", then the other two keys become "u i a o e" and "08697".

[***Step III***] The consonant 'Z' is never included in any encryption-decryption procedure.

[***Step IV***] The existence of keyword "**UP**" indicates that the consonants occurring previous to vowels are not included in matrix formation for encryption-decryption procedure, upon considering the alphabets to be connected in a circular manner. So "**UP**" means that consonants 'Y','D','H','N'&'T' are not included in matrix formations.

[***Step V***] The existence of keyword "**DOWN**" indicates that the consonants occurring next to vowels are not included in matrix formation for encryption-decryption procedure, upon considering the alphabets to be connected in a circular manner. So "**DOWN**" means that consonants 'B','F','J','P'&'V' are not included in matrix formations.

[***Step VI***] The matrices for vowel-to-consonant encryption/decryption are formed as under:

For keyword "UP"

A	**E**	**I**	**O**	**U**
B	C	F	G	J
K	L	M	P	Q
R	S	V	W	X

For keyword "DOWN"

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

[***Step VII***] The above matrices are used is a special way. To avoid same substitution for each vowel, the n^{th} occurrence of a vowel in a plain-text is replaced by n^{th} consonant lying beneath that vowel in its column. That means for keyword "UP", the 1^{st} occurrence of 'a' in plain-text is replaced by the consonant 'b'; 2^{nd} occurrence is replaced by the consonant 'k'; 15^{th} occurrence is replaced by the consonant 'r'; and so on. Similar procedure is followed for keyword "DOWN".

[***Step VIII***] For consonant-to-vowel encryption/decryption, the three keys mutually constructed, comprising of two sets of digits and one set of vowels are used in following way:

For keyword "UP"

B	**C**	**F**	**G**	**J**	**K**	**L**	**M**	**P**	**Q**	**R**	**S**	**T**	**W**	**X**
5	3	1	4	2	0	8	6	9	7	U	I	A	E	O

For keyword "DOWN"

C	**D**	**G**	**H**	**K**	**L**	**M**	**N**	**Q**	**R**	**S**	**T**	**W**	**X**	**Y**
5	3	1	4	2	0	8	6	9	7	U	I	A	E	O

4. RESULTS

Plain-text: meet me after the toga party
Secret key: 53142 UP
Encryption Phase:

Step 1: The occurrence of vowels is numbered as follows:

M	E	E	T	M	E	A	F	T	E	R	T	H	E	T	O	G	A	P	A	R	T	Y
	1	**2**			**3**				**4**				**5**									

M	E	E	T	M	E	A	F	T	E	R	T	H	E	T	O	G	A	P	A	R	T	Y
						1											**2**		**3**			

M	E	E	T	M	E	A	F	T	E	R	T	H	E	T	O	G	A	P	A	R	T	Y
															1							

Step 2: The vowels are converted to consonants as follows:

M	E	E	T	M	E	A	F	T	E	R	T	H	E	T	O	G	A	P	A	R	T	Y
	C	**L**			**S**				**C**				**L**									

M	E	E	T	M	E	A	F	T	E	R	T	H	E	T	O	G	A	P	A	R	T	Y
						B											**K**		**R**			

M	E	E	T	M	E	A	F	T	E	R	T	H	E	T	O	G	A	P	A	R	T	Y
															G							

Step 3: The consonants are converted to vowels as follows:

M	E	E	T	M	E	A	F	T	E	R	T	H	E	T	O	G	A	P	A	R	T	Y
6			**T**	**6**			**1**	**T**		**U**	**T**	**H**		**T**		**4**		**9**		**U**	**T**	**Y**

Step 4: So the cipher-text produced equivalent to plain-text is:

M	E	E	T	M	E	A	F	T	E	R	T	H	E	T	O	G	A	P	A	R	T	Y
6	**C**	**L**	**T**	**6**	**S**	**B**	**1**	**T**	**C**	**U**	**T**	**H**	**L**	**T**	**G**	**4**	**K**	**9**	**R**	**U**	**T**	**Y**

PLAIN-TEXT: MEET ME AFTER THE TOGA PARTY
CIPHER-TEXT: 6CLT 6S B1TCU THL TG4K 9RUTY

Step 5: The above generated cipher-text is sent to the receiver after omitting all the spaces.
Cipher-text sent: **6CLT6SB1TCUTHLTG4K9RUTY**

Decryption Phase:
Cipher-text: 6CLT6SB1TCUTHLTG4K9RUTY
Secret key: 53142 UP

Step 1: The cipher-text is analyzed for consonants first and consonants to vowel conversion is done as follows:

6	C	L	T	6	S	B	1	T	C	U	T	H	L	T	G	4	K	9	R	U	T	Y
	E	**E**			**E**	**A**			**E**				**E**		**O**		**A**		**A**			

Step 2: The cipher-text is the analyzed for vowels and digits to convert them to consonants as follows:

6	C	L	T	6	S	B	1	T	C	U	T	H	L	T	G	4	K	9	R	U	T	Y
M				**M**			**F**			**R**						**G**		**P**		**R**		

Step 3: The cipher-text is the analyzed for those consonants that are not allowed to participate in the encryption/decryption process as follows:

6	C	L	T	6	S	B	1	T	C	U	T	H	L	T	G	4	K	9	R	U	T	Y
			T					**T**			**T**	**H**		**T**							**T**	**Y**

Step 4: The results from above three steps are combined to obtain the plain-text as follows:

6	C	L	T	6	S	B	1	T	C	U	T	H	L	T	G	4	K	9	R	U	T	Y
M	**E**	**E**	**T**	**M**	**E**	**A**	**F**	**T**	**E**	**R**	**T**	**H**	**E**	**T**	**O**	**G**	**A**	**P**	**A**	**R**	**T**	**Y**

*Plain-text computed***:** meetmeafterthetogaparty

The above plain-text obtained at the receiver's end upon decryption can be easily read and understood by the human brain.

5. Conclusion

In this paper after analyzing the different existing symmetric substitution cipher techniques and their disadvantages, The proposed new cipher technique which is quite secure and efficient. The technique is indifferent to Brute Force attack and Relative Frequency attack. This algorithm may hold a wide usage for commercial and low scale purposes depending upon the cost of implementing the algorithm.

FUTURE WORK

With the literature review and study of all substitution techniques known to us, we found some problems that have been solved to some extent by our proposed algorithm. Certain aspects have still been left for future extension, as both the cipher-text and the plain-text skip on blank spaces, so a convention may be developed to handle the occurrence of spaces. The algorithm is based on the position wise occurrence of each vowel, so it may be difficult to keep track of their position in large text. So the usage of this algorithm is confined to small pieces of text. A convention for encrypting numbers and symbols may be developed, since no such mechanism has been implemented presently in this algorithm.

REFERENCES

[1] William Stallings, " Cryptography and Network Security", Fourth Edition, Pearson Education.
[2] http://math.ucsd.edu/~crypto/java/EARLYCIPHERS/Monoalphabetic.html
[3] http:// www.secretcodebreaker.com/scbsolvr.html
[4] http://www.murky.org/blg/2004/09/monoalphabeticsubstitution/everything2.com/title/monoalphabetic+substitution
[6] http:// www.cvni.net/radio/nsnl/nsnl010/nsnl10poly.html
[7] http:// en.wikibooks.org/wiki/Cryptography/Polyalphabetic_substitution
[8] http:// Aryabhata's Mathematics by Subhash Kak, RSA Conference, San Jose, Feb. 13-17, 2006
[9] Chris Savarese and Brian Hart, "The Caesar Cipher", Trinity College,Last updated: Mon, 26 Apr 2010 02:46:57 GMT

[10] Smith, Michael, "Station X: The Codebreakers of Bletchley Park", Channel 4 Books/Macmillan, London) ISBN 0 7522 2189 2, 1998.
[11] Lester S. Hill, "Cryptography in an Algebraic Alphabet",The American Mathematical Monthly, Vol. 36, No. 6. (Jun. - Jul., 1929), pp. 306-312.
[12] http://en.wikipedia.org/wiki/Hill_cipher
[13] http://en.wikipedia.org/wiki/Polyalphabetic_cipher
[14] Helen Fouché Gaines, "Cryptanalysis", Dover. ISBN 0-486-20097-3, 1939.
[15] Leon Battista Alberti, A. Zaccagnini, "A Treatise on Ciphers, trans",Torino 1997.
[16] Erskine, Ralph, "Enigma's Security: What the Germans Really Knew", in "Action this Day", pp 370–386, 2001.
[17] http://en.wikipedia.org/wiki/One-time_pad#Key_distribution
[18] F. A. Stahl, "A homophonic cipher for computational cryptography," afips, pp.565, 1973 Proceedings of the National Computer Conference, 1973
[19] http://www.simonsingh.net/The_Black_Chamber/homophoniccipher.htm
[20] valdemar c. rocha jr, cid b. Dearaújo, "Homophonic Substitution",Departamento de Eletrônica e Sistemas, UFPE, Recife, PE, Brazil.
[21] http://en.wikipedia.org/wiki/Brute_force_attack
[22] http://en.wikibooks.org/wiki/Cryptography/Frequency_analysis

BIOGRAPHY OF AUTHORS

Shobha Vatsa obtained her M.Tech in Computer Engineering from Shobhit University, Meerut (U.P). and MCA from U.P. Technical University, Lucknow, (U.P). She has conducted various events and seminars. She has attended several seminars, workshops and National conferences. Her area of research include Data mining, MANET (Mobile adhoc Network), Network security, Web Crawler, Software Engineering, DBMS.

Avimanyou Kumar Vatsa is working as Assistant Professor and Coordinator - CSE at Shobhit University, Meerut, (U.P.), INDIA. He obtained his M-Tech (Computer Engineering) with Hons. from Shobhit University and B-Tech(I.T.) from V.B.S. Purvanchal University, Jaunpur (U.P.). He has worked as software engineer in software industry. He has been in teaching from more than one decade. During this short period of time, he has been supervised several dissertation of M.Tech. students. He is on the editorial board and reviewers of several international and national journals in networks and security field. He has been member of several academic and administrative bodies. During his teaching he has been coordinated many Technical fests and National Conferences at Institute and University Level. He has attended several seminars, workshops and conferences at various levels. His many papers are published in various national, international journals and conferences. His area of research includes MANET (Mobile Ad-Hoc network), Network Security, Congestion Control and VOIP-SIP (Voice over IP).

Tanmeya Mohan was Bachelor of Technology in Computer Engineering at the School of Computer Engineering & Information Technology, Shobhit University, Meerut. Presently working as software engineer in software industry.