Design of an Effective Substitution Cipher Algorithm for Information Security using Fuzzy Logic

Sojwal S. Kulkarni
alias R.M. Jogdand
Associate Professor Department of Computer Science & Engineering
GIT, Belgaum

Dr. H.M.Rai
Ex-Professor- National Institute of Technology, Kurukshetra

Dr. Sanjay Singla
Associate. Professor, OITM, Hisar

Abstract :- The combination of cryptography and fuzzy logic has emerged as a promising component of information security. Acknowledge based model is introduced in the information security. This method includes a substitution cipher known as polyalphabetic cipher. The information that is trying to send is including the text that is alphabets. The existing Substitution cipher known as Vigenere cipher algorithm become impractical when we encounter with number of repeated ciphertext. So as to overcome this, a knowledge based model is constructed by using Fuzzy Logic. And a cipher text is generated which is close to the ideal line so that we have the frequencies of all alphabets equal or very close to each other then it would be impossible for an attacker to proceed with the frequency attack.

Keywords :- Knowledge based model, Fuzzy Logic, Substitution cipher, Polyalphabetic cipher.

I .INTRODUCTION

An Information security means protecting information and Information system from unauthorized access, use, disclosure, disruption, modification or destruction. An important aspect of Information security and risk management [14] is recognizing the value of Information and defining appropriate procedure and protection required for the information. Not all information is equal and so not all information required the same degree of protection. And it is closely related to the disciplines of cryptology and cryptanalysis. The main objective of the modern cryptography concerns itself with the following four objects i.e. Confidential, Integrity, Authentication and Non repudiation.

Hacking of the information is widely considered as one of the potential attacks on any secure system. So we have to secure the system from such type of vulnerabilities.

As substitution cipher has the disadvantage, so as to overcome this we are using Fuzzy logic in our paper to improve this. Fuzzy logic is the branch of logic in which the truth value of a logical proposition (or set membership) is represented as a real value on unit interval [0, 1]. Fuzzy logic provides means to represent approximate knowledge. It is a logic that arrives at a definite conclusion based on vague, ambiguous, or imprecise input information. When applying mathematical concepts to our daily lives, it is often difficult to adhere to the logical constraints of traditional set theory because of the vagueness of the real world. The logic enables an object to belong to a set with a certain degree; unlike traditional logic it addresses the complexity of the world. It also can be used practically to aid systems in decision making.

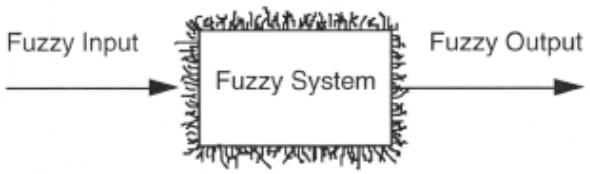

Figure 1.Fuzzy Systems

A. Fuzzy Logic in Cryptography:

Fuzzy logic[10][2] is a form of multi-valued logic derived from fuzzy set theory to deal with reasoning that is appropriate rather than precise. As mentioned above that the existing cipher has the disadvantage of frequency attack so here Fuzzy logic is used to generate the knowledge based model with the help of this a fuzzy encryption table is created by using this cipher text is generated which is close to the ideal line so that we have the frequencies of all alphabets equal or very close to each other then it would be impossible for an attacker to proceed with the frequency attack. Here the attempt has made to make the cipher more secure with the help of Fuzzy logic

II. BACKGROUND

Simple substitution cipher is a well-known crypto-system. It is the simplest form of substitution ciphers. Each symbol in the plaintext maps to a different symbol in the cipher text.

The existing Substitution Cipher Algorithm [17] is practically vulnerable to a technique known as frequency analysis. There are different Substitution cipher techniques are existing now such as- Ceaser cipher, Monoalphabetic cipher, Playfair cipher, Hill cipher. These all ciphers have their own disadvantages. To improve the monoalphabetic substitution cipher Polyalphabetic cipher is used. In this we are focusing on Vigenere cipher. However one can still use frequency analysis to crack the cipher if the message is long enough that is this becomes impractical when we encountered with repeated cipher text. Simple substitution cipher that makes it relatively easy to crypt analyze, is that the language statistics remain unchanged by the encryption process and hence frequency analysis presents a basic tool for breaking ciphers[13] and also some of the genetic algorithms[15] are used to crack the polyalphabetic ciphers. So as to overcome this we have to design a system which should withstand these types of attacks.

III. FREQUENCY ANALYSIS

In frequency attack [17] if the attacker knows the nature of the plain text (e.g.: non-compressed English text), then the analyst can exploit the regularities of the language.

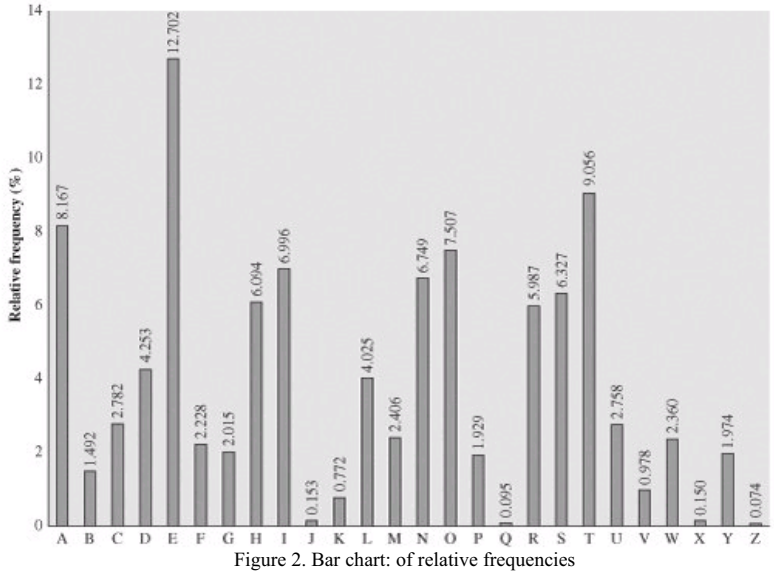

Figure 2. Bar chart: of relative frequencies

To see how such an attack might proceed, we give a partial example here. The cipher text to be solved is:
UZQSOVUOHZMOPVGPOZPEVSGZWSZOPFPESXUDBMETSXAIZ
VUEPHZHMDZSHZOWSFPAPPDTSVPQUZWYMXUZUHSX
EPYEPOPDZSZUFPOMBZWPFUPZHMDJUDTMOHMQ

The relative frequencies of the letters in the cipher text (in percentages) are as follows:

P 13.33	H 5.83	F 3.33	B 1.67	C 0.00
Z 11.67	D 5.00	W 3.33	G 1.67	K 0.00
S 8.33	E 5.00	Q 2.50	Y 1.67	L 0.00
U 8.33	V 4.17	T 2.50	I 0.83	N 0.00
O 7.50	X 4.17	A 1.67	J 0.83	R 0.00
M 6.67				

Table 1: Relative frequencies of the letters in the cipher text (in percentages)

Comparing this breakdown with fig, it seems likely that cipher letters P and Z are the equivalents of plain letters e and t, but it is not certain which is which. The letters S, U, O, M, and H are all of relatively high frequency and probably correspond to plain letters the set {a, h, i, n, o, r, s}. The letters with the lowest frequencies (namely, A, B, G, Y, I, J) are likely included in the set {b, j, k, q, v, x, z}. There are a number of ways to proceed at this point. The attacker could make some tentative assignments and start to fill in the plain text to see if it looks like a reasonable skeleton of a message. A more systematic approach is to look for other regularities. For example, certain words may be known to be in the text. Ideally if we have the frequencies of all alphabets equal or very close to each other then it would be impossible for an attacker to proceed with the frequency attack.

IV. SYSTEM ARCHITECTURE

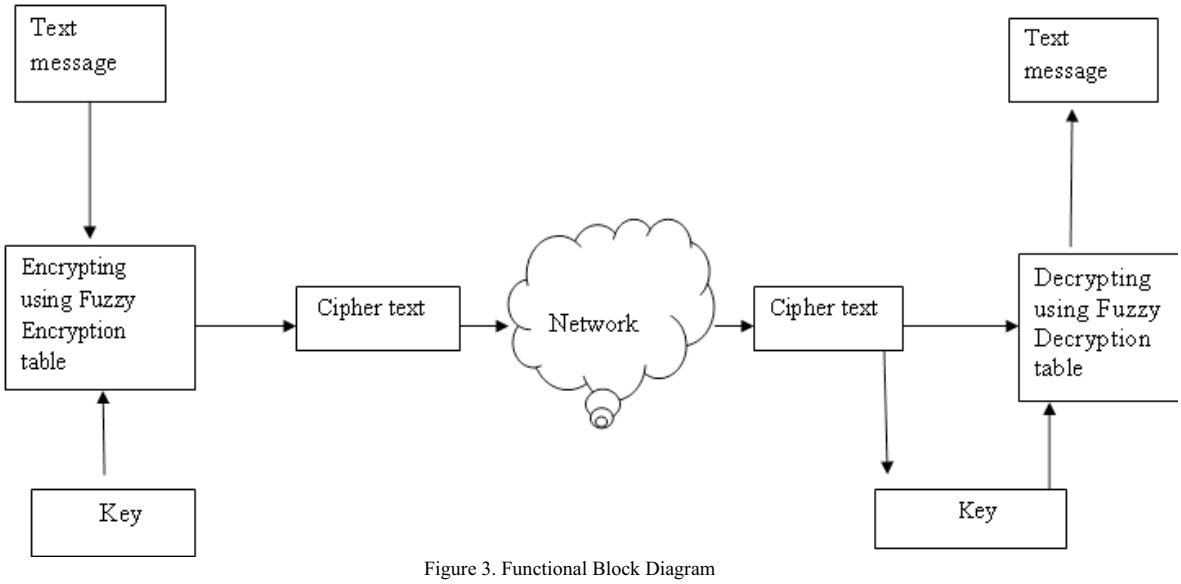

Figure 3. Functional Block Diagram

The original intelligible message or data that is fed into the algorithm as input. The secret key is also input to the encryption algorithm. The key is a value independent of the plaintext. The algorithm will produce a different output depending upon a specific key being used at the time. The exact substitutions and transformations performed by the algorithm i.e. using fuzzy encryption table. Then we get a ciphertext this is the scrambled

message produced as the output. It depends on the plaintext and the secret key. For a given message, two different keys will produce two different cipher texts. The cipher text is an apparently random stream of data and, as it stands, is unintelligible. This is send to the other side and is decrypted by using fuzzy decryption table.

A.Creating Fuzzy Knowledge based model:

Create a Fuzzy knowledgebase which uses the frequency of alphabets appearing in the plain text and assign some predefined values. The entries inside the table are rule based which are found by writing 60 different rules that gives Fuzzy values. This can be done by using Fuzzy Inference System (FIS).

ALPHABETS ↓	RANGE OF FREQUENCIES(PERCENTAGE OF FILE-SIZE) →									
	(0-5)%	(5-10)%	(10-15)%	(15-20)%	(20-25)%	(25-30)%	(30-35)%	(35-40)%	(40-45)%	(45-100)%
Ano % 11	3	8	2	9	4	7	6	5	1	9
Ano % 7	8	7	4	5	9	6	1	2	3	5
Ano % 5	1	2	3	4	5	8	7	7	9	1
Ano % 3	9	7	5	3	1	2	4	6	8	7
Ano % 2	5	4	9	8	6	5	3	1	2	3
Ano % 1	2	9	8	7	8	3	5	3	4	8

(Table 2: Fuzzy Knowledgebase)

B.Generation of Tables:

Generation of Encryption Table:

The Encryption table is a 26x26 integer matrix. Here each row is a mono-alphabetic cipher independent from the other rows and the column number represents the alphabet to be encrypted. Each row has values from 1 to 26 without repetition. The condition used to generate table is that each number occupies different positions in different rows (The most basic rules of Sudoku).

Generation of Decryption Table:

The Decryption table is also a 26x26 matrix. It satisfies all the conditions of the encryption table. Here the mono-alphabetic cipher decryption technique is used.

V. DESIGN AND IMPLEMENTATION

Fuzzy logic incorporates a simple, rule-based IF X AND Y THEN Z approach to a solving control problem rather than attempting to model a system mathematically. The fuzzy logic model is empirically-based, relying on an operator's experience rather than their technical understanding of the system. These terms are imprecise and yet very descriptive of what must actually happen. The design is based on Mamdani-style

inference system which is present in Fuzzy Inference system(FIS).It is very good for the representation of human reasoning and effective analysis. The implementation is done using MATLAB fuzzy logic tools. Fuzzy inference is the process of formulating the mapping from a given input to an output using fuzzy logic. The mapping then provides a basis from which decisions can be made, or patterns discerned. The process of fuzzy inference involves all of the pieces that are: Membership Functions, Logical Operations, and If- Then Rules.

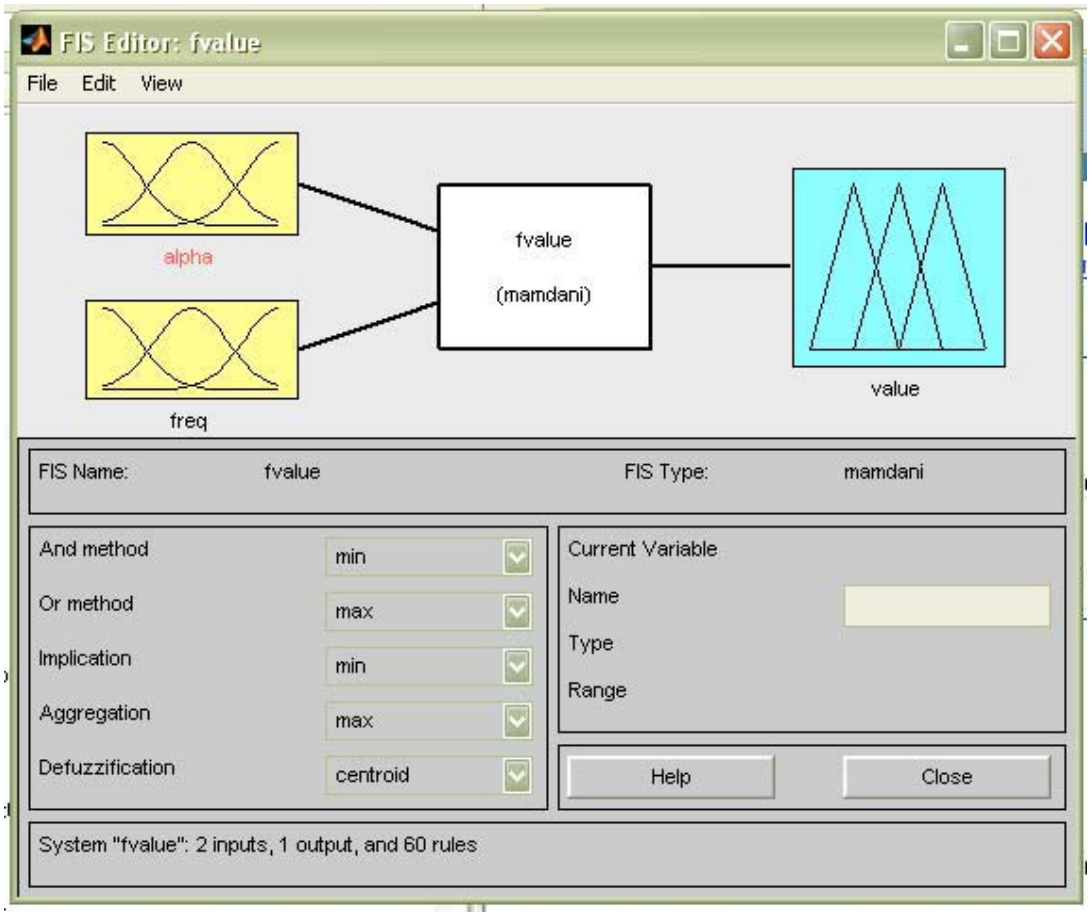

Figure 4. FIS Editor

A. FIS Editor (Figure 4):

This is the window through which a new FIS type with any particular model can be selected, variable can be added, and input or output variable names can be changed. In this case, the chosen model is Mamdani.

Fuzzy inference systems have been successfully applied in fields such as automatic control, data classification, decision analysis, expert systems, and computer vision. Because of its multidisciplinary nature, fuzzy inference systems are associated with a number of names, such as fuzzy-rule-based systems, fuzzy expert systems, fuzzy modeling, fuzzy associative memory, fuzzy logic controllers, and simply (and ambiguously) fuzzy systems.

VI. RESULT

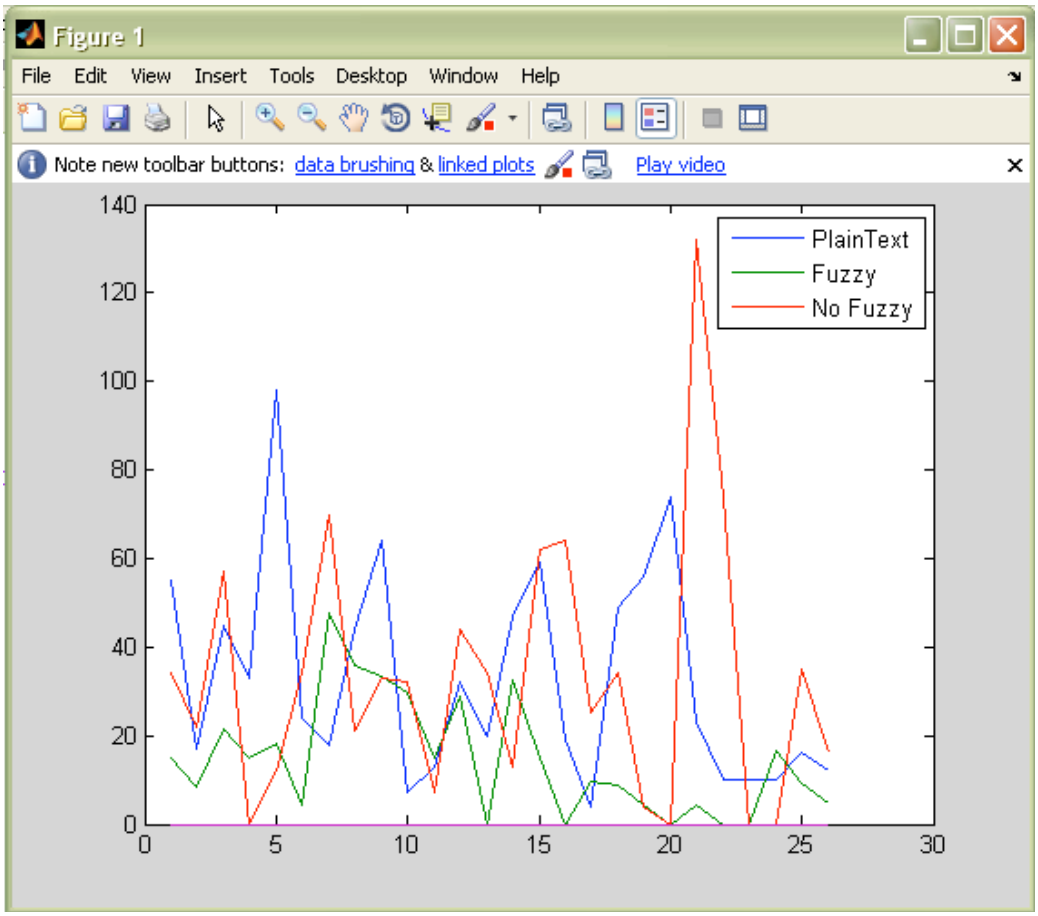

Figure 5. Encrypted Plain text with fuzzy and without fuzzy

The graph shown in figure 5 indicates the frequencies of each alphabet in various cases. Here the frequency range is from 0-140. The graph is plotted for frequency of occurrence of alphabets versus alphabets present in the text. The first graph shows the plaintext which is in blue and the second which is in green shows with fuzzy logic, the frequency distribution of alphabets in this case is less as compare to the plain text and also the algorithm which is generated without using fuzzy logic. The plaintext which is encrypted with Fuzzy logic is different from the original one. The third which is in red is without using fuzzy logic that is occurring somewhat same as that of original.

The result table in Table 3 contains the frequency distribution of plain text and cipher text with fuzzy logic and without fuzzy logic. The plaintext is ranging from 4-98. Next the cipher text is generated by using fuzzy as shown in graph it is different from the plaintext it is ranging from 0- 47 in this case. Here this we have done in MATLAB so the range for ciphertext it is taking automatically i.e. after encryption some of the letters frequencies are doubled. And the cipher text frequency which is for without fuzzy is ranging 0-132.

Plain\cipher text	Plain text frequency	Cipher text frequency without fuzzy	Cipher text frequency with fuzzy
a	55	34	15
b	17	22	8.4
c	45	57	21
d	33	0	15
e	98	12	18
f	24	34	4.4
g	18	70	47
h	44	21	35.6
i	64	33	33.4
j	7	32	29
k	13	7	15
l	32	44	29
m	20	34	0
n	47	13	32.6
o	59	62	15
p	19	64	0
q	4	25	9.7
r	49	34	8.8
s	56	4	4.4
t	74	0	0
u	23	132	4.4
v	10	74	0
w	10	0	0
x	10	0	16.7
y	16	35	9.2
z	12	16	4.8

Table 3. Result 1

Encryption type	Frequency distribution	
	Min	Max
Plain text	4	98
Without fuzzy	0	132
With fuzzy	0	47

Table 4. Result 2

VII. CONCLUSION

With the need for information security in today's world the importance of such encryption algorithms has increased tremendously. So we developed an algorithm which could withstand most of the cryptanalytic attacks. For encryption we used fuzzy logic, by creating a fuzzy encryption table, thus making it robust against cryptanalysis attacks. The frequency distribution range that we are getting by using fuzzy logic which is shown in result tables is good as compare to the frequency distribution of the plain text and also with the algorithm which is generated by without using fuzzy logic. The brute force attack on this algorithm seems to be impossible. Also in case of frequency attacks our results were very close to the ideal one, thus making it susceptible to frequency attacks of any kind.

VIII. FUTURE WORK

The Fuzzy logic in that the knowledge based model has now generated the cipher text which is susceptible to most of the cryptanalytic attack, then our future work will be to apply this model for sending images securely and also for different languages.

REFERENCES

[1]. Al-bahar ,J.F and K.C. Crandall,1990. Systematic risk management approach for construction projects. J. Const.Eng manage. 116: 533-546.
[2]. Amirudd Ismail , Abbas M.Abd and Zamri Bin Chik Approach to analyze Risk Factors for Construction Projects Utilizing Fuzzy logic [Journal of Applied Science 8(20):3738-3742,2008 ISSN 1812-5654 Asian Network for Scientific Information]
[3]. Andi, Santi and darmawan,2006. Identifying and managing important risks in building and infrastructure projects: Contractor perspective. International Civil engg. Practise .August 25-26,2006.
[4]. Ayyub, B. M., McGill, W. L., and Kaminskiy, M. P. (2007). "Critical Asset and Portfolio Risk Analysis: An All Hazards Framework." *Risk Analysis*, 27(4): 789-801.
[5]. Bezdek, J. C., A physical interpretation of fuzzy ISODATA, IEEE Transaction on Systems Man and Cyberentics, 1976, SME-6: 484_492.
[6]. Carr, V and J.H.M. Tah, 2001. A fuzzy approach to construction Project management system. J. Adv. Eng. Software, 32: 847-857
[7]. Chapman, C.B. and Cooper, D.F. (1983), "Risk analysis: testing some prejudices", European Journal of Operational Research, Vol.14, pp.238- 47
[8]. Chen Shouyu, Fuzzy recognition theoretical model, Journal of Fuzzy Mathematics, 1993, (2): 261_269
[9]. D.M. Zhao, Y.Q. Zhang, J.F. Ma, "Comprehensive risk assessment of the network security," *Computer Science*, vol. 31, pp. 66-69, 2004.
[10].Dong-mei Zhao 1, 2, Jing-Hong Wang1, Jian-Feng Ma2 Fuzzy Risk Assessment of the Network Security [Proceedings of the Fifth International Conference on Machine Learning and Cybernetics, Dalian, 13-16 August 2006]
[11]. JHM, Carr V.A proposal for construction project risk assessment using fuzzy logic[J].Construction Management and Econom ics.2000,18:491~500.
[12]. Liou, T.S. and Wang, M.J.J. (1992). Ranking fuzzy numbers with integral value. Fuzzy Sets and Systems, 50(3), 247-255.
[13]. Mohd Zaid Waqiyuddin Mohd Zulkifli "Attacks on Cryptography"April 2008.
[14]. Mustafa M A, FAI-Bahar J. Project risk assessment using the analytic hierarchy process[J].IEEE Transactions on Engineering Management,1991,38(1):46~52.
[15]. Ragheb Toemeh1 and Subbanagounder Arumugam2,Department of Computer Science and Engineering, Government College of Technology, India,Directorate of Technical Education, India."Applying Genetic Algorithms for Searching Key- Space of Polyalphabetic Substitution Ciphers".
[16]. Tuysuz F., Kahraman C., Project risk evaluation using a fuzzy analytic hierarchy process: an application to information technology projects, international journal of intelligent systems (12), 2006. 559-84
[17]. William Stallings- Cryptography and Network Security Principles and practices.
[18]. Zhang Y.Z, B.X. Fang and X.C. Yun, "A risk assessment approach forvnetwork information system," *Proceedings of the Third International*v*Conference on Machine Learning and Cybernetics*, Shanghai, pp. 2949-2952, Aug 26-29, 2004.