A MODIFIED HILL CIPHER ALGORITHM FOR ENCRYPTION OF DATA IN DATA TRANSMISSION

[1]A.V.N.Krishna, [2]Dr. A.Vinaya Babu.
[1]Professor, Indur Institute of Engg. & Tech. Siddipet, Medak Dst. Andhra Pradesh. , INDIA.
mobile : 9849520995. Email : hariavn@yahoo.com
[2]Director, Center for Continues & Distance Education, J.N.T.U, Hyderabad, INDIA

Abstract

One of the recent developments in telecommunication industry is the introduction of communications through Internet. The advantage of this approach is its application in different and diverse fields like e learning, e commerce, e-mail and so on.

*As popularity of electronic mail increases, the problem of privacy and security to the data transmitted also increases. As there are several techniques that provide security to the data transmitted, the most power full technique for data security is data encryption.There are several algorithms for encryption .A new algorithm is going to be discussed in this article will generate a sequence. The algorithm considers a matrix key and executes a sequence of steps, which generates the sequence. This sequence is represented by m*m matrix. The algorithm takes m*m successive plain text letters represent them in matrix form. This matrix plain text is multiplied with generated matrix key to substitute for m*m cipher text letters. Thus the cipher text obtained becomes impossible to break with out knowing the key.*

Key words:
Algorithm, example, security analysis, encryption, and decryption.

1. Introduction:

The use of Internet to provide communications among different levels of people is becoming very popular. As more and more people were using internet for communication transfer, the problem of security is also on the rise. Some one sitting or standing next to receiver can easily see the message. Any intruder can obstruct the message and may see the message.

Computers and Security may be classified as hardware security, software security, database security, network security, communication security and physical security. Different methods are available for computer security. Hardware security can be provided by hardware design. Physical security can be provided by locks, keys, guards etc. Software security can be provided by development standards, operating system protection, and internal program controls. However it is generally agreed that the more powerful tool in providing computer security is encryption . Thus converting the plain text to cipher text, the threat of an interception and the possibility of modification or fabrication are nullified.

2.1 Hill Cipher Operation :

Each letter is treated as a digit in base 26: A = 0, B =1, and so on. A block of n letters is then considered as a vector of n dimensions, and multiplied by a n × n matrix, modulo 26. The components of the matrix are the key, and should be random provided that the matrix is invertible in (to ensure decryption is possible). Consider the message 'ACT', and the key:

Key =[6 24 1;13 16 10;2 17 15];
Plain Text = A C T.

Since 'A' is 0, 'C' is 2 and 'T' is 19, the message is the vector:
Plain Text = 0 2 19

Thus the enciphered vector is given by:

Key *Plain Text=[67 222 319]mod 26 =
[15 14 7];
which corresponds to a ciphertext of 'POH'. Every letter has changed. The Hill cipher has achieved Shannon's diffusion, and an n-dimensional Hill cipher can diffuse fully across n symbols at once.

Decryption
In order to decrypt, we turn the ciphertext back into a vector, then simply multiply by the inverse matrix of the key matrix. We find that in the inverse matrix of the one in the previous example is:
[8 5 10;21 8 21;21 12 8];

Taking the previous example ciphertext of 'POH', we get:
= 0 2 19.
which gets us back to 'ACT'.

One complication is that not all matrices have an inverse. There is a (relatively) straightforward way to find this out, though. If the determinant of the matrix is 0, or has common factors with the modulus (i.e. factors of 2 or 13, in the case of modulus 26), then the matrix cannot be used in the Hill cipher; discard it and try another one. Fortunately, unless the basis has small factors, most matrices will have an inverse. Alas, because 2 is one of the factors of 26, quite a few matrices modulo 26 will not work.

Security

Unfortunately, the basic Hill cipher is vulnerable to a known-plaintext attack because it is completely linear. An opponent who intercepts n2 plaintext/ciphertext character pairs can set up a linear system which can (usually) be easily solved; if it happens that this system is indeterminate, it is only necessary to add a few more plaintext/ciphertext pairs. Calculating this solution by standard linear algebra algorithms then takes very little time.

The security could be greatly enhanced by combining with some non-linear step to defeat this attack. (Perhaps the simplest way to nonlinearise would be to use two different mixed alphabets when converting text to and from numerical vectors.) The combination of wider and wider weak, linear diffusive steps like a Hill cipher, with non-linear substitution steps, ultimately leads to a substitution-permutation network .

Key size

One might naïvely think that the key size, in bits, is n2log226 or about 4.7n2. In fact, it is slightly less than this because not all randomly selected matrices are usable. A slightly less naïve view might guess that 1/2 + 1/26 of candidate keys would be unusable, reducing the keyspace by about 54%. In fact, determinants are not uniformly distributed, and the key space reduction is closer to 70%. Additionally it seems to be prudent to avoid too many zeroes in the key matrix, since they reduce

diffusion. The net effect is that the effective keyspace of a basic Hill cipher is about 4.64n2 - 1.7. For a 5 × 5 Hill cipher, that is about 114 bits. Of course, key search is not the most efficient known attack.

3 A new Algorithm:

In this work a modified hill cipher is presented which executes a sequence of steps to generate the matrix key used in hill cipher algorithm. The new encryption algorithm is based on the concept of Polyalphabet cipher, which is an improvement over monoalphabetic technique.

3.1 The new algorithm has the following features…

1. A mono alphabetic substitution rule is used.
2. A random matrix key is used.
3. The matrix key generates a sequence.
4. This sequence is used to the characters in the plain text by an improvised hill cipher rule.

3.2 The new algorithm is combination of

a) Substitution cipher
b) Matrix key which generates a sequence
c) Modified Hill cipher algorithm
d) Coding method.

The steps that are proposed in the algorithm are

1. The decimal values & letters were given numerical values starting from 0.
2. A random matrix is used as a key. Let it be A.
3. Generate a ternary vector for 3*3 values.
4. Let this be B.
5. Multiply A*B.
6. Consider the mod function of values generated.
7. A sequence is generated.
8. This sequence is represented by m*m matrix form.
9. Individual numerical values of the message are represented by m*m matrix.
10. Matrix form of plain text is multiplied with matrix form of generated sequence.
11. Cipher text is obtained in m*m matrix form which provides necessary security to the data transmitted.

It can be seen that to extract the original information from the coded text is highly impossible for the third person who is not aware of encryption keys and the method of coding.

Even if the algorithm is known it is very difficult to break the code and generate key, given the strength of the algorithm. Thus given a short response time through Internet communication, the algorithm is supposed to be safe.

3. Some of the advantages of this algorithm can be summarized as follows.

a. High speed.It will make users to code the text into cipher text with in a few seconds.
b. It is almost impossible to extract the original information.
c. Even if the algorithm is known, it is difficult to extract the information.
d. Versatile to users: Different users of Internet can use different modified versions of the new algorithm. Since in this algorithm, the sign function is used, it is supposed to be strong enough.
e. As per the matrix the same character is substituted by different alpha numerical values, which provides more security for the message.

4. Example:

n= 0 : 8 ;

r = ternary vector 0:8

n = 0 1 2 3 4 5 6 7 8

r = 0 0 0 0 0 0 0 0 0
0 0 0 1 1 1 2 2 2
0 1 2 0 1 2 0 1 2

A= key at first level.

key= 3 2 -1
3 4 5
2 1 -2

r = mod (A*r,3).

r=r(3,1)+r(2,1)*3+r(1,1)*9;

thus the sequence generated for the ternary vector is

OUTPUT

N = 0 1 2 3 4 5 6 7 8
r = 0 13 11 6 2 12 10 5 1.
Key generated (K) = 0 13 11 6 2 12 10 5 1.
At second level
Encryption.
Plain Text SEND MONEY.
Numerical equivalent 19 5 14 4 13 15 14 5 25
Matrix multiplication. PT*KEY.

PT= |19 5 14|
| 2 0 0 |
|4 13 15 | *
| 0 13 11| mod 26.
KEY= |6 2 12 |
|10 5 1 ||

Numerical equivalent of cipher text
| 16 18 21 |

| 18 10 25 |
| 22 8 9 |

Cipher text P R U R J Y V H I.

Decryption.

Inverse Key : | 2 17 8 |
|10 16 19|
| 12 20 10|

Plain text : (Cipher Text *Key inverse)mod 26.
: [19 5 14; 4 13 15;14 5 25]:
: SEND MONEY.

5. Security analysis:

In the given algorithm a matrix key is used. This matrix key is multiplied with ternary vector. On the product values a mod function is used to generate the sequence. Thus this technique provides the necessary strength to the algorithm because even though the sequence is known, it is very difficult to generate the matrix key. Thus known the algorithm, known the plain text & cipher text pairs, it is quite impossible to generate the random matrix key(first level key). This algorithm provides matrix multiplication at two levels which provides more security to data transmitted. Thus this algorithm is supposed to be safe in real time environment.

Comparative Study.

In the new algorithm a ternary vector is considered. A mod function is used on the product of matrix key with the ternary vector. This mod function is used for the complete sequence. Since the algorithm involves matrix multiplication at two levels & a mod function, the security to encrypted data is very high & also the computational overhead is very low. Where as in public key algorithms like R.S.A. algorithm, large prime numbers are used for more security. An exponential function and a mod function are used to generate cipher text for every character. Thus the computational over head is much more in public key algorithms

Future work

In this work a ternary vector with super script 2 is used. So a 9-digit key is generated. By increasing the superscript to 3,4 and so on, the length of the key could be increased to 27,81 and so on which increases the security of the system still further.

Bibilography.

1. Henry Baker and Fred Piper: Cipher systems (North wood books, London 1982)
2. Thore R.S &Talange D.B. Security of Internet to pager E-mail messages (Internet for India-1997 IEEE Hyderabad section) page No.89-94.
3. William Stalling J.**:** Cryptography and network security (Pearson Education, ASIA1998)
4. Blum.M and Goldwasger.S: An efficient public key encryption scheme which hides all partial information, Advances in Cryptology: proc of crypto ‘ 84, 1985 (spring -verlag), PP 289-299.
5. Blum L., Blum M, Shub M.: A simple unpredictable pseudo random number generator, SIAM J. compute, 1986, 15, (2), pp 364-383.
6. Diffie W & Hellman M.E: New directions in cryptography, IEEE Trans. 1976, IT -22 , pp .644-654.
7. Brossard G.: Modern cryptology, a tutorial lecture Notes computer science, (325), (spring-verlas).
8. Robert W.Baldwin, Victor Chang C.: Locking the e-safe, Spectrum IEEE 1997(pp 40-46).
9. Bruce Scheneier: Applied cryptography (John Wiley & sons (ASIA) Pvt. Ltd.
10. Pandit S.N.N (1961): A New matrix Calculus, J Soc., Ind. And Appl. Math. Pp: 632-637.
11. Krishna A.V.N.: A new algorithm in network security, International Conference Proc. Of CISTM-05, 24-26 July 2005, Gurgoan, India.
12. Krishna A.V.N., Vishnu Vardhan.B.: Utility and Analysis of some Encryption algorithms in E learning environment, International Convention Proc. Of CALIBER 2006, 02-04 Feb. 2006, Gulbarga, India.
13. Maybec.J.S. (1981), Sign Solvability, Proceedings of first symposium on computer assisted analysis and model simplification, Academic Press, NY.
14. PanditS.N.N (1963): Some quantitative combinatorial search problems. (Ph.D. Thesis). 15. http://ardsrk.blogspot.com Provides excellent explanation on computing matrix inverses with regard to the hill cipher
15. 16.Lester S. Hill, Cryptography in an Algebraic Alphabet, The American Mathematical Monthly 36, June-July 1929, pp306–312.
16. 17.Lester S. Hill, Concerning Certain Linear Transformation Apparatus of Cryptography, The American Mathematical Monthly 38, 1931, pp135–154.
17. Jeffrey Overbey, William Traves, and Jerzy Wojdylo, On the Keyspace of the Hill Cipher, Cryptologia, 29(1), January 2005, pp59–72. (PDF)
18. 19.Shahrokh Saeednia, How to Make the Hill Cipher Secure, Cryptologia, 24(4), October 2000, pp353–360.

Article received: 2007-05-21