Hill Cipher algorithm with Self Repetitive Matrix for Secured Data Communication

[1] Prof.Sana F Amin [2] Prof. Nilofar S Hunnergi

[1] Assistant Professor
Ashokrao Mane Group of Institutions, Vathar

[2] Assistant Professor
Sanjay bhokre college of engineering, Miraj

ABSTRACT

The core of Hill-cipher is matrix manipulations It is a multi-letter cipher,for Decryption the inverse of matrix requires and Inverse of the matrix doesn't always exist. Then if the matrix is not invertible then encrypted text cannot be decrypted. However, a drawback of this algorithm is over come by use of self repetitive matrix.This matrix if multiplied with itself for a given mod value (i.e. mod value of the matrix is taken after every multiplication) will eventually result in an identity mat rix after N multiplications. So, after N+ 1 multiplication the matrix will repeat itself. Hence, it derives its name i.e. self repetitive matrix. It should be non singular square matrix.

KEYWORDS- Cryptography , encryption, Decryption, Hill-cipher

1. INTRODUCTION

Cryptography is defined as "the science or study of the techniques of secret writing, esp. code and cipher systems, methods, and the like." Cryptography is needed so that text can be kept secret. It is easy to imagine situations in ancient times where a writer who sent a message via courier would want to make sure that if the runner were intercepted, the interceptors could not read the message.

Recently, the uses of cryptography have grown drastically. Becuase advent of computers and with it the vast amount of information being shared on the internet, there has been a need to create better, more efficient encryption strategies to protect private information, such as credit card numbers, private communications, and so on.

2. PRESENT THEORY & PRACTICES

HILL CIPHER

It is a multi-letter cipher, developed by the mathematician Lester Hill in 1929.For encryption, algorithm takes m successive plaintext letters and instead of that substitutes m cipher letters. In Hill cipher each character is assigned a numerical value

Like:

a=0,

b=1,

…..

…..

z=25.

The substitution of cipher text letters in place of plaintext leads to m linear equations. For m=3, the system can be described as follows:

C_1= (K11P1+K12P2+K13P3) MOD26

C_2= (K21P1+K22P2+K23P3) MOD26

C_3= (K31P1+K32P2+K33P3) MOD26

This can be expressed in terms of column vectors and matrices:

C=KP

Where C and P are column vectors of length 3, representing the plaintext and the cipher text and K is a 3*3 matrix, which is the encryption key. All operations are performed mod 26 here. Decryption requires the inverse of matrix K. The inverse K^{-1} of a matrix K is defined by the equation. K K^{-1}= I where I is the Identity matrix.

K^{-1} is applied to the cipher text, and then the plain text is recovered. In general terms we can write as follows:

For encryption: C= Ek (P) = KP

For decryption: P= Dk (C) =K^{-1} C= K^{-1}KP=P

3. PROPOSED THEORY & PRACTICES

Modification to the Algorithm

As we have seen in Hill cipher decryption, it requires the inverse of a matrix. So while one problem arises that is: Inverse of the matrix doesn't always exist. Then if the matrix is not invertible then encrypted text cannot be decrypted. In order to overcome this problem we suggest the use of self repetitive matrix. This matrix if multiplied with itself for a given mod value (i.e. mod value of the matrix is taken after every multiplication) will eventually result in an identity matrix after N multiplications. So, after N+ 1 multiplication the matrix will repeat itself. Hence, it derives its name i.e. self repetitive matrix. It should be non singular square matrix.

MODIFIED HILL CIPHER ALGORITHM:

This algorithm generates the different key matrix for each block encryption instead of keeping the key matrix constant. This increases the secrecy of data. Also algorithm checks the matrix used for encrypting the plaintext, whether that is invertible or not. If the encryption matrix is not invertible, then the algorithm modifies the matrix such a way that it's inverse exist. The new matrix we obtain after modification of key matrix is called as Encryption matrix and with the help of this matrix encryption operation is performed. In order to generate different key matrix each time, the encryption algorithm randomly generates the seed number and from this key matrix is generated.

Key matrix,

$$K \quad \begin{pmatrix} K_{11} & K_{12} & K_{13} \\ K_{21} & K_{22} & K_{23} \\ K_{31} & K_{23} & K_{33} \end{pmatrix}$$

Where, K_{11} = seed number

$K_{12} =$ (seed number * m) mod n

K_{13} = (12 K * m) mod n

K_{21} = (13 K * m) mod n

. .

K_{33} = (32 *K* * *m)* mod *n*

Where *m* is successive numbers of plaintext letters taken at a time for encryption and *n* is length of the lookup table (total characters used for encryption and decryption) or we can set this *n* value as per requirement. Then with the help of key matrix, encryption matrix *E* is generated.

Steps for encryption matrix generation are as follows:

(1) Check whether the matrix *K* is invertible or not.

(2) If inverse of matrix *K* does not exist, then adjust the diagonal elements (Increment the values of diagonal elements, one element at a time) so that the inverse of the resultant matrix (matrix obtained after changing diagonal elements) is invertible. This matrix becomes the Encryption matrix *E.*

In this algorithm it takes *m* successive plaintext characters and substitutes for then *m* cipher text characters. The substitution is determined by *m* linear equations in which each character is assigned a numerical value (we can take the character's ASCII equivalent number or we can assign a lookup table like *a* = 0, *b* = 1 ... *z* = 25) Here for *m* = 3, the

System can be described as follows:

$C_1 = (E11\ P_1 + E_{12} P_2 + E_{13} P_3) \bmod n$

$C_2 = (E11\ P_1 + E_{12} P_2 + E_{13} P_3) \bmod n$

$C_3 = (E11\ P_1 + E_{12} P_2 + E_{13} P_3) \bmod n$

This case can be expressed in term of column vectors and matrices:

$$\begin{pmatrix} C_1 \\ C_2 \\ C_3 \end{pmatrix} = \begin{pmatrix} E_{11} & E_{12} & E_1 \\ E_{21} & E_{22} & E_{23} \\ E_{31} & E_{32} & E_{33} \end{pmatrix} \begin{pmatrix} P_1 \\ P_2 \\ P_3 \end{pmatrix} \bmod n$$

or $C = EP \bmod n$, where C and P are column vectors of length 3 , representing the Cipher text and plaintext respectively, and E is a 3×3 encryption matrix. All operations are performed mod n.

Steps for Decryption Matrix:

For decryption, from the seed number once again similar way E matrix is generated. Decryption required using the modulo inverse of the matrix E. The inverse E^{-1} of matrix E is defined by the equation

$$E.E^{-1} = E^{-1}.E = I$$

Where I is the matrix that is all zeros expect for ones along the main diagonal from upper left to lower right. Hence decryption matrix D is generated by doing modulo inverse of encryption matrix. Multiply decryption matrix D with received cipher text number vector C and then do modulo operation. Then operate on the output resultant vector, substitute its equivalent characters and which is the plaintext. We can explain this as

Plaintext = $P = D.C = E^{-1}C$. In general, the algorithm can be expressed as follows:

Cipher text = C = EP mod n

Plain text = P = E^{-1}C mod n = E^{-1}EP = P

Generation of a self repetitive Matrix A for a Given N:

The initial conditions for the existence of a self repetitive matrix are:

1.The matrix should be square.

2.It should be non-singular.

But trying to find out the value of N (the value where the matrix becomes a identity matrix) through the method of brute force may not be the best idea always; because the matrix is of dimension greater than 5*5 and with mod index (i.e.) greater than 91 then the brute force technique might take very long time and N value may be in the range of millions. A normal Pentium 4 machine might hang if asked to do the computations for 15*15 matrixes or more. Hence, it would be comfortable to know the value of N and then generate a random matrix accordingly.

This can be done as follows:

1 .First a diagonal matrix A is chosen, and then the values powers of each individual element when they reach unity is calculated and denoted as n1, n2, n3.... Now LCM of these values is taken to given the value of N.

2.Now the next step is generate a random square matrix whose N value is same as the N calculated in the previous step.

3.Pick up any random invertible square matrix B

4.Generate $C=B^{-1}AB$

5.The N value of C is also N

Mathematical proof:

$(B^{-1}AB)^N= (B^{-1})^N *(A)^N*(B)^N$ $A^N=I$
as calculated before as it is a diagonal matrix and N is the LCM of all elements

$(B^{-1}B)*(B^{-1}*B)$.......n times=I

4. RESULTS

Let us see the result with the case study for an array of 5 elements .

Let, m = 5, n = 97 and Seed number S = 141

Then,
K11 = 141
K22 = (K11*2) mod n
. .
. .
K55 = (K44*5) mod n

Hence Key Matrix:

$$K = \begin{matrix} 141 & 0 & 0 & 0 & 0 \\ 0 & 90 & 0 & 0 & 0 \\ 0 & 0 & 78 & 0 & 0 \\ 0 & 0 & 0 & 24 & 0 \\ 0 & 0 & 0 & 0 & 24 \end{matrix}$$

Consider the plaintext to be encrypted is “event”. Letters of the plaintext are represented by their equivalent number vector (30 47 30 39 45)

ENCRYPTION MATRIX

Then with the help of key matrix, encryption matrix is generated. Encryption matrix we get as

$$K = \begin{matrix} 62 & 0 & 0 & 0 & 0 \\ 0 & 6 & 0 & 0 & 0 \\ 0 & 0 & 22 & 0 & 0 \\ 0 & 0 & 0 & 35 & 0 \\ 0 & 0 & 0 & 0 & 35 \end{matrix}$$

Then, Cipher text for the plaintext is [17 88 78 7 23]

Decryption is done by doing inverse method of above and the cipher text is converted to the original as “event”

$$\begin{matrix} 30 \\ 47 \\ 30 \\ 39 \\ 45 \end{matrix} = \begin{matrix} 36 & 0 & 0 & 0 & 0 \\ 0 & 81 & 0 & 0 & 0 \\ 0 & 0 & 75 & 0 & 0 \\ 0 & 0 & 0 & 61 & 0 \\ 0 & 0 & 0 & 0 & 61 \end{matrix} \quad \begin{matrix} 17 \\ 88 \\ 78 \\ 7 \\ 23 \end{matrix} \quad \text{mod (97)}$$

Decrypted Plain text output.

Thus replacing the vector numbers (30 47 30 39 45) by their ASCII values we get the word “event”.

Avalanche Effect:

Here avalanche parameter is used to evaluate performance of the proposed system. The proposed system is built on Mat lab platform, Comparison of results performed between proposed algorithm and two existing algorithms. At the time of results evaluation, plain text and key value both were written randomly. To calculate total effect proposed system run so many times on

Different-different text file with the same key value and then final results were observed. Thus in the case of avalanche effect evaluation, after running proposed systems several times, the final results are the same i.e. in numeric form.

1. **Avalanche Effect Comparisons:** Avalanche Effect is the important property in cryptographic algorithms where, if an input is changed slightly (changing a single bit) the output changes significantly. In our case, we have chosen two different input plain text as "**welcometomycolle**" and "**welcometomycollewelcometomycolle**" Because of key length used in existing algorithm, "proposed algorithm" and "A Block Cipher Having a Key on One Side of the Plain Text Matrix and its Inverse on the Other Side" are using 128 bits key length where "A Modified Hill Cipher Involving a Pair of Keys and a Permutation" is using 256 bits key length. And proposed algorithm is using 56 bit key length.

a) A Block Cipher Having a Key on one side of the Plain Text Matrix and its Inverse on the other side

Plain Text: welcometomycolle

0000110000000111000001010101000100000010011000000000000000100110000000101000100011010110000010100000000000001111000010000000000000000

Change in Plain Text: welcometomycolla

00001000000010100000101010100110000000001010100010000000000110000000011000100111011100000110100000000000000111000011000000010000000

Avalanche Effect: 18

b) A Modified Hill Cipher Involving a Pair of Keys and a Permutation

Plain Text: welcometomycollewelcometomycolle

000000000000000000000000000000000000000000000000000000001000000000000000000000000000000010000000000000110100000000000000000000010011000000000010100110000000001010110000000000001000000010001101000110001000001000000001101000000101110010100101111100101100 10

Change in Plain Text: welcomemycollewelcometomycolla

1000000100000001000000010000000100001110000000000111000110000011100010100011000110
0000011111000010011111111001101000000000000000000000001001100000000000101001101 10
0001110101100000011111100000000111111101000110001011001011110001101001111101 1100
10111111111110010110010

C) Proposed Algorithm

Plain Text: welcometomycollegewe

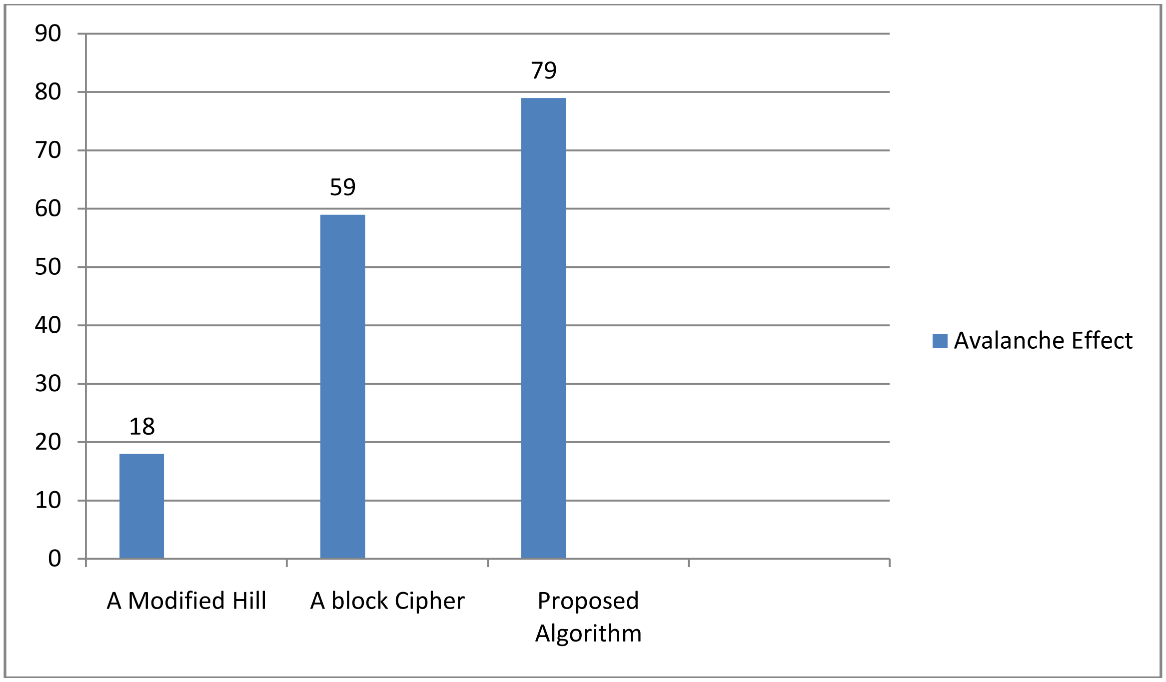
0110000101010110011100110001000000000101010010011001000000010001010000000000101
0110010001000101110001000000010100001100110101100000000101100000001010101000001
0100000000000000000000000000000000000000000000000000000000000000000000000000000000
00

Change in Plain Text: walcometomycollegewe

1000000100100001000000100011100010100010001110010000101010001000100010001000101
0100010011011001100001010000010001001000010100011001100001010000111001100100 01
0001100101010001010101000001010000000000000000000000000000000000000000000000000000
000000000000000000000000000000

Avalanche Effect: 79

ENCRYPTION TECHNIQUE	AVALANCHE EFFECT
"A Modified Hill Cipher Involving a Pair of Keys and a Permutation"	18
"A Block Cipher Having a Key on One Side of the Plain Text Matrix and its Inverse on the Other Side".	59
"Proposed Algorithm"	79



X Axis -Selected Algorithms, Y Axis - Bit Difference

Figure : Avalanche effect

5. CONCLUSION

This modified Hill Cipher method is easy to implement and difficult to crack.

- The cipher is considered secure, as it supports strong substitution techniques along with modular arithmetic.
- The block size which is specified as 64 bit is expandable as per requirement, thus gives flexibility in message string length.
- Possible ASCII printable character keys are 95^7 and key combinations are 2^{56}.
- As per our findings time required checking all possible keys at 50 billion keys per second for a 56 bit key: would approximately be 400 days.
  - The above performance will be appropriate for the following kind of applications
    1) In ATMs for pin numbers to maintain its secrecy and security of ATM card.

2) In Email applications for military and civilian purpose where security is of prime importance in terms of records and authentication of messages.
3) In SMS services, e-commerce, pay TV, computer passwords and touches many aspects of our daily lives.

REFERENCES

1. Blakley, G.R.; Twenty years of cryptography in the open literature Security and Privacy, 1999. Proceedings of the 1999 IEEE Symposium on Digital Object Identifier: 10.1109/SECPRI.1999.766903
2. Secure Hill cipher modifications and key exchange protocol ,Mahmoud, Ahmed Y.; Chefranov, Alexander G.; Automation Quality and Testing Robotics (AQTR), 2010 IEEE International Conference on Volume: 2 Digital Object Identifier.
3. The Design of Boolean Functions by Modified Hill Climbing Method ,Izbenko, Y.; Kovtun, V.; Kuznetsov, A.; Information Technology: New Generations, 2009. ITNG '09. Sixth International Conference on ,Digital Object Identifier: Publication Year: 2009 , Publication Year: 2010
4. A secure variant of the Hill Cipher ,Toorani, M.; Falahati, A.; on, Digital Object Identifier: 10.1109/ISCC.2009.5202241 ,Publication Year: 2009.
5. Invertible, Involutory and Permutation Matrix Generation Methods for Hill Cipher System ,Acharya, B.; Jena, D.; Patra, S.K.; Panda, G.; Advanced Computer Control, 2009. ICACC '09. International Conference on ,Digital Object Identifier: Publication Year: 2009.
6. Bruce Schneir: Applied Cryptography, 2nd edition, John Wiley & Sons, 1996
7. Piper,F "Encryption". Security and Detection, Ecos 97. European Conference;
8. Abrams, M., and Podell, H. "Cryptography" Potentials, IEEE Page No 36-38. Issue:1,Volume: 20, Feb-Mar,2001
9. Eskiciogiu,A. Litwin,L " Cryptography and Network Security" LOS Alamitos,CA:IEEE computer society press,1987
10. Garfinkel, S.L; "Public Key Cryptography" , Computer, IEEE, Volume: 29, Issue:6, June 1996.

11. W.Diffie; M.E.Hell man, “ New Directions in Cryptography” IEEE Transactions Information Theory, Nov, pp 644-654
12. E.Biham and A.Shmir; “Differential C for Cryptanalysis of the Encryption Standard”; Springer- Verilag,1993