NOVEL MODIFIED HILL CIPHER ALGORITHM

Bibhudendra Acharya, Girija Sankar Rath, and Sarat Kumar Patra
Department of Electronics and Communication Engineering
National Institute of Technology Rourkela, Orissa-769008, India
bibhudendra@gmail.com, {gsrath, skpatra}@nitrkl.ac.in

ABSTRACT
The Hill cipher algorithm is one of the symmetric key algorithms that have several advantages in data encryption. However, a drawback of this algorithm is that the inverse of the matrix used for encrypting the plaintext does not always exist. So, if the matrix is not invertible, the encrypted text cannot be decrypted. This paper presents a variant of the Hill cipher that overcomes this disadvantage. The proposed technique adjusts the encryption key to form a different key for each block encryption. The proposed variant yields higher security compared to the original one. Also in this paper, a method of generating self-invertible matrix for Hill Cipher algorithm has been proposed. In the self-invertible matrix generation method, the matrix used for the encryption is itself self-invertible. So, at the time of decryption, we need not to find inverse of the matrix. Moreover, this method eliminates the computational complexity involved in finding inverse of the matrix while decryption.

KEY WORDS
Hill Cipher, Encryption, Decryption, Self-invertible matrix, modified Hill Cipher.

1. Introduction

The desire to transmit messages securely is not new. For centuries, people have had a need to keep their communications private. Today, in the Information Age, as the Internet and other forms of electronic communication become more prevalent, electronic security is becoming increasingly important. Cryptography, the science of encryption, plays a central role in mobile phone communications, pay-TV, e-commerce, sending private emails, transmitting financial information, security of ATM cards, computer passwords and touches on many aspects of our daily lives [1]. Cryptography is the art or science encompassing the principles and methods of transforming an intelligible message (plaintext) into one that is unintelligible (ciphertext) and then retransforming that message back to its original form. In modern times, cryptography is considered to be a branch of both mathematics and computer science, and is affiliated closely with information theory, computer security, and engineering [2].

Cryptography systems can be broadly classified into: symmetric and asymmetric. Symmetric cryptosystems use the same key (the secret key) to encrypt and decrypt a message, and asymmetric cryptosystems use one key (the public key) to encrypt a message and a different key (the private key) to decrypt it. Asymmetric cryptosystems are also called public key cryptosystems. Symmetric encryption is referred to as conventional encryption or single key encryption. Conventional encryption can be further divided into categories of classical techniques and modern techniques. The hallmark of conventional encryption is that the cipher or key to the algorithm is shared, i.e., known by the parties involved in the secured communication. Substitution Cipher is one of the basic components of classical ciphers. A substitution cipher is a method of encryption by which units of plaintext are substituted with ciphertext according to a regular system; the units may be single letters (the most common), pairs of letters, triplets of letters, mixtures of the above, and so forth. The receiver deciphers the text by performing an inverse substitution [3]. The units of the plaintext are retained in the same sequence in the ciphertext, but the units themselves are altered. There are a number of different types of substitution cipher. If the cipher operates on single letters, it is termed a simple substitution cipher; a cipher that operates on larger groups of letters is termed polygraphic. A monoalphabetic cipher uses fixed substitution over the entire message, whereas a polyalphabetic cipher uses a number of substitutions at different times in the message— such as with homophones, where a unit from the plaintext is mapped to one of several possibilities in the ciphertext. Hill cipher is a type of monoalphabetic polygraphic substitution cipher.

Hill cipher is a block cipher that has several advantages such as disguising letter frequencies of the plaintext, its simplicity because of using matrix multiplication and inversion for enciphering and deciphering, its high speed, and high throughput [4]. But the drawback of this algorithm is that the inverse of the matrix used for encrypting the plaintext does not always exist. So, if the matrix is not invertible, the encrypted text cannot be decrypted. Moreover, Hill cipher can be easily broken with a known plaintext attack revealing weak security. This paper presents a variant of the Hill cipher that overcomes these disadvantages. This Modified Hill Cipher Algorithm, initially 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. To overcome the weak security of the Hill algorithm, the proposed technique adjusts the encryption key to form a different key for each block encryption. Also in this paper, a method of generating self-invertible matrix for Hill Cipher algorithm has been proposed. In the self-invertible matrix generation method, the matrix used for the encryption is itself self-invertible. So, at the time of decryption, we need not to find inverse of the matrix. Moreover, this method

eliminates the computational complexity involved in finding inverse of the matrix while decryption.

The paper is organized as follows. Following the introduction, the basic concept of Hill Cipher is outlined in section 2. Section 3 discusses about the modular arithmetic. In section 4, proposed Modified Hill Cipher Algorithm is presented. Finally, section 5 describes the concluding remarks.

2. Hill Cipher

Hill ciphers are an application of linear algebra to cryptology. It was developed by the mathematician Lester Hill. The Hill cipher algorithm takes m successive plaintext letters and substitutes m ciphertext letters for them. The substitution is determined by m linear equations in which each character is assigned a numerical value ($a = 0, b = 1, ..., z = 25$). Let m be a positive integer, the idea is to take m linear combinations of the m alphabetic characters in one plaintext element and produce m alphabetic characters in one ciphertext element. Then, an $m \times m$ matrix A is used as a key of the system such that A is invertible modulo 26 [5]. Let a_{ij} be the entry of A. For the plaintext block $x = (x_1, x_2, ..., x_m)$ (the numerical equivalents of m letters) and a key matrix A, the corresponding ciphertext block $y = (y_1, y_2, ..., y_m)$ can be computed as

Encryption:

$$(y_1, y_2, ..., y_m) = (x_1, x_2, ..., x_m)\, A \pmod{26}, \quad \ldots (1)$$

Where

$$A = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1m} \\ a_{21} & a_{22} & \ldots & a_{2m} \\ \ldots & \ldots & \ldots & \ldots \\ a_{m1} & a_{m2} & \ldots & a_{mm} \end{bmatrix}$$

The ciphertext is obtained from the plaintext by means of a linear transformation.

Decryption:

The reverse process, deciphering, is computed by

$$(x_1, x_2, ..., x_m) = (y_1, y_2, ..., y_m)\, A^{-1} \pmod{26}, \quad \ldots (2)$$

Where

$$A^{-1} = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1m} \\ a_{21} & a_{22} & \ldots & a_{2m} \\ \ldots & \ldots & \ldots & \ldots \\ a_{m1} & a_{m2} & \ldots & a_{mm} \end{bmatrix}^{-1} \pmod{26}$$

Since the block length is m, there are 26^m different m letters blocks possible, each of them can be regarded as a letter in a 26^m-letter alphabet. Hill's method amounts to a monoalphabetic substitution on this alphabet [6].

3. Modular Arithmetic

The arithmetic operation presented here are addition, subtraction, unary operation, multiplication and division. Based on this the self invertible matrix for Hill cipher algorithm is generated. The congruence modulo operator has the following properties [7]:

1. $a \equiv b \mod p$ if $n \mid (a - b)$
2. $(a \bmod p) = (b \bmod p) \Rightarrow a \equiv b \bmod p$
3. $a \equiv b \bmod p \Rightarrow b \equiv a \bmod p$
4. $a \equiv b \bmod p$ and $b \equiv a \bmod p \Rightarrow a \equiv c \bmod p$

Let $Z_p = [0, 1, ..., p - a]$ the set residues modulo p. If modular arithmetic is performed within this set Z_p, the following equations present the arithmetic operations:

1. Addition:
$(a + b) \bmod p = [(a \bmod p) + (b \bmod p)] \bmod p$
2. Negation:
$-a \bmod p = p - (a \bmod p)$
3. Subtraction:
$(a - b) \bmod p = [(a \bmod p) - (b \bmod p)] \bmod p$
4. Multiplication:
$(a * b) \bmod p = [(a \bmod p) * (b \bmod p)] \bmod p$
5. Division:
$(a / b) \bmod p = c$ when $a = (b * c) \bmod p$

The following Table I exhibits the properties of modular arithmetic.

TABLE I
PROPERTIES OF MODULAR ARITHMETIC

Commutative Law: $(\omega + x) \bmod p = (x + \omega) \bmod p$ $(\omega * x) \bmod p = (x * \omega) \bmod p$
Associative Law: $[(\omega + x) + y] \bmod p = [\omega + (x + y)] \bmod p$
Distribution Law: $[\omega * (x + y)] \bmod p = [\{(\omega * x) \bmod p\} * \{(\omega * y) \bmod p\}] \bmod p$
Identities: $(0 + a) \bmod p = a \bmod p$ $(1 * a) \bmod p = a \bmod p$
Inverses: For each $x \in Z_p$, there exists y such that $(x + y) \bmod p = 0$ then $y = -x$ For each $x \in Z_p$ there exists y such that $(x * y) \bmod p = 1$

4. Proposed 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.

$$\text{Key matrix, } K = \begin{bmatrix} K_{11} & K_{12} & K_{13} \\ K_{21} & K_{22} & K_{23} \\ K_{31} & K_{32} & K_{33} \end{bmatrix}$$

Where, K_{11} = seed number

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

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

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

. .

. .

. .

K_{33} = (K_{32} * 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 ciphertext 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 $a = 0, b = 1 \ldots z = 25$). Here for m = 3, the system can be described as follows:

$$\begin{aligned} C_1 &= (E_{11}P_1 + E_{12}P_2 + E_{13}P_3) \bmod n \\ C_2 &= (E_{21}P_1 + E_{22}P_2 + E_{23}P_3) \bmod n \quad \ldots (3) \\ C_1 &= (E_{31}P_1 + E_{32}P_2 + E_{33}P_3) \bmod n \end{aligned}$$

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

$$\begin{pmatrix} C_1 \\ C_2 \\ C_3 \end{pmatrix} = \begin{bmatrix} E_{11} & E_{12} & E_{13} \\ E_{21} & E_{22} & E_{23} \\ E_{31} & E_{32} & E_{33} \end{bmatrix} \begin{pmatrix} P_1 \\ P_2 \\ P_3 \end{pmatrix} \bmod n$$

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

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

$$E.E^{-1} = E^{-1}.E = I \quad \ldots (4)$$

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 ciphertext 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

$$P = D.C = E^{-1}C \quad \ldots (5)$$

In general, the algorithm can be expressed as follows:

$$C = EP \bmod n \quad \ldots (6)$$

$$P = E^{-1}C \bmod n = E^{-1}EP = P \quad \ldots (7)$$

The flowcharts for the encryption & decryption methods are represented in figures 1 & 2.

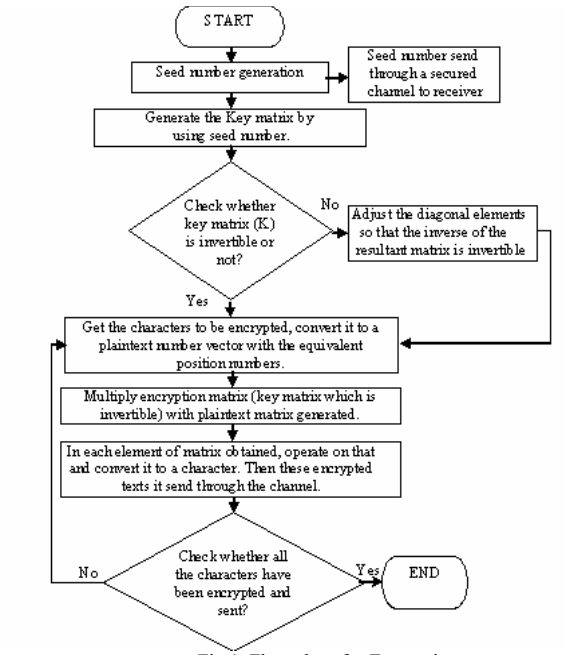

Fig 1. Flow chart for Encryption

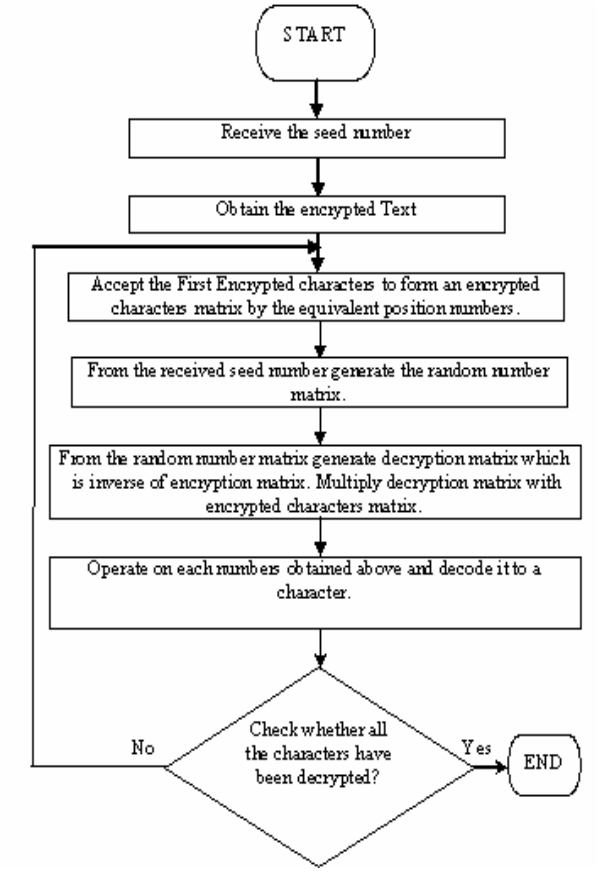

Fig 2. Flow chart for Decryption

Example: (For Modulo 257)

Let $m = 3$, $n = 257$and Seed number $S = 182$

Then,

$$\begin{aligned} K_{11} &= 182 \\ K_{12} &= 32 \\ K_{13} &= 96 \\ &\vdots \\ K_{33} &= 80 \end{aligned}$$

Hence key matrix:

$$K = \begin{bmatrix} 182 & 32 & 96 \\ 31 & 93 & 22 \\ 66 & 198 & 80 \end{bmatrix}$$

Consider the plaintext to be encrypted is "ram". Letters of the plaintext are represented by their ASCII equivalent number vector $(114 \quad 97 \quad 109)$.

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

$$E = \begin{bmatrix} 103 & 52 & 156 \\ 211 & 120 & 100 \\ 43 & 129 & 131 \end{bmatrix}$$

Then, $E(114 \quad 97 \quad 109) = (33790 \quad 46594 \quad 31694) \bmod 257$
$= (123 \quad 77 \quad 83)$ = {MS. Ciphertext for the plaintext is "{MS".

Decryption requires using the matrix D is generated by doing modulo inverse of encryption matrix.

$$D = \begin{bmatrix} 222 & 254 & 248 \\ 230 & 1 & 0 \\ 42 & 0 & 1 \end{bmatrix} \text{ and}$$

$$E = \begin{bmatrix} 103 & 52 & 156 \\ 211 & 120 & 100 \\ 43 & 129 & 131 \end{bmatrix} \begin{bmatrix} 222 & 254 & 248 \\ 230 & 1 & 0 \\ 42 & 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 41378 & 26214 & 25700 \\ 78642 & 53714 & 52428 \\ 44718 & 11051 & 10795 \end{bmatrix} \bmod 257 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Multiplying decryption matrix with encrypted number vector $(123 \quad 77 \quad 83)$ and then doing modulo operation, the resultant output vector we will get $(114 \quad 97 \quad 109)$. Then operating on the output resultant vector, substitute its equivalent characters and which is the decrypted plaintext output "ram".

5. Self-invertible Matrix Generation Method:

As Hill cipher decryption requires inverse of the matrix, we suggest the use of self-invertible matrix generation method while encryption in the Hill Cipher. In the self-invertible matrix generation method, the matrix used for the encryption is itself self-invertible. So, at the time of decryption, we need not to find inverse of the matrix. A general method of generating self-invertible matrix is

Let $A = \left[\begin{array}{c|ccc} a_{11} & a_{12} & \dots & a_{1n} \\ \hline a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{array}\right]$ be an $n \times n$ self-invertible matrix partitioned to $A = \left[\begin{array}{c|c} A_{11} & A_{12} \\ \hline A_{21} & A_{22} \end{array}\right]$

A_{11} is a 1×1 matrix $= [a_{11}]$,

A_{12} is a $1 \times (n-1)$ matrix $= [a_{12} \quad a_{13} \dots \quad a_{1n}]$

A_{21} is a $(n-1)\times 1$ matrix $= \begin{bmatrix} a_{21} \\ a_{31} \\ \dots \\ a_{n1} \end{bmatrix}$, and

A_{22} is a $(n-1)\times(n-1)$ matrix $= \begin{bmatrix} a_{22} & a_{23} & \dots & a_{2n} \\ a_{32} & a_{33} & \dots & a_{3n} \\ \dots & \dots & \dots & \dots \\ a_{n2} & a_{n3} & \dots & a_{nn} \end{bmatrix}$

So, $A_{12}\ A_{21} = I - {A_{11}}^2 = 1 - {a_{11}}^2$ … (8)

and $A_{12}\left(a_{11}I + A_{22}\right) = 0$ … (9)

Also, $a_{11} = -$ (of the Eigen values of A_{22} other than 1)

Since $A_{21}A_{12}$ is a singular matrix having the rank 1

and $A_{21}A_{12} = I - A_{22}^2$ … (10)

So, ${A_{22}}^2$ must have rank of $(n-2)$ with Eigen values +1 of $(n-2)$ multiplicity.

Therefore, A_{22} must have Eigen values ± 1.

It can also be proved that the consistent solution obtained for element A_{21} & A_{12} by equating (10) term by term will also satisfy the equation (8).

Algorithm:

i. Select A_{22}, a non-singular $(n-1)\times(n-1)$ matrix which has $(n-2)$ number of Eigen values of either +1 or −1 or both.
ii. Determine the other Eigen value λ of A_{22}.
iii. Set $a_{11} = -\lambda$.
iv. Obtain the consistent solution of all elements of A_{21} & A_{12} by using (10).
v. Formulate the matrix.

Example: (For Modulo 13)

Let $A_{22} = \begin{bmatrix} 9 & 6 & 10 \\ 12 & 10 & 2 \\ 5 & 3 & 4 \end{bmatrix}$ which has Eigen values $\lambda = \pm 1,\ 10$

So, $A_{11} = [3]$,

and one of the consistent solutions of $A_{12} = [11 \quad 9 \quad 4]$, and

$A_{21} = \begin{bmatrix} 10 \\ 2 \\ 5 \end{bmatrix}$.

So, $A = \begin{bmatrix} 3 & 11 & 9 & 4 \\ 10 & 9 & 6 & 10 \\ 2 & 12 & 10 & 2 \\ 5 & 5 & 3 & 4 \end{bmatrix}$.

Another consistent solution of $A_{12} = [1 \quad 2 \quad 11]$ and $A_{21} = \begin{bmatrix} 6 \\ 9 \\ 3 \end{bmatrix}$. So, $A = \begin{bmatrix} 3 & 1 & 2 & 11 \\ 6 & 9 & 6 & 10 \\ 9 & 12 & 10 & 2 \\ 3 & 5 & 3 & 4 \end{bmatrix}$.

6. Conclusion

This paper presents a symmetric cipher that is actually a variation of the Hill cipher. The proposed algorithm is called Modified Hill Cipher Algorithm. This algorithm eliminates the drawback of using a random key matrix in Hill cipher algorithm for encryption, where we may not be able to decrypt the encrypted message, if the matrix is not invertible. As this algorithm uses a different key for each block encryption thereby significantly increases its resistance to various attacks. Also this paper suggests efficient methods for generating self-invertible matrix for Hill Cipher algorithm. These methods encompass less computational complexity as inverse of the matrix is not required while decrypting in Hill Cipher. These proposed methods for generating self-invertible matrix can also be used in other algorithms where matrix inversion is required.

References

[1] G.R. Blakley, Twenty years of cryptography in the open literature, Security and Privacy 1999, *Proceedings of the IEEE Symposium*, 9-12 May 1999.

[2] H. Imai, G. Hanaoka, J. Shikata, A. Otsuka, A.C. Nascimento, Cryptography with information theoretic security", Information Theory Workshop, 2002, *Proceedings of the IEEE*, 20-25 Oct 2002.

[3] A. J. Menezes, P.C. Van Oorschot, S.A. Van Stone, *Handbook of applied cryptography* (CRC press, 1996).

[4] J. Overbey, W. Traves, J. Wojdylo, On the keyspace of the Hill cipher. *Cryptologia*, 29(1), 2005, 59-72.

[5] K. Petersen, Notes on number theory and cryptography, 2000. Http://www.math.unc.edu/ Faculty/petersen/Coding/cr2.pdf.

[6] Barr T.H., Invitation to cryptography (Prentice Hall, 2002)

[7] W. Stallings, *Cryptography and network security* (4th edition, Prentice Hall, 2005).