**Journal of Computer Science 10 (2): 251-254, 2014**
ISSN: 1549-3636
© 2014 Science Publications
doi:10.3844/jcssp.2014.251.254 Published Online 10 (2) 2014 (http://www.thescipub.com/jcs.toc)

# A HYBRID CRYPTOSYSTEM USING VARIABLE LENGTH SUB KEY GROUPS AND BYTE SUBSTITUTION

**[1]Adi Narayana Reddy Kaipa, [2]Vishnu Vardhan Bulusu,**
**[3]Ramakrishna Reddy Koduru and [4]Durga Prasad Kavati**

[1]Department of CSE, HITAM, Hyderabad, India
[2]Department of IT, JNTU College of Engineering, Jagityala, India
[3]Dpartment of Physics, QIS College of Engg. and Tech, Ongole, India
[4]Department of IT, BVRIT, Narsapur, India

Received 2013-10-07; Revised 2013-10-28; Accepted 2013-11-13

## ABSTRACT

Hill cipher is developed using simple linear transformation. It is vulnerable to known-plaintext attack and there exist several methods in the direction of linear transformation to overcome the problems. HCM-EE is an improved method using Eigen value but it is linear and too many mathematical operations. In this study an attempt has been made to introduce nonlinearity to the linear transformation based cryptosystem using byte substitution over GF ($2^8$) and a variable length sub key groups. The performance evaluation of the method is also studied and presented.

**Keywords:** Hill Cipher, Sub Key Groups, GF ($2^8$), Byte Substitution

## 1. INTRODUCTION

The secured transmission of information between the participants can be achieved using encryption and decryption techniques, which converts intelligible information into unintelligible form and vice versa. There are number of algorithms that are existed to provide secured transformation of data, but the efficiency of the algorithm is one of the most important aspects to be studied.

Hill (1929) is developed based on simple linear transformation. It is easy to implement and is of high speed and high throughput. But it is vulnerable to known plaintext attack and the inverse of every shared key matrix may not exist all the time. It is a simple traditional symmetric key cipher algorithm. In Hill cipher, the plaintext to be transmitted through the communication channel is partitioned into 'm' groups, each of size 'n' and these partitioned groups are called blocks. Assume that both 'n' and 'm' are positive integers and $M_i$ is the ith partitioned block. Transform each of the block $M_{i,}$ one at a time using secret key matrix. Map each character with unique numeric value like A = 0, B = 1 ... to produce the 'n' characters in each of the partitioned block. The ith cipher text block $C_i$ can be obtained by encrypting the ith plaintext block $M_i$ using Equation (1) as:

$$C_i = M_i K \bmod m \tag{1}$$

in which K is an n×n key matrix. The plain text can be obtained from the decrypted cipher text using Equation (2) as:

$$M_i = C_i K^{-1} \bmod m \tag{2}$$

In which $K^{-1}$ is the key inverse and it exist only if the GCD (det K (mod m), m) = 1. According to Overbey *et al.* (2005) the key space of the Hill cipher is precisely GL (n, $Z_m$)-the group of n x n matrices that are invertible over $Z_m$ for a predetermined modulus m and the key space of a prime modulus is larger than composite modulus.

Several researchers tried, to improve the security of Hill cipher. Yeh *et al.* (1991) used two prime numbers as bases that are secretly shared by the participants.

**Corresponding Author:** Adi Narayana Reddy Kaipa, Department of CSE, HITAM, Hyderabad, India

Although the algorithm thwarts the known-plaintext attack, it requires many mathematical calculations and it is not efficient to handle bulk data. Saeednia (2000) made Hill cipher more secure by encryption of each plaintext block with a new key matrix. The new keys are generated by the permutation of rows and columns of an original key matrix. It is vulnerable to known-plaintext attack since the permutated vector is encrypted with the original key matrix. Ismail *et al.* (2006) proposed a new scheme Hill Multiplying Rows by Initial Vector (HillMRIV). It is also vulnerable to known-plaintext attack. Rangel-Romeror *et al.* (2008) presented that If IV is not chosen carefully, some of the new keys to be generated by the algorithm, may not be invertible over $Z_m$, this make encryption/decryption process useless Lin *et al.* (2004) improved the security of Hill cipher by using several random numbers. It thwarts the known-plaintext attack but their scheme is not efficient and is vulnerable to the chosen-cipher text attack and is proved in Toorani and Falahati (2009; 2011) and improved the security with one-way hash function but Keliher and Delaney (2013) proved that it is still vulnerable to attacks. Mahmoud and Chefranov (2009; 2012; 2014) improved the algorithm by using eigen values but these are not efficient and too many seeds are exchanged. Reddy *et al.* (2012a; 2012b) improved the Hill cipher by generating session key using one-way function but the time complexity is more. In the study, an attempt is made to introduce nonlinearity using byte substitution of AES over GF ($2^8$) and sub key groups using pseudo random number sequence.

In the study, an attempt has been made to introduce the concept of byte substitution using S-box over GF ($2^8$) and sub key groups for the cipher text. The detailed algorithm is presented in section-2 and its performance and security analysis are studied and presented in section-3.

## 2. PROPOSED METHOD

### Algorithm

Let M be the plain text to be transmitted through the secured channel. Partition the plain text M into 'm' blocks each of size 'n' where 'n' is positive integer (greater than 1). Let $M_i$ be the $i^{th}$ partitioned block (i = 1, 2, ... m) and size of $M_i$ is n. Let $C_i$ be the ciphertext of the $i^{th}$ block corresponding to the $i^{th}$ block of plain text. Choose a prime number 'p' ($<2^8$).

Select a vector of 'n' relatively prime numbers ($k_1$, $k_2$,… $k_n$). Assume $k_i \in Z_p$. Rotate each row vector relatively right to the preceding row vector to generate a shared key matrix $K_{nxn}$. Let $r = \sum_{i=1}^{n} k_i \bmod p$. Then generate a sequence of pseudo-random numbers $S_i$ (i = 0, 1, … p-1) with initial seed value as r. Generate the sub-key as i from pseudo-random numbers as, $j = (i + S_i) \bmod b$, for all $i \in S_{Gj}$ and $b < \lfloor p/2 \rfloor$. (i.e., the sub-key groups are formed with the pseudorandom number sequence.)

Consider the table of S-box of size 16×16, contains the permutation of all possible values of 8 bits with the leftmost 4 bits as row value and the rightmost 4 bits as a column value. The row and column values serve as indices into the s-box to select a unique 8-bit output. Using the elements in the table of S-box transform the elements by replacing the vector of elements of Y = KM mod p by Z.

Now select the 'n' elements in sequence from the corresponding sub key group $S_G$ using $z_1$ mod b as index of the group and add to Z over mod $2^8$ to generate cipher text block C.

Receiver receives ciphertext ($C_1$, $C_2$). To get back Z subtract selected elements of the index group from C. Substitute the elements of Z using inverse S-box, then multiply the result with inverse key matrix to get the plaintext.

The algorithm is illustrated through the following xample.

### Example

Consider the set of three mutual prime numbers (5, 27, 13) and p = 29. Generate shared key matrix $K_{3x3}$. Assume the plaintext block M = [8, 13, 5]. Generate 'p' pseudo-random number sequence with seed value 45. Assume b = 3 and generate the three sub-key groups ($S_G$) from the random number sequence:

$S_G$ [0] = {1, 3, 4, 6, 7, 8, 11, 14, 16, 19, 23, 24, 26, 27}
$S_G$ [1] = {0, 2, 5, 9, 13, 13}
$S_G$ [2] = {10, 15, 17, 18, 20, 21, 23, 25, 28}

The encrypted plaintext C = [222, 190, 219] and the decrypted ciphertext M = [8, 13, 5].

## 3. ANALYSIS OF THE PROPOSED METHOD

### 3.1. Computational Cost

The time complexity of the algorithm to encrypt and to decrypt the text is O ($mn^2$), where 'm' is number of blocks and 'n' is size of each block, which is same as that of original Hill cipher.

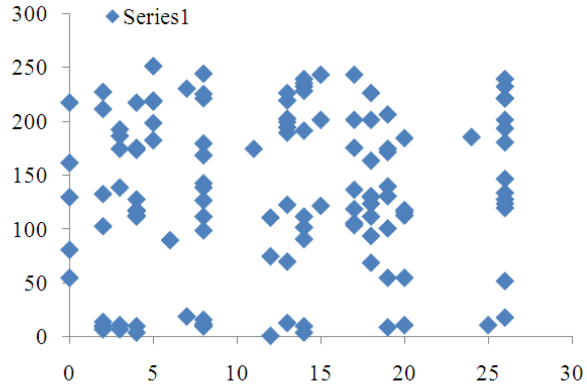

**Fig. 1.** Plaintext Vs ciphertext

It can be obtained as: let $T_{Enc}$ and $T_{Dec}$ denote the running time for encryption and decryption of each block of plaintext respectively Equation (3):

$$\begin{aligned} T_{Enc}(m) &\cong m(n^2+2n)T_{Mul}+m(n^2)T_{Add}+mnT_s \\ T_{Dec}(m) &\cong m(n^2+2n)T_{Mul}+m(n^2)T_{Add}+mnT_{is} \end{aligned} \quad (3)$$

where, $T_{Add}$, $T_{Mul}$, $T_s$ and $T_{is}$ are the time complexities for scalar modular addition, multiplication, substitution, inverse substitution respectively Equation (4):

$$\begin{aligned} T_{Enc}(m) &\cong m(n^2+2n)c_1+m(n^2)c_2+mnc_3 \cong O(mn^2) \\ T_{Dec}(m) &\cong m(n^2+2n)c_1+m(n^2)c_2+mnc_4 \cong O(mn^2) \end{aligned} \quad (4)$$

### 3.2. Result Analysis

The encrypted data is taken as an input to the runs test and correlation coefficient.

### 3.2.1. Runs Test

The runs test checks the randomness of the encrypted data. The data falls into two separate categories such as above and below to a median. The test result shows that the encrypted data is random since runs test gives 0.99 as a result.

### 3.2.2. Correlation Coefficient

Correlation coefficient is a number between -1 and 1 which quantifies the relation between two variables. The correlation coefficient is 1 in case an increasing linear relation, -1 in case of decreasing linear relation and some value in between in all other cases indicating the linear dependence between the variables. The **Fig. 1** shows that the correlation distribution of plaintext and ciphertext for the proposed algorithm. The correlation coefficient has calculated for the proposed algorithm and it had a value of (-0.06005) which represents the plaintext and ciphertext are independent.

### 3.3. Security Analysis

The proposed cryptosystem overcomes all the drawbacks of linear transformation based Hill cipher. The confusion and diffusion are added to the proposed cryptosystem by introducing byte substitution and variable length sub key groups. It overcomes known-plaintext attack, because the nonlinearity of the byte substitution and variable length of sub key groups. It is free from ciphertext-only attach, if the modulus is large prime number. This non-linear behavior made the proposed cryptosystem secure against linear cryptanalysis. It is also free from differential cryptanalysis because the sub key groups are variable length. If the modulus is small prime number then the byte substitution limited to only first 3 or 4 rows.

## 4. CONCLUSION

The structure of the proposed cryptosystem is similar to linear transformation based Hill cipher and each block of plaintext is encrypted with same session key but with different sub key groups. The confusion and diffusion are added to the proposed cryptosystem by introducing byte substitution and variable length sub key groups. The sub key groups are variable length, so the same plaintext blocks are encrypted to different cipher text block which eliminates the known-plaintext attack. It is free from linear and differential cryptanalysis because of byte substitution and sub key groups. The proposed system has reduced the memory size from $n^2$ to n, because key matrix is generated from the first row of the matrix. The correlation between plaintext and cipher text is too low because of byte substitution. This concludes that the plaintext and ciphertext are independent.

## 5. REFFERENCES

Hill, L.S., 1929. Cryptography in an algebraic alphabet. Am. Math. Monthly, 36: 306-312.

Ismail, I.A., M. Amin and H. Diab, 2006. How to repair the hill cipher. J. Zhej. Univ. Sci. A., 7: 2022-2030. DOI: 10.1631/jzus.2006.A2022

Keliher, L. and A.Z. Delaney, 2013. Cryptanalysis of the toorani-falahati hill ciphers. Mount Allison University.

Lin, C.H., C.Y. Lee and C.Y. Lee, 2004. Comments on Saeednia's improved scheme for the hill cipher. J. Chin. Instit. Eng., 27: 743-746. DOI: 10.1080/02533839.2004.9670922

Mahmoud, A.Y. and A. Chefranov, 2014. Hill cipher modification based on pseudo-random eigenvalues. Applied Math. Inform. Sci., 8: 505-516.

Mahmoud, A.Y. and A.G. Chefranov, 2009. Hill cipher modification based on eigenvalues hcm-EE. Proceedings of the 2th International Conference on Security of Information and Networks, Oct. 6-10, ACM Press, New York, USA., pp: 164-167. DOI: 10.1145/1626195.1626237

Mahmoud, A.Y. and A.G. Chefranov, 2012. Secure hill cipher modification based on generalized permutation matrix SHC-GPM. Inform. Sci. Lett., 1: 91-102.

Overbey, J., W. Traves and J. Wojdylo, 2005. On the keyspace of the hill cipher. Cryptologia, 29: 59-72. DOI: 10.1080/0161-110591893771

Rangel-Romeror, Y., R. Vega-Garcia, A. Menchaca-Mendez, D. Acoltzi-Cervantes and L. Martinez-Ramos *et al*., 2008. Comments on "How to repair the Hill cipher". J. Zhej. Univ. Sci. A., 9: 211-214. DOI: 10.1631/jzus.A072143

Reddy, K.A., B. Vishnuvardhan, Madhuviswanath and A.V.N. Krishna, 2012a. A modified hill cipher based on circulant matrices. Proceedings of the 2nd International Conference on Computer, Communication, Control and Information Technology, Feb. 25-26, Elsevier Ltd., pp: 114-118. DOI: 10.1016/j.protcy.2012.05.016

Reddy, K.A., B. Vishnuvardhan and Durgaprasad, 2012b. Generalized affine transformation based on circulant matrices. Int. J. Distribut. Parallel Syst., 3: 159-166.

Saeednia, S., 2000. How to make the hill cipher secure. Cryptologia, 24: 353-360. DOI: 10.1080/01611190008984253

Toorani, M. and A. Falahati, 2009. A secure variant of the hill cipher. Proceedings of the IEEE Symposium on Computers and Communications, Jul. 5-8, IEEE Xplore Press, Sousse, pp: 313-316. DOI: 10.1109/ISCC.2009.5202241

Toorani, M. and A. Falahati, 2011. A secure cryptosystem based on affine transformation. Sec. Commun. Netw., 4: 207-215. DOI: 10.1002/sec.137

Yeh, Y.S., T.C. Wu, C.C. Chang and. W.C. Yang, 1991. A new cryptosystem using matrix transformation. Proceedings of the 25th IEEE International Carnahan Conference on Security Technology, Oct. 1-3, IEEE Xplore Press, Taipei, pp: 131-138. DOI: 10.1109/CCST.1991.202204