Enhancing the Security of Hill Cipher using Columnar Transposition

Ashwin Ramesh
B-Tech Student: Dept. of Computer Science & Engineering
Manipal Institute of Technology
Manipal, India

***Abstract*—Hill cipher algorithm is based on poly-graphic substitution technique centered on linear algebra. The algorithm takes *m* successive plaintext letters and replaces them with *m* cipher text letters. This technique involves representing the plaintext as a matrix before undergoing encryption which produces another matrix, from which the cipher text is extracted. Through this paper the author wishes to propose a technique to deal with a loophole which compromises the security of the algorithm. The examination of the attack involves looking for repetitions of substrings in the cipher text. If the distance between the two substrings happen to be a multiple of *n*, where *n* is the size of the invertible square matrix representing the key, then their respective cipher texts would also be the same. This can predict the size of the key matrix used. The key can be then be discovered with more ease.**

***Keywords*—Encryption, Decryption, Hill cipher, Cryptography, Cryptanalysis, Plaintext, Cipher text, Security, Transposition, Substitution, Modified Hill Cipher.**

I. INTRODUCTION

With increasing emphasis on network security in the modern world, the need to keep and transmit data in an encrypted form has become really important. Cryptography, the discipline of securing important data [1] has become more of an inevitability than a luxury. A lot of information is sent via public network channels [2], and is important that the data maintains authenticity, integrity, confidentiality and non-repudiation [3] throughout. Cryptography can be divided into three categories [4] – secret key cryptography, public-key cryptography [5] and hash functions. Hill cipher falls under the first category.

In cryptography, substitution is a method where units of a plaintext are encoded by replacing them with new units. This is usually facilitated by a substitution cipher algorithm. A substitution process is much different from a transposition process, where the letters are merely shuffled [6]. Hill cipher is a poly-graphic [7] encryption substitution cipher which involves the usage of the concepts of linear algebra to produce the cipher text for a given plaintext. The basic technique behind the cipher algorithm is that of matrix multiplication. The first step involves representing each letter with a unique number. A simple representation is represented in *Table 1*.

TABLE I. ALPHABET MATRIX.

0	1	2	3	4	5	6	7	8	9
A	B	C	D	E	F	G	H	I	J
10	11	12	13	14	15	16	17	18	19
K	L	M	N	O	P	Q	R	S	T
20	21	22	23	24	25				
U	V	W	X	Y	Z				

The encryption process involves selecting a cipher key [8], and representing it as a square matrix any size *n*, by replacing the letters with their respective numbers. This square matrix should be an invertible matrix [9]. Filler elements can be used to fill the matrix in case the letters aren't enough. For instance, a key BDBBBCCDE, can be represented in a 3x3 matrix *K* as,

$$K = \begin{pmatrix} 1 & 3 & 1 \\ 1 & 1 & 2 \\ 2 & 3 & 4 \end{pmatrix}$$

Fig. 1. An example of the key matrix.

The plaintext to be encrypted is divided into blocks of *n* letters. Each represented as P_i, where *i* represents the index of a block. For instance, let the plaintext be ABCDEFABCXYZ. It can be divided into blocks of size 3, such that there are 4 blocks, ABC (P_0), DEF (P_1), ABC (P_2) and XYZ (P_3). The individual blocks are then multiplied with the key matrix. Modulo 26 of that product is the required cipher text. The mathematical representation is: $C_i = (P_i.K) \bmod 26$. The following would be the encryption procedure of the first 3 letters of the plaintext.

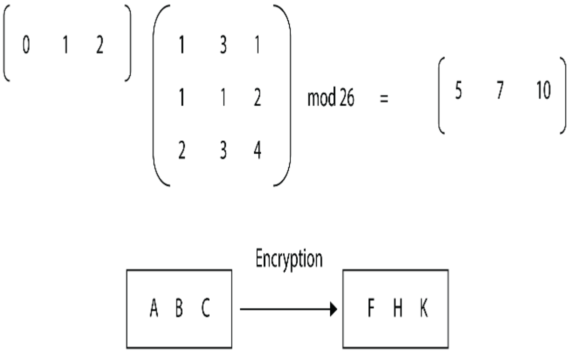
$\begin{pmatrix} 0 & 1 & 2 \end{pmatrix} \begin{pmatrix} 1 & 3 & 1 \\ 1 & 1 & 2 \\ 2 & 3 & 4 \end{pmatrix} \bmod 26 = \begin{pmatrix} 5 & 7 & 10 \end{pmatrix}$
Encryption
A B C
F H K

Fig. 2. The encryption procedure.

The decryption process requires the inverse of the key matrix, $K^{-1} = (K)^{-1}$ mod 26. The inverse matrix [9] is then multiplied with the cipher text. Modulo 26 of this product gives us back the plaintext. The mathematical representation is: $P_i = (C_i.K^{-1})$ mod 26. The following would be the decryption procedure of the first 3 letters of the cipher text.

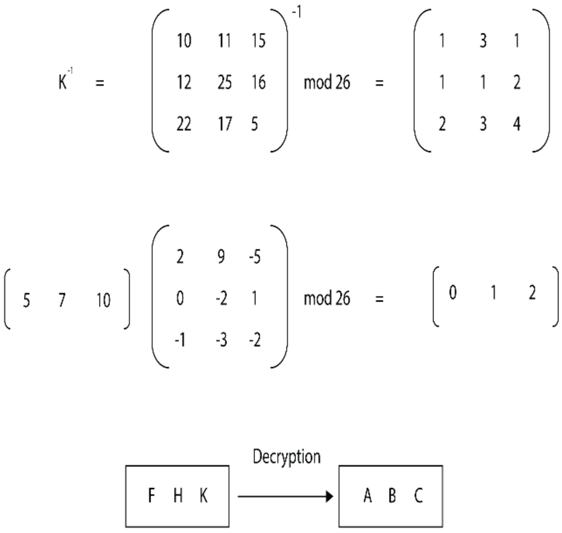
$K^{-1} = \begin{pmatrix} 10 & 11 & 15 \\ 12 & 25 & 16 \\ 22 & 17 & 5 \end{pmatrix}^{-1} \bmod 26 = \begin{pmatrix} 1 & 3 & 1 \\ 1 & 1 & 2 \\ 2 & 3 & 4 \end{pmatrix}$
$\begin{pmatrix} 5 & 7 & 10 \end{pmatrix} \begin{pmatrix} 2 & 9 & -5 \\ 0 & -2 & 1 \\ -1 & -3 & -2 \end{pmatrix} \bmod 26 = \begin{pmatrix} 0 & 1 & 2 \end{pmatrix}$
Decryption
F H K
A B C

Fig. 3. The decryption procedure.

The cipher text of the above mentioned plaintext would be FHKRCFFHKTMP. As one can notice that the cipher text of the same plaintext substrings at an interval of *n* are the same. This major security lapse can be used as a stepping stone by a cryptanalyst to guess the size of the key matrix taken during the encryption process and launch an attack on the security system.

In this paper, the author introduces a method to deal with such a situation by dealing it in an effective manner. This paper suggests 4 stages to remove this potential threat. The first stage involves a permutation process on the plain text, using columnar transposition. The second involves application of the Hill cipher algorithm. The third stage applies another round of columnar transposition using a key different to the one used in stage 1. In the last stage, the Hill cipher algorithm is applied again. The paper attempts to prove that the 4 stage encryption method removes the above mentioned threat and increase the security level of the algorithm. The two stages of columnar transposition protect the algorithm from a known-plaintext attack [10]. The original algorithm is vulnerable because of being a completely linear algorithm. Hence, a lot of research work has taken place to improve the security of the Hill cipher algorithm [11-15].

Columnar transposition involves writing the plaintext out in fixed sized rows. Then depending on a key value, it is read out column by column. The alphabetical order of the letters in the key decides the order in which the columns have to be read in. A double columnar transposition involves applying the same procedure twice using the same or different keys.

II. METHODOLOGY

A. *Encryption*

Stage 1: First Double Columnar Transposition.

Algorithm: Double Columnar Transposition.

The algorithm performs double columnar transposition on the plaintext.

Input: Plaintext (P), keys (K_1, K_2), size of key K_1 (m), size of K_2 (n).

Output: Intermediate cipher text (C).

Initialize: i ← 0.

1: Place the letters in P row-wise in the table with m elements in one row.

2: Add filler characters, if necessary, to fill all the elements of the last row.

3: Read the table column-wise depending on the order of the alphabets in K_1.

4: Store as P_{temp}.

5: Place the letters in P_{temp} row-wise in the table with n elements in one row.

6: Repeat step 2.

7: Read the table column-wise depending on the order of the alphabets in K_2.

8: Store as C.

9: end

Stage 2: First Application of Hill Cipher.

Algorithm: Hill Cipher Encryption.

The algorithm applies the Hill cipher encryption technique.

Input: Intermediate cipher text (P), key (key), size of key square matrix (n).

Output: Intermediate cipher text (C).

```
Initialize: i ← 0.
1: Represent letters of key with their respective numbers in the matrix K.
2: Divide P into blocks of size n. Add filler to complete the last block, if necessary.
3: while i < n:
4:        $C_i = (P_i.K) \bmod 26$
5:        $i \leftarrow i + 1$
6: Replace the numbers with their respective letters.
7: Store the cipher text as C.
8: end
```

Stage 3: Second Double Columnar Transposition.

Repeat the algorithm mentioned in stage 1 on the intermediate cipher text obtained after stage 2.

Stage 4: Second Application of Hill Cipher.

Repeat the algorithm mentioned in stage 2 on the intermediate cipher text obtained after stage 3. The result after the stage is final cipher text which is free from the hazard under consideration.

B. Decryption

The decryption process involves applying the reverse of all of the 4 stages in the reverse order.

Stage 1: First Application of Hill Cipher Decryption.

Algorithm: Hill Cipher Decryption.

The algorithm decrypts a cipher text encrypted by Hill cipher.

Input: Cipher text (C), inverse of the key (invkey), size of invkey square matrix (n).

Output: Intermediate plaintext (P).

```
Initialize: i ← 0.
1: Represent letters of invkey with their respective numbers in the matrix $K^{-1}$.
2: Divide C into blocks of size n.
3: while i < n:
4:        $P_i = (C_i.K^{-1}) \bmod 26$
5:        $i \leftarrow i + 1$
6: Replace the numbers with their respective letters.
7: Store the plaintext as P.
8: end
```

Stage 2: First Application of Columnar Transposition Decryption.

Algorithm: Columnar Transposition Decryption.

The algorithm decrypts a cipher text encrypted by columnar transposition.

Input: Intermediate plaintext (C), size of cipher text (size), keys (K_1, K_2), size of key K_1 (m), size of K_2 (n).

Output: Intermediate plaintext (P).

```
Initialize: i ← 0.
1: Place the letters in P column-wise in the table with (size/n) elements in one column.
2: Read the table row-wise depending on the order of the alphabets $K_2$.
3: Store as $C_{temp}$.
4: Place the letters in $C_{temp}$ column-wise in the table with (size/m) elements in one column.
5: Read the table row-wise depending on the order of the alphabets $K_1$.
6: Store as P.
7: end
```

Stage 3: Second Application of Hill Cipher Decryption.

Repeat the algorithm mentioned in stage 1 on the intermediate plaintext obtained after stage 2.

Stage 4: Second Application of Columnar Transposition Decryption.

Repeat the algorithm mentioned in stage 2 on the intermediate plaintext obtained after stage 3. The result after the stage is final plaintext.

III. RESULTS

Plaintexts taken during experiments:

TABLE 2. RESULTS OF TEST CASE I.

Plaintext	CATHERINEISACAT
K1 (Columnar Transposition)	BANGLE
K2 (Columnar Transposition)	SQUARE
K3 (Hill Cipher)	HYPNOTISE
Stage 1	ANACICRAZHIXESYTET
Stage 2	NANEOIHARZSLMAEZWV
Stage 3	EZZILVAAAOSWNHMNRE
Stage 4	HMLDWXAAAOWQSMUGYO

TABLE 3. RESULTS OF TEST CASE II.

Plaintext	MATHSDOESMATTER
K1 (Columnar Transposition)	PERSON
K2 (Columnar Transposition)	NOTICE
K3 (Hill Cipher)	TRIANGLES
Stage 1	AEEDTZSAYMOTTSRHMX
Stage 2	SQSUIQIMEVUCCBULDE
Stage 3	IUDQCEUVLSICQMNSEU
Stage 4	DSEKCEHHQACUDQKQWI

IV. ANALYSIS

TABLE 4. COMPARISON OF THE REULSTS OF THE HILL CIPHER ALGORITHM AND THE 4-STAGE ALGORITHM (TEST CASE I).

Hill Cipher Algorithm	HMLDWXAAAOWQSMUGYO
4 Stage Algorithm	KACDKPXETECUKAC

TABLE 5. COMPARISON OF THE REULSTS OF THE HILL CIPHER ALGORITHM AND THE 4-STAGE ALGORITHM (TEST CASE II).

Hill Cipher Algorithm	DSEKCEHHQACUDQKQWI
4 Stage Algorithm	VUWKBKWYSVUWCBO

In both the examples taken in the previous section, it can be clearly noticed that applying the Hill cipher encryption algorithm results in repetition of a substring (KAC in test case 1, and VUW in test case 2) at an interval which is equal to the length of the key used. But, applying the 4 stage algorithm on the given plaintexts shows that such a pattern is not generated in the cipher text. Hence, no information about the size of the key square matrix could be attained by just looking at the cipher text, unlike the old method.

V. CONCLUSION

The above results clearly show that 4 stage algorithm introduced in the paper eliminates the possibility of repetition of substrings at an interval which is a multiple of the size of the key square matrix. The two stages of transposition generate permutations of the text which makes cryptanalysis on the cipher text even more tedious. The potential loophole in the Hill cipher algorithm discussed in the paper is also eliminated by transposition of the text before applying the Hill cipher encryption algorithm.

REFERENCES

[1] Kahate, A. Cryptography and Network Security, 2nd Edition.

[2] Tanenbaum, A. S.; Wetherall, D. J.; Computer Networks, 5th Edition.

[3] Forouzan, B. A.; Cryptography and Network Security.

[4] Kessler, G. C.; An overview of cryptography. Retrieved on June 21, 2015, from http://www.garykessler.net/library/crypto.html

[5] Ayushi. A symmetric key cryptographic algorithm. International Journal of Computer Applications (0975 - 8887), 1, 15.

[6] Malay B. Pramanik; Implementation of Cryptography Technique using Columnar Transposition", International Journal of Computer Applications (0975 – 8887) Second National Conference on Recent Trends in Information Security, GHRCE, Nagpur, India, Jan-2014.

[7] Kumar, N.; Panduranga, H. T.; Advanced Partial Image Encryption using Two-stage Hill Cipher Technique. International Journal of Computer Applications (0975 – 8887), 60, 16.

[8] Stallings, W.; Cryptography and Network Security: Principles and Practice, 5th Edition

[9] Chirgaiya, S.; & Sharma, N.; A Novel Approach to Hill Cipher. International Journal of Computer Applications (0975 – 8887), 108, 11

[10] Borissov, Y.; Lee, M. H.; Bounds on Key Appearance Equivocation for Substitution Ciphers. IEEE transactions on information theory, 53(6), 2294-2296.

[11] Saeednia, S.; How to Make Hill Cipher Secure. Cryptologia, 24(4), 353-360.

[12] Bhowmick, A.; & Maiya, G.; Enhancing Resistance of Hill Cipher using Columnar and Myszkowski Transposition. International Journal of Computer Sciences and Engineering, 3(2), 2347-2693.

[13] Chefranov, A. G.; Secure Hill Cipher Modification. Proc. Of the First International Conference on Security of Information and Network (SIN2007) 7-10 May 2007, Gazimagusa (TRNC) North Cyprus, Elci, A., Ors, B., and Preneel, B (Eds) Trafford Publishing, Canada, 2008: pp 34-37, 2007.

[14] P. L. Sharma; M. Rehan; On Security of Hill Cipher using Finite Fields, International Journal of Computer Applications (0975 – 8887) Volume 71– No.4, May 2013.

[15] Kondwani Magamba,; Solomon Kadaleka, Ansley Kasambara,; Variable-length Hill Cipher with MDS Key Matrix, International Journal of Computer Applications (0975 – 8887) Volume 57– No.13, November 2012.