A Study on the Analysis of Hill's Cipher in Cryptography

N.Vijayaraghavan#1, S.Narasimhan#2, M.Baskar#3

#1, #2 Mathematics Division, Department of Science and Humanities,
KCG college of Technology, Chennai, India-600097

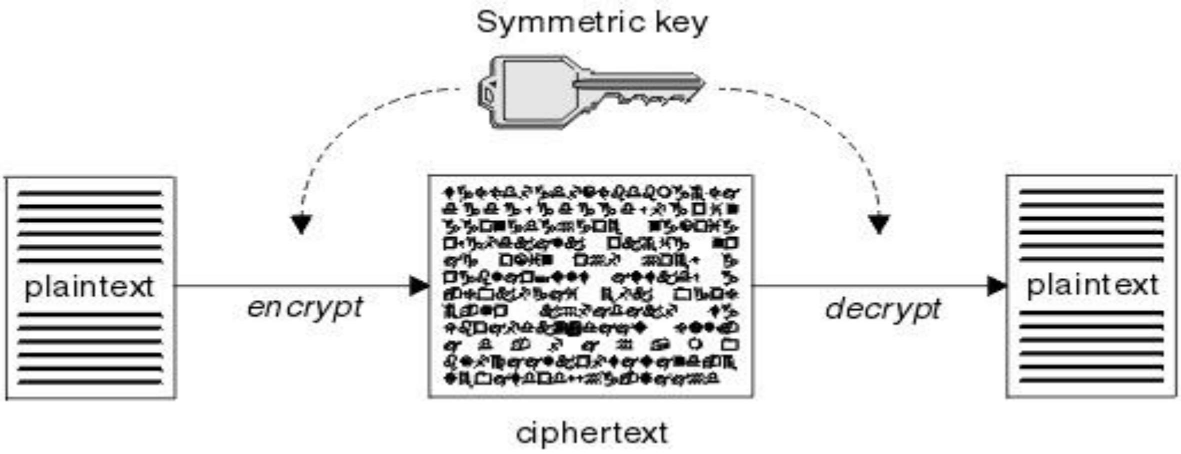
Abstract --- *Cryptography is the study of methods of sending a message in disguised form so that only the intended people can receive the original message. In data communication, the major issue now is the security of the data. To maintain its confidentiality, integrity and availability, we need to protect the data and its communications from rivals. Today in this e-world, the necessity of protecting the data is of higher importance. With the introduction of computers in this modern world, message sharing becomes very easy via e-mail, mobile phone communication etc.,. Also the information that is being sent via computers can be intercepted and read by an enemy. So the necessity of data hiding is obvious. In this research paper, we give a study on Hill's cipher in cryptography.*

Keywords— *Cryptography, Hill cipher, Encryption, Decryption, Linear Algebra*

I. Introduction

(In today's technology, the initial ciphers were traced, so new and much stronger ciphers introduced, which forced cryptographers to find better ciphers and so on. The significance of key is an enduring principle of cryptography. Moving on from this introduction, I will focus on a linear algebra based Cipher, the Hill cipher, which fixed the main problems associated with ciphers like the Caesar cipher.

HOW A PLAINTEXT GETS CONVERTED TO A CIPHERTEXT:



Caeser cipher

Until recently, encrypting secret messages was performed by hand using relatively trivial mechanisms to disguise information. One of the most well-known ciphers was named after Julius Caesar, namely, the Caesar cipher. The Caesar cipher is an example of a substitution cipher. Each letter of a given plaintext, the information to be encrypted, is substituted with another letter some given number of positions from it in the alphabet. For example, if we had an alphabet comprised of the standard 26 letters in the English alphabet and swapped each letter with the letter three places after it in the alphabet, we would have the following Caesar cipher

Plaintext:

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

Cipher text:

D E F G H I J K L M N O P Q R S T U V W X Y Z A B C

Using this cipher, the text “GOOD MORNING” would encode to “JRRG PRUQLQJ.” One of the main problems with the Caesar cipher is that if an individual intercepts the Ciphertext and guesses that the Caesar cipher was used for the encryption, he or she could easily go through the 25 shift values until they come upon a shift that decodes the ciphertext into a meaningful plaintext. On a broader scale, simple substitution ciphers all fall victim to simple analysis. Given any particular alphabetic language, certain letters are used more or less frequently than other letters. In a simple substitution cipher, these frequencies with which each letter occurs in an average sentence or paragraph are maintained. For example, if a substitution cipher encoded “e” to “w,” “w” would occur in the cipher text with the same frequency as “e” in the original language, allowing for a relatively simple analysis to break the substitution cipher.

2. Hill cipher

The Hill cipher is based on linear algebra and overcomes the frequency distribution problem of the Caesar cipher that was previously discussed. The rest of this paper will be devoted to an explanation of the Hill cipher, its shortcomings, and one way to secure the cipher further.

For both encryption and decryption, the Hill cipher assigns numerical values to each letter of an alphabet. Throughout this paper, we will use the standard 26 character English alphabet and define the following associations between letters in our alphabet and numbers. The core of Hill-cipher is matrix manipulations. 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=1,
b=2,
…..
…..
z=26.

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:

C1=(K11P1+K12P2+K13P3)MOD26---------1
C1=(K21P1+K22P2+K23P3)MOD26---------2
C1=(K31P1+K32P2+K33P3)MOD26---------3

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 ciphertext 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. The inverse of a matrix doesn’t always exist, but when it does it satisfies the proceeding equation. 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-1KP=P

2.1 Working of a Hill Cipher:

One way to destroy the value of frequency analysis is to encrypt a string of letters as one block. Here is an additive cipher that encrypts a block of four letters. Our plaintext messages split into blocks of four is

GOOD MORN INGT OYOU

Numbers are substituted for letters by a = 1, b = 2, … , z = 26.

G	O	O	D	M	O	R	N	I	N	G	T	O	Y	O	U
7	15	15	4	13	15	18	14	9	14	7	20	15	25	15	21

The key adds to each component of the block – thought of as a column vector – component-wise. For example, if
the keys were

3
12
21
17

Then

$$\begin{array}{cc} G & 7 \\ O & 15 \\ O & 15 \\ D & 4 \end{array} + \begin{array}{c} 3 \\ 12 \\ 21 \\ 17 \end{array} = \begin{array}{cc} J & 10 \\ A & 1 \\ J & 10 \\ U & 21 \end{array} \mod 26$$

The block GOOD encrypts as JAJU.

Therefore our message encrypts as:

GOOD \ MORN \ INGT \ OYOU
JAJU \ PAME \ LZBK \ RKJL

Decryption is accomplished by adding the additive inverse of the key to the ciphertext.

That is Plaintext to ciphertext

$$\begin{array}{c} 3 \\ 12 \\ 21 \\ 17 \end{array}$$

So for ciphertext to plaintext, the inverse of the above column matrix,

$$\begin{array}{c} 23 \\ 14 \\ 5 \\ 9 \end{array}$$

To encrypt a four-letter block, the key is a 4 × 1 matrix. There are 264 =456076 possible keys – one of which produces plaintext. If we know one plaintext/ciphertext block correspondence, we can solve for the key.

2.3 Cracking of a Hill Cipher:

How do you crack a Hill cipher? In other words, how do you discover the inverse key when you do not know the key? If you have "captured" enough plaintext along with the corresponding ciphertext, then you may be able to use the following theorem.

Theorem 1 (Cracking Theorem). Suppose the length m of the alphabet is a prime. Let p1, p2,..., pn be n plaintext vectors for a Hill n-cipher having (unknown) key matrix A, and let c1, c2,..., cn be the corresponding ciphertext vectors. Suppose these plaintext vectors are linearly independent over Zm. Form the matrix

P = [p1| p2| ..., |pn]

having the plaintext vectors as its columns, and the matrix

C = [c1 |c2 |... |cn]

having the ciphertext vectors as its columns.

Then the same sequence of elementary row operations that reduces CT to the identity matrix I reduces PT to the transpose (A^{-1})T of the inverse key matrix A^{-1}.

Conclusion :

The HILL-cipher method being discussed here is a powerful method and the first general method for successfully applying algebra -specifically linear algebra. The applications of algebra in cryptography is a lot and hill cipher is just an example of it. Unfortunately, the basic Hill cipher is vulnerable to a known-plaintext attack because it is completely linear. While matrix multiplication alone does not result in a secure cipher it is still a useful step when combined with other non-linear operations, because matrix multiplication can provide diffusion.

References :

[1] Bibhudendra Acharya, Girija Sankar Rath, Sarat Kumar Patra, Saroj Kumar Panigrahy. 2007. Novel Methods of Generating Self-Invertible Matrix for Hill Cipher Algorithm, International Journal of Security, Vol 1, Issue 1, 2007, pp. 14-21.
[2] Imai H., Hanaoka G., Shikata J., Otsuka A., Nascimento A.C. 2002. Cyptography with Information Theoretic Security. Information Theory Workshop, 2002, Proceedings of the IEEE, 20-25 Oct 2002.
[3]Menezes, A. J., P.C. Van Oorschot, S.A. Van Stone. 1996. Handbook of Applied Cryptography. CRC press.
[4] Overbey, J., Traves, W., Wojdylo, J., 2005. On the keyspace of the Hill cipher. Cryptologia, 9(l):59-72.
[5] Petersen, K., 2000. Notes on Number Theory and Cryptography.