A MODIFIED HILL CIPHER FOR BETTER SECURITY

Johni S Pasaribu
Department of Informatics Engineering, Politeknik Piksi Ganesha
Jl. Gatot Subroto no. 301 Bandung
johni.pasaribu@piksi-ganesha-online.ac.id, johni_0106@yahoo.com

Abstract. Security is the most challenging aspects in the internet and network application. These days the applications like Internet and networks are growing very fast, thereby the importance and the value of the exchanged data over the internet or other media types are increasing. While transmitting any confidential information across network, some unauthorized user tend to steal or corrupt the data. To prevent this interruption, security of data is needed. As data security is needed for protecting data from unauthorized users and destructive forces, some security techniques are used. Encryption is one such technique which is used to protect the database from unauthorized accesses and unwanted actions of external clients. It is used to secure the confidential information from unauthorized user by converting the data in to that format that is only understandable by the authorized receiver that knows the respective decryption to obtain the original data or information. In this paper, our focus will be on Database Security, as databases are considered as the storehouses of data. Generally, data of an organization or a company is stored in databases and is very crucial to the organization. Today, most of the organizations allow their clients to use their services (online banking, online shopping etc) by accessing their databases. This leads to a requirement of high level security to deal with information attackers. An information attacker tries to illegally acquire or modify the highly confidential data of the organization. An innovative approach for database security using Hill Cipher encryption & decryption method is proposed. This cryptography based approach provides an additional level of security to the database systems. Important data or information crucial to the organization should be secured from illegal theft and attacks. This encryption scheme completely is based on key matrix that a organization decides before encryption and decryption. In this proposed scheme a common key is used to encrypt the data items.

Keywords: Database, Data security, Encryption, Decryption, Hill Chiper Algorithm.

INTRODUCTION

Data Security means to protect a database from destructive forces and the unwanted actions of unauthorized clients. Encryption is technique that is basically used to secure the confidential information from unauthorized user by converting the data in to the understandable form that is only understandable by the authorized user that knows the respective decryption to obtain the original data or information. In their work, they used Attribute-based encryption (ABE) public-key based one to many encryption which will allow users to encrypt and decrypt data based on user selected attributes. Existing scheme would be enhanced by using Hill cipher. Hill cipher is a multi-letter cipher and the encryption technique takes m successive plaintext letters and substitutes for m cipher text letters.

In cryptography technique, encryption is the method of encoding messages (or data) in such a way that third user cannot read it, but only authenticated user can do it. Encryption doesn't control hacking but it can control the hacker from reading the data that is encrypted. In encryption scheme, the message or information is encrypted using an encryption algorithm, converting it into an unreadable ciphertext. This is done with the use of an encrypted key, which will specify how the message has to be encoded. An adversary that can see the ciphertext should not be able to determine anything about the original input message. An authorized user is able to decode the ciphertext using a decryption algorithm, that usually requires a secret

decryption key that adversaries do not have access to. Decryption is the reverse process of encryption to get back the original data.

Cryptography is used to secure the data. 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. Cryptography and encryption have been particularly important. Throughout history in times of war when a general would not want the enemy to figure out the plans he was distributing amonghis troops. Recently, the uses of cryptography have grown drastically. Cryptography is still important in times of war, but with the 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. If protection of confidential information is to be done then cryptography is provide high level of privacy of individuals and groups. Cryptography is the methods that allow information to be sent in a secure from in such a way that the only receiver able to retrieve this information.

ENCRYPTION AND DECRYPTION

Encryption is the process of converting information into an encrypted form, so that it is intelligible only to someone who knows how to 'decrypt' it to obtain the original message. It is used in connection with electronic data, whether stored on a computer or transmitted over an unsecured network such as the Internet. Encryption tools are available and can be used to secure:

- Stored data, from single files to entire hard disks;
- Computer code such as computer operating systems;
- Information transmitted over the Internet, including e-mails and internet telephony (Voice over Internet Protocol or VoIP);
- Entire communications infrastructures, such as wireless networks (including mobile telephony).

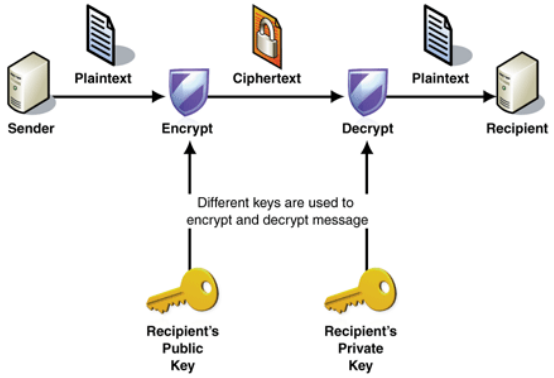

Figure 1. Cryptography (Encryption & Decryption)

There are basically two types of encryption techniques used. Both the encryption techniques are described below.

Symmetric Encryption

In Symmetric-key method, the encryption and decryption keys are identical. Therefore communicating users must agree on a secret key before they wish to communicate. Symmetric

encryption uses a single key to encrypt and decrypt the message. When user encrypt the message they provide key to the recipient user before decrypt it for using symmetric encryption, the sender encode the message and, if the recipient does not already have a key, then sender sends the key and cipher text separately to the recipient user. The message then decode by recipient key.

This method is very easy and fast to implement but has weaknesses. This means that the algorithm that is used to encode the message is easier for attackers to understand, enabling them to more easily decode the message.

Asymmetric Encryption

In public-key encryption schemes, the encryption key is published for anyone to use and encrypt messages. However, only the receiving user has access to the decryption key and is capable of reading the encrypted messages. Public-key encryption is a relatively recent invention, all encryption schemes have been symmetric-key (also called private-key) schemes. Asymmetric encryption is also known as Public-Key encryption that uses two different keys -a public key to encrypt the message and a private key to decrypt it. The public key is used to encrypt and private key to decrypt.

One can easily distribute the public key to communicate because only with private key one can decrypt it. To protect the message between users the sender encrypts it by public key. Then receiver decrypts it by private key. Only recipient can decrypt the message in this technique of encryptions.

Decryption is the process of converting the cipher text into plain text or getting back the original data. A reverse process of encryption is called as Decryption. The process of decryption requires two things-a decryption algorithm and a key. A decryption algorithm means the technique that has been used in Decryption. Generally the encryption and decryption algorithm are same.

Hill Cipher Algorithm

In the field of **cryptography**, **Hill cipher** is a polygraphic substitution, linear algebra based cipher technique. It was invented by Lester S. Hill in 1929, it was the first polygraphic cipher in which it was practical to operate on more than three symbols at once. The basic concept behind this algorithm is that each letter (Alphabet) is allotted a number generally starting from 0 in a continuous sequence one after the other as shown in figure 2.

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
↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25

Figure 2. Alphabet Numbering

As shown in figure 2, Alphabets are numbered as A = 0, B =1, ... , Z=25, but this is not a fixed requirement of the cipher. The encryption of plain text takes **n** successive plain text letters and substitutes them for **n** cipher text letters. In case n = 3, the encryption can be expressed in terms of the matrix multiplication as follows:

$$C = K.\ P \pmod{26} \qquad \ldots [1]$$

$$\begin{pmatrix} C_1 \\ C_2 \\ C_3 \end{pmatrix} = \begin{pmatrix} k_{11} & k_{12} & k_{13} \\ k_{21} & k_{22} & k_{23} \\ k_{31} & k_{32} & k_{33} \end{pmatrix} \begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix} \bmod 26$$

Mathematical Model:
Abbreviations used are:
C = Cipher Text
P = Plain Text
K = Key used for Encryption of plain text to cipher text.
K^{-1} = Inverse of key K used for Decryption of cipher text to plain text.
K is the key matrix and must be invertible so there is K^{-1} (K^{-1} is the matrix inverse). The inverse matrix can be calculated as:

$$K.\ K^{-1} = I \qquad \ldots [2]$$

where I is the identity matrix. The key matrix, K can be written like this:

$$K = \begin{pmatrix} K_{11} & K_{12} & \ldots & \ldots & K_{1n} \\ K_{21} & K_{22} & \ldots & \ldots & K_{2n} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ K_{n1} & K_{n2} & \ldots & \ldots & K_{nn} \end{pmatrix}$$

PROPOSED METHOD

In this paper, an innovative approach for database security using *Hill Cipher* encryption & decryption method is proposed. This cryptography based approach provides an additional level of security to the database systems. Important data or information crucial to the organization should be secured from illegal theft and attacks.

This encryption scheme completely is based on key matrix that a organization decides before encryption and decryption. In this proposed scheme a common key is used to encrypt the data items.

Encryption Algorithm

- Fetch the source text present in the database that is going to be encrypted.
- Fetch their position values (Table ID, Table Name, Row number, Column Number).
- Choose an Encryption key matrix (say K) which would be used in Hill Cipher Algorithm for encryption of text.
- Obtain the cipher text using Hill Algorithm by applying the same key matrix which was derived in the previous step.
- Place this encrypted data in the corresponding field.
- Execute the query obtain cipher text values.

Decryption Algorithm

- Obtain the cipher text for the purpose of decryption.
- Obtain Inverse of the key matrix (K^{-1}) used for encryption of plain text.
- Apply this inverse key matrix in the decryption of cipher text using Hill Cipher Algorithm.

- The above step will give plain text as a result.

This is a simple algorithm which provides an effective solution for data security problem because encryption of text encrypts the plain text to cipher text using key matrix and decryption of text decrypts the cipher text to plain text using inverse key matrix using hill algorithm. Therefore, this increases the level of database security which helps in eliminating database data item attacks to a large extent. Hill cipher algorithm is a linear algebra based substitution algorithm which is fast and easy to apply. It is a good encryption scheme which is computationally secure. This encryption scheme will provide database security without affecting the system performance. Encrypted data is always slightly high in volume as compared to the original data. Decryption process is the inverse of the encryption process in which the inverse key matrix is used to decrypt the cipher text.

EXPERIMENTAL RESULTS

Fetch Source text from the database

In figure 3, the database data items that were considered in the experiment are shown.

	NPM	NAMA MAHASISWA	JURUSAN
1	0511068	SUPRIADI DANI	IF
2	0211044	ABDUL SUKUR	SI
3	0711031	ASEP DIAN	MI
4	0511038	TINA TALISA	IF

Figure 3. Database Data Items

Encryption of fetched data values from the database

For Encryption of the data item that is fetched from the database, in our case: SUPRIADI DANI. *Hill Cipher* encryption algorithm is applied along with the key matrix K. Plaintext: "SUPRIADI DANI"

Encryption Key (k):

$$K = \begin{pmatrix} 3 & 7 \\ 5 & 12 \end{pmatrix}$$

Cipher Text for first two letters of the plain text:

$$C_{1,2} = \begin{pmatrix} 3 & 7 \\ 5 & 12 \end{pmatrix}\begin{pmatrix} 18 \\ 20 \end{pmatrix} = \begin{pmatrix} 194 \\ 330 \end{pmatrix} (\text{mod } 26) = \begin{pmatrix} 12 \\ 18 \end{pmatrix} = \begin{pmatrix} M \\ S \end{pmatrix}$$

Similarly, cipher text for remaining letters (P,R,I,A,D,I,D,A,N,I) is calculated. Cipher Text for second two letters of the plain text:

$$C_{3,4} = \begin{pmatrix} 3 & 7 \\ 5 & 12 \end{pmatrix}\begin{pmatrix} 15 \\ 17 \end{pmatrix} = \begin{pmatrix} 164 \\ 279 \end{pmatrix} (\text{mod } 26) = \begin{pmatrix} 8 \\ 19 \end{pmatrix} = \begin{pmatrix} I \\ T \end{pmatrix}$$

After encryption to first record plain text *P* in the database to get result cipher text *C*:

P = S U P R I A D I D A N I

C = M S I T Y O N H J P R F

Similarly, to other record in that database each plain text *P* will get results cipher text *C*:

P = A B D U L S U K U R

C = H M T V D L A M X E
P = A S E P D I A N
C = W I N S N H N A
P = T I N A T A L I S A
C = J J N N F R L V C M

In a recent investigation, Hill Cipher results each cipher text and this cipher text has not the same pattern as the plain text.

Decryption of encrypted database values

For decryption use the *inverse* key matrix P = K^{-1}. C mod 26. Inverse of K which is K^{-1} is used for decrypting the cipher text into plain text.

$C = K \cdot P$
$K^{-1}. C = K^{-1}. K \cdot P$
$K^{-1}. C = 1 \cdot P$
$P = K^{-1}. C \pmod{26}$ **… [3]**

With the same key matrix $K = \begin{pmatrix} 3 & 7 \\ 5 & 12 \end{pmatrix}$, the process of decryption started to calculate invertion of matrix *K*. Suppose the matrix 2 x 2, $K = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ if ad-bc ≠ 0, then: $K^{-1} = \frac{1}{ad-bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix} = \begin{pmatrix} \frac{d}{ad-bc} & \frac{-b}{ad-bc} \\ \frac{-c}{ad-bc} & \frac{a}{ad-bc} \end{pmatrix}$. Determinant from $\begin{pmatrix} 3 & 7 \\ 5 & 12 \end{pmatrix}$ is = 3×12-7×5 = 1 (mod 26) = 1. So the invertion of key matrix $K = \begin{pmatrix} 3 & 7 \\ 5 & 12 \end{pmatrix}$ is $K^{-1} = \begin{pmatrix} 3 & 7 \\ 5 & 12 \end{pmatrix}^{-1} = \begin{pmatrix} 12 & -7 \\ -5 & 3 \end{pmatrix} = \begin{pmatrix} 12 & 19 \\ 21 & 3 \end{pmatrix} \bmod 26 = \begin{pmatrix} 12 & 19 \\ 21 & 3 \end{pmatrix}$.

This inverse key matrix K^{-1} will be used to process decryption because fulfill the equation (1), that is identity matrix:

$$K. K^{-1} = \begin{pmatrix} 3 & 7 \\ 5 & 12 \end{pmatrix}\begin{pmatrix} 12 & 19 \\ 21 & 3 \end{pmatrix} = \begin{pmatrix} 183 & 78 \\ 312 & 131 \end{pmatrix} \pmod{26} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = I$$

For decrypting a cipher text *C* = M S I T Y O N H J P R F encoded utilizing the Hill Cipher, we need the inverse of the key matrix K^{-1} to the equation number 3. The process of decryption is same as the done for the encryption. The process of decryption is applied to each record in our database as in the encryption process.

Firstly conversion the alphabets in cipher text to numbers.
C = M S I T Y O N H J P R F
C = 12 18 8 19 24 14 13 7 9 15 17 5

The process of decryption encrypted database is shown in this section:
Plain Text for the first two letters of the cipher text in first record:

$$P_{1,2} = K^{-1} \cdot C_{1,2} = \begin{pmatrix} 12 & 19 \\ 21 & 3 \end{pmatrix}\begin{pmatrix} 12 \\ 18 \end{pmatrix} = \begin{pmatrix} 486 \\ 306 \end{pmatrix} \pmod{26} = \begin{pmatrix} 18 \\ 20 \end{pmatrix} = \begin{pmatrix} S \\ U \end{pmatrix}$$

Plain Text for the second two letters of the cipher text in first record:

$$P_{3,4} = K^{-1} \cdot C_{3,4} = \begin{pmatrix} 12 & 19 \\ 21 & 3 \end{pmatrix}\begin{pmatrix} 8 \\ 19 \end{pmatrix} = \begin{pmatrix} 457 \\ 225 \end{pmatrix} \pmod{26} = \begin{pmatrix} 15 \\ 17 \end{pmatrix} = \begin{pmatrix} P \\ R \end{pmatrix}$$

After all the letters in first record finish decrypting, so will get results of plain text:

P = 18 20 15 17 8 0 3 8 3 0 13 8
P = S U P R I A D I D A N I

Similarly, plain text for remaining letters in other record in this database, will get results of plain text:
P = 0 1 3 20 11 18 20 10 20 17
P = A B D U L S U K U R
P = 0 18 4 15 3 8 0 13
P = A S E P D I A N
P = 19 8 13 0 19 0 11 8 18 0
P = T I N A T A L I S A

Cryptanalysis to Hill Cipher Algorithm

In Cryptography, the study of the cryptanalysis is the crucial step, which enables us to decide the strength of a cipher. In the literature of cryptography, the various well known attacks for finding the strength of a cipher are

1. ciphertext only attack (brute force attack),
2. known plaintext attack,
3. chosen plaintext attack, and
4. chosen ciphertext attack.

In all these attacks, the primary objective is to determine either the key or a function of the key so that the cipher can be broken. This is the desire of the cryptanalyst. Let us now consider, firstly, the ciphertext only attack.

Cryptanalysis to Hill Cipher Algorithm is very hard to do with ciphertext-only attack, especially if the key matrix K have the big size. This difficulty caused cipher text Hill Cipher does not have the pattern and each character in one block influencing the other character.

The technique that can be used to in cryptanalysis to Hill Cipher Algorithm is known-plaintext attack. If cryptanalysis have pairs plaintext and ciphertext which connected, then Hill Cipher Algorithm can be broken. But the very hard process is to determine the length of key matrix. This is become the one of strength in Hill Cipher Algorithm. Through the method to know the length of key matrix or with estimation and trial.

The worst possibility in Hill Cipher Algorithm is when a cryptanalysis have pairs plaintext and ciphertext which connected and then know the length of key matrix. With this information, cryptanalysis can break Hill Cipher easily. For example, cryptanalysis know the length of key matrix is 2 and have pairs plaintext and ciphertext as shown below:
P = S U P R I A D I D A N I
C = M S I T Y O N H J P R F

From before information, we know that the character SU in plaintext corresponding with the character MS, and the character PR corresponding with the character IT. To break can be done with linear equation. Suppose the key matrix can be written in the form:
$K = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$
So Plaintext P is:

$P = \begin{pmatrix} S & P \\ U & R \end{pmatrix} = \begin{pmatrix} 18 & 15 \\ 20 & 17 \end{pmatrix}$
And then Ciphertext C is:
$C = \begin{pmatrix} M & I \\ S & T \end{pmatrix} = \begin{pmatrix} 12 & 8 \\ 18 & 19 \end{pmatrix}$

Refer the equation (1), then linear equation will be given by the example is:
$C = K.P$
$\begin{pmatrix} 12 & 8 \\ 18 & 19 \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 18 & 15 \\ 20 & 17 \end{pmatrix}$
18a + 20b = 12 (i)
15a + 17b = 8 (ii)
18c + 20d = 18 (iii)
15c + 17d = 19 (iv)

With solve through four equations and using arithmetic modulo 26, so we get the values of a, b, c, and d:
a = 3, b = 7, c = 5, d = 12

With these values a, b, c, dan d then the length of key matrix K is given by:
$K = \begin{pmatrix} 3 & 7 \\ 5 & 12 \end{pmatrix}$

With the key matrix K, cryptanalysis only need to decryption ciphertext totality in order to get whole plaintext.

CONCLUSIONS

From the above analysis, the writer conclude:

1. Hill Cipher Algorithm is classical cryptography algorithm with very strong in its security.
2. The key matrix hill cipher must be invertible matrix.
3. Hill cipher strong deal to ciphertext-only attack but weak deal to known-plaintext attack.
4. Cryptanalysis using linear equation is fastest technique, easy and accurate to break hill cipher compared to technique of multiplication matrix.
5. Computation in hill cipher enough complicated when calculated to big text.

Modification that writer have done to hill cipher become chaining hill cipher enough effective to add the strength of this classical cryptography algorithm that is using 29 characters (add the character space, point/dot and comma) and encryption process more complicated.

REFERENCES

Munir, Rinaldi. (2006) *Diktat Kuliah IF5054 Kriptografi*, Program Studi Teknik Informatika, Sekolah Teknik Elektro dan Informatika ITB: Bandung.

Scheneier, Bruce. (1996) *Applied Cryptography: Protocols, Algorithms, and Source Code in C 2nd Edition*, John Wiley & Sons: New York.

Forouzan, Behrouz. (2006) *Cryptography and Network Security*, McGraw-Hill: New York.

Silberschatz, A., Korth, H. F. and Sudarshan, S. (2002) *Database System Concepts, 4th Edition*, McGraw – Hill: New York.

Anton, Howard and Rorres, C. (2000) *Elementary Linear Algebra*, John Wiley & Sons: New York.

Fathansyah. (1999) *Basis Data*, Informatika: Bandung.

http://en.wikipedia.org/wiki/Hill_cipher

Marcus, T., Prijono, A. and Widiadhi, J. (2004) *Delphi Developer dan SQL Server 2000*, Informatika: Bandung.

Sadikin, Rifki. (2012) *Kriptografi untuk Keamanan Jaringan*, Andi Offset: Yogyakarta.