Journal of Computer Science 7 (5): 785-789, 2011
ISSN 1549-3636

A Novel Enhancement Technique of the Hill Cipher for Effective Cryptographic Purposes

[1]A.F.A. Abidin, [1]O.Y. Chuan and [2]M.R.K. Ariffin
[1]Department of Computer Science, Faculty of Informatics,
University Sultan Zainal Abidin,
21300 Kuala Terengganu, Terengganu, Malaysia
[2]Institute for Mathematical Research, University Putra Malaysia,
43400 UPM Serdang, Selangor, Malaysia

Abstract: Problem statement: The Hill cipher is the first polygraph cipher which has a few advantages in data encryption. However, it is vulnerable to known plaintext attack. Besides, an invertible key matrix is needed for decryption. It may become problematic since an invertible key matrix does not always exist. **Approach:** In this study, a robust Hill algorithm (Hill++) has been proposed. The algorithm is an extension of Affine Hill cipher. A random matrix key, RMK is introduced as an extra key for encryption. An algorithm proposed for involutory key matrix generation is also implemented in the proposed algorithm. **Results:** A comparative study has been made between the proposed algorithm and the existing algorithms. The encryption quality of the proposed algorithm is also measured by using the maximum deviation factor and correlation coefficient factor. **Conclusion/Recommendations:** The proposed algorithm introduced a random matrix key which is computed based on the previous ciphertext blocks and a multiplying factor. A modified of Hill Cipher is free from the all-zero plaintext blocks vulnerability. Usage of involutory key for encryption and decryption managed to solve the non invertible key matrix problem. It also simplify the computational complexity in term of generating the inverse key matrix.

Key words: Affine cipher, cryptography theory, hill cipher, involutory matrix, polygraph cipher, data encryption, symmetric cryptosystem, Advanced Encryption Standard (AES), Data Encryption Standard (DES), Initial Vector (IV), proposed algorithm

INTRODUCTION

Today, information is one of the most valuable intangible assets. Due to this fact, information security had become an important issue. Cryptography is one of the methods used to protect data from unauthorized access and being stolen. There are two types of cryptosystem, which are symmetric cryptosystem and asymmetric cryptosystem. In Symmetric cryptosystem, the sender and recipient share the same key. It means the same key is used for encryption and decryption. In Asymmetric cryptosystem, different keys are used. A public key is used by sender to encrypt the message while the recipient used a private key to decrypt it. Both of these cryptosystem have their own pros and cons. For instance, Symmetric cryptosystem consume less computing power but it is less secure than Asymmetric cryptosystem. Currently, there are a few cryptosystem which are widely implemented such as Advanced Encryption Standard (AES), Twofish, River Cipher 4 (RC4) and Data Encryption Standard (DES). However, these modern cryptosystem have their origins. The classical cipher such as Caesar Cipher, Hill Cipher, Vigenère Cipher act as the foundation for the cryptology's world today.

This study focuses on Hill Cipher which was first described in 1929 by its inventor, the mathematician Lester S. Hill, in the journal The American Mathematical Monthly (Eisenberg, 1998). Although its vulnerability to cryptanalysis has rendered it unusable in practice, it still serves an important pedagogical role in cryptology and linear algebra (Toorani and Falahati, 2009). Hill Cipher is the first polygraph cipher. It has a few advantages in data encryption such as resistant towards frequency analysis, high speed and high throughput. The core of Hill Cipher is matrix manipulation (Bibhudendra, 2006; Al-Saidi and Said, 2009; Pour *et al*., 2009; Sastry and Shankar, 2007). It is

Corresponding Author: A.F.A. Abidin, Faculty of Informatics, University Sultan Zainal Abidin, 21300 Kuala Terengganu, Terengganu, Malaysia

linear algebra equation is C = K × P (mod m), where C represents cipher text row vector, K represents key matrix and P represents plaintext row vector respectively. For decryption, an inverse key matrix, K^{-1} is needed. This is the major drawback of Hill cipher since not every key matrix is invertible. Another vulnerability of the Hill cipher is it compromised to the known-plaintext attacks. This means the message encrypted can be broken if the attacker gains enough pairs of plaintexts and ciphertexts.

In this study, a modified version of Hill cipher is proposed to overcome all the drawbacks mentioned above. The proposed algorithm is an extension from Affine Hill cipher. The rest of this study is organized as follows.

Literature review: Since the Hill Cipher serves as an important pedagogical role in both cryptology and linear algebra, several researches have been done to improve the Hill cipher. Rushdi and Mousa (2009) had designed a robust cryptosystem algorithm for non-invertible matrices based on Hill cipher. The noninvertible key matrix problem is solved by converting each plaintext character into two cipher text characters. So with the decryption, the process involves the conversion of two cipher text characters into one plaintext character. Although this algorithm solved the non-invertible key problem, it is time-consuming as the decryption process involved the computation of an inverse key matrix. It will definitely delay the decryption process especially when it involved high dimensional key matrix.

Ismail *et al*. (2006) proposed a modified Hill cipher which used one-time-one key matrix to encrypt each plaintext blocks. In this algorithm, each plaintext block is encrypted by using its own key. This unique key is computed by multiplying the current key with a secret Initial Vector (IV). The multiplying operation is carried out row by row, thus the algorithm is named as Hill Multiplying Rows by Initial Vector (Hill MRIV). This algorithm is proved to yield better encryption quality. However, it is proved by Rangel-Romero *et al*. (2006) that this proposed algorithm is still vulnerable towards known-plaintext attack. Besides, Ismail *et al*. (2006) does not tackle the non-invertible key matrix problems which may lead to failure in decryption.

Bibhudendra (2006) also proposed an advanced Hill cipher algorithm (AdvHill) which is able to solve the non-invertible key matrix problem. To make sure every ciphertext block can be decrypted, an involutory key matrix is used for encryption. An involutory key matrix is a key which can be used for both encryption and decryption. It means an inverse key matrix is not needed for decryption and this definitely simplify the computational complexity and save the computational time. However, this algorithm still contains some of the major drawbacks of the original Hill cipher such as the vulnerability to known-plaintext attack. Besides, this algorithm is also not suitable to encrypt all-zeroes plaintext as C will always equals zero when P equals zero (Rangel-Romero *et al*., 2006).

MATERIALS AND METHODS

The proposed algorithm: The proposed algorithm is based on Affine Hill cipher, which is the combination of Hill cipher and affine cipher. Affine Hill cipher mixes the Hill cipher with a nonlinear affine transformation (Stindon, 2006). Differing from the Hill cipher, the plaintext is encrypted as C = PK + V (mod m). To produce a robust cryptosystem, we extend this encryption core. This extension will solve the non-invertible key matrix problem and increase the randomization of the algorithm to enhance its resistance towards common attacks. With the proposed algorithm, the plaintext is encrypted as C = PK + RMK (mod m) with three parameters, α, β and γ. The non-invertible key matrix problem is solved by implementing an algorithm proposed by Bibhudendra (2006). This algorithm produces an involutory key matrix which can be used for both encryption and decryption process. Obviously, it reduces the computational complexity in decryption. Besides, a random matrix key, RMK is needed for encryption apart from the involutory key matrix. This RMK is computed based on the previous ciphertext blocks and a multiplying factor. It enhances the security of the proposed algorithm as it increased the resistance of the algorithms to known plaintext attack.

Mathematical notation:

C = Ciphertext
P = Plaintext
K = n × n matrix key
RMK = Random n × n matrix key
α = 1st seed number
β = 2nd seed number
γ = Multiplying factor
n = Length of matrix rows and columns
m = Value of modulus, m > 1
b = Number of plaintext blocks

Proposed algorithm: The encryption algorithm is as follows:

- Randomly select α_i, β_i and γ_i

- Generate $\{x_1, x_2, \ldots, x_n\}$ based on α_i
- Produce a $\frac{n}{2}\times\frac{n}{2}$ matrix based on the set of numbers $\{x_1, x_2, \ldots, x_n\}$
- With this $\frac{n}{2}\times\frac{n}{2}$ matrix, generate an involutory matrix K_i based on the algorithm proposed by Bibhudendra *et al*. (2009)
- Generate $\{y_1, y_2, \ldots, y_{n\times n}\}$ based on β
- Produce a n × n matrix, $Ktemp_1$ based on the set of numbers $\{y_1, y_2, \ldots, y_{n\times n}\}$
- Compute RMK_1 by multiplying each rows of $Ktemp_1$ matrix key with γ_1
- For the first plaintext block P_1, the encryption formula is:

$$C_1 = P_1K_1 + RMK_1 \pmod m$$

Repeat Step 1 to 4 to produce the remaining involutory matrix key, $\{K_2, K_3, \ldots, K_b\}$. To compute the remaining n × n random matrix key, $\{RMK_2, RMK_3, \ldots, RMK_b\}$, the formula is: $RMK_i = C_{i-1} \times \gamma_i$, where $i = \{2, 3, \ldots, b\}$

The proposed encryption algorithm is as follows:

$$C_i = \begin{cases} C_i = P_iK_i + RMK_i \pmod m, & i = 1 \\ C_i = P_iK_i + RMKi \pmod m \end{cases}$$

where, $RMK_i = C_{i-1} \times \gamma_i$, $i = \{2, 3, \ldots, b\}$.

The decryption algorithm is as follows:

- Generate $\{x_1, x_2, \ldots, x_n\}$ based on α_i
- Produce a $\frac{n}{2}\times\frac{n}{2}$ matrix based on the set of numbers
- $\{x_1, x_2, \ldots, x_n\}$
- With this $\frac{n}{2}\times\frac{n}{2}$ matrix, generate an involutory matrix K_i based on the algorithm proposed by Bibhudendra *et al*. (2009).
- Generate $\{y_1, y_2, \ldots, y_{n\times n}\}$ based on β_1
- Produce a n × n matrix, $Ktemp_1$ based on the set of numbers $\{y_1, y_2, \ldots, y_{n\times n}\}$
- Compute RMK_1 by multiplying each rows of $Ktemp_1$ matrix key with γ_1
- For the first plaintext block P_1, the decryption formula is:

$$P_1 = (C_1 - RMK_1) \times K_1 \pmod m$$

To compute the remaining n × n random matrix key, $\{RMK_2, RMK_3, \ldots, RMK_b\}$, the formula is: $RMK_i = C_{i-1} \times \gamma_i$, where $i = \{2, 3, \ldots, b\}$.

The proposed decryption algorithm is as follows:

$$P_i = \begin{cases} P_i = (C_i - RMK_i) \times K_i \pmod m, & i = 1 \\ P_i = (C_i - RMK_i) \times K_i \pmod m \end{cases}$$

where, $RMK_i = C_{i-1} \times \gamma_i$, $i = \{2, 3, \ldots, b\}$.

RESULTS AND DISCUSSION

Comparison has been done between the proposed encryption algorithm, Hill++ and a few others previous Hill algorithms. Table 1 shows that Hill++ has features which obviously overcome some of the vulnerabilities in the existing Hill algorithms.

Obviously, Table 1 shows that Hill++ is the better than all the previous Hill algorithms as it fulfilled the two comparison factors. Since the algorithm (Bibhudendra, 2006) is implemented in Hill++, thus an inverse key matrix is not needed for the decryption process. Apart from this, all the algorithms including the algorithm proposed by Bibhudendra (2006) suffered a same problem. When all the characters in a plaintext block are zeroes, all these algorithms may become problematic as if P equals zero, C will also become 0. It means the encryption process will totally failed. However, Hill++ still have the ability to encrypt the plaintext effectively even if P equals zero.

Measuring factors of encryption's quality: The comparison done on previously had showed that Hill++ is better compared to the existing Hill algorithms. But, depending on this comparison only is not enough in judging the encryption quality of the proposed algorithm. Thus, two measuring techniques have been used to evaluate the encryption quality of the proposed algorithm. These techniques are the maximum deviation measures Ziedan *et al*. (2003) and the correlation coefficient measures.

Table 1: Comparison between Hill++ and existing hill algorithm

	Comparison factor	
Cipher	Need inversed key matrix	Vulnerable if there are all-zero blocks
Original hill	Yes	Yes
Ismail et al.'s algorithm	Yes	Yes
Rushdi et al.'s algorithm	Yes	Yes
Bibhudendra et al.'s algorithm	No	Yes
Hill++	No	No

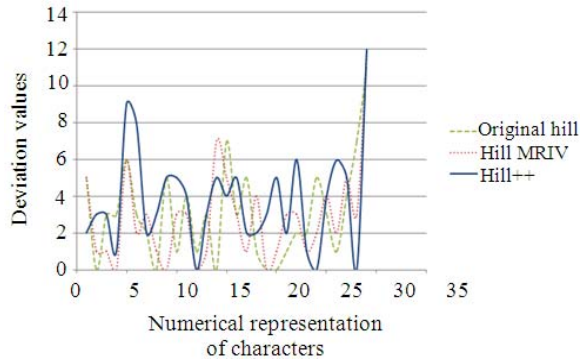

Fig. 1: Deviation graph between plaintext and ciphertext encrypted using different algorithms

Table 2: The evaluation for encryption quality of original Hill cipher, HillMRIV and Hill++

	Measuring factor	
Cipher	MF1	MF2
Original hill	80.0	0.07553
Hill MRIV	73.5	0.06059
Hill++	103.0	-0.00081

The maximum deviation factor measures the quality of encryption based on the deviation between the plaintext and ciphertext. The more is the ciphertext deviated from the plaintext, the better is the encryption algorithm. The correlation coefficient factor measures the quality of encryption based on the relationship between two variables, which are plaintext and ciphertext in this case. If the correlation coefficient is one, it means that the plaintext and ciphertext are highly dependent. If the correlation coefficient is zero, it means that the ciphertext and the plaintext are not correlated (independent between these two variables). Thus, the smaller is the values of the correlation coefficient, the better is the quality of encryption.

Results and analysis: In our experiment, a simple message with a total of one hundred and twelve characters is encrypted by using three different version of Hill cipher algorithm. These algorithms are the original Hill, HillMRIV and Hill++. The results are presented both in graph and table. Graph in Fig. 1 showed three curves which represent the absolute difference between plaintext and ciphertext (encrypted by using three different algorithms, original Hill, HillMRIV and Hill++ respectively). Based on the graph, the area under the absolute difference curve can be computed. The wider is the area under the absolute difference curve, the better is the quality of encryption. The sum of deviation (area) computed from this graph is presented under the MF1 column in Table 2.

Table 2 showed the comparisons of the results from two different measuring factors. In Table 2, MF1 indicates the maximum deviation measure while MF2 represent the correlation coefficient measure. From the table, it showed that Hill++ has the greatest MF1 and its MF2 is the closest to zero. A negative value of correlation coefficient means that the two variables (plaintext and ciphertext) are negatively correlated. The MF2 of Hill++ which is nearly zero showed that Hill++ can encrypt plaintext into totally different ciphertext. With the greatest MF1 and MF2 which is the closest to zero, it can be concluded that Hill++ is a better algorithm compared to the original Hill and HillMRIV.

CONCLUSION

We have presented a modified version of Hill cipher which is an extension of Affine Hill cipher. We called it as Hill++. Hill++ introduces a random matrix key which is computed based on the previous ciphertext blocks and a multiplying factor. This significantly increased the resistance of the algorithm to the known plaintext attack. Hill++ also implemented an involutory key generation algorithm where the same key matrix can be used for both encryption and decryption. It means Hill++ does not require any additional operation to compute an inverse key, which is definitely more time-saving. The comparison showed that Hill++ is the only algorithm which full filled both comparison factors, which are no inverse key needed and not vulnerable if there are all-zeroes block. Statistical analysis presented also showed satisfactory results. Hill++ has better encryption quality compared to the original Hill cipher and HillMRIV.

REFERENCES

Al-Saidi, N.M.G. and M.R.M. Said, 2009. A new approach in cryptographic systems using fractal image coding. J. Math. Stat., 5: 183-189. DOI: 10.3844/jmssp.2009.183.189

Bibhudendra, A., 2006. Novel methods of generating self-invertible matrix for hill cipher algorithm. Int. J. Secur., 1: 14-21. http://dspace.nitrkl.ac.in:8080/dspace/handle/2080/620

Bibhudendra, A., K.P. Saroj, K.P. Sarat and P. Ganapati, 2009. Image encryption using advanced hill cipher algorithm. Int. J. Recent Trends Eng., 1: 663-667. http://www.ijrte.academypublisher.com/vol01/no01/ijrte0101663667.pdf

Eisenberg, M., 1998. Hill ciphers and modular linear algebra. Mimeographed notes. University of Massachusetts. http://www.apprendre-en-ligne.net/crypto/hill/Hillciph.pdf

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

Pour, D.R., M.R.M. Said, K.A.M. Atan and M. Othman, 2009. The new variable-length key symmetric cryptosystem. J. Math. Stat., 5: 24-31. DOI: 10.3844/jmssp.2009.24.31

Rangel-Romero, Y., G. Vega-García, A. Menchaca-Méndez, D. Acoltzi-Cervantes and L. Martínez-Ramos *et al*., 2006. Comments on How to repair the Hill cipher. J. Zhejiang Univ. Sci. A., 9: 211-214. DOI: 10.1631/jzus.A072143

Rushdi, A.H. and F. Mousa, 2009. Design of a robust cryptosystem algorithm for non-invertible matrices based on hill cipher. Int. J. Comput. Sci. Network Secur., 9: 11-16. http://paper.ijcsns.org/07_book/200905/20090502.pdf

Sastry, V.U.K. and N.R. Shankar, 2007. Modified hill cipher with interlacing and iteration. J. Comput. Sci., 3: 854-859. DOI: 10.3844/jcssp.2007.854.859

Stindon, D.R., 2006. Cryptography Theory and Practice. 3rd Edn., Chapman and Hall/CRC, ISBN: 1584885084, pp: 593.

Toorani, M. and A. Falahati, 2009. A secure variant of the hill cipher. Proceedings of the 40th IEEE Symposium on Computers and Communications Sousse, July 5-8, Tunisia, pp: 313-316. ISSN: 1530-1346

Ziedan, I.E., M.M. Fouad and D.H. Salem, 2003. Application of data encryption standard to bitmap and JPEG images. Proceedings of the 20th National Radio Science Conference, Mar. 18-20, Cairo, Egypt, pp: 1-8. http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=1217349