Al-Quds University
Deanship of Graduate Studies
Electronic and Computer Engineering Master Program

Thesis Approval

**Self Generating Multi Key Cryptosystem
For Non-Invertible Matrices based on Hill Cipher**

Prepared By

Student Name: Mousa Mohammad Farajallah
Registration No: 20714094

Supervisor: Dr. Rushdi Hamamreh

Master thesis submitted and accepted Date: 31-01-2010

The names and signatures of the examining committee members are as follows:

1- Dr. Rushdi Hamamreh	: Head of Committee	Signature: .....
2- Dr. Raid Yousef Zaghal	: Internal Examiner	Signature: .....
3- Dr. Mohammed Aldasht	: External Examiner	Signature: .....

Jerusalem- Palestine

1431 - 2010

Abstract

Information security is an important issue. Recently many social aspects have developed reliance on networking infrastructures. Health, finance, and many other sectors are using local networks and the global Internet. The security of this infrastructure has been called into a question over the last decade. In particular, how to send data securely.

The Hill cipher model is one of the famous symmetric cryptosystems since invented till now, which can be used to protect information from unauthorized access. It was invented by Lester S. Hill in 1929, and it was the first polygraph cipher which was practical to operate on more than three symbols at once. But, in spite of that timeworn invention, only a few systems used it, and the reason is due to that Hill cipher requires the inverse of the key matrix for decryption. This inverse is dependant to the suggested number, and not a normal inverse, since the matrix that has not a prime determinant relative to the previous suggested number has not an inverse. So not any matrix has such an inverse. Therefore, a limited range of matrices are eligible to be used in the Hill cipher scheme. Furthermore, due to its linear nature, the basic Hill cipher succumbs to known-plaintext attacks [13, 22, 26].

On the other hand, many advantages of Hill Cipher model encourage working on it. First, it is resistant to the frequency letter analysis. It's also very simple since it uses matrix multiplication. Finally, it provides high speed and high throughput. However, non-invertible key matrix is the main disadvantage of Hill Cipher, this problem leads to many other sub problems such as the encrypted text can't be decrypted. Second, limitation on selecting the key matrix of Hill Cipher, and easy to discover key by known-plaintext attacks.

The key contribution of this thesis is the development of new technique to modify Hill Cipher algorithm to overcome its major problem, none-invertible key matrix. This new technique, depends on changing the method of finding the inverse of the matrix into a normal mathematic inverse, and modifying other steps in the original Hill cipher to accommodate the changes in calculating the inverse of the matrix. These modifications include the way of processing plaintext vector, which is modified so that the plaintext vector used in encryption side must be double of the matrix size, and this vector is divide

into two vectors, one is multiplied by the modular value, and is added to the other vector. A lot of changes are made on the original Hill cipher, after all these changes the new algorithm of Hill cipher called Mousa-Rushdi-Hill Cipher (MRHC).

A second important contribution is the improvement of Hill Cipher security against known pair of plaintext-cipher text attack. This improvement in security was possible by using key generation idea, which is the multi key generation process, depending on one of secure hash functions, which is secure hash algorithm 512 bit (***SHA-512***), it is used to produce 128 different numbers as an input for the key matrix generation code.

ملخص الرسالة

يعد امن البيانات من أهم القضايا في العصر الحديث، حيث أن كثيرا من المؤسسات تعتمد على الشبكات المحلية وشبكة الانترنت في كثير من جوانب عملها، لذلك تطور خلال الآونة الأخيرة سؤال يطرح نفسه! كيف نحمي البيانات من الاختراق.

خوارزمية (Hill Cipher) هي من أشهر خوارزميات التشفير المتماثل منذ اختراعها وحتى يومنا هذا والتي تستخدم لتشفير البيانات لمنع الأشخاص غير المرخص لهم من الاطلاع على البيانات المرسلة عبر الانترنت، اخترعت عام 1929 من قبل (Lester S. Hill) ، حيث أنها أول خوارزمية تعمل على تشفير أكثر من ثلاثة رموز دفعة واحدة، ولكن على الرغم من قدم اختراعها إلا أنها قليلة الاستخدام والسبب يعود إلى أنها بحاجة إلى معكوس كل مصفوفة مستخدمة في تشفير البيانات من اجل فك تشفير تلك البيانات. وليس جميع المصفوفات لها معكوس، والمعكوس هنا ليس المعكوس العادي، وإنما المعكوس نسبة إلى أي رقم، حيث انه ليس لهذه الخوارزمية معكوس عندما تكون محددة المصفوفة ليست أولية نسبة إلى ذلك الرقم، وبالتالي احتمالات فشل المصفوفات كمفتاح لهذه الخوارزمية كبير جدا. بالإضافة إلى هذا الخلل الكبير فانه ولطبيعة خوارزمية (Hill Cipher) المعتمدة على الجبر الخطي فمن السهل اكتشاف مفتاح التشفير في حال كان لدى المخترق جزء من النص الأصلي والجزء المشفر المقابل لهذا النص.

أمور كثيرة شجعتني للعمل على حل هذه الخوارزمية، حيث أنها تحتوي على الكثير من الحسنات والايجابيات، فهي مقاومة لتحليل تكرارية الحروف ،إضافة إلى أنها بسيطة جدا في الحسابات والتعامل لاعتمادها على الجبر الخطي البدائي، وأخيرا فان إنتاجيتها عالية جدا وتشفير البيانات من خلالها سريع أيضا، وحيث إن هذه الخوارزمية تحوي خللا كبيرا لبعض المصفوفات أدى إلى عدم القدرة على استرجاع البيانات المشفرة من خلال هذا المفتاح، كما أدى هذا الخلل إلى مشاكل وسلبيات فرعية كثيرة ، مما حفزني على محاولة إيجاد حل لهذا الخلل الرئيسي والذي من خلاله تمكنت من حل المشاكل الفرعية الأخرى، حيث أن النتائج التي حصلنا عليها في هذه الرسالة أثبتت صحة هذا النموذج الرياضي الجديد المتين.

Contents

1.1 Overview

The networks come by the need of sharing resources and exchanging data between computers. Indeed; we need to ensure that this exchange is secure, this security can be achieved by using one or more encryption software, and since this age called information age, we need to protect this information from unauthorized access or changes, and only authorized users can access this information at any time, which means critical data need sufficient protection and security.

Two major changes affecting the need of security in our word. First, with the introduction of computer, we need software for protecting data and files, especially within the shared systems, and the need of security is even more urgent when computers can be accessed via public telephone network. Second, with the introduction of distributed systems, and using the network to carry data between terminal computers, between computer and computer, thus security is needed.

In order to have a secure network, cryptography techniques such as symmetric key encryption and asymmetric one, are being used. Symmetric and asymmetric encryption techniques will be discussed later in chapter two, but in brief words, symmetric key encryption is good developed and efficient, but the main problem is how to share secrete key. Asymmetric or public key encryption, needs higher computational, and also less efficient in encrypting large messages. Nevertheless, public key cryptography is used to solve sharing secrete key problem produced by symmetric key encryption.

Over the time, the cryptanalysis on internet and internet-attached systems rapidly grown, and attack techniques become more automated causing large damage, also crackers have been able to penetrate systems with less information, so the designer of cryptosystems try to increase the resistance of these systems to cryptanalysis, the penalty is the increased time and complexity of these cryptosystems.

1.2 Motivation

Information security is an important issue, Critical data needs sufficient protection and security, and with computer systems becoming widely spread and complex, the importance of data security has increased. Cryptology is the science of codes and ciphers, which contain many techniques to transmit data in a way that makes their content unreadable to anyone who doesn't have a permission to read or write on these data, and the cryptographic algorithms are the basis for protecting computer systems, through the network data transmissions [1]. Today's many social aspects have sensitive information, they are relying on networking infrastructures in data transmission and the security of this infrastructure is the most important challenge.

The challenge of data security has become more complex since the exchange of data has increased significantly in the last decade. Moreover, the number of data hackers who entered this area has increased significantly. Finally, the owners of competence in data encryption eager to design simple cryptographic algorithms in calculation. Also, excellent and powerful resistant to the hackers. Thus, these challenges itself are the best motivation to work in encryption and decryption algorithm, exactly in Hill-Cipher model, as mentioned before, the owners of competence in data encryption eager to design simple cryptographic algorithms, and Hill-Cipher contains this property, since it depends on a simple linear algebra, but the complex task is *how to make Hill-Cipher has a powerful resistant to hackers*, and *how to make all matrices eligible to act as key in Hill-Cipher*. These two challenges form a good motivation to work on this problem.

1.3 Problem Statement

Hill cipher is an application of modular linear algebra to cryptology. Many researches and papers tried to use Hill Cipher algorithm to build a comprehensive cryptosystem, since Hill cipher has many advantages. It's simple and easy since it uses multiplications of matrices. It's also fast and highly productive, also Hill cipher a very strong substitution technique against a cipher-only attack [32, 43].

However, Hill cipher has two compound problems, in which the second one indirectly depends on the first one. The first problem is that Hill Cipher requires an inverse of each matrix used in encryption side. This inverse not the normal mathematical inverse, it is an inverse relative to a modular value, this means not only the matrix that has zero determinant, but also all matrices have determinant value not prime relative to the modular value do not have such an inverse. So, many matrices have no such an inverse. Therefore, the secret key cannot be neither randomly nor mathematically produced. Because there will be uncertainty of the key validity, also when the key matrix not invertible, two different plaintext vectors will be transferred into the same cipher text vector, which adds another difficulty, since the recipient can't specify from which plaintext vector this cipher text vector come. The second problem is due to the key remains constant during the encryption process, it will be easy for the hacker to get it, simply by getting a pair of plaintext and cipher text, when the hacker has plaintext vector s, and s cipher text vector, where s is the key matrix size, it is easy to formulate plaintext matrix, cipher text matrix, and using matrix solution to get the secrete key [18].

1.4 Proposed Solution

The suggested solution of the problem statement, is a new encryption cryptosystem called ***MRHC*** presents a lot of modification to the original Hill cipher, by solving the major problem in Hill cipher (none-invertible key matrix), since ***nobody*** before could solve it, since when all matrices can be invertible. ***MRHC*** method provides higher degree of security, and gets benefit from simplicity of the original Hill cipher, except the method of finding inverse of the matrix, the Hill cipher will take very small time comparing to other systems. Actually, ***MRHC*** not only a modification of original Hill cipher, but also it is a new system based on the way of simplified encryption, with a new mathematical model that change the mechanism to deal with key matrix.

1.5 Research Methodology

During this thesis, we have analyzed the original Hill Cipher models, and pointed out the problems of original Hill Cipher, three techniques are developed, which add a lot of modifications on the original Hill Cipher to overcome these problems. The three techniques are implemented using proposed mathematical model, then show some restriction of second technique on the modular value of the system. Finally, one numerical example on each technique is introduced.

Assimilation is build for the three techniques of our system MRHC, AES and original Hill cipher in matlab version 7.6.0.324, after simulate these codes we analysis the results of simulations, from the viewpoint of encryption time, and from viewpoint of security level. We use a very huge sample of data to make our result as correct as possible, and to complete simulation result before deadline, we use forty homogenous PCs at ***Palestine polytechnic University*** from 11-October-2009 until 14-December-2009.

In this thesis, simulation consists of two steps. First, compare the five encryption algorithms. Second, exclude the original Hill Cipher, since we found that the original Hill Cipher takes more time than other algorithms, no matter the change of values of system parameters, and the time curve increasing exponentially, while it increases slowly at other algorithms. Also, a second reason of excluding original Hill Cipher, but not a core reason, that the key matrix can't be randomly selected. So, restricting other algorithms with only keys that are eligible for the original Hill Cipher, and as a result sample of data never can be huge enough to reach credible and reliable results.

To analyze results of matlab simulation, we have used the mean equation degrees, to produce tables and figures comparing between five algorithms. also we have defined a new parameters to measure the factor of change matrix key on encryption time, and another factor that measure the number of calls ***SHA-512***.

1.6 Objectives

This section we summarizes the objectives of the thesis as follow:

1. Study characteristics of the original Hill Cipher, analyze the mathematical model, and give one example on each analysis. Finally, identify points of failure in this algorithm.

2. Develop new cryptosystem depending on the original Hill Cipher, that benefits from all original Hill Cipher advantages, and remove disadvantages, this new cryptosystem has three techniques, where each one is suitable for some range of parameters, parameters are:

   2.1. Plaintext length.
   2.2. Key matrix size.
   2.3. Modular value of the system.

3. Unlike communication systems, that accept some error rate during data or image transfer, in cryptosystems no error can be tolerated, this is due to decrypting cipher text must exactly same as plaintext, this idea is taken into account in our system to ensure no bit error detection.

4. Our new cryptosystem will be available via internet, for security applications such as Kerberos software, for authentication tickets application, and for many other security based environment, this is one of the future work.

5. Compare the performance of ***MRHC*** three techniques, ***AES*** and Original Hill Cipher using a high-performance language for technical computing (Matlab), in order to test our techniques.

6. Define powerful, work area of each algorithm, to allow dynamic switching between techniques, when developing comprehensive and complete standalone software as future work.

1.7 Contributions.

Many studies tried to overcome none-invertible matrix of Hill Cipher. Bibhudendra Acharya and his colleagues were tried to use self-invertible matrix, for image encryption, but this technique is restricted only for self-invertible matrix, also Chu-Hsing Lin and his colleagues were tried to enhance the security of Hill Cipher using one-way hash function which was a very good idea, but still cannot overcome the none-invertible matrix problem, in this thesis we have tried to get benefit from all previous works. At the end of this work we can summarize many contributions:

- Develop new technique to modify Hill Cipher algorithm, to overcome its major problem-noninvertible key matrix, this new technique; depends on changing the method of finding the inverse of the matrix, into normal mathematical inverse. Then, change other steps in the original Hill cipher to accommodate the changes in calculating the inverse of the matrix. One of these changes is the way of processing plaintext vector, where the plaintext vector used in encryption side is double of the matrix size, and this vector is divided into two vectors, one is multiplied by the modular value of the system that selected, and adding it to the other vector, a lot of other changes are made in the original Hill cipher.

- Another important contribution, which improves the security of Hill Cipher against known plaintext attack. This enhancement in security is possible by using key generation idea based on multi key generation process, which depends on one of secure hash functions, the powerful one, ***SHA-512***, which is used it in a way to produce 128 different numbers as an input of key matrix generation code.

- Remove known plaintext attack, since the number of equations and data available to the hackers are half number of unknowns, this means it is not possible for the crackers to calculate the key matrix using mathematical model, and the cracker on average needs 5.43e+12 encryption process this means the probability of know only one key matrix of size 3×3 when

modular value equals to 26 is 1.8416e-013, this number is less than AES, but if we change the modular value into 52 (only capital and small letters in English) and key matrix into 5× 5, then the cracker needs 7.95e+42 encryption process and the probability is 1.2579e-043 , and this more than security of AES by six million times.

- Get benefit from the simplicity of encryption process of Hill cipher to show that ***MRHC*** algorithm takes less time than ***AES*** and original Hill cipher.

- Develop three techniques of ***MRHC***, and define the critical points of these techniques. The critical points means which techniques are suitable for certain key matrix size, range of plaintext length, and modular value of the system.

6.1 Conclusion

Cryptography consider one of the top hot research subjects, and exactly the security of encryption algorithm, it is well known, that there is a tradeoff between the quality and the time of encryption algorithm, in this thesis Hill cipher is modified to increase the security to reach ***AES*** in some cases, and increase over ***AES*** in other cases, while the time of our algorithm maintain a value less than ***AES*** when the key matrix size is smaller than the critical point ($s \leq 9$) mentioned at the simulation analysis, consequently we summarize a number of conclusions from this work:

- Due to the simplicity of Hill Cipher algorithm, the time of encryption and decryption algorithm, will smaller than other algorithms.

- In this thesis we introduce a new method to overcome the non-invertible matrices problem, and as a result three techniques are obtained, where each one is modification of others.

- From the three techniques implemented in our system, second one is the best with respect to the time consumption for ($s \leq 4$), while third one after this point is converted to the best.

- By solving none-invertible matrix problem, the space of the key matrix will be open include all permutations of s integer, where s is the key matrix size.

- Also, this thesis provides a very good solution to overcome the will-known plaintext attack problem, this solution, also depends on the key matrix size in a huge range while on modular value in small range.

- Security level of third technique is the best for all ranges, and when($s \geq 9$), ***AES*** consumes less time than all ***MRHC*** techniques.

6.2 Future Work

- From this work, exactly on cryptanalysis of ***MRHC***, we have learned and note that the major problem of original Hill cipher (none-invertible key matrix) can be used, to prevent the cracker from breaking Hill cipher, more work is needed to perform this work.

- From this thesis we noted that when increase key matrix size, the level of security will be increased notably, regardless of the modular value of the system, so if we can increase key matrix size and decrease number of iteration by minimize number of plaintext vector, it is likely that the modified algorithm time will be enhanced.

- Develop a complete cryptosystem web based application, based on ***MRHC*** algorithm.