Short Paper

COMMENTS ON SAEEDNIA's IMPROVED SCHEME FOR THE HILL CIPHER

Chu-Hsing Lin*, Chia-Yin Lee, and Chen-Yu Lee

ABSTRACT

In 2000, Saeednia proposed a new scheme to make the Hill cipher secure. The author makes use of permutations of columns and rows of a matrix to get a different key for encrypting each message. This paper shows that the cipher key $\boldsymbol{H}_t$ can be obtained by parameter $\boldsymbol{u}$. Besides, the Saeednia's scheme costs a lot of time in matrix computation. To overcome the drawbacks of Saeednia's scheme, a more secure cryptosystem with a one-way hash function is proposed.

Key Words: Hill cipher, known-plaintext attack, cryptosystem, one-way hash function.

I. INTRODUCTION

The Hill cipher is a famous symmetric cryptosystem from the early days, which was invented by Lester S. Hill (1929; 1931). The cryptosystem is a simple linear transformation $\boldsymbol{XH}$ (mod p), where the key is an $m\times m$ nonsingular matrix $\boldsymbol{H}$ with $h_{ij}\in Z_p$, for a fixed $p>1$, and such that gcd(det $\boldsymbol{H}$ (mod p), p)=1, $\boldsymbol{X}$ is a $1\times m$ plaintext message, and p is a selected positive integer. The plaintext $\boldsymbol{X}$ is encrypted as $\boldsymbol{Y}=\boldsymbol{XH}$ (mod p), and the ciphertext $\boldsymbol{Y}$ is obviously decrypted as $\boldsymbol{X}=\boldsymbol{YH}^{-1}$ (mod p).

The following example shows how the Hill cipher works.

Let p=7, the plaintext message $\boldsymbol{X}$=[3 5], the cipher key $\boldsymbol{H}=\begin{bmatrix}4 & 4\\ 6 & 3\end{bmatrix}$ and $\boldsymbol{H}^{-1}=\begin{bmatrix}5 & 5\\ 4 & 2\end{bmatrix}$.

The ciphertext

$$\boldsymbol{Y}=\boldsymbol{XH} \pmod p=[0\ 6],$$

the plaintext message can be decrypted by

$$\boldsymbol{X}=\boldsymbol{YH}^{-1} \pmod p=[3\ 5].$$

The weakness of the Hill cipher is that the cryptosystem can be broken under the known-plaintext attack (Denning, 1982; Evertse, 1987; Yeh *et al.*, 1991). An analyzer knows only m pairs of plaintext-ciphertext, the cipher key $\boldsymbol{H}$ can be determined by solving the equations

$$\boldsymbol{H}=\boldsymbol{X}^{-1}\boldsymbol{Y}.$$

For example, assume the key $\boldsymbol{H}=\begin{bmatrix}4 & 4\\ 6 & 3\end{bmatrix}$ and let p=26, if we have two pairs of plaintext-ciphertext, $\boldsymbol{X}_1$=[2 9], $\boldsymbol{Y}_1$=[10 9] and $\boldsymbol{X}_2$=[3 5], $\boldsymbol{Y}_2$=[16 1], then we can compute the cipher key $\boldsymbol{H}$ by

$$\begin{bmatrix}\boldsymbol{X}_1\\ \boldsymbol{X}_2\end{bmatrix}^{-1}\begin{bmatrix}\boldsymbol{Y}_1\\ \boldsymbol{Y}_2\end{bmatrix} \pmod p$$

$$=\begin{bmatrix}15 & 25\\ 17 & 6\end{bmatrix}\begin{bmatrix}10 & 9\\ 16 & 1\end{bmatrix} \pmod{26}=\begin{bmatrix}4 & 4\\ 6 & 3\end{bmatrix}.$$

To overcome the weakness, Saeednia (2000) proposed a method, which uses random permutations of columns and rows of the key matrix. But the matrix multiplications are used many times in this method; it costs a lot of time to compute the matrix multiplication when the size of the matrix is too large.

*Corresponding author. (Tel: 886-4-23590121 ext. 3287; Fax: 886-4-23591567; Email: chlin@mail.thu.edu.tw)

C. H. Lin and C. Y. Lee are with the Department of Computer Science and Information Engineering, Tunghai University, Taichung, Taiwan 407, R.O.C.

C. Y. Lee is with the Department of Computer Science and Information Engineering, National Chiao Tung University, Hsinchu, Taiwan 300, R.O.C.

We propose a new scheme that uses a one-way hash function to solve existing problems in Saeednia's scheme.

II. SAEEDNIA'S SCHEME

When Alice and Bob want to communicate securely, first, they share an $m\times m$ nonsingular matrix $\boldsymbol{H}$ as the cipher key with $h_{ij}\in Z_p$, for a fixed $p>1$, and such that gcd(det $\boldsymbol{H}$ (mod p), p)=1.

If Alice wants to encrypt a plaintext message $\boldsymbol{X}$, she chooses a vector $\boldsymbol{t}$ ($t_i\in Z_p$) at random and using a predetermined permutation algorithm performs simultaneous permutations of the rows and columns of $\boldsymbol{H}$, according to $\boldsymbol{t}$, to produce the new key $\boldsymbol{H}_t$ (that may be seen as $\boldsymbol{H}_t=\boldsymbol{P}_t^{-1}\boldsymbol{H}\boldsymbol{P}_t$, where $\boldsymbol{P}_t$ is the $m\times m$ permutation matrix associated to $\boldsymbol{t}$). Using the key $\boldsymbol{H}_t$ to encrypt the message $\boldsymbol{X}$ as

$$\boldsymbol{Y}=\boldsymbol{H}_t\boldsymbol{X} \pmod p,$$

besides, computing a parameter $\boldsymbol{u}$ by

$$\boldsymbol{u}=\boldsymbol{H}\boldsymbol{t} \pmod p,$$

then sends the pair $(\boldsymbol{Y}, \boldsymbol{u})$ to Bob.

In order to decrypt the ciphertext, Bob starts to compute the permutation vector $\boldsymbol{t}$ by

$$\boldsymbol{t}=\boldsymbol{H}^{-1}\boldsymbol{u} \pmod p,$$

and uses $\boldsymbol{t}$ to obtain $(\boldsymbol{H}^{-1})_t$ form $\boldsymbol{H}^{-1}$. Then he can recover the plaintext message $\boldsymbol{X}$ by computing

$$\boldsymbol{X}=(\boldsymbol{H}^{-1})_t\boldsymbol{Y} \pmod p.$$

It is easy to see that $(\boldsymbol{H}^{-1})_t=(\boldsymbol{H}_t)^{-1}$, because $(\boldsymbol{H}_t)^{-1}$ is existent. Here we would note that since $\boldsymbol{H}_t=\boldsymbol{P}_t^{-1}\boldsymbol{H}\boldsymbol{P}_t$, we have

$$(\boldsymbol{H}_t)^{-1}=(\boldsymbol{P}_t^{-1}\boldsymbol{H}\boldsymbol{P}_t)^{-1}=\boldsymbol{P}_t^{-1}\boldsymbol{H}^{-1}\boldsymbol{P}_t \tag{1}$$

on the other hand,

$$(\boldsymbol{H}^{-1})_t=\boldsymbol{P}_t^{-1}\boldsymbol{H}^{-1}\boldsymbol{P}_t \tag{2}$$

from (1) and (2), we can see that $(\boldsymbol{H}_t)^{-1}=(\boldsymbol{H}^{-1})_t$.

III. TWO COMMENTS ON SAEEDNIA'S SCHEME

In the following, two comments on Saeednia's scheme are presented. The first comment shows that Saeednia's scheme has a weakness of the parameter $\boldsymbol{u}$. The second comment shows that Saeednia's scheme is not efficient enough.

Comment 1

From the parameter $\boldsymbol{u}=\boldsymbol{H}\boldsymbol{t}$ (mod p), a cryptanalyst is able to determine the matrix $\boldsymbol{H}$ with known-plaintext attack. This is the same problem as in the original Hill's method. By collecting m pairs of $(\boldsymbol{t}, \boldsymbol{u})$, a cryptanalyst can obtain the key $\boldsymbol{H}$. Further, the cryptanalyst can obtain the permutation matrix $\boldsymbol{P}_t$ associated to $\boldsymbol{t}$. Therefore, he can compute the cipher key $\boldsymbol{H}_t$ by

$$\boldsymbol{H}_t=\boldsymbol{P}_t^{-1}\boldsymbol{H}\boldsymbol{P}_t.$$

If $\boldsymbol{t}$ can not be obtained, then the cryptanalyst can collect m pairs of $(\boldsymbol{X}, \boldsymbol{Y})$ to obtain $\boldsymbol{H}_t$, where $\boldsymbol{Y}=\boldsymbol{H}_t\boldsymbol{X}$ (mod p). Besides, the cryptanalyst knows that $\boldsymbol{u}=\boldsymbol{H}\boldsymbol{t}$ (mod p) and $\boldsymbol{H}_t=\boldsymbol{P}_t^{-1}\boldsymbol{H}\boldsymbol{P}_t$, so he can obtain the following relations:

$$\boldsymbol{H}=[\boldsymbol{U}][\boldsymbol{T}]^{-1}, \text{ where } [\boldsymbol{U}]=[\boldsymbol{u}_1\ \boldsymbol{u}_2\ \cdots\ \boldsymbol{u}_m] \text{ and}$$

$$[\boldsymbol{T}]=[\boldsymbol{t}_1\ \boldsymbol{t}_2\ \cdots\ \boldsymbol{t}_m] \text{ are } m\times m \text{ matrices} \tag{3}$$

$$\boldsymbol{H}_t=\boldsymbol{P}_t^{-1}\boldsymbol{H}\boldsymbol{P}_t \Leftrightarrow \boldsymbol{H}=\boldsymbol{P}_t\boldsymbol{H}_t\boldsymbol{P}_t^{-1} \tag{4}$$

from (3) and (4), the equations can be rewritten as

$$[\boldsymbol{U}][\boldsymbol{T}]^{-1}=\boldsymbol{P}_{t_1}\boldsymbol{H}_{t_1}\boldsymbol{P}_{t_1}^{-1}$$

$$[\boldsymbol{U}][\boldsymbol{T}]^{-1}=\boldsymbol{P}_{t_2}\boldsymbol{H}_{t_2}\boldsymbol{P}_{t_2}^{-1}$$

$$\vdots$$

$$[\boldsymbol{U}][\boldsymbol{T}]^{-1}=\boldsymbol{P}_{t_m}\boldsymbol{H}_{t_m}\boldsymbol{P}_{t_m}^{-1}$$

Assume that the predefined function $\boldsymbol{t}\Rightarrow\boldsymbol{P}_t$ is known and $[\boldsymbol{T}]^{-1}$ exists, and then the parameter $\boldsymbol{t}$ can be obtained by solving the above m equations. It means that the cryptanalyst can collect m pairs of parameters to solve the equations $[\boldsymbol{U}][\boldsymbol{T}]^{-1}=\boldsymbol{P}_{t_m}\boldsymbol{H}_{t_m}\boldsymbol{P}_{t_m}^{-1}$ and m pairs to obtain each $\boldsymbol{H}_t$ from $\boldsymbol{Y}=\boldsymbol{H}_t\boldsymbol{X}$ (mod p). Finally, we can obtain the key $\boldsymbol{H}$ by m^2 know-plaintext pairs $(\boldsymbol{u}, \boldsymbol{X}, \boldsymbol{Y})$.

Comment 2

Saeednia uses many matrix multiplications to encrypt and to decrypt a message in his scheme; like the cipher key $\boldsymbol{H}_t$ is produced by $\boldsymbol{H}_t=\boldsymbol{P}_t^{-1}\boldsymbol{H}\boldsymbol{P}_t$. When the size of matrix $\boldsymbol{H}$ is too large, it requires a lot of time to compute the matrix multiplication and inversion. We will analyze the complexity in Section VI.

IV. THE PROPOSED SCHEME

To overcome the weakness of Saeednia's

Table 1 The time comparison of the cryptosystem

	Saeednia's scheme	Our scheme
Encryption	$1T_{M\_INV}+2(m^3+m^2)T_{MUL}+2(m^3-m)T_{ADD}$	$(m^3+m)T_{MUL}+(m^3-m^2+m)T_{ADD}+(m+1)T_{hash}$
Decryption	$2T_{M\_INV}+2(m^3+m^2)T_{MUL}+2(m^3-m)T_{ADD}$	$1T_{M\_INV}+(m^3+m)T_{MUL}+(m^3-m^2+m)T_{ADD}+(m+1)T_{hash}$

scheme, we use two encryption parameters (h_{ij}, $\boldsymbol{V}$) in the proposed scheme, where h_{ij} is picked up in random, and $\boldsymbol{V}$ is generated from h_{ij} with a one-way hash function.

Suppose that two people, Alice and Bob, want to communicate securely. First, they share a common cipher key $\boldsymbol{H}$, which is a $m\times m$ nonsingular matrix with $h_{ij}\in Z_p$, for a fixed $p>1$, and satisfies gcd (det $\boldsymbol{H}$ (mod p), p)=1. In order to encrypt a plaintext message $\boldsymbol{X}$, $\boldsymbol{X}=[\boldsymbol{x}_1\ \boldsymbol{x}_2\ \cdots\ \boldsymbol{x}_m]$, Alice chooses a random integer a let $0<a<p$, and uses a one-way hash function $f(x)$, e.g. SHA (FIPS 180-2, 2002), to computes the parameter b by $b=f(a||h_{11}||h_{12}||\cdots||h_{ij}||\cdots||h_{mm})$, where h_{11}, h_{12}, $\cdots$, h_{ij}, $\cdots$, h_{mm} are the elements of $\boldsymbol{H}$.

Using b to pick up the ij^{th} element h_{ij} from $\boldsymbol{H}$ (that may be seen as $i=(\left\lfloor\frac{b-1}{m}\right\rfloor \pmod m)+1$, $j=b-\left\lfloor\frac{b-1}{m}\right\rfloor\times m$, where m is the dimension of the matrix $\boldsymbol{H}$). Then she uses h_{ij} to generate an element of vector $\boldsymbol{V}=[v_1\ v_2\ \ldots\ v_m]$ for m times, where the elements are

$$v_1=f(h_{ij}) \pmod p$$

$$v_2=f(v_1) \pmod p=f^2(h_{ij}) \pmod p$$

$$v_3=f(v_2) \pmod p=f^3(h_{ij}) \pmod p.$$

$$\vdots$$

$$v_m=f(v_{m-1}) \pmod p=f^m(h_{ij}) \pmod p.$$

Then, she encrypts the plaintext message $\boldsymbol{X}$ as

$\boldsymbol{Y}=h_{ij}\boldsymbol{XH}+\boldsymbol{V}$ (mod p), where p is a prime number, and sends the pair ($\boldsymbol{Y}$, a) to Bob.

In order to decrypt the ciphertext $\boldsymbol{Y}$, Bob first computes b by $b=f(a||h_{11}||h_{12}||\cdots||h_{ij}||\cdots||h_{mm})$, and uses b to pick up the ij^{th} element h_{ij} from $\boldsymbol{H}$. He also generates each element v_k ($1\le k\le m$) of $\boldsymbol{V}$ from h_{ij} the same way as in the encryption scheme.

Finally, he can recover the plaintext message by computing

$$\boldsymbol{X}=h_{ij}^{-1}\boldsymbol{H}^{-1}(\boldsymbol{Y}+\boldsymbol{V}) \pmod p.$$

V. ANALYSIS OF KNOWN-PLAINTEXT ATTACK

In the original Hill cipher, an analyzer can use known-plaintext attack to obtain the cipher key $\boldsymbol{H}$. However, in our system, it is hard to use known-plaintext attack for the following reason:

Due to the ciphertext $\boldsymbol{Y}=(\boldsymbol{CXH}+\boldsymbol{V})$ (mod p), where the parameters $C_i=(h_{ij})_i$ described in previous section, the equations can be written as

$$\boldsymbol{Y}_1=\boldsymbol{C}_1\boldsymbol{X}_1\boldsymbol{H}+\boldsymbol{V}_1 \pmod p$$

$$\boldsymbol{Y}_2=\boldsymbol{C}_2\boldsymbol{X}_2\boldsymbol{H}+\boldsymbol{V}_2 \pmod p$$

$$\vdots$$

$$\boldsymbol{Y}_m=\boldsymbol{C}_m\boldsymbol{X}_m\boldsymbol{H}+\boldsymbol{V}_m \pmod p.$$

Although the analyzer knows m pairs of ($\boldsymbol{X}_i$, $\boldsymbol{Y}_i$) ($1\le i\le m$), the cipher key $\boldsymbol{H}$ and two encryption parameters C and $\boldsymbol{V}$ are unknown. It means that m equations can't be used to solve an $m\times m$ nonsingular matrix $\boldsymbol{H}$ and $2m$ unknown parameters. Therefore, the analyzer can't use the known-plaintext attack to break our scheme.

VI. PERFORMANCE ANALYSIS

We define some notations as follows.

T_{MUL} : the time for the scalar modular multiplication.

T_{ADD} : the time for the scalar modular addition.

T_{M_INV} : the time for the modular inversion of a $m\times m$ matrix.

T_{hash} : the time taken by the hash function $f(x)$

We use an $m\times m$ nonsingular matrix as the ciphering key and obtain the results shown in Table 1. In Table 1, those operations of the modular inversion of a $m\times m$ matrix, which are equal to ($2m^3$) times the scalar modular multiplication and ($2m^3-2m^2$) times the scalar modular addition (Using the Gaussian elimination method). On the other hand, we know that operations of the one-way hash function are much faster than modular matrix inversion, so our scheme is more efficient.

VII. CONCLUSION

The Hill cipher is a famous cryptosystem, which is efficient and easy to implement. However, it is easy to break by known-plaintext attack. In this paper, we have presented an improved scheme to

make the Hill cipher secure. The characteristics of our scheme are more security and efficiency than Saeednia's scheme.

NOMENCLATURE

$a\|\|b$	the concatenation of a and b
det $\boldsymbol{H}$	the determinant of a matrix $\boldsymbol{H}$
SHA	secure hash algorithm
Z_p	the set of positive integers: $\{0, 1, 2, ..., p-1\}$

REFERENCES

Denning, D. E., 1982, *Cryptography and Data Security*, Addison-Wesley, MA, USA.

Evertse, J. H., 1987, "Linear Structures in Blockciphers," *Advances in Cryptology – EUROCRYPT '87*, Lecture Notes in Computer Science (LNCS) Vol. 304, Springer-Verlag, pp. 249-266, The Netherlands.

Federal Information Processing Standard (FIPS) 180-2, 2002, "Secure Hash Standard," NIST, U. S. Department of Commerce.

Hill, L. S., 1929, "Cryptography in an Algebraic Alphabet," *American Mathematical Monthly*, Vol. 36, No. 6, pp. 306-312.

Hill, L. S., 1931, "Concerning Certain Linear Transformation Apparatus of Cryptography," *American Mathematical Monthly*, Vol. 38, No. 3, pp. 135-154.

Saeednia, S., 2000, "How to Make the Hill Cipher Secure," *Cryptologia*, Vol. 24, No. 4, pp. 353-360.

Yeh, Y. S., Wu, T. C., Chang, C. C., and Yang, W. C., 1991, "A New Cryptosystem Using Matrix Transformation," *Proceedings of 25th Annual IEEE International Carnahan Conference*, Taiwan, pp. 131-138.

Manuscript Received: Dec. 27, 2002
Revision Received: Oct. 17, 2003
and Accepted: Nov. 12, 2003