



Conventional HILL Algorithm:

From Classical Cryptography to Modern Cryptography

Matar Niane*

Information systems security expert, (Dakar, Senegal)

Email: nianemakhou99@gmail.com

Abstract

With the development of modern computer tools using standard character encoding which assigns each graphic character a number, in particular the written characters of human language, HILL's cipher, which, in its conventional version, uses the alphabet of 26 letters reaches its limits. It has also been demonstrated that the algorithm is fallible at a *frequency analysis attack* but also to a *known plaintext attack*.

For this purpose, in order to adapt the HILL's cipher to modern communication systems and to strengthen its cryptographical security, we are proposing in this article an improvement of the algorithm. The study will mainly focused on an expansion of the alphabet and the secret key. Thus, the standard character-encoding (example of the ASCII code) will be used instead of a 26-letter alphabet and the secret key extended to three parameters with the additional use of two randomly chosen numbers p_1 and p_2 from the set real numbers(\mathbb{R}).

Keywords: HILL Algorithm; Classic Cryptography; Modern Cryptography.

1. Introduction

HILL's cipher [1,2,3] is a polygram substitution cipher. The algorithm, in its conventional version, operates on groups of letters, belonging to an alphabet of 26 letters, by a linear system of equations. However, despite its purely mathematical nature, it has certain weaknesses, in particular against a *known plaintext attack* [4] but also against a *block frequency analysis attack*.

* Corresponding author.

In addition, because it operates by linear transformations, it does not satisfy the principle of confusion stated by Claude Shannon, which imposes the use of highly nonlinear relation [5,6].

Thus, in order to overcome these vulnerabilities and get it out of its classic grip, we present, in this article, an improvement of the conventional algorithm of the Hill Cipher so that it can meet the requirements of modern cryptography. The first improvement is to expand the secret key with the additional use of two randomized real numbers (p_1, p_2), in addition to the matrix G [3], to allow the algorithm to be more resistant to a *block frequency analysis attack* but also to better respond to the cryptographic principles of confusion and dissemination. The second improvement concerns the modernization of the algorithm by linking the alphabet to the standard character encoding of moderate digital media.

2. Example of Encryption and Decryption

The HILL cipher is a symmetric key encryption system [6]. It is then considered that the two correspondents *Alice* and *Bob* have already exchanged the secret key K for the encryption and decryption of their communication. It is also accepted that the value of the secret key is $K = (p_1 = 45, p_2 = 2258, G = \begin{pmatrix} 3 & 1 \\ 5 & 2 \end{pmatrix})$, $A = \begin{pmatrix} 2 & -1 \\ -5 & 3 \end{pmatrix}$ and that *Alice* account send with complete confidentiality to *Bob* the message $M = (\text{maman})$ using as alphabet the ASCII code of the characters.

- a. The dimension of the matrix G is $n = 2$ and the size of the plaintext is $l = 5$. Since the size of the plaintext (M) is not a multiple n , it is considered that *Alice* randomly chose the character s to complete the size of the plaintext at $L = 6$ which is a multiple of n .
- b. *Alice* calculates the values of the $X[i]$ from the relation (1.1)
- c. She then calculates the values of the $Y[i]$ by applying the relation (1.2) :
- d. The concatenation of $Y[i]$ obtained previously constitutes the antigram.

$$\begin{cases} \text{pour } i = 1 \text{ à } L \\ X[i] = p_1 * x[i] \\ \text{fin;} \end{cases} \quad (1.1)$$

The **table 1** gives the result of the calculations of the $X[i]$ and $Y[i]$.

Table 1: Determination of vectors $V[i]$ of the antigram

Message (x)	m	a	m	a	n	s
Code ASCII ($x[i]$)	109	97	109	97	110	115
$X[i] = p_1 * x[i]$	4905	4365	4905	4365	4950	5175
$Y[1] = X[1] \oplus p_2$ and $Y[i] = X[i] \oplus Y[i - 1]$	7163	2806	6623	2258	7044	4019

- e. *Alice* cuts the chain formed by all the ($Y[i]$) in k Vectors (V) of dimension $n = 2$. She then obtains the vectors $V[1], V[2]$ and $V[3]$ next:

$$V[1] = (7163 \quad 2806);$$

$$V[2] = (6623 \quad 2258);$$

$$V[3] = (7044 \quad 4019);$$

f. Alice applies the matrix product of the relation (1.3) on each of the vectors $V[j]$. The result gives the vectors $C[1]$, $C[2]$ and $C[3]$ represented on the Table 2 below.

$$\begin{cases} \text{pour } j = 1 \text{ à } k \\ C[j] = G \cdot V[j]; \\ \text{fin;} \end{cases} \quad (1.3)$$

$$C[1] = G \cdot V[1] = \begin{pmatrix} 3 & 1 \\ 5 & 2 \end{pmatrix} \cdot \begin{pmatrix} 7163 \\ 2806 \end{pmatrix} = \begin{pmatrix} 24295 \\ 41427 \end{pmatrix}$$

$$C[2] = G \cdot V[2] = \begin{pmatrix} 3 & 1 \\ 5 & 2 \end{pmatrix} \cdot \begin{pmatrix} 6623 \\ 2258 \end{pmatrix} = \begin{pmatrix} 22127 \\ 37631 \end{pmatrix}$$

$$C[3] = G \cdot V[3] = \begin{pmatrix} 3 & 1 \\ 5 & 2 \end{pmatrix} \cdot \begin{pmatrix} 7044 \\ 4019 \end{pmatrix} = \begin{pmatrix} 25151 \\ 43258 \end{pmatrix}$$

Table 2: Determination of vectors $C[j]$ of the cryptogram

$C[j]$	$C[1]$		$C[2]$		$C[3]$	
$C[j] = (G \cdot V[j])$	24295	41427	22127	37631	25151	43258

Table 3 shows that the two groups of letters are identical in the plaintext have different cryptograms. This shows that the algorithm makes it difficult to attack by frequency analysis of letter blocks, which is one of the weaknesses of HILL’s conventional algorithm.

Table 3: Comparison on the crypto values of two blocks of identical plaintext

Clear message	m	a	m	a	n	s
Code ASCII	109	97	109	97	110	115
cryptogram	24295	41427	22127	37631	25151	43258

g. Alice finally sends the cryptogram to Bob

h. At the reception, Bob divides the cryptogram into k Vectors C of dimension $n = 2$

i. Bob calculation of vectors $V[j]$ in applying the relation (1.4) below.

$$\begin{cases} \text{pour } j = 1 \text{ à } k \\ V[j] = A * C[j]; \\ \text{fin;} \end{cases} \quad (1.4)$$

$$V[1] = A * C[1] = \begin{pmatrix} 2 & -1 \\ -5 & 3 \end{pmatrix} * \begin{pmatrix} 24295 \\ 41427 \end{pmatrix} = \begin{pmatrix} 7163 \\ 2806 \end{pmatrix}$$

$$V[2] = A * C[2] = \begin{pmatrix} 2 & -1 \\ -5 & 3 \end{pmatrix} * \begin{pmatrix} 22127 \\ 37631 \end{pmatrix} = \begin{pmatrix} 6623 \\ 2258 \end{pmatrix}$$

$$V[3] = A * C[3] = \begin{pmatrix} 2 & -1 \\ -5 & 3 \end{pmatrix} * \begin{pmatrix} 25151 \\ 43258 \end{pmatrix} = \begin{pmatrix} 7044 \\ 4019 \end{pmatrix}$$

j. He concatenates V and gets the string of Y . Bob proceeds then to calculation of $X[i]$ from relation (1.5):

$$\begin{cases} X[1] = Y[1] \oplus p_2 \\ \text{pour } i = 2 \text{ à } L \\ X[i] = Y[i] \oplus Y[i - 1] \\ \text{fin;} \end{cases} \quad (1.5)$$

k. Bob finally finds the *plaintext* from the relation (1.6) following.

$$\begin{cases} \text{pour } i = 1 \text{ à } L \\ x[i] = \frac{1}{p_1} * X[i] \\ \text{fin;} \end{cases} \quad (1.6)$$

The **table 4** shows the set of decryption calculation results.

$$X[1] = Y[1] \oplus p_2$$

Table 4: Decryption Procedure

$V[j] = A \cdot C[j]$	$V[1]$		$V[2]$		$V[3]$	
	7163	2806	6623	2258	7044	4019
$Y[i]$	$Y[1]$	$Y[2]$	$Y[3]$	$Y[4]$	$Y[5]$	$Y[6]$
$X[i] = Y[i] \oplus Y[i - 1]$	4905	4365	4905	4365	4950	5175
$x[k] = 1/p_1 \cdot X[k]$	109	97	109	97	110	115
plaintext	m	a	m	a	n	s

The entire encryption and decryption process is summarized in the table below.

$$Y[1] = X[1] \oplus p_2$$

$$X[1] = Y[1] \oplus p_2$$

Table 5: Summary encryption procedures and decryption

plaintext	m	a	m	a	n	s
Serial number (<i>i</i>)	1	2	3	4	5	6
Code ASCII (<i>x</i> [<i>i</i>])	109	97	109	97	110	115
$X[i] = p_1 * x[i]$	4905	4365	4905	4365	4950	5175
$Y[i] = X[i] \oplus Y[i - 1]$	7163	2806	6623	2258	7044	4019
ENCRYPTION						
cryptogram	24295	41427	22127	37631	25151	43258
DECRYPTION						
$Y[i]$	7163	2806	6623	2258	7044	4019
$X[i] = Y[i] \oplus Y[i - 1]$	4905	4365	4905	4365	4950	5175
$x[k] = \frac{1}{p_1} X[k]$	109	97	109	97	110	115
plaintext	m	a	m	a	n	s

3. Description of the Algorithm

The algorithm is mainly composed of three parts: determination of the antigram, encryption and decryption.

3.1. Antigram generation

Either $K = (G, p_1, p_2)$ the secret key. The calculation of $X[i]$ is given by the relation (2.1) following.

$$\begin{cases} \text{pour } i = 1 \text{ à } L \\ X[i] = p_1 * x[i] \\ \text{fin;} \end{cases} \quad (2.1)$$

Depending on the encryption mode *Cipher Block Chaining (CBC)* values $Y[i]$ are calculated from the relation (2.2) below.

$$\begin{cases} Y[1] = X[1] \oplus p_2 \\ \text{pour } i = 2 \text{ à } L \\ Y[i] = X[i] \oplus Y[i - 1] \\ \text{fin;} \end{cases} \quad (2.2)$$

$$\begin{aligned} Y[1] &= X[1] \oplus p_2 \\ Y[2] &= X[2] \oplus Y[1] \\ Y[3] &= X[3] \oplus Y[2] \\ &\vdots \\ &\vdots \\ Y[i] &= X[i] \oplus Y[i - 1] \\ Y[L] &= X[L] \oplus Y[L - 1] \end{aligned}$$

The antigram is the chain formed by the set of $Y[i]$ obtained from the relation (2.2) as presented below.

$$\underbrace{Y[1], Y[2], Y[3], Y[4] \dots \dots \dots, Y[i], \dots \dots \dots Y[L]}_{Antigramme}$$

The relation (2.2) shows that each value of $Y[i]$ depends on that of $Y[i - 1]$ but also that of $X[i]$. In addition, each value of $X[i]$ is a function of the values of $x[i]$ and p_1 (see relation(2.2)). Therefore, any change to a value $x[i]$ of the plaintext causes a change in the corresponding value $X[i]$ and all the values of $Y[i]$ at $Y[L]$ and the corresponding cryptogram values: *it is the avalanche effect*.

3.2. Encryption

The calculation of all $Y[i]$ gives a string of numbers that must be divided into k vector V of dimension n .

The encryption algorithm is then as follows:

$$\underbrace{Y[1,1], Y[1,2], \dots \dots Y[1, n]}_{v[1]} \underbrace{Y[2,1], Y[2,2], \dots \dots Y[2, n]}_{v[2]} \dots \dots \dots \underbrace{Y[k, 1], Y[k, 2], \dots \dots Y[k, n]}_{v[k]}$$

$$\begin{cases} \text{pour } i = 1 \text{ à } k \\ C[i] = G * V[i] \\ \text{fin;} \end{cases} \quad (2.3)$$

$$\begin{bmatrix} c_{i1} \\ c_{i2} \\ c_{i3} \\ \vdots \\ c_{in} \end{bmatrix} = \begin{bmatrix} \alpha_{11} & \dots & \alpha_{1n} \\ \vdots & \ddots & \vdots \\ \alpha_{n1} & \dots & \alpha_{nn} \end{bmatrix} \begin{bmatrix} Y_{i1} \\ Y_{i2} \\ Y_{i3} \\ \vdots \\ Y_{in} \end{bmatrix} \quad (2.4)$$

$c_{i1}, c_{i2}, \dots \dots, c_{in}$ are the elements of the vector $C[i]$ and $Y_{i1}, Y_{i2}, \dots \dots, Y_{in}$ are the elements of the vector $V[i]$ with i ranging from 1 à k .

The concatenation of the $C[i]$ vectors, gives the cryptogram. The process of equation gives rise to one of the systems of equation with as unknown the elements of the matrix G as well as the elements of the vectors $V[i]$ whose calculation depends on the parameters p_1 et p_2 of the secret key K .

3.3. Decryption

The recipient has the symmetric secret key $K = (G, p_1, p_2)$. Since G is invertible, its inverse A is easily calculable. As with encryption, the cryptogram is divided into k vector C of dimension n . Decryption begins with the calculation of vectors $V[i]$ from the matrix A and vectors $C[i]$ as shown by the relation (2.5). The antigram corresponds to the concatenation of all vectors $V[i]$.

$$\underbrace{c[1,1], c[1,2], \dots \dots c[1, n]}_{c[1]} \underbrace{c[2,1], c[2,2], \dots \dots c[2, n]}_{c[2]} \dots \dots \dots \underbrace{c[k, 1], c[k, 2], \dots \dots c[k, n]}_{c[k]}$$

$$\begin{cases} \text{pour } i = 1 \text{ à } k \\ V[i] = A * C[i] \\ \text{fin;} \end{cases} \quad (2.5)$$

After the calculation of the antigram, the values of the $X[i]$ are obtained by applying the relation(2.6).

$$\begin{cases} X[1] = Y[1] \oplus p_2 \\ \text{pour } i = 2 \text{ à } L \\ X[i] = Y[i] \oplus Y[i - 1] \\ \text{fin;} \end{cases} \quad (2.6)$$

The decryption operation ends with the calculation of the values some $x[i]$ corresponding to the plaintext from the relation (2.7).

$$\begin{cases} \text{pour } i = 1 \text{ à } L \\ x[i] = \frac{1}{p_1} * X[i] \\ \text{fin;} \end{cases} \quad (2.7)$$

4. Cryptological Security

The secret key $K = (G, p_1, p_2)$ is known only to the two correspondents *Alice* and *Bob*. As stated by Auguste Kerckhoffs [7] [8]:

- a. The security of the cryptographic system must be based on the secrecy of the key and not on that of the algorithm.
- b. Decryption without the key must be impossible with the means of the moment, because it requires astronomical times.
- c. If the plaintext message and the encrypted message are known, it should not be possible to extract the key in a reasonable time.

The design of a mechanism for transforming the plaintext message into a cryptogram must also obey the principles of dissemination and confusion set out by Claude Shannon [5].

4.1. The Principle of Dissemination

According to this principle, the statistics of the clear message must spread over the entire cryptogram, forcing the cryptanalyst to intercept and analyze a very large amount of data to conduct its decryption. For example, the frequency of a bigram in the clear must not have an effect to a bigram in the cryptogram leading to a frequency analysis. Compliance with the principle of diffusion should ideally lead to a statistically indistinguishable cryptogram from a random sequence of symbols.

Indeed, the diffusion is provided by the avalanche effect in the calculation of $Y[i]$. The repetitions of blocks of characters in the plaintext are hidden in the cryptogram. In addition, any modification of a character in the

plaintext results in a modification of the cryptogram.

4.2. The Principle of Confusion

According to this principle, the relations between the plaintext and the cryptogram must be complex. The equation of the process must result in very large systems of equations, where all the variables depend on all the others in relation so confused that the work of solving is practically impossible.

Confusion is brought by secret elements p_1 and p_2 that increase the number of unknown variables in the process of equation knowing the plaintext and the cryptogram. The $X[i]$ and $Y[i]$ are unknown variables for the cryptanalyst. The relation between the plaintext message and the cryptogram then leads to a system of equations that is almost impossible to solve.

4.3. Resistance to attacks

A cryptographic algorithm is considered safe when it resists different types of attacks [9] such as:

- **Ciphertext-only attack**

It should be noted that a *block frequency analysis attack* is made impossible by the use of the additional parameters p_1 and p_2 and of the *avalanche effect*. Thus, having only the ciphertext (C), it is impossible for an attacker to find the corresponding plaintext (M)

- **known-plaintext attack**

The attacker has plaintext and its cryptogram. Due to the unknown parameters in the process of equation between the elements of the plaintext and those of the cryptogram, it will be very difficult to find the key knowing the plaintext (M) and ciphertext (C).

- **chosen-plaintext attack**

Knowing cryptogram (C), the attacker chooses a plaintext (M) it encrypts it and compares cryptograms. Due to the *avalanche effect* of $Y[i]$, if $M' \neq M$, then $C' \neq C$.

Choosing a plaintext (M) whose encryption gives the cryptogram (C) can be an exhaustive task as long as the key $K = (p_1, p_2, G)$ remains unknown.

- **adaptive-plaintext attack**

The attacker does not know the key, but he can have what he wants decrypted by the decryption method and see the plaintext. It has access to the decryption algorithm but cannot disassemble it to get the key.

5. Conclusion

In this article, we have presented an algorithm that brings an improvement on the one hand on the cryptography security of the conventional HILL algorithm on the other hand on the modernization of its field of action. The example studied in this article showed that frequency analysis has been concealed in the cryptogram and that the number of unknown parameters has been increased in linear equation systems connecting *plaintext* to cryptogram.

6. Abbreviations

For the purpose of this article, the following symbols will be used:

- **Alice and Bob** : Names of the two entities running the algorithm
- **M**: the *plaintext*
- **C** : the *ciphertext*
- **l**: the initial size of the *plaintext*
- **L** : the size of the *plaintext* in a multiple **n**
- **x[i]** : the code point of the character at the position **i** in the *plaintext*
- **X[i] = p₁ * x[i]**
- **K = (G, p₁, p₂)** : the secret key of the system
- **p₁ et p₂**: randomly chosen from the set of real numbers (\mathbb{R}) (**p₁ et p₂ > 1**)
- **G** : an invertible square matrix of dimension **n** whose elements belong to the whole \mathbb{R} real numbers
- **A** : the inverse of **G**
- **n** : dimension of the matrix **G**
- **Y[i] = X[i] ⊕ Y[i - 1]** with **Y[1] = X[1] ⊕ p₂**
- **V[k]** : vector of dimension **n** whose elements are the **Y[i]**
- **C[k]**: the crypto of vector **V[k]** (**C[k] = (G * V[k])**)

References

- [1] L. S. Hill. "Cryptography in an Algebraic Alphabet". *The American Mathematical Monthly*, vol. 36, n°16, pp. 306-312, Juin-Juillet, 1929.
- [2] N. HADJ-SAID, A. ALI-PACHA, A. M'HAMED, A. BELGORAF. " Sécurité des Données en se basant sur le Chiffre de Hill"..
- [3] F. Gmira, S. Hraoui, A. Saaidi, A. W. Jarar, K. Sator. " L'Amélioration de la Sécurité du Chiffrement Algébrique Modulaire par les Générateurs de Fibonacci". *Mediterranean Telecommunication Journal*. vol. 4, n°1, 2014.
- [4] S. William. *Cryptography and network security*. Pearson Education India, 2006.
- [5] C. Channon. "Communication theory of secrecy systems". *Bell System Technical Journal*, vol. 328, n°14,

pp. 656-715, 1949.

- [6] P. Guillot. *La cryptologie : L'art des codes secret*. EDP Sciences, 2013.
- [7] A. Kerckhoffs. “ La Cryptographie Militaire”. *Journal des sciences militaires*. vol. IX, pp. 5-38, Janvier 1983, pp. 161–191, Février 1883.
- [8] C. Ngô. *Énergie, Entropie, Information, Cryptographie et Cybersécurité*. EDP Sciences, 2019.
- [9] M. Videau. *Critères de sécurité des algorithmes de chiffrement à clé secrète*. Paris VI, Thèse de Doctorat, Université Pierre et Marie Curie, 2005.