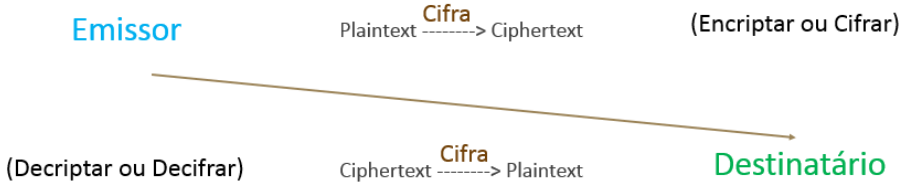
UNIVERSIDADE DE LISBOA

INSTITUTO SUPERIOR TÉCNICO

Aplicações da Álgebra Linear: Criptografia

Cifra de Hill

A **Criptografia** é o estudo dos princípios e técnicas pelas quais a informação pode ser transformada da sua forma original (plaintext), para outra ilegível (ciphertext), de forma que possa ser conhecida apenas por seu destinatário.
Ao acto de criar a mensagem ilegivel a partir da legível, chamamos de **encriptar** a mensagem, e ao acto inverso chamamos de **decriptar**.
Para que tal possa acontecer, é necessário que tanto o emissor como o destinatário possuam conhecimento do código/método de correspondência que permite cifrar/decifrar a mensagem, pelo que a esse código dá-se o nome de **Cifra**.



A história da Criptografia começa nos grandes Impérios da Antiguidade Clássica, donde vêm os primeiros exemplos de cifras. A mais antiga remonta ao Antigo Egipto, onde o emissor trocava alguns hieroglifos, codificando de uma maneira basica a sua mensagem. Este é um exemplo de **Cifra de Substituição.**
A chamada "Cifra de César" é um outro, mais conhecido, exemplo de Cifra de substituição. O autor da cifragem trocava cada letra por outra situada a três posições à frente no alfabeto. Segundo o autor, esse algoritmo foi responsável por enganar muitos inimigos do Império Romano; no entanto, após ter sido descoberta a chave, como todas, perdeu sua funcionalidade.
Ao longo da Idade Média e da Idade Moderna, verificou-se um desenvolvimento das cifras de Substituição, até à 2ªGuerra Mundial, onde o desenvolvimento da Criptografia obteve o seu grande BOOM, através do desenvolvimento de máquinas como Enigma, que encriptava, por processos mecânicos, a mensagem, ao mesmo tempo que esta era escrita. Durante a Guerra Fria foram desenvolvidos outras Cifras, e de outras técnicas de cifragem.

Cifra de Substítuição

Em criptografia, uma cifra de substituição é um método de criptografia que opera de acordo com um sistema pré-definido de substituição. Para encriptrar uma mensagem, as unidades do texto - que podem ser letras isoladas, pares ou outros grupos de letras - são substituídas para formar a cifra. As cifras de substituição são decifradas pela substituição inversa.

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	Alfabeto
D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	Cifra

A título de exemplo, frase “Olá sou fixe”, codificada, ficará “ROD VRX ILAH”.

A grande desvantagem da utilização das Cifras de Substituição (Pré-Hill), é o facto de serem fáceis de serem quebradas através de métodos estatísticos, onde, através do estudo da frequência da ocorrência das letras na mensagem, se pode fazer corresponder às letras mais frequentes do alfabeto.
Ora, para ultrapassar esta desvantagem, Lester S. Hill, em 1929, recorrendo a Álgebra Linear, os conceitos de Matriz, Vector, Transformações Matriciais, desenvolveu a Cifra de Hill, cujo método será apresentado a seguir. Para se compreender o Método de Cifragem/Decifragem de Hill, é necessário umm conhecimento prévio de Aritmética Modular(módulo 26, o número de letras do alfabeto).

Cifra de Hill

Cifrar a Mensagem

Para cifrar a mensagem é necessário, primeiro, dividir o texto em grupos de **n** letras. (Se o número de letras no texto não for divisível por **n,** no final da mensagem (plaintext), adicionar letras arbitrarias, que não dificultem a compreensão da mensagem, até que seja verificada a dívisão completa por **n**)

Ex: Para n=2 pretende-se cifrar a mensagem:

“As Tropas dos Aliados invadiram a Normandia com Sucesso” – 49 letras
“(As) (Tr)(op)(as) (do)(s A)(li)(ad)(os) (in)(va)(di)(ra)(m a) (No)(rm)(na)(di)(a c)(om) (Su)(ce)(ss)(oo)” – 50 letras

Para um determinado n, chamamos a n-cifra de Hill a classe da cifra, neste caso temos **2-cifra de Hill**.

Tendo a classe da cifra e a mensagem dividida em **k** grupos, atribui-se a cada grupo o vector (matriz **n**x1), ou juntando tudo n uma matriz **n**x**k** com os **k** ventores em coluna,

de numeros correspondente, utilizando a seguinte equivalência (a letra Z equivale ao número 0, por ser a 26ª letra):

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*
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

Alguns autores atribuem à letra A o número 0, a B, 1 e assim sucessivamente. A correspondencia não altera de maneira efetiva o método, mas há, naturalmente, algumas diferenças.

De seguida, é necessário escolher uma matriz secreta **nxn** invertível em Z_{26} (Conjunto de numeros inteiros positivos, formado por resíduo (residue) da divisão desse número por 26 ={0,1,2,…,25}).

TEOREMA:

- Matriz Anxn com entradas em Z_{26} é invertível mod 26 sse det A tenha inverso, ou seja existir um a^{-1} tal que

$$\det A \times a^{-1} = 1 \pmod{26}$$

COROLÁRIO:

- Matriz Anxn com entradas em Z26 é invertível mod 26 sse det A for igual a:

a	1	3	5	7	9	11	15	17	19	21	23	25
a^{-1}	1	9	21	15	3	19	7	23	11	5	17	25

Escolhida a matriz, é só preciso multiplicar a Matriz Secreta pelos **k** vectores, ou pela matriz **nxk**, e assim se obtem a mensagem cifrada.

"**As** Tropas dos **Ali**ados inva**dir**am **a** Norman**dia** com Sucessoo"

$$\begin{pmatrix} 1 & 20 & 15 & 1 & 4 & 19 & 12 & 1 & 15 & 9 & 22 & 4 & 18 & 13 & 14 & 18 & 1 & 4 & 1 & 15 & 19 & 3 & 19 & 15 \\ 19 & 18 & 16 & 19 & 15 & 1 & 9 & 4 & 19 & 14 & 1 & 9 & 1 & 1 & 15 & 13 & 14 & 9 & 3 & 13 & 21 & 5 & 19 & 15 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 2 \\ 0 & 3 \end{pmatrix} \begin{pmatrix} 13 & 4 & 21 & 13 & 8 & 21 & 4 & 9 & 1 & 11 & 24 & 22 & 20 & 15 & 18 & 18 & 3 & 22 & 7 & 15 & 9 & 13 & 5 & 19 \\ 5 & 2 & 22 & 5 & 19 & 3 & 1 & 12 & 5 & 16 & 3 & 1 & 3 & 3 & 19 & 13 & 16 & 1 & 9 & 13 & 11 & 15 & 5 & 19 \end{pmatrix} \text{ Mod } 26$$

"**M**EDBUV**M**EHSU**C**DA**I**LAEKPX**C**VAT**C**OCRSRM**C**PVA**G**IOMIKMOEESS"

Nota1: A matriz correspondente à mensagem codificada não é a matriz que se obtém diretamente da multiplicação da Matriz secreta com a Matriz plaintext, mas sim a correspondente a essa matriz, em módulo 26. Os números que existem na Matriz da mensagem codificada, correspondem ao resto da divisão dos numeros obtidos por mutiplicação direta das matrizes, por 26, este é o método utilizado na aritmética modular para os numeros inteiros positivos.

Como podemos ver na mensagem codificada, as letras realçadas correspondem, na mensagem à letra 'A' no texto legível. Verificamos que a letra 'A', é codificada e várias

letras diferentes, sendo impossível, ou pelo menos muito complicado quebrar a Cifra através dos métodos estatísticos.

Fácilmente se conclui que o método de cifragem, corresponde à operação:

$$Ap=c$$

Em que:

- A é a matriz secreta;
- p são os vários vectores plaintext;
- c são os vectores cyphertext;

Decifrar a Mensagem

Logo, para decifrar, sabendo a matriz A, e os vectores c, temos:

$$p=A^{-1}c$$

Assim tomando para exemplo a matriz A:

$$\det A=3, \text{ pela tabela em cima } \det A^{-1} = 9$$

$$\text{logo } A^{-1} = \quad 9 \begin{pmatrix} 3 & -2 \\ 0 & 1 \end{pmatrix} \quad = \quad \begin{pmatrix} 27 & -18 \\ 0 & 9 \end{pmatrix}$$

"MEDBUVMEHSUCDAILAEKPXCVATCOCRSRMCPVAGIOMIKMOEESS"

$$\begin{pmatrix} 1 & 8 \\ 0 & 9 \end{pmatrix} \begin{pmatrix} 13 & 4 & 21 & 13 & 8 & 21 & 4 & 9 & 1 & 11 & 24 & 22 & 20 & 15 & 18 & 18 & 3 & 22 & 7 & 15 & 9 & 13 & 5 & 19 \\ 5 & 2 & 22 & 5 & 19 & 3 & 1 & 12 & 5 & 16 & 3 & 1 & 3 & 3 & 19 & 13 & 16 & 1 & 9 & 13 & 11 & 15 & 5 & 19 \end{pmatrix} = \begin{pmatrix} 1 & 20 & 15 & 1 & 4 & 19 & 12 & 1 & 15 & 9 & 22 & 4 & 18 & 13 & 14 & 18 & 1 & 4 & 1 & 15 & 19 & 3 & 19 & 15 \\ 19 & 18 & 16 & 19 & 15 & 1 & 9 & 4 & 19 & 14 & 1 & 9 & 1 & 1 & 15 & 13 & 14 & 9 & 3 & 13 & 21 & 5 & 19 & 15 \end{pmatrix} \text{ Mod } 26$$

"As Tropas dos Aliados invadiram a Normandia com Sucessoo"

Nota2: A matriz quadrada multiplicada a esquerda corresponde à matriz inversa de A, novamente em modulo 26, efectuando as mesmas operações descritas na Nota1, com a particularidade dos numeros negativos (-18): nesta situação, para os números negativos, a sua equivalência em aritmética modular, corresponde a 26 menos o resto da divisão do módulo do número negativo, por 26.

Quebrar a Cifra de Hill

A cifra de Hil só pode ser quebrada se:

- For conhecida a classe da Cifra (saber n);
- For conhecido um segmento da mensagem;

Assim, o principio sugere que, se tivermos p1, p2, … pn vectores(plaintext) do segmento da mensagem, e sendo estes linearmente independentes, tendo também c1, c2, … cn vectores (cyphertext correspondentes), podemos chegar a matriz A.

TEOREMA

Se p1, p2, … pn linearmente independentes e c1, c2, … cn .

Se $P = \begin{bmatrix} \mathbf{p}_1^T \\ \mathbf{p}_2^T \\ \vdots \\ \mathbf{p}_n^T \end{bmatrix}$ corresponde matriz nxn com os vetores $p1^T$, $p2^T$ … pn^T em linha

e $C = \begin{bmatrix} \mathbf{c}_1^T \\ \mathbf{c}_2^T \\ \vdots \\ \mathbf{c}_n^T \end{bmatrix}$ matriz nxn com os vetores $c1^T$, $c2^T$ … cn^T correspondentes em linha,

aplicando uma sequência de operações elementares, é possivel reduzir a matriz **C** a **I**, e, aplicando as mesmas operações, transformar **P** em $\mathbf{(A^{-1})^T}$.

Ou seja:

$$[C|P] \longrightarrow [I|(A^{-1})^T]$$

O que será o mesmo que:

$$C^{-1}.P = (A^{-1})^T$$

Isto tudo porque, sendo $P=p^T$ e $C=c^T$:

Partindo de Ap=c (a igualdade que temos como base da cifra de Hill)

$$AP^T=C^T$$

$$P^T=A^{-1}.C^T$$

$$P=C.(A^{-1})^T$$

$$\mathbf{C^{-1}.P= (A^{-1})^T}$$

Apesar de tudo, nem sempre é este o método mais fácil, pelo que por vezes, é necessário recorrer à igualdade mais direta e fazer: $A=C^T.(P^T)^{-1}$

Para dar um exemplo, suponhamos que os Aliados intercetam uma mensagem entre dois capitães de tropas do Eixo. A mensagem codificada era:

KBBOICWHMEOOVVYICPCOWKEMSXSUKBQHGCGKSOARIMISBOYAWOZBKFKBQKOCE
MFSYIILCCPCSSYVWLEMAENOETKBSXSUKBMRVSBHWATC

Também se descobriu, através de espionagem, que a cifra era de classe 2 e que a mensagem começava por: “Arre…”

Assim, utilizando o último teorema enunciado, e tendo:

$$P=\begin{bmatrix}1 & 18\\ 18 & 5\end{bmatrix} \qquad C=\begin{bmatrix}11 & 2\\ 2 & 15\end{bmatrix}$$

$$\left[\begin{array}{cc|cc}11 & 2 & 1 & 18\\ 2 & 15 & 18 & 5\end{array}\right] \xrightarrow[\text{(Anexo)}]{\text{Op Elementares}} \left[\begin{array}{cc|cc}1 & 0 & 1 & 0\\ 0 & 1 & 8 & 9\end{array}\right]$$

$(A^{-1})^T$

As forças dos aliados conseguiram assim descodificar a mensagem dizendo:

"Arre gaita Se eles conseguem chegar ate aqui e o fim do regima Prepara uma emboscadaa Nao os podemos deixar chegar a fronteira"

E assim se deu a volta à 2ª Guerra Mundial…

Anexo:

19.L1 -> L1

$$\left[\begin{array}{cc|cc}11 & 2 & 1 & 18\\ 2 & 15 & 18 & 5\end{array}\right] \longrightarrow \left[\begin{array}{cc|cc}209 & 38 & 19 & 342\\ 2 & 15 & 18 & 5\end{array}\right] = \left[\begin{array}{cc|cc}1 & 12 & 19 & 4\\ 2 & 15 & 18 & 5\end{array}\right] \text{ Mod 26}$$

-2.L1 + L2 -> L2

$$\longrightarrow \left[\begin{array}{cc|cc}1 & 12 & 19 & 4\\ 0 & -9 & -20 & -3\end{array}\right] = \left[\begin{array}{cc|cc}1 & 12 & 19 & 4\\ 0 & 17 & 6 & 23\end{array}\right] \text{ Mod 26}$$

23.L2 -> L2

$$\longrightarrow \left[\begin{array}{cc|cc} & 12 & 19 & 4\\ 0 & 391 & 138 & 529\end{array}\right] = \left[\begin{array}{cc|cc}1 & 12 & 19 & 4\\ 0 & 1 & 8 & 9\end{array}\right] \text{ Mod 26}$$

-12.L2 +L1 -> L1

$$\longrightarrow \left[\begin{array}{cc|cc}1 & 0 & -77 & -104\\ 0 & 1 & 8 & 9\end{array}\right] = \left[\begin{array}{cc|cc}1 & 0 & 1 & 0\\ 0 & 1 & 8 & 9\end{array}\right] \text{ Mod 26}$$

Autor:
João Francisco Formigão e Santos
Nº81510
Curso MEAer