

## 0.1 Mensagens Cifradas

Neste campo estudam-se dois problemas fundamentais: codificar mensagens de modo a que estranhos não entendam o seu significado e descobrir qual o código que foi utilizado numa determinada mensagem de modo a descodificar uma mensagem que não nos é destinada. Quem viu pelo menos um filme de espionagem já se deparou com cada uma destas situações.

As codificações de mensagens podem seguir diversos padrões sendo a mais simples feita caracter a caracter. A cifra usada pelo exército de Júlio César era deste tipo. A ideia base consiste em atribuir a cada letra do alfabeto um número de acordo com a ordenação do alfabeto. Assim,

letra	A	B	C	D	E	F	G	H	I	J	K	L	M
número	0	1	2	3	4	5	6	7	8	9	10	11	12
letra	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
número	13	14	15	16	17	18	19	20	21	22	23	24	25

A cifra usada por Júlio César consistia em substituir cada letra pela que fica 3 posições mais à frente. Isto é, designando por  $P$  o número correspondente à letra original e por  $C$  o número correspondente à letra cifrada:  $C \equiv P+3(\text{mod } 26)$ , sendo obviamente  $0 \leq C \leq 25$ .

Pode-se assim construir uma tabela de correspondência entre texto original e texto cifrado:

texto inicial	A	B	C	D	E	F	G	H	I	J	K	L	M
texto cifrado	3	4	5	6	7	8	9	10	11	12	13	14	15
texto inicial	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
texto cifrado	D	E	F	G	H	I	J	K	L	M	N	O	P
texto inicial	13	14	15	16	17	18	19	20	21	22	23	24	25
texto cifrado	16	17	18	19	20	21	22	23	24	25	0	1	2
	Q	R	S	T	U	V	W	X	Y	Z	A	B	C

Deste modo, para codificar uma mensagem basta substituir cada letra pela que corresponde na tabela acima. Por exemplo, para codificar a mensagem NÃO MATEM O MENSAGEIRO, deve-se escrever QDR PDWHP R PHQVDJHLUR. Para dificultar mais a descodificação pelo inimigo, as letras não são agrupadas na sua ordem natural mas em grupos de tamanho fixo, normalmente de 5 letras. Assim, dever-se-ia escrever: QDRPD WHPRP HQVDJ HLUR.

Para decifrar a mensagem procede-se em ordem inversa. Se  $C \equiv P + 3(\text{mod } 26)$ , então  $P \equiv C - 3(\text{mod } 26)$ ,  $0 \leq P \leq 25$ . Este código é um caso particular de  $C \equiv P + k(\text{mod } 26)$ ,  $0 \leq P \leq 25$ .

É claro que este código é muito ingénuo e facilmente decifrável, pois basta encontrar a transformada de uma letra para imediatamente ter o código completamente decifrado. Além disso, o facto de só haver 26 transformações deste tipo facilita ainda mais a tarefa de descodificação.

Esta cifra é um caso particular de  $C \equiv aP + b \pmod{26}$ ,  $0 \leq C \leq 25$ . Codificações deste tipo têm o nome de cifras afins. Se  $\text{mdc}(a, 26) = 1$ , então, quando  $P$  percorre os elementos de um sistema completo de restos módulo 26,  $C$  também o faz. Como há 12 hipóteses de escolha para  $a$  (1, 3, 5, 7, 9, 11, 15, 17, 19, 21, 23, 25) e 26 para  $b$ , obtêm-se  $12 \times 26 = 312$  transformações deste tipo. Com recurso a qualquer pequeno computador, quebrar um código destes é de extrema facilidade. Para decodificar a mensagem basta observar que se  $C \equiv (aP + b) \pmod{26}$ ,  $0 \leq C \leq 25$ , então  $P \equiv \bar{a}(C - b) \pmod{26}$ ,  $0 \leq P \leq 25$ , em que  $\bar{a}$  é o inverso de  $a$  módulo 26.

**Exemplo 1** Se  $C \equiv (7P + 10) \pmod{26}$ , então  $P \equiv 15(C - 10) \pmod{26} \equiv (15C + 6) \pmod{26}$ ,  $0 \leq P \leq 25$ . Assim, por exemplo a letra  $L$ , a que corresponde o valor 11, será codificada por  $J$  visto que  $7 \times 11 + 10 = 87 \equiv 9 \pmod{26}$  e a 9 corresponde a letra  $J$ .

Para decodificar este tipo de código é já necessário conhecer a codificação de duas letras diferentes.

Para ajudar a decodificação de mensagens escritas usando uma cifra afim é conveniente conhecer uma tabela de frequência de ocorrências das letras do alfabeto num texto vulgar na língua que se supõe estar a ser utilizada. As tabelas mais estudadas são as de língua inglesa, mas existem tabelas para quase todas as línguas, algumas bem guardadas pelos serviços secretos de certos países. Por exemplo, em língua inglesa a tabela de frequência em percentagem do texto total é a que se segue:

letra	A	B	C	D	E	F	G	H	I	J	K	L	M
%	7	1	3	4	13	3	2	3	8	<1	<1	4	3
letra	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
%	8	7	3	<1	8	6	9	3	1	1	<1	2	<1

É claro que, é mais fácil decodificar textos mais longos do que pequenas mensagens, uma vez que quanto maior for a mensagem mais provável é que a taxa de ocorrência das letras nessa mensagem siga o padrão médio.

Como exemplo considere-se a seguinte mensagem, que se sabe ter sido codificada usando uma cifra do tipo  $C \equiv P + b \pmod{26}$  e ter sido escrita em inglês:

YFXMP CESPZ CJTDF DPQFW QZCPY NTASP CTYRX PDDLRL PD

Contando as letras, verifica-se que a letra com mais ocorrências é o P (7 ocorrências). Se, neste texto, se mantiver o padrão médio dos textos em língua inglesa, P deve ser a transformada de E, que é a letra com maior frequência de ocorrência. Como a P corresponde 15 e a E corresponde 4, teríamos:  $15 \equiv 4 + b \pmod{26}$ , donde  $b = 11$ . Ou seja  $C \equiv P + 11 \pmod{26} \implies P \equiv C - 11 \pmod{26}$ . Constrói-se agora a tabela de correspondência texto cifrado  $\longrightarrow$  texto original, usando esta relação:

texto	A	B	C	D	E	F	G	H	I	J	K	L	M
cifrado	0	1	2	3	4	5	6	7	8	9	10	11	12
texto	15	16	17	18	19	20	21	22	23	24	25	0	1
original	P	Q	R	S	T	U	V	W	X	Y	Z	A	B
texto	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
cifrado	13	14	15	16	17	18	19	20	21	22	23	24	25
texto	2	3	4	5	6	7	8	9	10	11	12	13	14
original	C	D	E	F	G	H	I	J	K	L	M	N	O

Com a ajuda desta tabela facilmente se escreve a mensagem descodificada:

NUMBE RTHEO RYISU SEFUL FOREN CIPHE RINGM ESSAG ES

Falta agrupar as letras de modo a formar palavras com sentido:

NUMBER THEORY IS USEFUL FOR ENCIPHERING MESSAGES

Suponhamos agora que uma transformação afim da forma  $C \equiv (aP + b)(\text{mod } 26)$ ,  $0 \leq P \leq 26$ , foi usada para codificar a seguinte mensagem:

USLEL JUTCC YRTPS URKLT YGGFV ELYUS LRYXD JURTU ULVCU URJRK  
 QLLQL YXSRV LBRYZ CYREK LVEXB RYZDG HRGUS LJLLM LYPDJ LJ TJU FALGU  
 PTGVT JULYU SLDAL TJRWU SLJFE OLPV

Neste caso é necessário conhecer o transformado de pelo menos duas letras para decifrar a mensagem. Tal como no caso anterior, começa-se por fazer a contagem do número de vezes que cada letra aparece na mensagem. Conclui-se ser a letra L a mais frequente, com 22 ocorrências, sendo a letra U a segunda mais frequente com 16 ocorrências. Atendendo às tabelas de frequência de ocorrência das letras em texto normal, somos levados a pensar que L corresponde a E e U corresponde a T. Ou seja:

$$\begin{cases} 4a + b \equiv 11(\text{mod } 26) \\ 19a + b \equiv 20(\text{mod } 26) \end{cases}$$

A solução deste sistema é  $a \equiv 11(\text{mod } 26)$  e  $b \equiv 19(\text{mod } 26)$ . Se as conjecturas estiverem correctas a codificação foi feita usando  $C \equiv (11P + 19)(\text{mod } 26)$  e a descodificação pode ser feita usando  $P \equiv 19(C - 19)(\text{mod } 26)$ , uma vez que 19 é o inverso de 11 módulo 26. Pode-se assim construir a tabela de descodificação:

texto	A	B	C	D	E	F	G	H	I	J	K	L	M
cifrado	0	1	2	3	4	5	6	7	8	9	10	11	12
texto	3	22	15	8	1	20	13	6	25	18	11	4	23
original	D	W	P	I	B	U	N	G	Z	S	L	E	X
texto	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
cifrado	13	14	15	16	17	18	19	20	21	22	23	24	25
texto	16	9	2	21	14	7	0	19	12	5	24	17	10
original	Q	J	C	V	O	H	A	T	M	P	Y	R	K

Finalmente, obtém-se:

THEBE STAPP ROACH TOLEA RNNUM BERTH EORYI STOAT TEMPT TOSOL  
VEEVE RYHOM EWORK PROBL EMBYW ORKIN GONTH ESEEX ERCIS ESAST  
UDENT CANMA STERT HEIDE ASOFT HESUB JECT

e, separando as palavras convenientemente:

THE BEST APPROACH TO LEARN NUMBER THEORY IS TO ATTEMPT TO  
SOLVE EVERY HOMEWORK PROBLEM BY WORKING ON THESE EXERCISES A  
STUDENT CAN MASTER THE IDEAS OF THE SUBJECT.

Apesar de um pouco mais complexas, estas codificações continuam a ser facilmente decifráveis.

## 0.2 Algarismos de Controlo

As congruências também podem ser utilizadas para verificar erros nas sequências de algarismos. Por exemplo, no número de um bilhete de identidade, de um passaporte ou de um cartão VISA, além do número de ordem existe um algarismo que permite verificar se não houve erro na transmissão do número. Normalmente esse algarismo é o último ou o primeiro da sequência.

Vejamus como o algarismo de controlo funciona para o número de ISBN (International Standard Book Number) de uma edição de um livro. Este número é formado por várias sequências de caracteres com um significado preciso. Por exemplo, 0-201-06561-4. Nesta sequência o primeiro grupo informa qual o idioma em que o livro foi escrito (0 representa inglês), o segundo grupo identifica o editor (201 é a editora Addison-Wesley), o terceiro grupo é a numeração interna que o livro recebeu no editor, isto é, cada editor numera sequencialmente os livros que publica, atribuindo um número diferente a diferentes edições do mesmo livro. Deste modo, para se encomendar um livro basta fornecer ao livreiro o número de ISBN do livro. Ora é frequente haver erros de transmissão desse número, sendo os mais frequentes a troca de algarismos. O último algarismo serve para verificar se o ISBN é válido, detectando grande parte dos erros de transmissão que possam ocorrer.

Seja  $x_1x_2\cdots x_{10}$  a sequência de algarismos de um ISBN, então  $\sum_{i=1}^{10} ix_i \equiv 0 \pmod{11}$ , ou seja  $\sum_{i=1}^9 ix_i + 10x_{10} \equiv 0 \pmod{11}$ , ou ainda  $\sum_{i=1}^9 ix_i - x_{10} \equiv 0 \pmod{11}$ , ou, finalmente,  $x_{10} \equiv \sum_{i=1}^9 ix_i \pmod{11}$ , isto é, o algarismo de controlo é o resto da divisão por 11 de uma soma ponderada dos restantes algarismos do número. Como também se pode ter um resto 10 e só se pode usar um caracter para representar o resto, nesse caso usa-se  $X$ .

**Exemplo 2** Para um ISBN 0-201-06561 o algarismo de controlo é 4, pois:

$$x_{10} \equiv 1 \times 0 + 2 \times 2 + 3 \times 0 + 4 \times 1 + 5 \times 0 + 6 \times 6 + 7 \times 5 + 8 \times 6 + 9 \times 1 = 136 \equiv 4 \pmod{11}$$

Vejamus que qualquer troca de algarismos pode ser detectada através do algarismo de controlo. Suponhamos que a sequência  $x_1x_2\cdots x_{10}$  é um ISBN válido e que houve

um engano no algarismo de ordem  $j$ . O algarismo que efectivamente foi escrito em vez de  $x_j$  pode ser representado por  $x_j + a$ , com  $-10 < a < 10$ . Então a nova soma será:  $\sum_{i=1}^{10} ix_i + ja \equiv ja \not\equiv 0(\text{mod } 11)$ , uma vez que  $11 \nmid j$  e  $11 \nmid a$  e 11 é primo. Pode-se assim concluir que não se tem um ISBN válido.

Suponhamos agora que se trocam os algarismos  $x_j$  e  $x_k$  entre si. A nova soma será  $\sum_{i=1}^{10} ix_i + (jx_k - jx_j) + (kx_j - kx_k) = \sum_{i=1}^{10} ix_i + (j - k)(x_k - x_j)$ . Ora  $11 \nmid (j - k)$  e  $11 \nmid (x_k - x_j)$ . Como  $\sum_{i=1}^{10} ix_i \equiv 0(\text{mod } 11)$ , então  $\sum_{i=1}^{10} ix_i + (j - k)(x_k - x_j) \not\equiv 0(\text{mod } 11)$ . Assim detecta-se que o número que se escreveu não é um ISBN válido.

O algarismo de controlo é obtido sempre de modo idêntico, mas, conforme os casos, a soma pode ter outra ponderação e o módulo utilizado ser diferente. Por exemplo, os números dos passaportes de alguns países são formados por 7 algarismos, sendo o último algarismo tal que:  $x_7 \equiv 7x_1 + 3x_2 + x_3 + 7x_4 + 3x_5 + x_6(\text{mod } 10)$ .

**Exemplo 3** *Um passaporte tem o número 211894, o último algarismo é determinado por  $x_7 \equiv 7 \times 2 + 3 \times 1 + 1 + 7 \times 8 + 3 \times 9 + 4 = 105 \equiv 5(\text{mod } 10)$  e o número impresso no passaporte seria 2118945.*

Note-se que este processo permite sempre verificar se um algarismo foi trocado, mas se dois algarismos forem trocados entre si o último algarismo pode ser igual, bastando que os algarismos trocados sejam tais que  $|x_i - x_j| = 5$ .