OBJETIVOS

- **Sistemas monoalfabéticos.**
  - **Escrituras sagradas, Cifra de Cesar**
- **Sistemas polialfabéticos**
  - **Vigenére**
- **Transposições.**
  - **Transposições regulares (Bastão de Licurgo, Transposição colunar)**
- **Dispositivos criptográficos.**
  - **Enigma.**
- **Criptoanálise.**
  - **Transposições, Sistemas monoalfabéticos.**
- **Codificação.**
  - **ASCII, Base64.**

SISTEMAS MONOALFABÉTICOS

- Uma tabela de substituição contém os caracteres que serão substituídos e os caracteres de substituição. Esta tabela também é conhecida como cifrante ou alfabeto cifrante. Quando apenas um cifrante é aplicado para transformar um texto claro num criptograma, a substituição é chamada de monoalfabética.

SISTEMAS MONOALFABÉTICOS (CIFRAS HEBRAICAS)

- Atbash, Albam e Atbah são as três cifras hebraicas mais conhecidas. Datam de 600-500 a.C. e eram usadas principalmente em textos religiosos - escribas hebreus usaram a cifra Atbash para escrever algumas palavras no Livro de Jeremias.
- Estas cifras baseiam-se no sistema de substituição simples monoalfabética. As três são reversíveis porque na primeira operação obtém-se o texto cifrado e, aplicando-se o mesmo método ao texto cifrado, obtém-se o texto original.

SISTEMAS MONOALFABÉTICOS (CIFRAS HEBRAICAS)

- Características
  - Origem: Usada pelos escribas hebreus em 600-500 a.C.
  - Classe: Substituição Simples.
  - Tipo: Monoalfabética (apenas um alfabeto cifrante) Monogrâmica (cada caracter substituído por apenas um outro) ou Substituição Simples.
  - Características: Reversível, ou seja, uma cifragem dupla devolve a mensagem original
  - Segurança: Baixíssima
  - Uso: Apenas em textos muito curtos
  - Criptoanálise: Uma simples análise da frequência de ocorrência das letras baseada na característica estatística da língua é suficiente para decifrar o texto.

SISTEMAS MONOALFABÉTICOS (CIFRA DE CESAR)

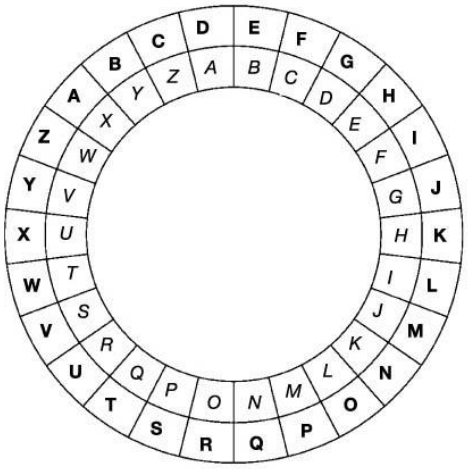
- Características
  - Origem: Usada pelo imperador romano Júlio César em 50 a.C.
  - Classe: Substituição Simples.
  - Tipo: Monoalfabética (porque usa apenas um alfabeto cifrante) Monogrâmica (porque trata cada um dos caracteres individualmente).
  - Segurança: Baixíssima
  - Uso: Aplicável apenas em textos muito curtos.
  - Criptoanálise: Uma simples criptoanálise baseada na característica estatística da língua é suficiente para decifrar o texto.

SISTEMAS MONOALFABÉTICOS (CIFRA DE CESAR)

- Júlio César usava na sua correspondência particular um código de substituição no qual cada letra da mensagem original era substituída pela letra que a seguia em três posições no alfabeto: a letra A era substituída por D, a B por E, e assim sucessivamente.

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	Z
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	C

- Criptografia
  - $l' = (l + ch) \bmod 26$
- decriptografia:
  - $l = (l' - ch) \bmod 26$
- Texto original: CRIPTOGRAFIA
- Texto cifrado: FULSXRJUDILD



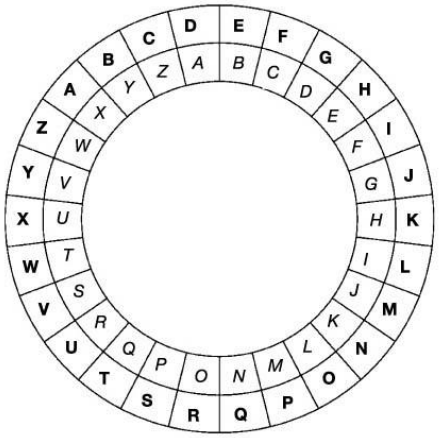
SISTEMAS MONOALFABÉTICOS (CIFRA DE CESAR)

- Cifra de Cesar
  - De acordo com o método de substituição monoalfabética geral, cada letra do texto original pode ser substituída por uma outra qualquer;
  - A chave é a relação entre as letras das mensagens;

a	b	c	d	e	f	g	h	i	j	k	l	m
0	1	2	3	4	5	6	7	8	9	10	11	12

n	o	p	q	r	s	t	u	v	w	x	y	z
13	14	15	16	17	18	19	20	21	22	23	24	25

- Exemplo: seja uma chave *ch*=3 (k=3) e a mensagem de texto puro
  - Texto original: COMPLEXO
  - Texto cifrado: FRPSOHAR



SISTEMAS POLIALFABÉTICOS (VIGENÈRE)

- Características
  - Criado pelo francês Blaise de Vigenère (1523 - 1596)
  - A cifra de Vigenère pertence à classe de substituições com palavra-chave.
  - O tipo da substituição é polialfabética monogrâmica (ou monográfica) porque faz uso de vários alfabetos cifrantes (polialfabética) aplicados individualmente (monogrâmica) aos caracteres da mensagem clara.
  - O método faz uso de chaves, que podem ser palavras ou frases.
  - A segurança da cifra era alta para a época - hoje é considerada baixa.

SISTEMAS POLIALFABÉTICOS (VIGENÈRE)

- O sistema de Vinegère é um sistema polialfabético. Ele adota como chave um conjunto de *p* letras:
  - *ch=(l1,l2,l3,...,lp)*
- A mensagem deve então ser dividida em blocos de *p* letras. Chama-se *p* de período do sistema polialfabético. Seja um bloco de texto puro *tp:*
  - *tp=(a1,a2,a3,...,ap)*
- A saída para este blobo será um outro, também com *p* letras. Este será a substituição de *tp* usando-se *ch*:
- *tc=((a1+l1)mod26,(a2+l2)mod26,...,(ap+lp)mod26)*

SISTEMAS POLIALFABÉTICOS (VINEGÈRE)

- Considerando uma mensagem M=VAMOSATACARAMANHASEMFALTA
- Usando uma chave ch=GRITO=(G,R,I,T,O)=(6,17,8,19,14)
- Como se tem p=5 deve-se separar a mensagem em blocos de 5 letras:
  - M=VAMOS.ATACA.RAMAN.HASEM.FALTA
- Fazendo a substituição. A mensagem criptografada M' será:
  - M'=BRUHG.GKIVO.XRUTB.NRAXA.LRTMO

SISTEMAS POLIALFABÉTICOS (VINEGÈRE)

- M=VAMOS.ATACA.RAMAN.HASEM.FALTA
- ch=GRITO. GRITO. GRITO. GRITO. GRITO
- M'=BRUHG.GKIVO.XRUTB.NRAXA.LRTMO

		Plaintext																								
		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**	**z**
Key	*a*	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	Z
	b	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A
	c	C	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
	d	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	C
	e	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D
	f	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E
	g	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F
	h	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G
	i	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H
	j	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I
	k	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J
	l	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K
	m	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L
	n	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M
	o	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N
	p	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
	q	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
	r	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
	s	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
	t	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
	u	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
	v	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
	w	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
	x	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
	y	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
	z	Z	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

BASTÃO DE LICURGO

- A cifra de transposição mais antiga é a do bastão de Licurgo, um bastão de madeira ao redor do qual se enrolava firmemente uma tira de couro ou pergaminho, longa e estreita. Escrevia-se a mensagem no sentido do comprimento do bastão e depois a tira era desenrolada, contendo a mensagem cifrada.

T	E	S	T	A	N	D	O		O
B	A	S	T	A	O		D	E	
L	I	C	U	R	G	O			

- TBLEAISSCTTUAARNOGDOODEO

BASTÃO DE LICURGO

- Características
  - Origem: Usada pelos gregos de Esparta em 475 a.C. (?)
  - Classe: Transposição
  - Segurança: Baixíssima
  - Uso: Apenas interesse histórico por ser o primeiro "dispositivo mecânico" de criptografia.
  - Criptoanálise: Uma simples criptoanálise estatística baseada na frequência de ocorrência das letras da língua é suficiente para decifrar o texto.

TRANSPOSIÇÕES REGULARES

- A grade é uma figura geométrica
- São baseadas em unidades ou ciclos que se repetem
- Em cada ciclo, o número de caracteres manipulados é sempre o mesmo
- O número total de caracteres da mensagem clara ou cifrada por um múltiplo do número de caracteres do ciclo
- As figuras geométricas usadas são os quadrados, retângulos e triângulos, entre outros.

TRANSPOSIÇÕES REGULARES

- Tipos de inserção
  - Análise com uma grade quadrada 3X3 (três linhas e três colunas) com uma seqüência de caracteres ABCDEFGHI.
    - Inserção dos caracteres na horizontal, da esquerda para a direita e de cima para baixo. Retirando as colunas da esquerda para direita e de cima para baixo.
      - Resultado: ADG BEH CFI.

A	B	C
D	E	F
G	H	I

TRANSPOSIÇÕES DA ATUALIDADE

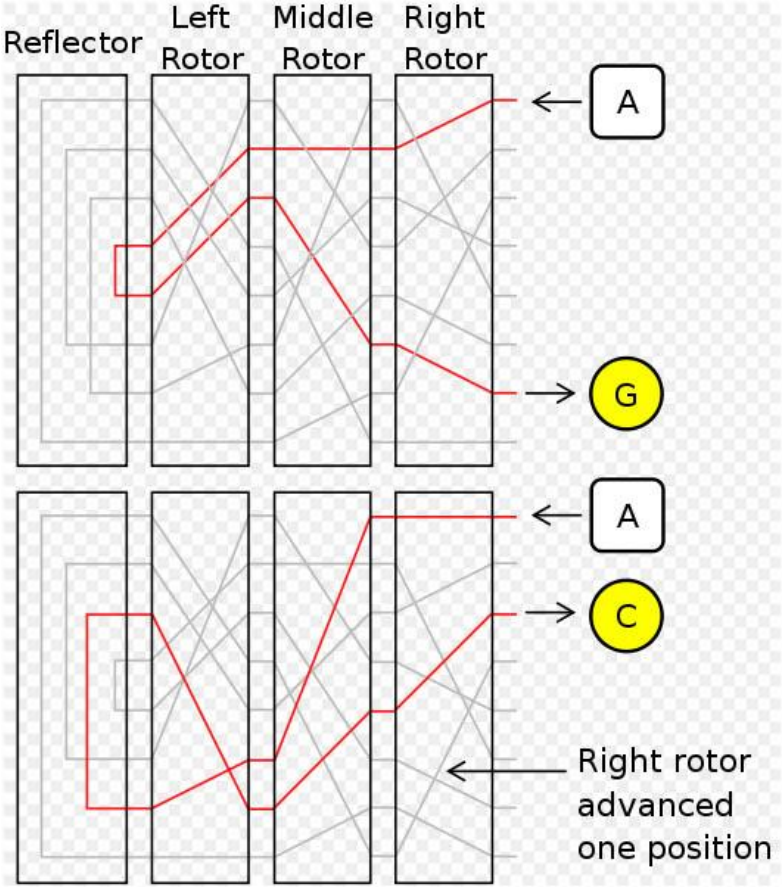
- DES (Data Encryption Standard). O primeiro passo desse algoritmo é transformar a chave principal em 16 subchaves. O método usado é uma transposição de bits de acordo com uma tabela de permutação (PC-1). A seguir, os bits de cada uma das 16 subchaves sofrem uma rotação, ou seja, um deslocamento a esquerda.
- Exemplo de rotação de bits.

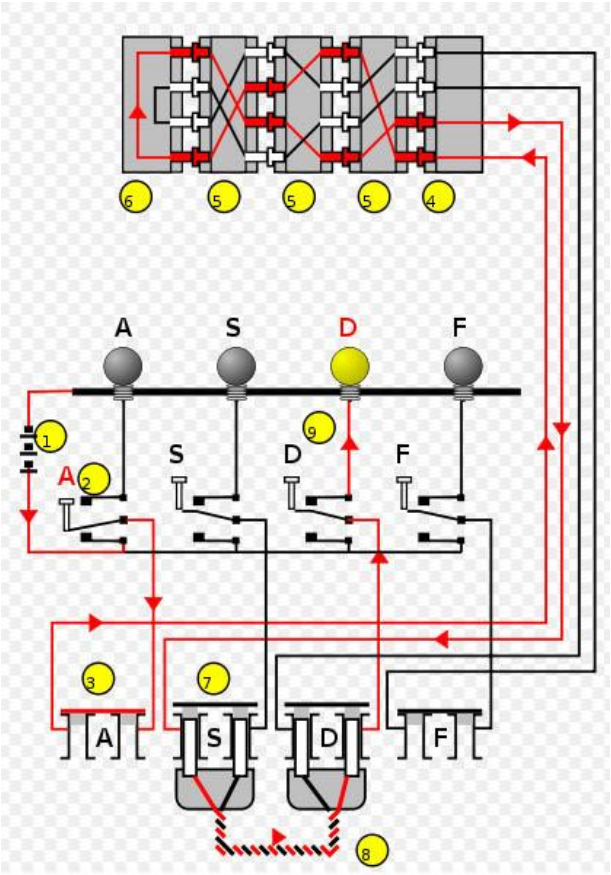
Caractere	ASCII	Valor binário	Rotação para a direita	Novo caractere	ASCII
B	66	01000010	00100001	!	33
Z	90	01011010	00101101	-	45
B	98	01100010	00110001	1	49
z	122	01111010	00111101	=	61

DISPOSITIVOS CRIPTOGRÁFICOS (ENIGMA)

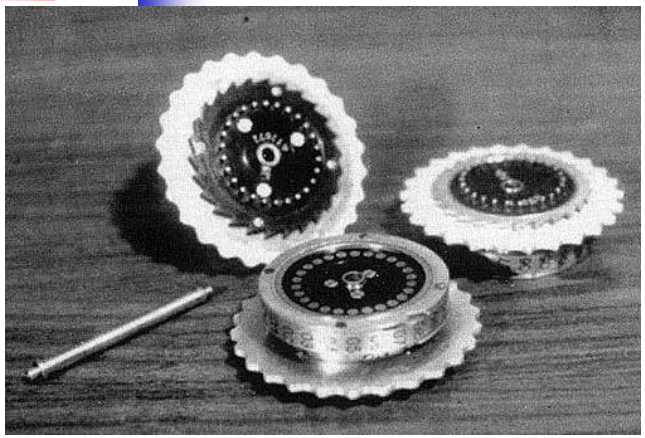
- Enigma
- Primeiro cifrador mecânico comercial;
  - Adotado pela Alemanha em 1933 e usado pelo exército e pela marinha;
- Segunda Guerra Mundial
  - – Enigma (Alemão): Algoritmo Conhecido;
  - » Usado na coordenação de ataques de submarinos no Atlântico Norte.
- Simulador do Enigma:
  - http://enigmaco.de/_fs/index-enigma.html
  - http://users.telenet.be/d.rijmenants/en/enigmasim.htm

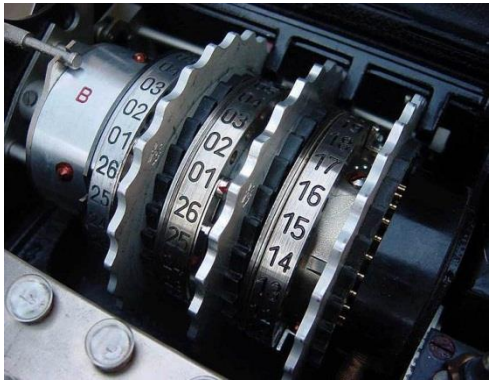
DISPOSITIVOS CRIPTOGRÁFICOS (ENIGMA)

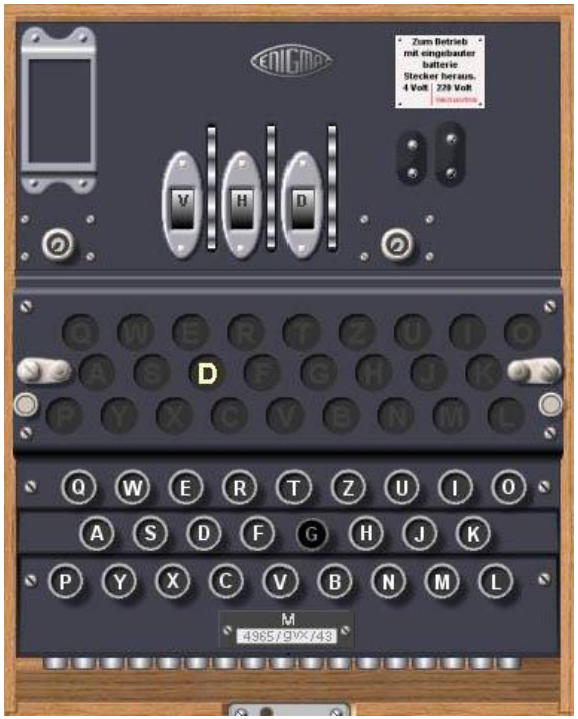




DISPOSITIVOS CRIPTOGRÁFICOS (ENIGMA)







DISPOSITIVOS CRIPTOGRÁFICOS (ENIGMA)

- Cada vez que a máquina era usada para fazer uma cifragem o último misturador era movido de uma posição.
  - Cada vez que o último completava uma volta, o penúltimo misturador era movido de uma posição.
  - Cada vez que o penúltimo completava uma volta, o antepenúltimo misturador era movido de uma posição (como num odômettro de carro).
  - Deste modo, a posição dos misturadores só se repetia a cada 26x26x26 cifragens (17.576).
  - Os 3 rotores poderiam ser colocados em seqüências diferentes (123, 132, 213, 231, 312, 321), dessa forma 6 X 17.576 = 105.456.
  - Com cruzamento de ligações no entre os rotores, pode-se ter em torno de 100 bilhões de possibilidades.

DISPOSITIVOS CRIPTOGRÁFICOS (ENIGMA)

- Cálculo da segurança do Enigma:

Elemento	Contribuição	Cálculo
(1)Posições do 3 rotores	17.576	26x26x26
(2)Sequência do 3 rotores	6	3!
(3)Substituir 6 letras entre 26	100.391.791.500	26!/26x6!x14!
Total da primeira versão	10.586.916.764.424.000	(1)X(2)x(3)
(4)Três entre cinco rotores	10	5!/3!(5-3)!
Total	105.869.167.644.240.000	(1)X(2)x(3)x(4)

DISPOSITIVOS CRIPTOGRÁFICOS (ENIGMA)

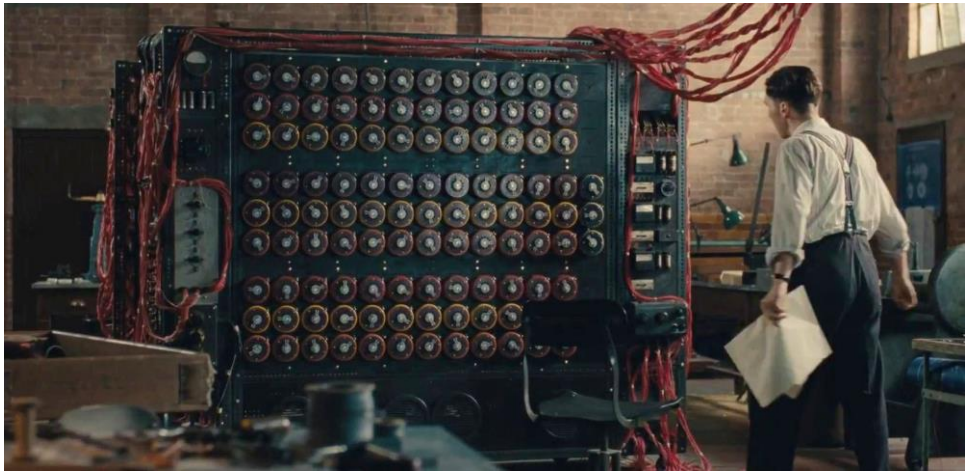
- Em 1938, Turing se uniu ao GC&CS, o braço de decodificação de mensagens da inteligência britânica, para efetuar a Criptoanálise da Máquina Enigma.
- O Enigma era uma máquina de codificação que mudava seus códigos diariamente, obrigando a que o projeto de decifração se tornasse bastante rápido.
- Após o Reino Unido iniciar a Segunda Guerra Mundial ao declarar guerra à Alemanha em 1939, Turing foi direcionado para o quartel da GC&CS em Bletchley Park.

DISPOSITIVOS CRIPTOGRÁFICOS (ENIGMA)

- A partir de uma máquina decodificadora polonesa, Turing projetou a Bomba eletromecânica ("Bombe"), um equipamento eletromecânico que ajudaria a decriptar as mensagens do Enigma e foi montada em 1940.
- Novas Bombas foram construídas após Turing e sua equipe pedirem apoio à Winston Churchill, e mais de duzentas operavam ao fim da Guerra em 1945.
- Turing também introduziu sua equipe em Bletchley Park ao matemático Tommy Flowers, que em 1943 projetou o Colossus, um computador primitivo que ajudou a decodificar outra máquina criptográfica alemã, o Lorenz.

DISPOSITIVOS CRIPTOGRÁFICOS (ENIGMA)

- Recomendável para assistir: O Jogo da Imitação.

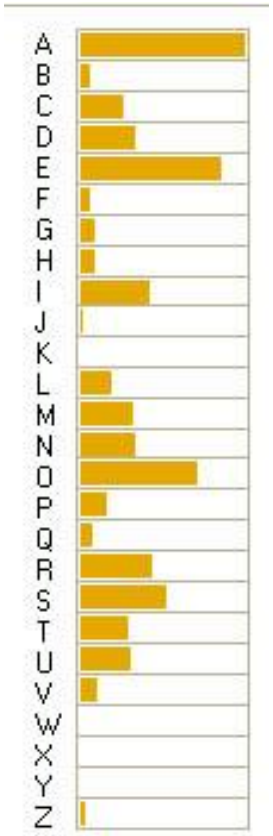


CRIPTOANÁLISE

- A criptologia é uma ciência que se ocupa da criptografia e da criptoanálise. A criptografia é o estudo dos métodos e das técnicas de cifragem. A criptoanálise, por sua vez, se ocupa do estudo das formas de se "quebrar" sistemas.
- Criptoanálise é a ciência de quebrar uma mensagem cifrada. Quebrar não é o mesmo que decifrar. Decifrar é obter a mensagem original quando se conhece o sistema e usando a chave também conhecida. Quebrar é hackear o sistema e descobrir a chave.

CRIPTOANÁLISE

- Freqüência de incidência de letras no português



Letra	Freq.%	Letra	Freq.%
A	14.63	N	5.05
B	1.04	O	10.73
C	3.88	P	2.52
D	4.99	Q	1.20
E	12.57	R	6.53
F	1.02	S	7.81
G	1.30	T	4.34
H	1.28	U	4.63
I	6.18	V	1.67
J	0.40	W	0.01
K	0.02	X	0.21
L	2.78	Y	0.01
M	4.74	Z	0.47

CRIPTOANÁLISE

- Características do Português do Brasil
  - O comprimento médio das palavras em Português do Brasil é de 4.53 letras.
  - Quando as letras são ordenadas pela freqüência, formam grupos bem definidos:
    - A, E, O
    - S, R, I
    - N, D, M, U, T, C
    - L, P, V, G, H, Q, B, F
    - Z, J, X, K, W, Y

Letras	Freq.
6 vogais: A, E, I, O, U, (Y)	48.75 %
20 consoantes	
5 de frequência alta: S, R, N, D, M	49.12 %
10 de frequência média: T, C, L, P, V, G, H, Q, B, F	21.03 %
6 de frequência baixa: Z, J, X, K, W	1.10 %
	100.00 %

  - As vogais A, E, I, O, U e as consoantes S, R, N, D, M formam mais de 3/4 dos textos em Português.
  - A média de vogais a cada 10 letras é de 4.88

CRIPTOANÁLISE

- Métodos estatísticos:
  - A ponte entre a estatística e a criptoanálise foi construída pelo casal Friedman, a partir da década de 1920. Os testes usados são o teste kappa, o teste phi e o índice de coincidência de Friedman.
    - **Teste kappa**: diferencia resultados aleatórios, ocorridos ao acaso, de resultados decorrentes de regras fixas. Supondo uma caixa contendo 26 letras do alfabeto. A chance de tirar a letra M da caixa é de 1 em 26, ou seja, 1/26 ou 0,0385. Indica que se fizermos 100 retiradas, pode-se tirar a letra M é de 3 ou 4 vezes.
    - O kappa pode ser usado para determinação do idioma usado. Se comparar dois textos diferentes em português, ambos com o mesmo número de letras e colocados um sob o outro, quantas colunas teriam a mesma letra? Calculando a soma das probabilidades de ocorrência de cada uma das letras, encontra-se a resposta. Por exemplo, Se a chance da letra A aparecer é de 15 em 100 (15/100), a chance de encontrá-la repetida em uma mesma coluna é de 15/100x15/100=0,0225. Calculando a probabilidade de repetição de cada letra e somar os resultados, resulta em 0,0781 para o português, ou seja, para 100 colunas, existe a chance de encontrar 7 à 8 letras repetidas (0,0781x100=7,81).
    - Para o francês é 0,0778, alemão é 0,0754, italiano é 0,0738, espanhol é 0,0775, inglês é 0,0667.

CRIPTOANÁLISE

- Métodos estatísticos:
  - **Índice de Coincidência**: o I.C. de Friedman é a relação entre o quanto existe de predeterminado e quanto de existe de randômico numa determinada amostra.
  - O valor kappa da língua portuguesa é igual a 0,0781 e o valor randômico das 26 letras do alfabeto latino é 0,0385, portanto o IC=0,0781/0,0385=2,03.
  - No francês, o IC=0,0778/0,0385=2,02.
  - No espanhol, o IC=0,0775/0,0385=2,01.
  - No inglês, o IC=0,0667/0,0385=1,73.
  - Um texto totalmente randômico tem um IC=0,0385/0,0385=1.
  - **Teste phi**: em 1935, Solomon Kullback, assistente de Friedman, criou o teste phi. Este teste foi baseado no princípio da coincidência de Friedman e facilita a determinação do método de cifragem utilizado (transposição, substituição monoalfabética ou substituição polialfabética). O teste phi é realizado em várias etapas: inicialmente se determinam os valores limites de phi randômico e de phi do texto claro para determinado idioma, em seguida, calcula-se o valor de phi do criptograma e se compara o resultado com os valores limites.
  - Exemplo: conforme o criptograma analisado: IBEIQJ MHKIQJ JNQ AXOXMBXJ AB ABMXOKNK PQKLHB N OKBLHBRMXN AB QMQKKBRMXN ANJ TBIKNJ RNQ NWHAN SHXIQ.

CRIPTOANÁLISE

- Métodos estatísticos (Teste phi):
  - Identificação do idioma para obter o valor kappa adequado. Será chamada de kappa do idioma de k(i). O português possui k(i)=0,0781.
  - Contagem do número de letras no criptograma, N=84.
  - Cálculo do phi randômico (de acordo com o número de letras no criptograma), phi(r)=k(r)xNx(N-1)=0,0385x84x83=268.
  - Cálculo do phi do texto claro phi(t). Phi(t)=k(i)xNx(N-1)=0,0781x84x83=545.
  - Frequência de ocorrência de cada letra no texto criptografado:

f	6	10	0	0	1	0	0	5	5	6	8	2	6	10	3	1	8	3	1	1	0	0	1	7	0	0
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	z
f(f-1)	30	90	0	0	0	0	0	20	20	30	56	2	30	90	6	0	56	6	0	0	0	0	0	42	0	0

  - Cálculo phi de cada letra. Letra A tem f(f-1)=6(6-1)=30.
  - Somas do phi de cada letra para obter o phi do criptograma: phi(c)=30+90+20+20+30+56+2+30+90+6+56+6+42=478
  - Comparação: o phi(c)=478 está mais próximo do phi(t)=545 do que do phi(r)=268, isso mostra uma grande probabilidade do criptograma não ser uma coleção de letras tomadas ao acaso.
  - Resposta do criptograma: “Textos curtos são difíceis de decifrar porque a frequência de ocorrência das letras não ajuda muito”

CRIPTOANÁLISE

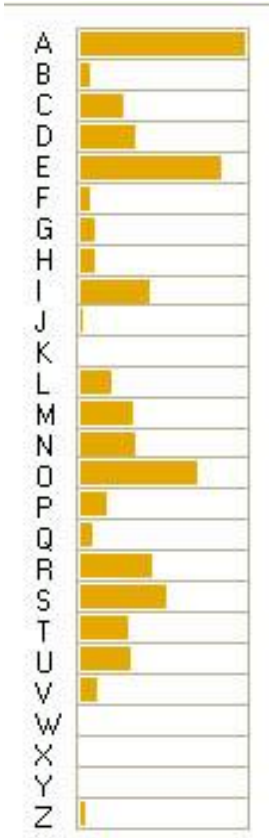
- Métodos estatísticos:
  - A ponte entre a estatística e a criptoanálise foi construída pelo casal Friedman, a partir da década de 1920. Os testes usados são o teste kappa, o teste phi e o índice de coincidência de Friedman.
    - Exemplo: conforme o criptograma analisado: IBEIQJ MHKIQJ JNQ AXOXMBXJ AB ABMXOKNK PQKLHB N OKBLHBRMXN AB QMQKKBRMXN ANJ TBIKNJ RNQ NWHAN SHXIQ.
    - Identificação do idioma para obter o valor kappa adequado. Será chamada de kappa do idioma de k(i). O português possui k(i)=0,0781.
    - Contagem do número de letras no criptograma, N=84.
    - Cálculo do phi randômico (de acordo com o número de letras no criptograma), phi(r)=k(r)xNx(N-1)=0,0385x84x83=268.

CRIPTOANÁLISE

- Sistemas de transposição (transposição geométrica simples)
  - Texto criptografado:
    - ORSLN ENRQU AIOEH UDCPU SDRHL AEIOE EHOAA ERQAN RQEAQ OSAUL TFUAI UPQEA OREEV GEEUU LGAIR INETE RQORT STOUR ROUDA OEORM OEPEE MUPRA PLMAR MEMRI NREMD ROMHO ACEOE EECBO VDITT SSSRU MAOAE AMFQA SERFN XMOAU THMFR OTBEZ IIOFO APOEH ECCLR RCOCM URIOI ATADL MMAOE DCESE AAIOD EEOBS R

CRIPTOANÁLISE

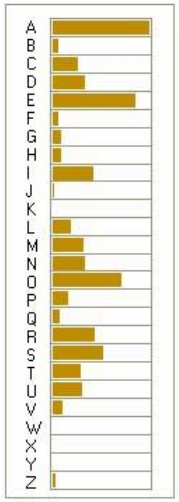
f	25	3	8	8	34	5	2	6	12	0	0	7	14	6	27	6	7	24	10	9	14	10	0	1	0	1
letra	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z
f(f-1)	600	6	56	56	1122	20	2	30	132	0	0	42	182	30	702	42	42	552	90	72	182	90	0	0	0	0



- Identificar o idioma do texto (texto curto).
- Frequência de ocorrência de letras:
- Total de 231 caracteres, a porcentagem de vogais (A E I O U): 48.5% ou 112 em 231.
- A letra mais frequente é o E (15%).
- As consoantes de alta frequência (28%) são R M S T D, as de média frequência (21%) são C Q L P N H F B e as de baixa frequência (3%) são V G Z X W K J.
- Pelos primeiros resultados, e principalmente pela forte presença da **letra E**, o idioma tem uma grande probabilidade de ser **Português** ou Espanhol.

CRIPTOANÁLISE

- Português

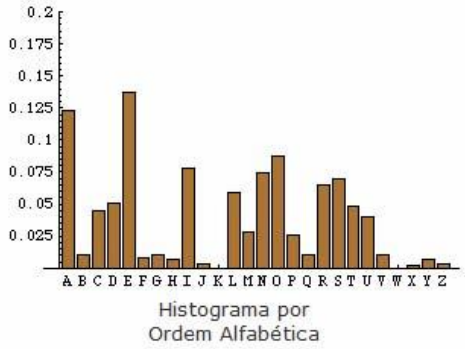


Histograma por
Ordem Alfabética

Letra	Freq.%	Letra	Freq.%
A	14.63	N	5.05
B	1.04	O	10.73
C	3.88	P	2.52
D	4.99	Q	1.20
E	12.57	R	6.53
F	1.02	S	7.81
G	1.30	T	4.34
H	1.28	U	4.63
I	6.18	V	1.67
J	0.40	W	0.01
K	0.02	X	0.21
L	2.78	Y	0.01
M	4.74	Z	0.47

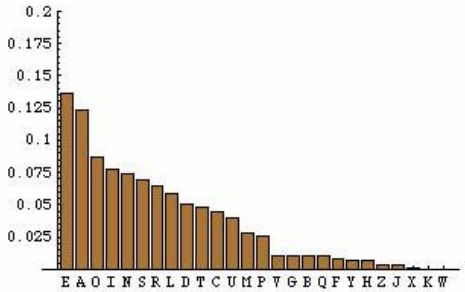
Histograma por
Ordem de Frequência

- Espanhol



Histograma por
Ordem Alfabética

Letra	Freq.%	Letra	Freq.%
A	12.30	N	7.41
B	1.03	O	8.68
C	4.49	P	2.63
D	5.04	Q	1.02
E	13.69	R	6.44
F	0.77	S	6.97
G	1.04	T	4.82
H	0.65	U	3.99
I	7.78	V	1.04
J	0.28	W	0.02
K	0.02	X	0.16
L	5.84	Y	0.66
M	2.84	Z	0.34



Histograma por
Ordem de Frequência

CRIPTOANÁLISE

- Teste phi:
- Contagem de letras: N=231
- Phi(r)=k(r)xNx(N-1)=0,0385x231x230=2.046
- Phi(t)=k(i)xNx(N-1)=0,0781x231x230=4.150
- Frequência de ocorrência de cada letra.

f	25	3	8	8	34	5	2	6	12	0	0	7	14	6	27	6	7	24	10	9	14	10	0	1	0	1
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	z
f(f-1)	600	6	56	56	1122	20	2	30	132	0	0	42	182	30	702	42	42	552	90	72	182	90	0	0	0	0

- Cálculo do phi de cada letra: f(f-1)
- Phi do criptograma: phi(c)=600+6+56+56+.....90=4.050.
- Comparação: o phi(c) está mais próximo do phi(t), que indica que o criptograma segue regras.

CRIPTOANÁLISE

- Determinando o tipo de cifra
  - O Índice de Coincidência pode nos revelar se se trata de uma cifra de transposição/substituição monoalfabética ou de uma substituição polialfabética.

$$I = \frac{\sum_{i=1}^{26} n_i \cdot (n_i - 1)}{n \cdot (n-1)}$$

f	25	3	8	8	34	5	2	6	12	0	0	7	14	6	27	6	7	24	10	9	14	10	0	1	0	1
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	z
f(f-1)	600	6	56	56	1122	20	2	30	132	0	0	42	182	30	702	42	42	552	90	72	182	90	0	0	0	0

- Essa fórmula permite calcular a probabilidade da ocorrência de cada uma das letras do alfabeto num texto cifrado. *n1* corresponde a A e *n* ao número total de letras do texto analisado. Repete-se o cálculo com *n2* para B até *n26* para Z.
- Inicialmente tomemos o A como exemplo: esta letra ocorre 25 vezes no texto, portanto basta calcular 25 x (25 - 1) = 25 x 24 = 600.
- Calcular n(n-1) = 231 x (231 - 1) = 231 x 230 = 53130. Conforme a fórmula, basta dividir um pelo outro, ou seja, 600/53130 = 0,01129

CRIPTOANÁLISE

- Determinando o tipo de cifra
  - O Índice de Coincidência pode nos revelar se se trata de uma cifra de transposição/substituição monoalfabética ou de uma substituição polialfabética.
  - Calculando o índice de coincidência obtem-se o valor **Kappa = 0.076228**, que corresponde a um **IC = 1.98** (0.076228/0.0385), muito distante do IC de um texto randômico que seria 1.0 (0.0385/0.0385). Portanto, praticamente confirma-se a hipótese de transposição.
  - Como a distribuição da frequência de ocorrência das letras do texto cifrado segue um padrão quase idêntico ao encontrado no Português, podemos afirmar que se trata de uma **transposição**, pois uma substituição alteraria esta distribuição.
  - Um IC alto, como o encontrado, é típico de idiomas como o Português, o Francês e o Espanhol, porém apenas o Português e o Espanhol apresentam a letra E com uma frequência tão elevada.

CRIPTOANÁLISE

- Determinando o formato da matriz
  - Como se trata de uma transposição, deve-se determinar é uma matriz total ou parcialmente preenchida. Pode ter um formato quadrado ou retangular. Se todas as células foram preenchidas para se obter o texto cifrado, então deve-se encontrar uma divisão exata para as 231 letras:
    - 231|3
    - 77|7
    - 11|11
    - 0
  - Esta pode ser 3x77, 7x33, 11x21, 21x11, 33x7 e 77x3.
  - Os formatos 11x21 e 21x11 parecem os melhores, se bem que não excluem as outras possibilidades. Serão analisados primeiro.
  - Além disto, existe uma importante característica da língua portuguesa: quase 50% dos caracteres são vogais. Analisando a porcentagem das vogais em cada linha obtém-se o seguinte resultado:

CRIPTOANÁLISE

- Determinando o formato da matriz

```
ORSLNENRQUA4 (36%)
IOEHUDCPUSD5 (46%)
RHLAEIOEEHO7 (64%)
AAERQANRQEA6 (54%)
QOSAULTFUAI6 (54%)
UPQEAOREEVG6 (54%)
EEUULGAIRIN7 (64%)
ETERQORTSTO4 (36%)
URROUDAOEOR7 (64%)
MOEPEEMUPRA6 (54%)
PLMARMEMRIN3 (27%)
REMDROMHOAC4 (36%)
EOEEECBOVDI7 (64%)
TTSSSRUMAOA4 (36%)
EAMFQASERFN4 (36%)
XMOAUTHMFRO4 (36%)
TBEZIIOFOAP6 (54%)
OEHECCLRRCO4 (36%)
CMURIOIATAD6 (54%)
LMMAOEDCESE5 (46%)
AAIODEEOBSR7 (64%)
```

```
ORSLNENRQUAIOEHUDCPUS     9 (43%)
DRHLAEIOEEHOAAERQANRQ     11 (52%)
EAQOSAULTFUAIUPQEAORE     13 (62%)
EVGEEUULGAIRINETERQOR     11 (52%)
TSTOURROUDAOEORMOEPEE     12 (57%)
MUPRAPLMARMEMRINREMDR     6 (29%)
OMHOACEOEEECBOVDITTSS     10 (48%)
SRUMAOAEAMFQASERFNXMO     9 (43%)
AUTHMFROTBEZIIOFOAPOE     11 (52%)
HECCLRRCOCMURIOIATADL     8 (38%)
MMAOEDCESEAAIODEEOBSR     12 (57%)

A matriz 21x11 parece ter
uma distribuição de vogais
ligeiramente melhor, com
valores mais próximos de
48% e com uma distribuição
mais uniforme.
```

CRIPTOANÁLISE

- Os métodos mais comuns de transposição são entrada por linha/saída por coluna, entrada por coluna/saída por linha e transposição de colunas numeradas (chave). Em todos estes métodos, a mensagem clara é obtida pelo processo inverso, ou seja, uma entrada por linha/saída por coluna é decifrada com uma entrada por coluna/saída por linha.
- Testando o primeiro processo nas duas matrizes, obtemos as seguintes matrizes transpostas:

```
ODEETMOSAHM
RRAVSUMRUEM
SHQGTPHUTCA
LLOEOROMHCO
NASEUAAAMLE
EEAURPCOFRD     OIRAQUEEUMPRETEXTOCLA
NIUURLEARRC     ROHAOPETROLEOTAMBEMMA
ROLLOMOEOCE     SELESQUEREMMESMOEHUMI
QETGUAEATOS     LHARAEUROPADESFAZERAO
UEFADREMBCE     NUEQUALQUERRESQUICIOD
AHUIAMEFEMA     EDIALOGODEMOCRATICOEE
IOAROECQZUA     NCONTRARAMEMBUSHOLIDE
OAIIEMBAIRI     RPERFEITOUMHOMEMFRACO
EAUNOROSIIO     QUEQUERSEPROVARFORTEB
HEPERIVEOOD     USHEAVITORIADOFRACASS
URQTMNDRFIE     ADOAIGNORANCIANOPODER
DQEEORIFOAE
CAAREETNATO
PNOQPMTXPAB
URROEDSMODS
SQERERSOELR
```

CRIPTOANÁLISE

- Solução:
  - Sem acentos: O IRAQUE E UM PRETEXTO CLARO HA O PETROLEO TAMBEM MAS ELES QUEREM MESMO E HUMILHAR A EUROPA DESFAZER A ONU E QUALQUER RESQUICIO DE DIALOGO DEMOCRATICO E ENCONTRARAM EM BUSH O LIDER PERFEITO UM HOMEM FRACO QUE QUER SE PROVAR FORTE BUSH E A VITORIA DO FRACASSADO A IGNORANCIA NO PODER
  - Com acentos e sinais de pontuação: **O Iraque é um pretexto. Claro, há o petróleo também, mas eles querem mesmo é humilhar a Europa, desfazer a ONU e qualquer resquício de diálogo democrático. E encontraram em Bush o líder perfeito: um homem fraco que quer se provar forte. Bush é a vitória do fracassado, a ignorância no poder.**

CRIPTOANÁLISE

- Sistemas monoalfabéticos

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	Z
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	C

  - Análise de frequencia de ocorrência das letras
  - Método de “correr o alfabeto” (somente para casos da cifra de César)

Criptograma	E	G	O	A	E	P
1	F	H	P	B	F	Q
2	G	I	Q	C	G	R
3	H	J	R	D	H	S
4	I	K	S	E	I	T
5	J	L	T	F	J	U
6	K	M	U	G	K	V
7	L	N	V	H	L	W
8	M	O	W	I	M	X
9	N	P	X	J	N	Y
10	O	Q	Y	K	O	Z
11	P	R	Z	L	P	A
12	Q	S	A	M	Q	B
13	R	T	B	N	R	C
14	**S**	**U**	**C**	**O**	**S**	**D**

CRIPTOANÁLISE

- Sistemas monoalfabéticos
  - Mensagem criptografada: “**R VAFHSVPVRAGR ABF CEBGRTRE PBZ YRVF. GRZBF DHR ABF CEBGRTRE PBZ ZNGRZNGVPN**”
  - Observações
    - Os espaços entre as palavras foram conservados.
    - Letras acentuadas e letras especiais (como ç) foram substituídas pelas letras originais do alfabeto.
    - A mensagem está em português.

CRIPTOANÁLISE

- Mensagem criptografada: “**R VAFHSVPVRAGR ABF CEBGRTRE PBZ YRVF. GRZBF DHR ABF CEBGRTRE PBZ ZNGRZNGVPN**”

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 Z
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 C

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

CRIPTOANÁLISE

- Mensagem criptografada: “**R VAFHSVPVRAGR ABF CEBGRTRE PBZ YRVF. GRZBF DHR ABF CEBGRTRE PBZ ZNGRZNGVPN**”
- Mensagem criptografada: “**E INSUFICIENTE NOS PROTEGER COM LEIS. TEMOS QUE NOS PROTEGER COM MATEMATICA**” → **Cifra de Cesar (K=13)**

CODIFICAÇÃO BASE64

- **Base64** é um método para codificação de dados para transferência na Internet (*codificação MIME para transferência de conteúdo*). É utilizado frequentemente para transmitir dados binários por meios de transmissão que lidam apenas com texto, como por exemplo para enviar arquivos anexos por e-mail.
- É constituído por 64 caracteres ([A-Z],[a-z],[0-9], "/" e "+") que deram origem ao seu nome. O carácter "=" é utilizado como um sufixo especial e a especificação original (RFC 989) definiu que o símbolo "*" pode ser utilizado para delimitar dados convertidos, mas não criptografados, dentro de um *stream*.

CODIFICAÇÃO BASE64

- Esse método consiste em pegar **três bytes**, dividir estes bytes em quatro grupos de seis bits. Cada conjunto de seis bits pode ser de 0 a 63, que é, 64 possíveis valores (daí a origem do nome do método). Então, a codificação em base64 utiliza a seguinte tabela:

A	B	C	D	E	F	G	H	I	J	K	L	M
0	1	2	3	4	5	6	7	8	9	10	11	12
N	O	P	Q	R	S	T	U	V	W	X	Y	Z
13	14	15	16	17	18	19	20	21	22	23	24	25
a	b	c	d	e	f	g	h	i	j	k	l	m
26	27	28	29	30	31	32	33	34	35	36	37	38
n	o	p	q	r	s	t	u	v	w	x	y	z
39	40	41	42	43	44	45	46	47	48	49	50	51
0	1	2	3	4	5	6	7	8	9	+	/	
52	53	54	55	56	57	58	59	60	61	62	63	

CODIFICAÇÃO BASE64

- É possível que ao codificar, não se consiga obter três bytes. Neste caso, é necessário preencher o conteúdo da mensagem. Isso pode ser realizado usando o símbolo de preenchimento ==.
- **Exemplo: CRIPTOGRAFIA**

A	B	C	D	E	F	G	H	I	J	K	L	M
65	66	67	68	69	70	71	72	73	74	75	76	77
N	O	P	Q	R	S	T	U	V	W	X	Y	Z
78	79	80	81	82	83	84	85	86	87	88	89	90

Texto	C	R	I	P	T	O	G	R	A	F	I	A
ASCII	67	82	73	80	84	79	71	82	65	70	73	65
Binário	01000011	01010010	01001001	01010000	01010100	01001111	01000111	01010010	01000001	01000110	01001001	01000001

CODIFICAÇÃO BASE64

- 010000 110101 001001 001001 010100 000101 010001 001111 010001 110101 001001 000001 010001 100100 100101 000001

A	B	C	D	E	F	G	H	I	J	K	L	M
0	1	2	3	4	5	6	7	8	9	10	11	12
N	O	P	Q	R	S	T	U	V	W	X	Y	Z
13	14	15	16	17	18	19	20	21	22	23	24	25
a	b	c	d	e	f	g	h	i	j	k	l	m
26	27	28	29	30	31	32	33	34	35	36	37	38
n	o	p	q	r	s	t	u	v	w	x	y	z
39	40	41	42	43	44	45	46	47	48	49	50	51
0	1	2	3	4	5	6	7	8	9	+	/	
52	53	54	55	56	57	58	59	60	61	62	63	

Binário	**010000**	**110101**	**001001**	**001001**	**010100**	**000101**	**010001**	**001111**	**010001**	**110101**	**001001**	**000001**
Decimal	16	53	9	9	20	5	17	15	17	53	9	1
Base64	Q	1	J	J	U	F	R	P	R	1	J	B

Binário	**010001**	**100100**	**100101**	**000001**
Decimal	17	36	37	1
Base64	R	k	l	B

CODIFICAÇÃO BASE64

- **Texto Original: TESTE DE CRIPTOGRAFIA**
- **Base64: VEVTVEUgREUgQ1JJUFRPR1JBRklBCg==**

CODIFICAÇÃO BASE64

1. Exemplo: Codificando com base64:

- root@kali:~# echo -n "TESTE DE CRIPTOGRAFIA COM CIFRAS" | base64
- VEVTVEUgREUgQ1JJUFRPR1JBRklBIENPTSBDSUZSQVM=

2. Revertendo um Base64 e decodificando:

2.1. Revertendo o Base64

- root@kali:~# echo -n =MVQSZUSDBSTPNEIBlkRBJ1RPRFUJJ1QgUERgUEVTVEV | rev

2.2. Decodificando:

- root@kali:~# echo -n VEVTVEUgREUgQ1JJUFRPR1JBRklBIENPTSBDSUZSQVM= | base64 -d
- Resultado: TESTE DE CRIPTOGRAFIA COM CIFRAS

BIBLIOGRAFIA

- STALLINGS, W. Criptografia e Segurança de Redes - Princípios e Práticas - 6ed., Pearson, 2015.
- TKOTZ, V. Criptografia Numa Boa. Disponível em: <http://numaboa.com.br/criptografia>. Acesso em: 03.08.2024.
- TURING, A. O Jogo da Imitação - <https://www.youtube.com/watch?v=hOsme_E8qzc > Acesso em: 03.08.2024.
- ENIGMA.<https://www.ciphermachinesandcryptology.com/en/enigmasim.htm > Acesso em: 03.08.2024.
- CESAR. <https://cryptii.com/pipes/caesar-cipher>. Acesso em: 03.08.2024.
- VIGENERE. <https://www.dcode.fr/vigenere-cipher>. Acesso em: 03.08.2024.
- BASE64. <https://www.base64encode.org/>. Acesso em: 03.08.2024.