CiberSegurança

Módulo 1 - 01. INTRODUÇÃO. CIFRAS CLÁSSICAS

MEET, MEIC, MEIM

2021–2022

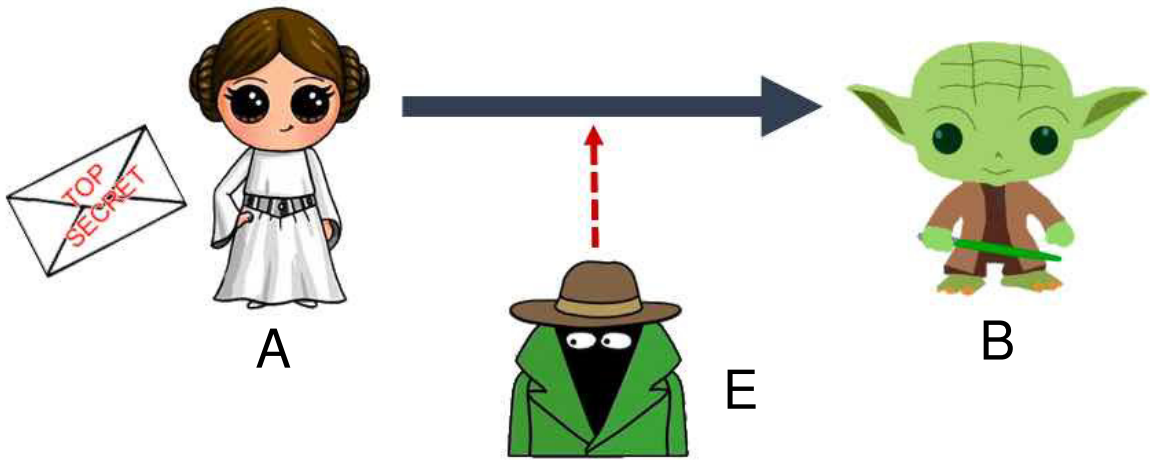
Módulo I

Mecanismos criptográficos para proteção da informação

- Conceitos de aritmética modular e curvas elípticas
- Cifras simétricas e assimétricas.
- Funções de Hash criptográficas
- Aplicação: assinaturas digitais e cifra autenticada

No modelo mais simples de comunicação encontramos três **entidades** (*entity, party*):

- **remetente** (*sender*) da informação, designado por *A* (*Alice*);
- **destinatário** (*receiver*), designado por *B* (*Bob*);
- **adversário** (*adversary, eavesdropper*), designado por *E* (*Eve*).



Criptografia versus esteganografia

- **Esteganografia**, do grego *steganos* (encoberto) e *graphos* (escritura).
  Técnicas que permitem ocultar a existência mesma da mensagem (tintas invisíveis, microfilmes ...).

- **Criptografia**, do grego *krypto* (oculto) e *graphos* (escritura).
  Técnicas que ocultam o significado real da mensagem, tornando-o ininteligível a terceiros não autorizados.

Cifra versus codificar

Termos usados indistintamente na linguagem corrente, no entanto, formalmente:

- **cifra**: a base da encriptação é uma letra ou blocos de letras, que se transformam em outras letras ou blocos de letras através de um procedimento que pode ser replicado para todas as combinações possíveis de letras ou de blocos de letras;
- **codificação**: a unidade de encriptação são palavras ou frases, que são encriptadas através de livros de códigos (dicionários) e não é possível encriptar palavras que não figurem no livro.

$\rightsquigarrow$ O sistemas de cifra são mais versáteis que os sistemas de codificação pelo que, atualmente, devido à proliferação das comunicações na nossa era digital, o termo **cifrar** é usado quase como sinónimo de **encriptar**.

- **texto limpo** (*plaintext*): mensagem a transmitir em segurança, normalmente escrito em minúsculas;
- **cifra**: operação que transforma o texto limpo numa mensagem *"com significado oculto"*;
- **texto cifrado** ou **criptograma** (*ciphertext*) à mensagem obtida com a cifra a partir do texto limpo, normalmente escrito em maiúsculas;
- **criptoanálise**: estudo dos procedimentos necessários para tentar comprometer as técnicas criptográficas;
- **criptologia**: conjunto de técnicas criptográficas e criptoanalíticas (área transversal, entre a matemática e as ciências da computação).

Objetivos principais da criptografia

O objetivo inicial da criptografia era tornar ininteligível a terceiros não autorizados a informação enviada entre duas entidades.

Atualmente, como o consequência da revolução digital, são considerados objetivos principais da criptografia :

1. a **confidencialidade** (*confidentiality, secrecy, privacy*);
2. a **integridade da informação** (*data integrity*);
3. a **autenticação** (*authentication*);
4. o **não repúdio** (*non-repudiation*).

Descrição dos objetivos

1. a **confidencialidade**: manter o conteúdo da informação secreto para todos excepto para o destinatário da mesma;
2. a **integridade da informação**: assegurar que não há alteração da informação por entidades não autorizadas;
3. a **autenticação**: autenticar a identidade das entidades que comunicam entre si e da informação (origem, conteúdo, data de envio ...);
4. o **não repúdio**: assegurar que as entidades participantes não podem negar a autoria das suas ações ou compromissos.

As ferramentas criptográficas básicas usadas para conseguir os objetivos anteriores são chamadas de **primitivas criptográficas** (e.g. **sistemas de cifra**, **funções de Hash**).

Um **sistema de cifra** é constituído por:

- **alfabetos** $\mathcal{A}$ e $\mathcal{A}'$, em que estão escritos, respetivamente, o **texto limpo** m e o **texto cifrado** c;
- um conjunto $\mathcal{K}$ chamado **espaço de chaves**;
- um conjunto com as **transformações de cifragem**, $\{e_k\}_{k\in\mathcal{K}}$;
- um conjunto com as **transformações de decifragem**, $\{d_k\}_{k\in\mathcal{K}}$,

tais que, para cada chave $k \in \mathcal{K}$, as transformações e_k e d_k são inversas, isto é,

$$c = e_k(m) \text{ sse } m = d_k(c) \qquad \text{onde} \quad \begin{cases} m: & \text{texto limpo} \\ c: & \text{texto cifrado} \\ k: & \text{chave} \\ e_k: & \text{método para cifrar} \\ d_k: & \text{método para decifrar} \end{cases}$$

Que significa *chave* em criptografia?

- Formalmente, uma **chave** k de um sistema de cifra é um elemento do espaço de chaves $\mathcal{K}$ que determina quais as transformações de cifragem e de decifragem que devem ser usadas (uma espécie de "etiqueta").
- Na prática, usa-se o termo **chave** para designar **uma peça de informação necessária** para cifrar e/ou para decifrar.

$\rightsquigarrow$ No intuito de facilitar a compreensão dos conceitos de chaves, espaço de chaves e chaves simétricas e assimétricas, apresentamos de seguida as **cifras clássicas** (anteriores à época digital) mais relevantes.

Amostras rudimentares de encriptação em todas as grandes civilizações da antiguidade (egipcia, babilónica, chinesa ...)

Por exemplo, a cifra de **Atbash** foi usada aprox. 600 a.C., no Antigo Testamento para "ocultar" nomes de cidades:

$\rightsquigarrow$ A cifra Atbash consiste em sustituir a primeira letra do alfabeto pela última, a segunda pela penúltima ...

Os primeiros registos históricos conhecidos de uso consciente de um sistema de encriptação, com fins militares, referem o uso da *cítala* (do grego *skytale*, bastão), também chamada o *bastão de Licurgo*, no século V. a.C.

⇝ O sistema consiste em duas varas da mesma espessura, uma na posse do remetente e outra na posse do destinatário. A mensagem era escrita numa tira de tecido, após ter enrolado em espiral a tira na vara. Uma vez escrita a mensagem, a tira era desenrolada e enviada ao destinatário que a enrolava na vara semelhante para ler a mensagem original.

Consideremos uma cítala em que é possível escrever 5 caracteres em cada volta :

texto limpo: acitalaeumsistemadecifrausadonaantigagrecia

distribuição dos caracteres na cítala:

a c i t a l a e u
m s i s t e m a d
e c i f r a u s a
d o n a a n t i g
a g r e c i a

texto cifrado:
AMEDACSCOGIIINRTSFAEATRACLEANIAMUTAEASIUDAG

Cifrando com outra cítala espartana ...

Considerando uma cítala mais estreita, em que só seja possível escrever três caracteres a volta:

texto limpo: acitalaeumsistemadecifrausadonaantigagrecia

distribuição dos caracteres na cítala:

a c i t a l a e u m s i s t e
m a d e c i f r a u s a d o n
a a n t i g a g r e c i a

texto cifrado:
AMACAAIDNTETACILIGAFAERGUARMUESSCIAISDATOEN

- Consideramos como **alfabetos** do texto limpo e do texto cifrado respetivamente:

$$\mathcal{A} = \{a, b, c, \ldots, x, y, z\},\ \mathcal{A}' = \{A, B, \ldots, X, Y, Z\}$$

  $\rightsquigarrow$ estamos a usar a convenção estabelecida para a maioria dos sistemas de cifra clássicos, na cifra original devia ser o alfabeto grego ...

- A **chave** desde sistema de cifra é o bastão, mais precisamente, **a largura do bastão**, que determina os "saltos" na mensagem.
- Transformações de **cifragem/decifragem**: reordenar as letras que formam a mensagem seguindo os "saltos" definidos pela chave.

  $\rightsquigarrow$ A mensagem encriptada contém as mesmas letras que a mensagem original.

A cifra de César

O seguinte registo histórico de uso sistemático de cifra com fins militares deve-se ao emperador romano Júlio César.

O médodo consiste em sustituir cada letra do alfabeto por aquela situada três posições à frente:

texto limpo: a c i f r a d e c e s a r

texto cifrado: D F L I U D G H F H V D U

Cifras de translação ou deslocação

A Cifra de César é um exemplo de cifra de **translação** ou **deslocação**: o alfabeto é deslocado um número fixo de posições e cada letra é susbtituída pela letra correspondente à deslocação.

Exemplo 1 Translação de 6 posições:

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
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

texto limpo: e x e m p l o
texto cifrado: K D K S V R U

Exemplo 2 Translação de 24 posições

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
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

texto limpo: e x e m p l o
texto cifrado: C V C K N J M

O sistema de cifra de translação (ou deslocação)

- Os **alfabetos** do texto limpo e do texto cifrado, são, respetivamente:

$$\mathcal{A} = \{a, b, c, \ldots, x, y, z\},\ \mathcal{A}' = \{A, B, \ldots, X, Y, Z\}$$

- A **chave** é a translação/deslocação a realizar: $\{0, 1, \ldots, 25\}$. O **espaço de chaves** contem 26 elementos.
- As **transformações de cifragem/decifragem**, consistem em substituir cada letra pela letra correspondente na translação realizada.

$\rightsquigarrow$ A Cifra de César é um sub-sistema da cifra por translação, com os mesmos alfabetos do texto limpo e do texto cifrado, mas com uma **única chave**.

$\rightsquigarrow$ As cifras por translação podem ser quebradas facilmente: basta experimentar as 26 chaves possíveis (**brute-force attack**)

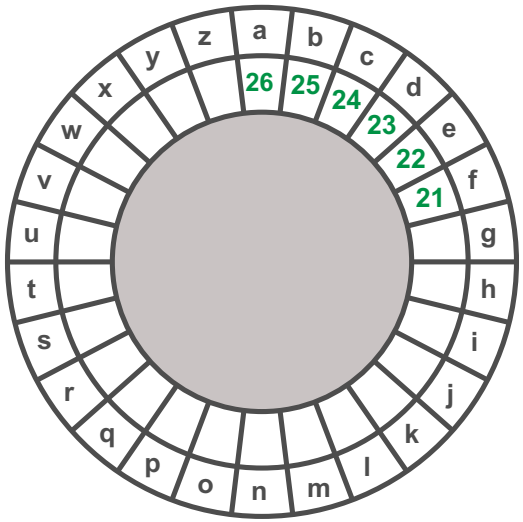
Os métodos de encriptação em que cada letra é substituída por outra (sempre pela mesma) são chamados **cifras de substituição mono-alfabéticas**.

Nas cifras de substituição mono-alfabéticas, a **chave** necessária para cifrar ou decifrar a mensagem é a correspondência bijetiva entre as letras e as suas substitutas.

- Existem $26! \sim 4,03 \times 10^{26}$ possibilidades de substituição de um alfabeto com 26 símbolos

  isto é,

- O **espaço de chaves** para a cifra de substituição mono-alfabética contém $\sim 4,03 \times 10^{26}$ chaves



- Os **alfabetos** do texto limpo e do texto cifrado, são, respetivamente:

$$\mathcal{A} = \{a, b, c, \ldots, x, y, z\},\ \mathcal{A}' = \{A, B, \ldots, X, Y, Z\}$$

- A **chave** é uma das 26! permutações possíveis dos elementos do alfabeto (e o **espaço de chaves** é o conjunto das 26! permutações possíveis);

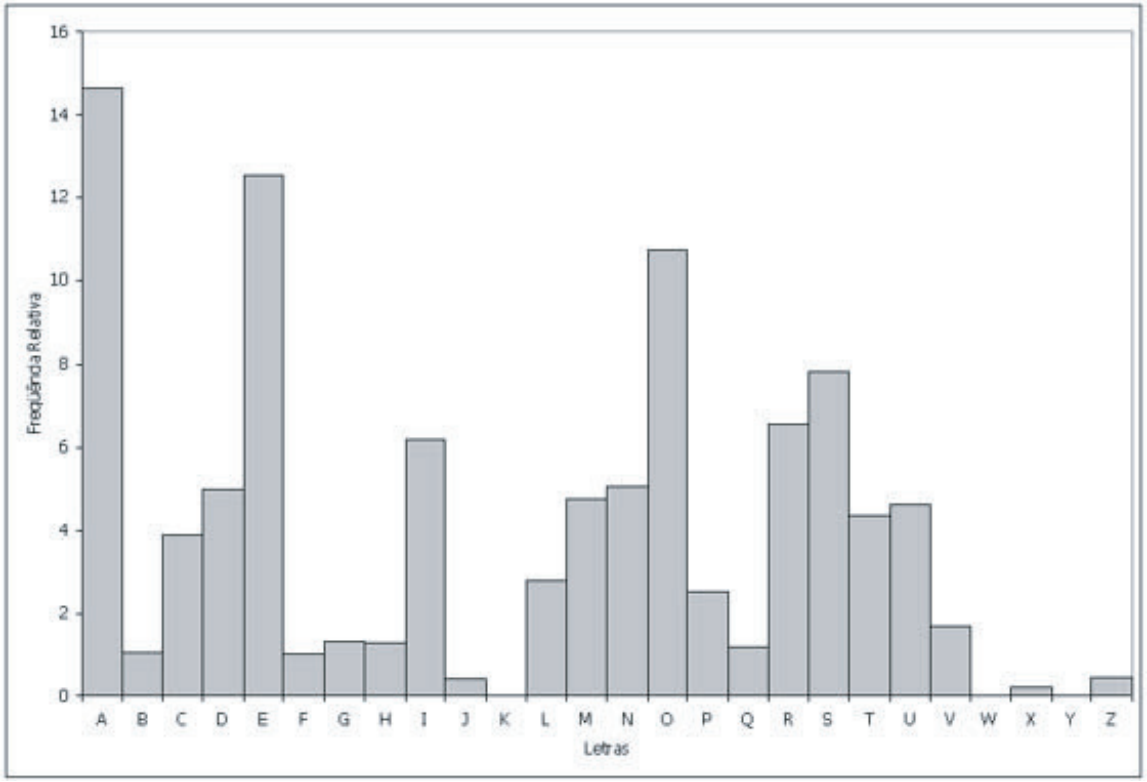
  $\rightsquigarrow$ A cifra por translação é por sua vez um sub-sistema da cifra de substituição mono-alfabética geral, considerando como **chaves** unicamente as 26 permutações definidas pelas translações do alfabeto.

- As **transformações de cifragem/decifragem**, consistem em substituir cada letra pela letra correspondente na permutação escolhida.

A pesar do elevadíssimo número de chaves, os sistemas de cifra mnono-alfabéticos são muito simples de cripto-analisar, porque os símbolos usados na substituição replicam a frequência de letras, de pares de letras .... da língua original.

$\rightsquigarrow$ A primeira descrição realizada da análise de frequências como método de cripto-análise de cifras de substituição mono-alfabética foi feita no século IX, no livro *"Manuscrito para desencriptar mensagens criptográficas"* de Abu Yusuf Al Kindi, um dos grandes sábios da Idade do Ouro do Islão.

Análise de frequências



(Frequência das letras em português)

Sequência: a.e.o.s.i.r.d.n.t.c.m.u.p.l.v.g.b.f.q.h.j.z.x.k.w.y

Exemplo

HVKLWVHLOSZIEVHVKLWVHEVIIVKZIZOREILWLHXLMHVOSLHLWRHXLZN
ZIVOLROFNRMLFHVWLRHWLHZFGLNLEVRHWZUIVMGVZXVOVIZIZNZMG
VHJFVLHRMZOEVINVOSLZKZIVXVHHVMZKZHHZWVRIZWVKVLVHHFITRF
LWVHVMSLWLSLNVNEVIWVZTVMGVJFVVHKVIZEZXLNVXLFZZGIZEVHHZI
ZIFZKRHZMWLZHUZRCZHYIZMXZHKRMGZWZHMZXZKZMVTIZWLZHUZOGL

"Z" - 41 ocorrências
"V" - 39 ocorrências
"H" - 30 ocorrências
"L" - 27 ocorrências

Letras duplas : II, HH, ZZ,

Sequência: Z.V.H.L.I.W.M.R.K.O.F.E.X.N.G.S.U.T.J.C.Y

⇝ Os caracteres A,B,D,P,Q não aparecem

Exemplo

Caracteres ordenados por frequências

Ciphertext:	Z.V.H.L.I.W.M.R.K.O.F.E.X.N.G.S.U.T.J.C.Y
Português:	a.e.o.s.i.r.d.n.t.c.m.u.p.l.v.g.b.f.q.h.j.z.x.k.w.y
texto limpo	a.e.s.o.r.d.n.i.p.l.u.v.c.m.t.h.f.g.q.x.b

Candidatos a letras duplas: I, H, Z

Com um pouco de paciência (ou computadores ...)

sepodesolharvesepodesverreparalivrodosconselhosodiscoamareloiluminouse
doisdosautomoveisdafrenteaceleraramantesqueosinalvermelhoaparecesse
napassadeiradepeoessurgiuodesenhodohomemverdeagentequeesperavacomecou
aatravessararuapisandoasfaixasbrancaspintadasnacapanegradoasfalto

O Disco de Alberti

Uma **cifra de substituição poli-alfabética** é um sistema de encriptação que usa mais do que um alfabeto, alternando entre eles durante o processo de cifra, para dificultar o cripto-análise por frequências de letras ou grupos de letras.

O **primeiro sistema de cifra poli-alfabética** foi inventado pelo sábio renacentista Leon Battista Alberti em 1467 e usa dois discos articulados.

A cifra descrita por Alberti funciona do modo seguinte:

1. fixar previamente, entre o remetente e o destinatário, uma letra do disco pequeno como "letra-chave" de cifra;
2. o primeiro caracter no texto cifrado, em maiúsculas, indica a posição inicial da letra-chave de cifra;
3. cada letra maiúscula no texto cifrado indica uma alteração na posição dos discos.

Por exemplo, considerando k como letra-chave:

chave:	k:C					k:A					k:M				k:R						k:B	
texto limpo:		O	D	I	S		C	O	D		E	A	L		B	E	R	T		I		
texto cifrado:	C	x	l	t	m	A	n	s	p	M	b	i	g	R	s	q	k	n	B	v		

O sistema da cifra de Alberti

- Os **alfabetos** do texto limpo e do texto cifrado, são, respetivamente:

$$\mathcal{A} = \{A, B, C, D, E, F, G, I, L, M, N, O, P, Q, R, S, T, V, X, Z, 1, 2, 3, 4\},$$

$$\mathcal{A}' = \{a, b, c, d, e, f, g, h, i, k, l, m, n, o, p, q, r, s, t, v, y, x, z, \&\}$$

- A **chave**: cada uma das 24 posições iniciais possíveis do disco (identificadas por uma letra-chave do alfabeto $\mathcal{A}'$);
- As **transformações de cifragem/decifragem**, o processo indicado anteriormente, iniciado na letra chave.

A cifra de Alberti é resistente à análise de frequências, mas é um método rígido cuja segurança depende do desconhecimento do cripto-analista do método usado.

Este pressuposto não é aceitável na criptografia moderna, que considera como um dos princípios *"sagrados"* a lei enunciada pelo criptógrafo holandês August Kerkchoffs no século XIX :

Princípio de Kerkchoffs

Para avaliar a segurança de uma técnica criptográfica, devemos assumir que esta é do conhecimento de eventuais inimigos.

$\rightsquigarrow$ O disco de Alberti pode ser usado de inúmeros modos para ir trocando alfabetos, para além do descrito no *"De cifris"* mas a versatilidade que oferece o mecanismo para ir alternando entre diferentes alfabetos não foi explorada em profundidade até séculos depois (por exemplo, nas famosas máquinas Enigma).

A *Tábula Recta* de J. Trithemius

	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
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

A cifra proposta por G.B. Bellaso no século XVI consiste em mudar entre os alfabetos de susbtituição da *Tabula Recta* em função de uma palavra **chave**.

Por exemplo, dada a palavra chave "BELLASO", e o texto limpo "acifradebellaso", obtemos:

Chave :	B	E	L	L	A	S	O	B	E	L	L	A	S	O	B
Texto limpo :	a	c	i	f	r	a	d	e	b	e	l	l	a	s	o
Texto cifrado :	B	G	T	Q	R	S	R	R	F	F	W	L	S	G	P

Um adversário pode saber que se está a usar a Cifra de Bellaso-Vigenère mas, sem a palavra chave, a cripto-análise é complicada.

$\rightsquigarrow$ O primeiro cripto-análise desta cifra foi realizado por volta de 1854 por Charles Babbage, que só divulgou os seus resultados em 1863, após o militar e criptógrafo Friedrich Kasiski publicar um livro sobre criptografia em que se explicava cómo atacar cifras polialfabéticas.

O método de Kasiski-Babbage consiste em estimar, primeiro, o comprimento possível da palavra-chave e aplicar, depois, uma análise de frequência a cada sub-alfabeto para encontrar a cifra correspondente.

A estimativa do possível comprimento da palavra chave faz-se procurando no texto grupos de 3 ou 4 letras repetidos, analisando-se os intervalos cada repetição.

A ideia subjacente é que há grupos de 3 ou 4 letras muito frequentes, cuja cifra calha sempre nas mesmas posições da palavra chave. Se for o caso, os intervalos entre cada repetição serão, em muitos casos, múltiplos do comprimento da palavra chave.

Exemplo: o método de Kasiski-Babbage

Considere-se, por exemplo, o texto seguinte, cifrado com uma cifra de Bellaso-Vigenère com palavra-chave desconhecida:

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

Observamos que existem grupos de 4 letras repetidos, por exemplo, o grupo "TXLP" aparece nas posições 32 e 116, o grupo "PVOQ", aparece nas posições 213, 225, 387 e 495 ...

Exemplo: o método de Kasiski-Babbage

Realizando a análise completa dos grupos de 4 letras repetidos, no texto cifrado, observamos as seguintes ocorrências:

Grupos	**Posições**	**Intervalos**
TXLP	[32, 116]	84
XLPQ	[33, 117]	84
LPQR	[34, 118]	84
RLDW	[139, 505]	366
RKPH	[187, 451]	264
ZOPV	[211, 493]	282
OPVO	[212, 494]	282
PVOQ	[213, 225, 387, 495]	12,162,108
MHXF	[230, 404]	174
HXFX	[231, 405]	174
XFXZ	[232, 406]	174
FXZK	[233, 407]	174
XZKD	[234, 408]	174
ZKDL	[235, 409]	174
UTGH	[260, 374]	114
UMXG	[349, 445]	96
ICMS	[416, 488]	72

Exemplo: o método de Kasiski-Babbage

Os intervalos mais frequentes entre grupos de 4 letras repetidos são 174 e 84 pelo que é provável, que o comprimento da chave seja um divisor de ambos os números. Como o seu máximo comum divisor é 6, vamos supor que o comprimento da palavra chave é 6.

Se o comprimento da palavra chave for efetivamente 6, os caracteres que aparecem no texto nas posições $0, 6, 12, ..., 6n, ...$ foram cifrados usando o mesmo alfabeto de translação. Os caracteres que aparecem no texto nas posições $1, 7, 13, ..., 6n + 1, ...$ foram cifrados também o mesmo alfabeto (eventualmente diferente do anterior) ...

Reagrupamos então os caracteres do texto em seis sub-alfabetos e realizamos a análise de frequências de cada um deles.

Exemplo: o método de Kasiski-Babbage

A lista de caracteres nas posições $0, 6, 12, ..., 6n, ...$ é:

QVRFPFQGUEPWTCGCEKNGQWOUCOKPKQQOTKCVQGQXEF
VCGTKWKEUTTGQOQXUGEUXQRQPKXCUCPCTXKKKCWUNQ

Tratando-se de um texto breve, poderá não haver diferenças significativas que permitam, por exemplo, distinguir qual a letra que substiui o "A" ou "E". Mas ordenando os caracteres anteriores do mais frequente ao menos frequente obtemos:

Q.K.C.G.U.T.P.E.X.W.O.V.F.R.N

e observamos que os caracteres A.B.C.H.I.J.L.M.S.Y.Z não aparecem na lista.

Exemplo: o método de Kasiski-Babbage

No caso de um texto em português, a sequência das letras ordenadas da mais frequente à menos é (sequência AEOSRI):

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

Assim, por exemplo, um texto cifrado com uma cifra de deslocação "B", terá aproximadamente a sequência correspondente:

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

e um texto cifrado com uma cifra de deslocação "Z" será:

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

Provavelmente, a ordem da sequência obtida anteriormente ao analizar o texto cifrado não vai coincidir exatamente com nenhuma das sequências associadas aos alfabetos de deslocação.

Exemplo: o método de Kasiski-Babbage

No entanto, podemos estudar, por exemplo, quais são as 6 letras mais frequentes e quais são as 6 letras menos frequentes (ou que não aparecem) em cada sequência, e observamos que então

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

apresenta um máximo de coincidencias (10) com a sequência da letra "C":

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

pelo que é possível que o primeiro sub-alfabeto seja o alfabeto de deslocação "C".

Repetindo o processo anterior para cada sub-alfabeto, obtemos "CRIPTO" como candidata a palavra chave da cifra de Bellaso Vigenère.

Exemplo: o método de Kasiski-Babbage

Decifrando então o texto cifrado usando como palavra chave "CRIPTO", obtemos o texto limpo:

```
o i n s t i t u t o s u p e r i o r d e e n g e n h a r i a d e l i s b o a i s e l e u m a i n s t
i t u i c a o d e e n s i n o s u p e r i o r n a a r e a d a e n g e n h a r i a e d a t e c n o l
o g i a l o c a l i z a d a e m l i s b o a p o r t u g a l c o m u m p a s s a d o d e a n o s c o
m a c t i v i d a d e s n o s d o m i n i o s d o e n s i n o d a f o r m a c a o p r o f i s s i o
n a l d a i n v e s t i g a c a o e d a p r e s t a c a o d e s e r v i c o s a c o m u n i d a d e
e s t a a c t u a l m e n t e i n t e g r a d o n o i n s t i t u t o p o l i t e c n i c o d e l i
s b o a a g r e g a v a r i a s a r e a s d o c o n h e c i m e n t o c o m a c t i v i d a d e s d
e i n v e s t i g a c a o e d e s e n v o l v i m e n t o e m c o o p e r a c a o c o m a i n d u s
t r i a e s e r v i c o s g a r a n t e s i m u l t a n e a m e n t e u m c a r a c t e r d e i n o
v a c a o e i n t e r d i s c i p l i n a r i d a d e c o m u m a c o n s t a n t e l i g a c a o a
```

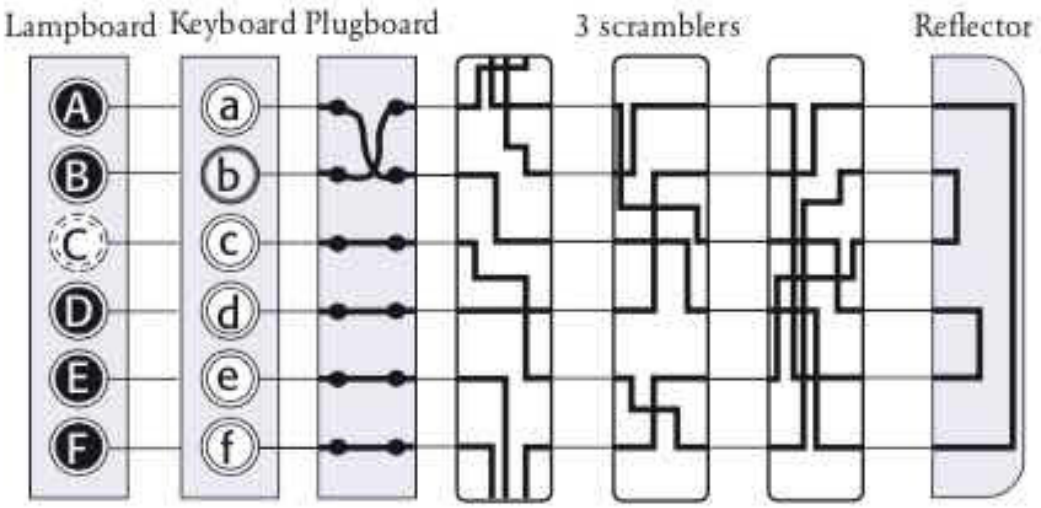
⇝ A comparação das letras mais/menos frequentes em textos pequenos pode não conseguir distinguir entre diferentes possibilidades de alfabetos de deslocação

⇝ Se a estimativa do comprimento da palavra chave não está correta, o processo anterior não funciona.

Mecanização dos sistemas de cifra

No século XX iniciou-se a mecanização dos métodos de encriptação, com a invenção de máquinas que permitiam combinar diferentes sistemas (transposição, cifras poli-alfabéticos), como a a máquina Enigma:

Inventada em 1918 pelo engenheiro alemão Arhur Serbius, combinava um dispositivo que permitia realizar transposições de pares de letras (mecanismo exterior, com cabos e tomadas) com um sistema interno que continha três *rotores* (*scramblers*) intercambiáveis (em versões posteriores 5 rotores) e um *reflector*.



$\rightsquigarrow$ Um *rotor* de uma máquina Enigma é simplesmente uma versão eléctrica de um Disco de Alberti.

Nas primeiras versões da máquina Enigma:

- o sistema de rotores voltava à posição inicial após $26 \times 25 \times 26 = 16900$ alfabetos (periodo da cifra poli-alfabética);

  $\rightsquigarrow$ Periodo da cifra poli-alfabética muito grande o que dificultava os métodos de Babbage-Kasiski para descobrir periodo e realizar análise de frequências.

- as máquinas admitiam $26 \times 26 \times 26 \times 6 = 105456$ posições iniciais para os rotores (os três rotores eram intercambiáveis) e tinham ainda um sistema de transposição de 6 pares de letras que adicionava 165765600 posições iniciais (aproximadamente 10^9 chaves possíveis);

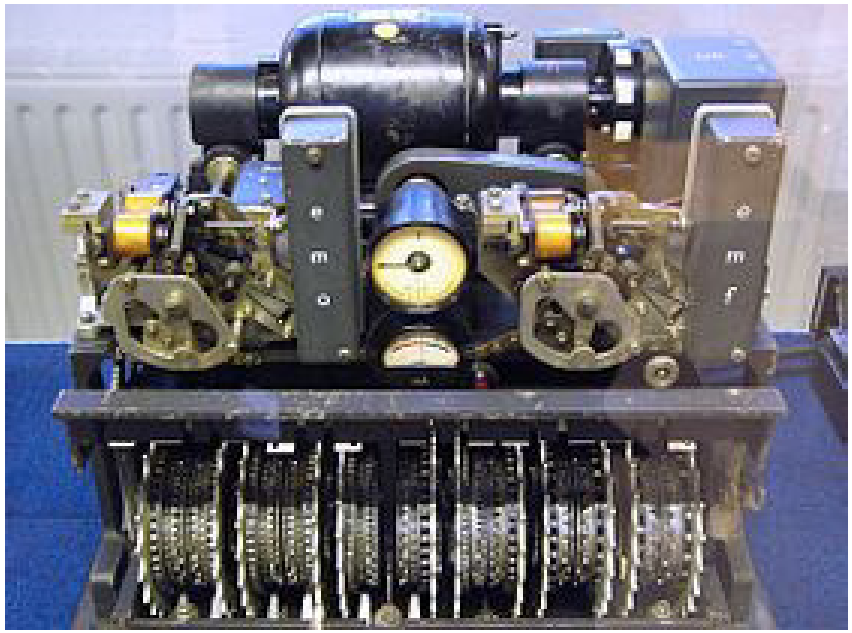
  $\rightsquigarrow$ Espaço de chaves grande que dificultava os ataques **por força bruta**.

Cripto-análise da cifra de Enigma

- A primeira cripto-análise da cifra de Enigma (3 rotores) foi realiza nos início da década de 1930 por três matemáticos dos serviços secretos polacos (o *Biuro Szyfrow*): Rejewski, Rózycki e Zygalski.
- O método de Rejewski, Rózycki e Zygalski aproveitava uma redundância do protocolo de transmissão de mensagens implementado pelos nazis (cifravam duas vezes seguidas, no início, uma mesma sequência de três caracteres).
- Em 1938, pouco antes da invasão nazi da Polónia, o Biuro Szyfrow partilhou os seus avanços com os serviços secretos franceses e britânicos.
- Durante a WWII, em Bletchey Park, uma equipa liderada por Alan Turing, implementou métodos novos para cripto-analisar máquinas Enigma mais sofisticadas (5 rotores).

A máquina de Lorenz

Usada nas comunicações de Hitler com a cúpula militar, o sistema também foi cripto-analisado pelos serviços secretos britânicos.

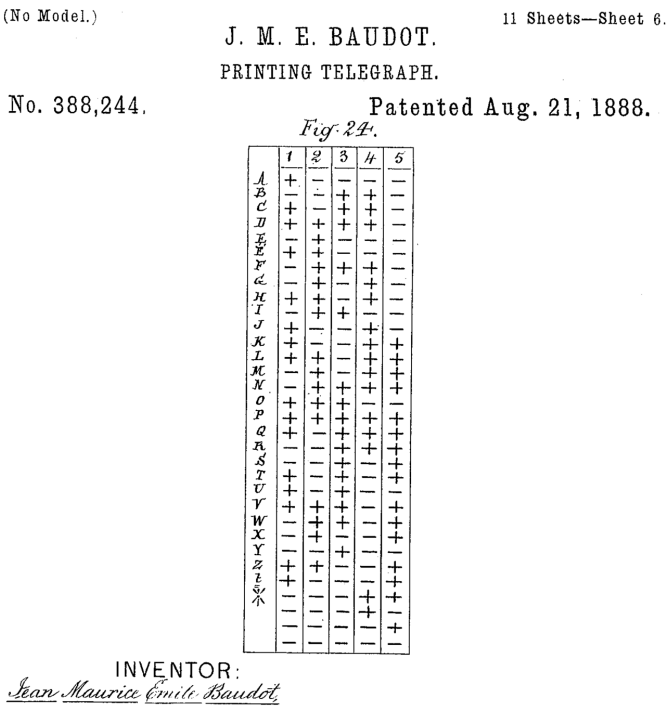


$\rightsquigarrow$ A cifra das máquinas de Lorenz era extraordinariamente segura e o seu cripto-análise foi possível por dois motivos: (1) um erro humano na na transmissão de uma mensagem, que forneceu informação sobre a chave das máquinas de Lorenz; (2) a construção do Colossus, o primeiro computador programável da história, desenhado por Max Newman, em 1943 (embora o mundo só soube da existência do Colossus nos anos 70).

A cifra de Vernam

As máquinas de Lorenz cifravam mensagens usando a cifra inventada em 1917 pelo engenheiro americano Gilbert Vernam para as transmissões por telégrafo.

A cifra de Vernam é um sistema de cifra baseado numa versão do código Morse chamada alfabeto Baudot, **diretamente adaptável às comunições digitais**.



(No Model.) 11 Sheets—Sheet 6.

J. M. E. BAUDOT.

PRINTING TELEGRAPH.

No. 388,244. Patented Aug. 21, 1888.

Fig. 24.

	1	2	3	4	5
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	+	−	−	−	+
↕	−	−	−	+	−
↔	+	+	+	−	−

INVENTOR:

Jean Maurice Emile Baudot

No alfabeto de Baudot, cada caracter é representado usando 5 *bits*, por exemplo:

A	10000
B	00110
⋮	
P	11111
⋮	
Z	11001

A cifra de Vernam usa uma **sequência de chaves (keystream)**, isto é, uma sequência de 0 e 1 (originalmente uma sequência - e + no telégrafo) que se combina com a sequência do texto limpo usando o XOR *bit* a *bit* (*XORwise*):

XOR	0	1
0	0	1
1	1	0

Por exemplo:

texto limpo:	e	x	e	m	p	l	o
sequência limpa:	01000	01001	01000	01011	11111	11011	11100
keystream:	00100	00011	11000	10100	11001	10010	00000
sequência cifrada:	01100	01010	10000	11111	00110	01001	11100
texto cifrado:	l	G	A	P	B	X	O

Para decifrar a mensagem, basta realizar a operação XOR da mensagem cifrada com a mesma *keystream*:

texto cifrado:	I	G	A	P	B	X	O
sequência cifrada:	01100	01010	10000	11111	00110	01001	11100
keystream:	00100	00011	11000	10100	11001	10010	00000
sequência limpa:	01000	01001	01000	01011	11111	11011	11100
texto limpo:	e	x	e	m	p	l	o

A cifra ONE-TIME-PAD (OTP)

A cifra de Vernam é chamada **ONE-TIME-PAD (OTP)** quando:

1. a *keystream* é escolhida aleatoriamente;
2. a *keystream* e do mesmo comprimento que a mensagem a enviar;
3. a *keystream* é utilizada uma única vez.

C. Shannon provou em 1949 que a OTP é um cifra **inquebrável** ou, formalmente, **incondicionalmente segura**.

Uma cifra é dita **incondicionalmente segura** quando não pode ser cripto-analisada por um adversário com acesso a recursos (computacionais) ilimitados.

$\rightsquigarrow$ É fácil convencer-se da segurança incondicional de OTP: dada uma sequência de texto limpo de *n-bits*, se a *keystream* é aleatória e com comprimento *n-bits*, podemos obter QUALQUER sequência de *n-bits* como texto cifrado.

Exemplo 1

texto limpo:	e	x	e	m	p	l	o
sequência limpa:	01000	01001	01000	01011	11111	11011	11100
keystream 1:	00100	00011	11000	10100	11001	10010	00000
sequência cifrada:	01100	01010	10000	11111	00110	01001	11100
texto cifrado:	l	G	A	P	B	X	O

Exemplo 2

texto limpo:	e	x	e	m	p	l	o
sequência limpa:	01000	01001	01000	01011	11111	11011	11100
keystream 2:	00111	11101	00011	00011	11000	00111	11001
sequência cifrada:	01111	10100	01011	01000	00111	11100	00101
texto cifrado:	N	U	M	E	R	O	S

O problema da distribuição das chaves

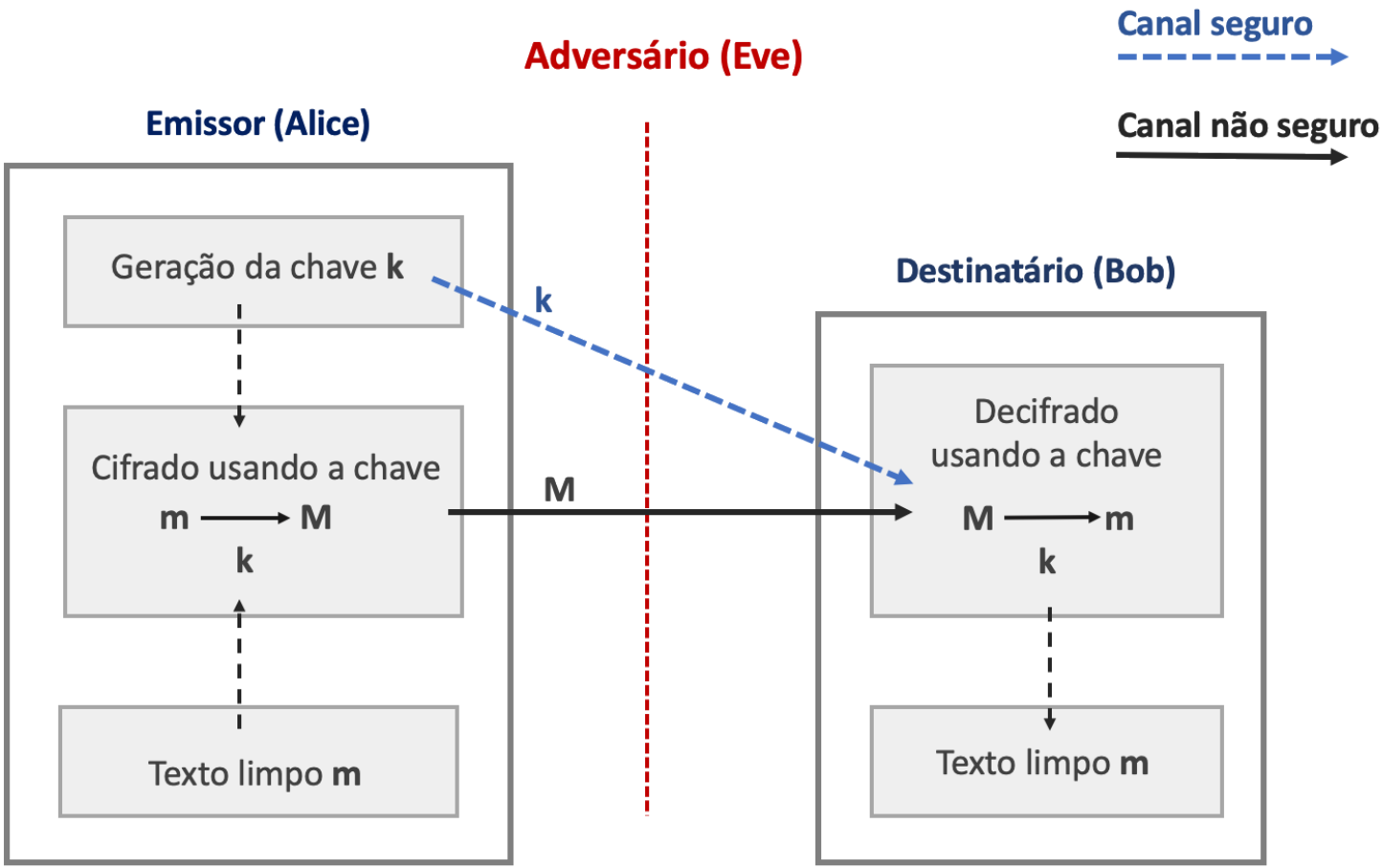
No caso de todas as cifras históricas a informação necessária para realizar o cifra, a **chave para cifrar**, é a mesma que é necessária para decifrar: são cifras denominadas atualmente **simétricas**.

As cifas simétricas, como a cifra de Vernam podem ser satisfatoriamente seguras desde que:

- a chave para cifrar seja escolhida aleatoriamente;
- a chave para cifrar seja alterada frequentemente.

No entanto, como é preciso manter a **chave para cifrar secreta**, cada vez que se realize uma **troca de chave**, a informação sobre a nova chave deve ser veiculada entre Alice e Bob através de um **canal seguro**, isto é, um canal não acessível ao adversário.

Distribuição das chaves em cifras simétricas

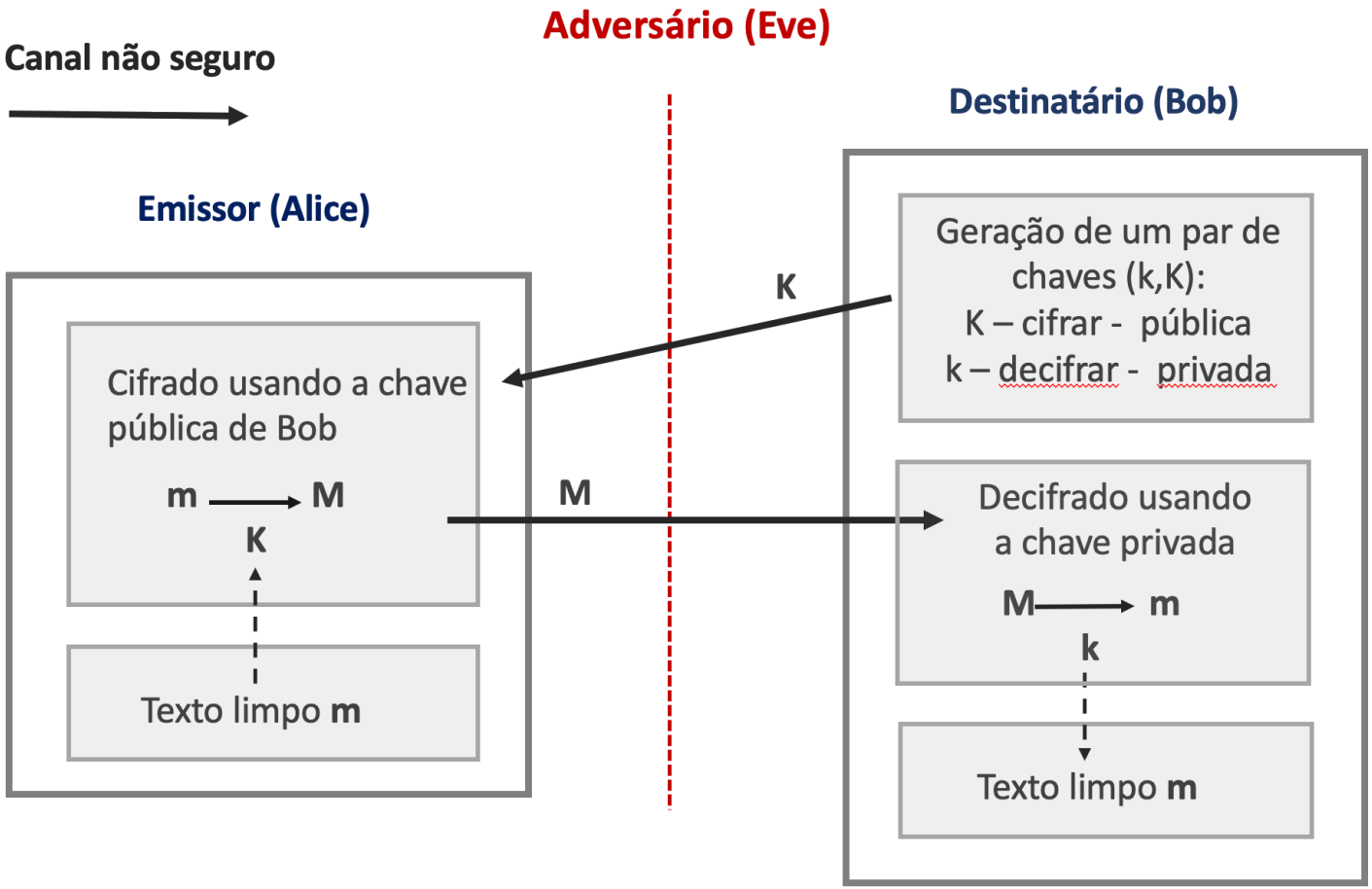


O problema de distribuição da chave foi considerado insolúvel até que Whitfield Diffie, Martin Hellmann e Merkle idealizaram, em 1975, um novo sistema de encriptação chamado de **chave assimétrica** ou **chave pública**.

A ideia consistia em encontrar um sistema de encriptado baseado em funções matemáticas de tipo *one way functions*, que são funções fáceis de calcular cujas inversas são impossíveis (ou pelo menos computacionalmente impossíveis) de calcular, a menos que se tenha algum tipo de informação extra (uma *trap-door*).

Diffie, Hellmann e Merkle imaginaram que este tipo de funções *one-way* podiam ser usadas para construir cifras com a propriedade que o conhecimento da chave para cifrar não permitisse obter a chave para decifrar.

Cifras assimétricas (chaves públicas)



Os primeiros sistemas de cifra de chave pública

Em 1977, Ron Rivest, Leonard Adleman e Adir Shamir publicaram pela primeira vez uma *one-way funcion*, com *trap-door*, que permitia implementar a ideia de Diffie, Hellman e Merkle, usando aritmética modular (o algoritmo RSA).

De referir que existe uma história paralela, ligeiramente anterior, no desenvolvimento da criptografia de chave pública, que decorreu no secretismo dos serviços de informação britânicos e só foi revelada em 1992.

Por volta de 1969, James Ellis, um especialista em criptografia do GHCQ (Government Communications Headquarters), esboçou o conceito de funções one-way que, posteriormente, em 1974, dois especialistas em teoria de números do GHCQ, Clifford Cocks e Malcom Williamson, usaram para obter sistemas baseados na fatorização de números primos (RSA) e no logaritmo discreto (Protocolo Diffie-Hellmann).

Duas abordagens significativamente diferentes da segurança:

- assumir que o adversário possui acesso a recursos computacionais ilimitados (análise da **segurança incondicional**);
- assumir que tem recursos computacionais limitados em termos, por exemplo, de memória ou tempo de cálculo (análise da **segurança computacional**).

A segurança computacional, para além dos recursos computacionais, depende também dos algoritmos matemáticos conhecidos em cada momento, que podem mudar drasticamente o nível de segurança de uma cifra.

$\rightsquigarrow$ Um exemplo ilustrativo em cifras clássicas: antes do método de "análise de frequências" de Al-Kindi, considerava-se que uma cifra mono-alfabética só podia ser quebrada usando um ataque por força-bruta, i.e., testando os possíveis 26! alfabetos de substituição.

A capacidade de um adversário de obter informação sobre texto limpo a partir de texto cifrado depende também do tipo de ataque que o adversário consegue efetuar:

1. ataque unicamente com texto cifrado (*ciphertext only attack*);

   o adversário tenta obter a chave de decifrado e/ou o texto limpo a partir unicamente do texto cifrado;

2. ataque com texto limpo conhecido (*known–plaintext attack*);

   o adversário possui alguns textos limpos e os seus correspondentes textos cifrados;

3. ataque com texto limpo escolhido (*chosen-plaintext attack*);

   o adversário pode escolher qual o texto limpo cujo texto cifrado consegue obter;

4. ataque com texto cifrado escolhido (*chosen-ciphertext attack*);

   o adversário pode escolher qual o texto cifrado cujo texto limpo consegue obter.

As cifras mono-alfabéticas mostram como um espaço de chaves enorme **não é suficiente** para obter uma cifra segura.

Contudo, um espaço de chaves grande é uma condição **necessária** para uma cifra ser computacionalmente segura. Efetivamente, se o espaço de chaves é "pequeno", o adversário pode sempre montar um ataque por força-bruta (*brute-force attack*), que consiste em experimentar todas as chaves possíveis.

Atualmente, os sistemas de cifra considerados seguros, como AES (*Advanced Encryption Standard*), trabalham com espaços de chaves de tamanho 2^{128} ou 2^{256} (cada chave é uma sequencia de 128 ou 256 *bits*).