Capítulo 2

Criptografia simétrica

Este capítulo apresenta uma introdução às técnicas clássicas de criptografia frequentemente usadas em sistemas operacionais e redes de computadores. Este texto não tem a mínima pretensão de ser completo sobre esse vasto tema; leitores em busca de uma abordagem mais profunda e completa devem procurar livros específicos sobre criptografia.

2.1 Terminologia

O termo "criptografia" provém das palavras gregas *kryptos* (oculto, secreto) e *graphos* (escrever). Assim, a criptografia foi criada para codificar informações, de forma que somente as pessoas autorizadas pudessem ter acesso ao seu conteúdo. Conceitualmente, a criptografia faz parte de um escopo mais amplo de conhecimento:

Criptografia: técnicas para codificar/decodificar informações, ocultando seu conteúdo de pessoas não autorizadas;

Criptanálise: conjunto de técnicas usadas para "quebrar" uma criptografia, expondo a informação ocultada por ela;

Criptologia: área geral, englobando criptografia e criptanálise.

Criptossistema: conjunto de algoritmos/mecanismos para realizar um tipo específico de criptografia.

As técnicas criptográficas são extensivamente usadas na segurança de sistemas, para garantir a confidencialidade e integridade dos dados. Além disso, elas desempenham um papel importante na autenticação de usuários e recursos.

Alguns conceitos fundamentais para estudar as técnicas criptográficas são [Menezes et al., 1996]:

Texto aberto: a mensagem ou informação a codificar (x);

Texto cifrado: a informação codificada de forma a ocultar seu conteúdo (x');

Chave: informação complementar, necessária para cifrar ou decifrar as informações (k);

Cifrar: transformar o texto aberto em texto cifrado ($x \xrightarrow{k} x'$) ;

Decifrar: transformar o texto cifrado em texto aberto ($x' \xrightarrow{k} x$) ;

Cifrador: mecanismo responsável por cifrar/decifrar as informações;

No restante deste texto, a operação de cifragem de um conteúdo aberto x usando uma chave k e gerando um conteúdo cifrado x' será representada por $x' = \{x\}_k$ e a decifragem de um conteúdo x' usando uma chave k será representada por $x = \{x'\}_k^{-1}$.

2.2 Cifradores, chaves e espaço de chaves

Uma das mais antigas técnicas criptográficas conhecidas é o *cifrador de César*, usado pelo imperador romano Júlio César para se comunicar com seus generais. O algoritmo usado nesse cifrador é bem simples: cada caractere do texto aberto é substituído pelo k-ésimo caractere sucessivo no alfabeto. Assim, considerando $k = 2$, a letra "A" seria substituída pela letra "C", a letra "R" pela "T", e assim por diante.

Usando esse algoritmo, a mensagem secreta "Reunir todos os generais para o ataque" seria cifrada da seguinte forma:

mensagem aberta:	`REUNIR TODOS OS GENERAIS PARA O ATAQUE`
mensagem cifrada com $k = 1$:	`SFVOJS UPEPT PT HFOFSBJT QBSB P BUBRVF`
mensagem cifrada com $k = 2$:	`TGWPKT VQFQU QU IGPGTCKU RCTC Q CVCSWG`
mensagem cifrada com $k = 3$:	`UHXQLU WRGRV RV JHQHUDLV SDUD R DWDTXH`

Para decifrar uma mensagem no cifrador de César, é necessário conhecer a mensagem cifrada e o valor de k utilizado para cifrar a mensagem, que é a *chave criptográfica*. Caso essa chave não seja conhecida, ainda é possível tentar "quebrar" a mensagem cifrada testando todas as chaves possíveis, o que é conhecido como análise exaustiva ou "ataque de força bruta". Considerando o cifrador de César e somente letras maiúsculas, a análise exaustiva é trivial, pois há somente 26 valores possíveis para a chave k (as 26 letras do alfabeto).

O número de chaves possíveis em um algoritmo de cifragem é conhecido como o seu **espaço de chaves** (*keyspace*). Em 1883, muito antes dos computadores eletrônicos, o criptólogo francês Auguste Kerckhoffs enunciou um princípio segundo o qual "*o segredo de uma técnica criptográfica não deve residir no algoritmo em si, mas no espaço de chaves que ela provê*". Obedecendo esse princípio, a criptografia moderna se baseia em algoritmos públicos, extensivamente avaliados pela comunidade científica, para os quais o espaço de chaves é extremamente grande, tornando inviável qualquer análise exaustiva, mesmo por computador.

Um bom exemplo de aplicação do princípio de Kerckhoffs é dado pelo algoritmo de criptografia AES (*Advanced Encryption Standard*), adotado como padrão pelo governo americano. Usando chaves de 128 bits, esse algoritmo oferece um espaço de chaves com 2^{128} possibilidades, ou seja, 340.282.366.920.938.463.463.374.607.431.768.211.456 chaves diferentes... Se pudéssemos testar um bilhão (10^9) de chaves por segundo, ainda assim seriam necessários 10 sextilhões de anos para testar todas as chaves possíveis!

2.3 O cifrador de Vernam-Mauborgne

O cifrador de Vernam-Mauborgne foi proposto em 1917 por Gilbert Vernam, engenheiro da ATT, e melhorado mais tarde por Joseph Mauborgne. Neste criptossistema, a cifragem de um texto aberto x com b bits de comprimento consiste em realizar uma operação XOR (OU-exclusivo, $\oplus$) entre os bits do texto aberto e os bits correspondentes de uma chave k de mesmo tamanho:

$$x' = x \oplus k, \text{ ou seja, } \forall i \in [1...b], x'_i = x_i \oplus k_i$$

Como a operação XOR é uma involução (pois $(x \oplus y) \oplus y = x$), a operação de decifragem consiste simplesmente em reaplicar essa mesma operação sobre o texto cifrado, usando a mesma chave:

$$x = x' \oplus k, \text{ ou seja, } \forall i \in [1...b], x_i = x'_i \oplus k_i$$

O exemplo a seguir ilustra este cifrador, usando caracteres ASCII. Nele, a mensagem x ("TOMATE") é cifrada usando a chave k ("ABCDEF"):

x (texto)	T	O	M	A	T	E
k (chave)	A	B	C	D	E	F
x (ASCII)	84	79	77	65	84	69
k (ASCII)	65	66	67	68	69	70
x (binário)	01010100	01001111	01001101	01000001	01010100	01000101
k (binário)	01000001	01000010	01000011	01000100	01000101	01000110
$x' = x \oplus k$	00010101	00001101	00001110	00000101	00010001	00000011
x' (ASCII)	21	13	14	5	17	3

Para decifrar a mensagem x', basta repetir a operação $\oplus$ com a mesma chave:

x'	00010101	00001101	00001110	00000101	00010001	00000011
k	01000001	01000010	01000011	01000100	01000101	01000110
$x' \oplus k$	01010100	01001111	01001101	01000001	01010100	01000101
(ASCII)	84	79	77	65	84	69
(texto)	T	O	M	A	T	E

A Figura 2.1 ilustra a aplicação do cifrador de Vernam-Mauborgne sobre uma imagem. A chave é uma imagem aleatória com as mesmas dimensões da imagem a cifrar; a operação de OU-exclusivo deve ser aplicada entre os pixels correspondentes na imagem e na chave.

Apesar de extremamente simples, o cifrador de Vernam-Mauborgne é considerado um criptossistema inquebrável, sendo comprovadamente seguro se a chave utilizada for realmente aleatória. Entretanto, ele é pouco usado na prática, porque exige uma chave k do mesmo tamanho do texto a cifrar, o que é pouco prático no caso de mensagens longas. Além disso, essa chave deve ser mantida secreta e deve ser usada uma única vez (ou seja, um atacante não deve ser capaz de capturar dois textos distintos

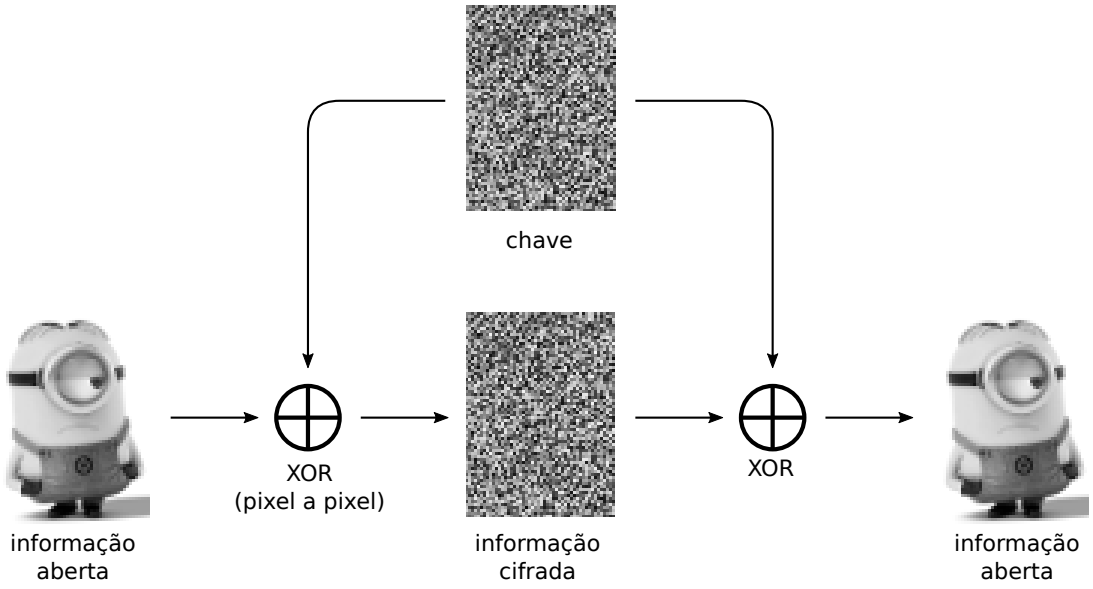

Figura 2.1: Aplicação do cifrador de Vernam sobre uma imagem.

cifrados com a mesma chave). O requisito de uso único da chave levou este cifrador a ser chamado também de *One-Time Pad* [1].

[1] O cifrador conhecido como *One-Time Pad* consiste em uma melhoria do algoritmo original de G. Vernam, proposto por J. Mauborgne pouco tempo depois.

2.4 Criptografia simétrica

De acordo com o tipo de chave utilizada, os algoritmos de criptografia se dividem em dois grandes grupos: *algoritmos simétricos* e *algoritmos assimétricos*. Nos **algoritmos simétricos**, a mesma chave k é usada para cifrar e decifrar a informação. Em outras palavras, se usarmos uma chave k para cifrar um texto, teremos de usar a mesma chave k para decifrá-lo. Essa propriedade pode ser expressa em termos matemáticos:

$$\{ \{ x \}_k \}_{k'}^{-1} = x \iff k' = k$$

Os cifradores de César e de Vernam são exemplos típicos de cifradores simétricos simples. Outros exemplos de cifradores simétricos bem conhecidos são:

- DES (*Data Encryption Standard*): criado pela IBM nos anos 1970, foi usado amplamente até o final do século XX. O algoritmo original usa chaves de 56 bits, o que gera um espaço de chaves insuficiente para a capacidade computacional atual. A variante 3DES (*Triple*-DES) usa chaves de 168 bits e é considerada segura, sendo ainda muito usada.
- AES (*Advanced Encryption Standard*): algoritmo simétrico adotado como padrão de segurança pelo governo americano em 2002. Ele pode usar chaves de 128, 192 ou 256 bits, sendo considerado muito seguro. É amplamente utilizado na Internet e em programas de cifragem de arquivos em disco.
- A5/1, A5/2, A5/3: algoritmos de criptografia simétrica usados em telefonia celular GSM, para cifrar as transmissões de voz.

A Figura 2.2 ilustra o esquema básico de funcionamento de um sistema de criptografia simétrica para a troca segura de informações. Nesse esquema, a chave simétrica deve ter sido previamente compartilhada entre quem envia e quem recebe a informação.

Os criptossistemas simétricos são muito rápidos e bastante eficientes para a cifragem de grandes volumes de dados, como arquivos em um disco rígido ou o tráfego em uma conexão de rede. Entretanto, se a informação cifrada tiver de ser enviada a outro usuário, a chave criptográfica secreta usada terá de ser transmitida a ele através de algum meio seguro, para mantê-la secreta. Esse problema é conhecido como *o problema da distribuição de chaves*, e será discutido na Seção 3.1.

2.4.1 Cifradores de substituição e de transposição

De acordo com as operações usadas para cifrar os dados, os cifradores simétricos podem ser baseados em operações de *substituição* ou de *transposição*. Os **cifradores de substituição** se baseiam na substituição de caracteres por outros caracteres usando tabelas de substituição (ou *alfabetos*). Esses cifradores podem ser **monoalfabéticos**, quando usam uma única tabela de substituição, ou **polialfabéticos**, quando usam mais de uma tabela.

O cifrador de César é um exemplo trivial de cifrador de substituição monoalfabético. Outro exemplo dessa família, bem conhecido na cultura popular, é a linguagem *Alien*, usada em episódios da série de TV *Futurama*, cuja tabela é apresentada na Figura 2.3.

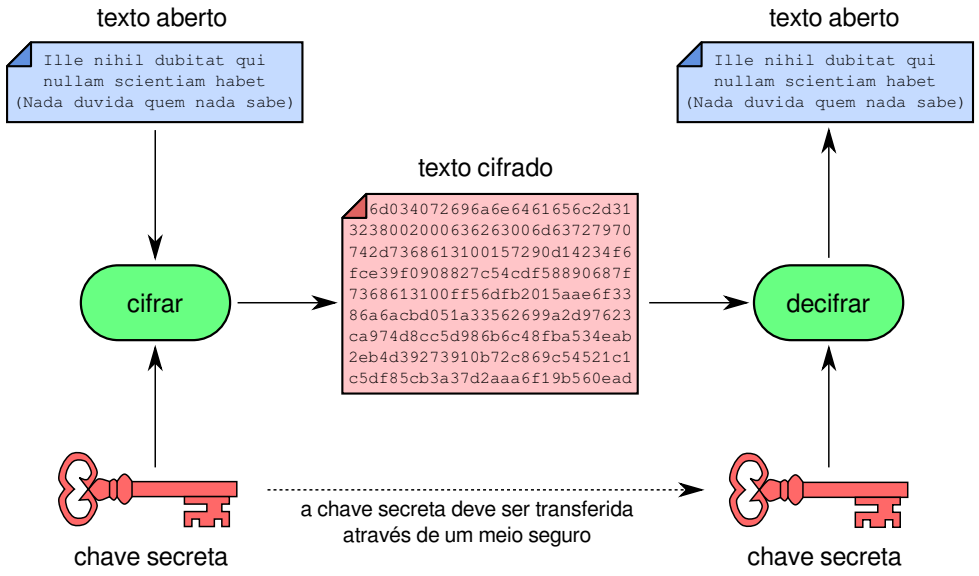

Figura 2.2: Criptografia simétrica.

Figura 2.3: Linguagem *Alien* da série *Futurama*.

Os cifradores de substituição polialfabéticos operam com mais de uma tabela de substituição de caracteres. Um exemplo clássico de cifrador polialfabético é o cifrador de Vigenère, que foi inicialmente proposto por Giovan Battista Bellaso em 1553 e refinado por Blaise de Vigenère no século XIX. Trata-se de um método de cifragem que combina vários cifradores de César em sequência. As operações de cifragem/decifragem usam uma tabela denominada *tabula rasa*, apresentada na Figura 2.4.

Nesse cifrador, para cifrar uma mensagem, primeiro se escolhe uma palavra-chave qualquer, que é repetida até ter o mesmo comprimento da mensagem. Em seguida, cada caractere da mensagem original é codificado usando um cifrador de substituição específico, definido pela linha da *tabula rasa* indicada pela letra correspondente da palavra-chave.

Um exemplo de cifragem usando a palavra-chave “bicicleta” é indicado a seguir. Nele, pode-se observar que a letra “M” da mensagem aberta, combinada à letra “T” da palavra-chave, gera a letra “F” na mensagem cifrada.

Mensagem aberta	ATACARE**M**OS AO AMANHECER DE SEXTA-FEIRA
Palavra-chave	BICICLE**T**AB IC ICLETABIC IC LETAB ICICL
Mensagem cifrada	BBCKCCI**F**OT IQ IOLRAEDMT LG DIQTB-NGQTL

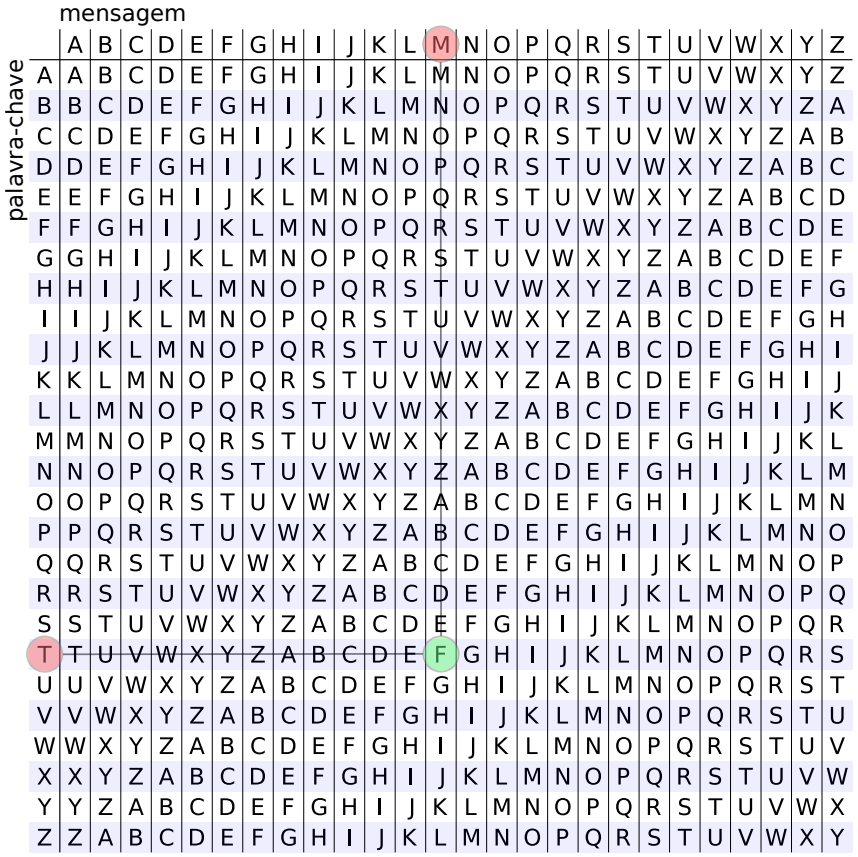

Figura 2.4: Tabula rasa do cifrador de Vigenère.

No cifrador de Vigenère, uma senha com n caracteres distintos irá usar n linhas distintas da *tabula rasa*. Considerando um alfabeto com 26 letras, teremos um espaço de chaves de 26^n, o que é bastante respeitável para a época em que o esquema foi usado.

Os cifradores de substituição podem ainda ser **monográficos** ou **poligráficos**, conforme cifrem caracteres individuais ou grupos de caracteres. Os cifradores de César e de Vigenère são monográficos. O cifrador *Playfair*, que trata o texto aberto em grupos de duas letras, é um bom exemplo de cifrador poligráfico.

Por outro lado, os **cifradores de transposição** (ou permutação) têm como mecanismo básico a troca de posição (ou embaralhamento) dos caracteres que compõem uma mensagem, sem substituí-los. O objetivo básico da operação de transposição é espalhar a informação aberta em toda a extensão do texto cifrado.

Um exemplo clássico de cifrador de transposição é o algoritmo *Rail Fence*, no qual os caracteres da mensagem aberta são distribuídos em várias linhas de uma "cerca" imaginária. Por exemplo, considerando a mesma mensagem do exemplo anterior ("Atacaremos ao amanhecer de sexta-feira") e uma cerca com 4 linhas ($k = 4$), teríamos a seguinte distribuição de caracteres na cerca:

```
A . . . . . E . . . . . A . . . . . C . . . . . E . . . . . I . .
. T . . . R . M . . . O . M . . . E . E . . . S . X . . . E . R .
. . A . A . . . O . A . . . A . H . . . R . E . . . T . F . . . A
. . . C . . . . . S . . . . . N . . . . . D . . . . . A . . . . .
```

A mensagem cifrada é obtida ao percorrer a cerca linha após linha: `AEACEITRMOMEESXERAAOAAHRETFACSNDA`. É interessante observar que os caracteres da

mensagem cifrada são os mesmos da mensagem aberta; uma análise de frequência dos caracteres poderá auxiliar a inferir qual a língua usada no texto, mas será pouco útil para identificar as posições dos caracteres na mensagem original.

Algoritmos de cifragem por substituição e por transposição raramente são utilizados isoladamente. Cifradores simétricos modernos, como o 3DES, AES e outros, são construídos usando vários blocos de substituição e de transposição aplicados de forma alternada, para aumentar a resistência do criptossistema [Stamp, 2011].

2.4.2 Cifradores de fluxo e de bloco

De acordo com a forma de agrupar os dados a cifrar, os cifradores simétricos podem ser classificados em dois grandes grupos: os *cifradores de fluxo* e os *cifradores de bloco*. Os **cifradores de fluxo** (*stream ciphers*) cifram cada byte da mensagem aberta em sequência, produzindo um byte cifrado como saída. Por essa característica sequencial, esses cifradores são importantes para aplicações de mídia em tempo real, como VoIP (voz sobre IP) e comunicações em redes celulares. Exemplos típicos de cifradores de fluxo incluem o RC4, usado até pouco tempo atrás nas redes sem fio, e o A5/1, usado para cifrar fluxo de voz em telefones GSM.

A maioria dos cifradores de fluxo funciona de forma similar: um fluxo contínuo de bytes aleatórios (*keystream*) é gerado por um bloco PRNG (*Pseudo-Random Number Generator*) a partir de uma semente que é a chave simétrica, ou calculada a partir dela. Cada byte desse fluxo é combinado com um byte do fluxo de dados aberto, para produzir um byte do fluxo cifrado. A combinação entre os fluxos geralmente é feita usando a operação XOR, de forma similar ao cifrador de Vernam. No lado do receptor, o mesmo bloco PRNG, inicializado com a mesma semente, refaz a operação XOR para decifrar o fluxo de dados. A figura 2.5 ilustra esse funcionamento.

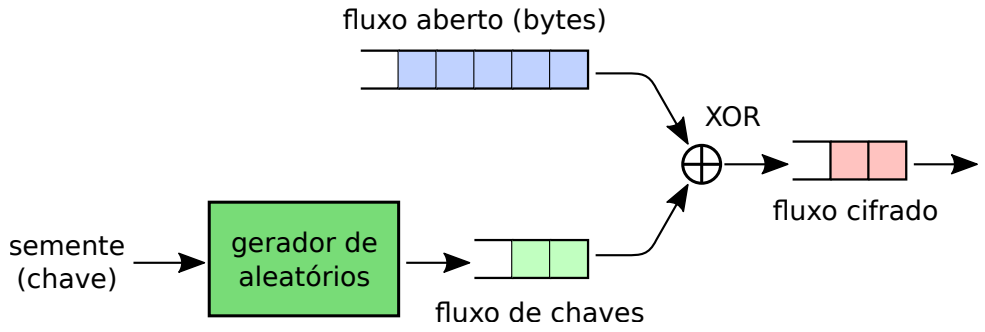

Figura 2.5: Funcionamento básico de um cifrador por fluxo

Como seu próprio nome diz, os **cifradores de bloco** (*block ciphers*) cifram os dados em blocos de mesmo tamanho, geralmente entre 64 e 128 bits. Os dados a serem cifrados são divididos em blocos e o algoritmo de cifragem é aplicado a cada bloco, até o final dos dados. Caso o último bloco não esteja completo, bits de preenchimento (*padding*) são geralmente adicionados para completá-lo. A operação em blocos provê a estes algoritmos uma maior eficiência para cifrar grandes volumes de dados, como tráfego de rede e arquivos em disco. Exemplos comuns de cifradores de bloco incluem o AES (*Advanced Encryption Standard*) e o DES/3DES (*Data Encryption Standard*).

O *modo de operação* de um cifrador de blocos define a forma como algoritmo percorre e considera os blocos de dados a cifrar. Esse modo de operação pode ter

um impacto significativo na segurança do algoritmo. O modo de operação mais simples, chamado ECB (de *Electronic Codebook*), consiste em aplicar o mesmo algoritmo sobre os blocos de dados abertos em sequência, obtendo os blocos de dados cifrados correspondentes. Esse modo de operação está ilustrado na Figura 2.6.

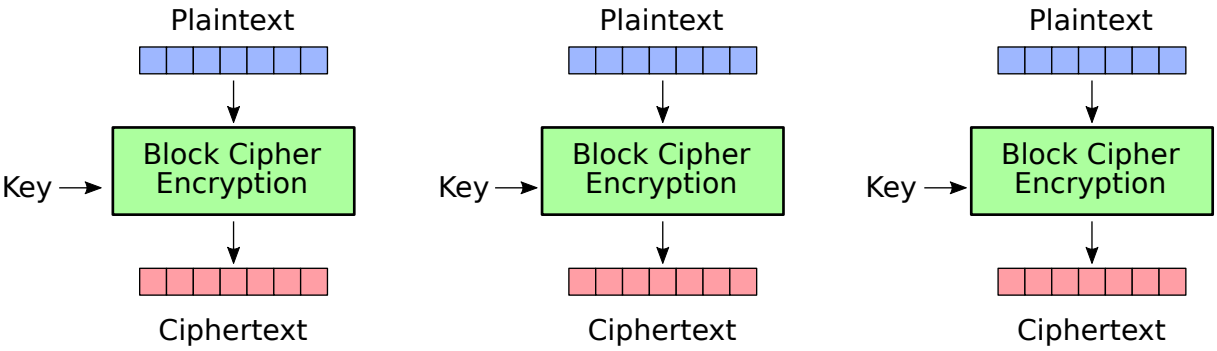

Figura 2.6: Cifragem por blocos em modo ECB [Wikipedia, 2018]

O modo de operação ECB preserva uma forte correlação entre os trechos da mensagem aberta e da mensagem cifrada, o que pode ser indesejável. A figura 2.7 demonstra o efeito dessa correlação indesejada na cifragem por blocos em modo ECB aplicada aos pixels de uma imagem.

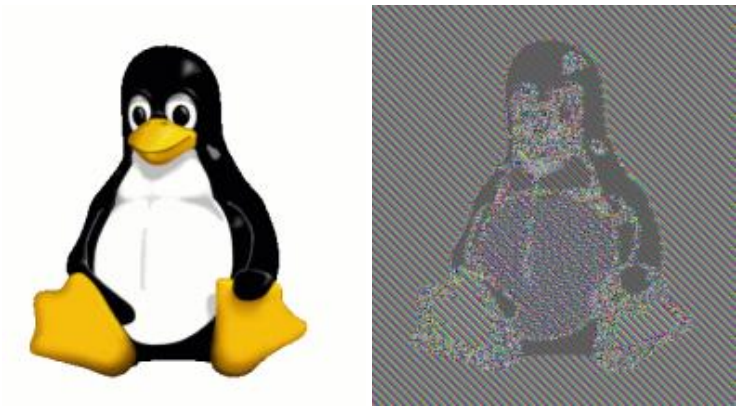

Figura 2.7: Cifragem por blocos em modo ECB: imagem aberta (à esquerda); imagem cifrada (à direita) [Wikipedia, 2018]

Para evitar a correlação direta entre blocos da entrada e da saída, cifradores de bloco modernos usam modos de operação mais sofisticados, que combinam blocos entre si. Um modo de operação bem simples com essa característica é o CBC (*Cipher Block Chaining*), no qual a saída do primeiro bloco é combinada (XOR) com a entrada do segundo bloco e assim por diante, como ilustrado na Figura 2.8. Nessa figura, o IV (*Initialization Vector*) corresponde a um bloco aleatório adicional, que deve ser conhecido ao cifrar e decifrar a mensagem. A Figura 2.9 demonstra o efeito do modo de operação CBC sobre a cifragem em blocos de uma imagem.

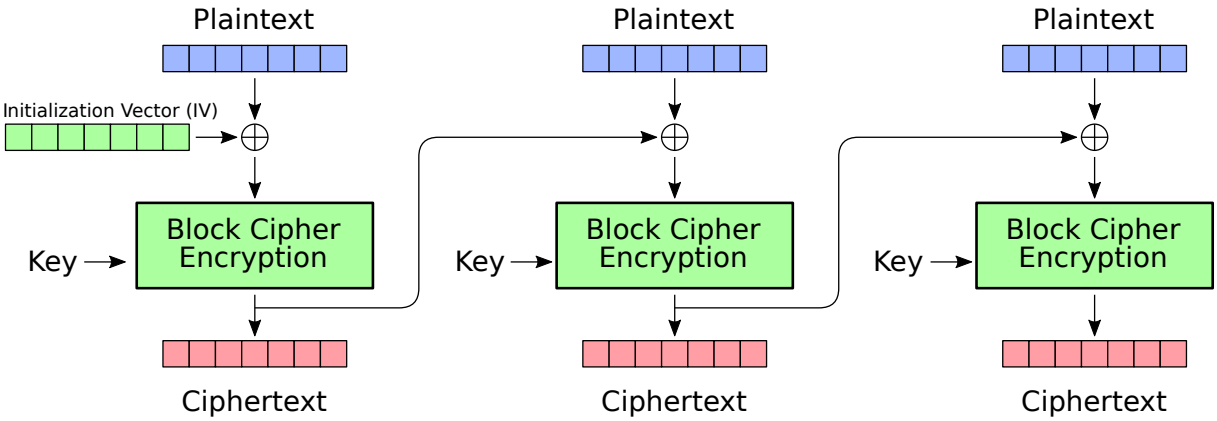

Figura 2.8: Cifragem por blocos em modo CBC [Wikipedia, 2018]

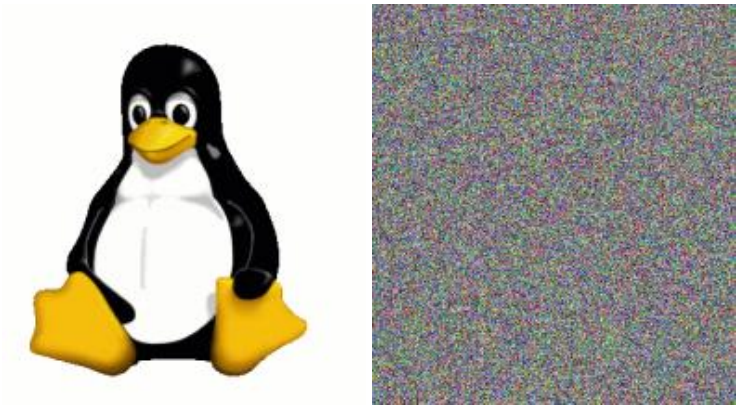

Figura 2.9: Cifragem por blocos em modo CBC: imagem aberta (à esquerda); imagem cifrada (à direita) [Wikipedia, 2018]

Exercícios

1. Na série de TV Futurama (escrita por Matt Groening, o mesmo autor dos Simpsons) é usada uma escrita cifrada denominada *Alien Language*, mostrada na Figura 2.3. Explique qual o tipo de criptografia empregado na *Alien Language* e indique qual o tamanho do espaço de chaves da mesma.

2. O texto em português a seguir foi cifrado usando o cifrador de César. Encontre o **texto original** e a **chave** usada para cifrá-lo; explique seu procedimento.

   ```
   Kjqne fvzjqj vzj ywfsxkjwj t vzj
   xfgj j fuwjsij t vzj jsxnsf.
   Htwf Htwfqnsf.
   ```

   Para facilitar seu trabalho, a tabela a seguir traz a frequência de caracteres típica de textos na língua portuguesa:

letra	freq%	letra	freq%	letra	freq%	letra	freq%	letra	freq%
A	14,6	B	1,04	C	3,88	D	4,99	E	12,6
F	1,02	G	1,30	H	1,28	I	6,18	J	0,40
K	0,02	L	2,78	M	4,74	N	5,05	O	10,7
P	2,52	Q	1,20	R	6,53	S	7,81	T	4,34
U	4,63	V	1,67	W	0,01	X	0,21	Y	0,01
Z	0,47								

3. Use o cifrador de Vigenère para cifrar a mensagem secreta "Encontramos aliens" usando a palavra-chave "missao".

Referências

A. Menezes, P. Van Oorschot, and S. Vanstone. *Handbook of Applied Cryptography*. CRC Press, 1996.

M. Stamp. *Information Security - Principles and Practice, 2nd edition*. Wiley, 2011.

Wikipedia. Wikipedia online enciclopedia. http://www.wikipedia.org, 2018.