Construção de um Módulo de Criptologia

Sidnei Ramos da Cruz

DEPARTAMENTO DE MATEMÁTICA
FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE DE COIMBRA

Construção de um Módulo de Criptologia

Sidnei Ramos da Cruz

Dissertação para a obtenção do Grau de **Mestre em Matemática**

Área de Especialização em **Computação**

Júri

Presidente: Professor Doutor Jorge Picado

Orientador: Professor Doutor Pedro Quaresma

Vogais: Professora Doutora Cristina Caldeira

Data: 25 de Agosto de 2008

Resumo

Com este trabalho pretende-se fazer um estudo da criptologia clássica, nomeadamente, os métodos criptográficos: *Cifra de Deslocamento Simples*, *Cifra de Deslocamento Linear*, *Cifra de Vigenère*, e criptoanalíticos *Procura Exaustiva no Espaço das Chaves*, *Análise de Frequências*.

Desenvolveu-se ainda uma página da Rede que permita aos seus utilizadores a compreensão, a experimentação, e a exploração dos métodos referidos.

Palavras Chave: Criptografia, Criptoanálise.

Abstract

The goal of this work was to present the cryptography and the cryptoanalysis of the classical ciphers, in particular the shift chiper, the affine cipher and the Vigenère chipher.

The cryptoanalysis methods used where the exaustive search in the key space and the frequencies analysis of the Portuguese language.

A Web-page was developed in order to give to its users a workbench to the mentioned chipers.

Keywords: Cryptography, Cryptanalysis.

Agradecimentos

Agradeço ao meu orientador, Doutor Pedro Quaresma todo o apoio na elaboração deste trabalho e pelo seu apoio amigo ao longo destes anos no Departamento de Matemática da Faculdade de Ciências e Tecnologia da Universidade de Coimbra.

Quero também agradecer à minha namorada pelo apoio e incentivo ao longo destes meses, à minha família, distante, mas ao mesmo tempo perto, aos meus amigos e à equipa do gabinete de Apoio aos Estudantes da CPLP, da Faculdade de Ciências e Tecnologia da Universidade de Coimbra.

Um obrigado ao Departamento de Matemática da Faculdade de Ciências e Tecnologia da Universidade de Coimbra pela oportunidade concedida de estudar num dos Departamento de reconhecimento Internacional.

A todos, o meu muito obrigado.

Dedico este trabalho aos meus queridos pais.

Sidnei Cruz

Conteúdo

Capítulo 1

Introdução

Desde os tempos em que se inventou a escrita, o homem sentiu a necessidade de esconder a informação, ao mesmo tempo surgiu a necessidade de descobrir os segredos que os outros pretendem manter em sigilo.

Esta necessidade conduziu ao aparecimento e desenvolvimento da criptologia, ciência que se ocupa da escrita secreta em todas as suas formas, abrangendo, por um lado, quer a criptografia quer, por outro a criptoanálise.

Até recentemente a criptologia era vista como uma arte, foi oficialmente considerada uma ciência, há cerca de vinte anos [8]. A figura 1.1 mostra-nos algumas das áreas de estudo desta ciência.

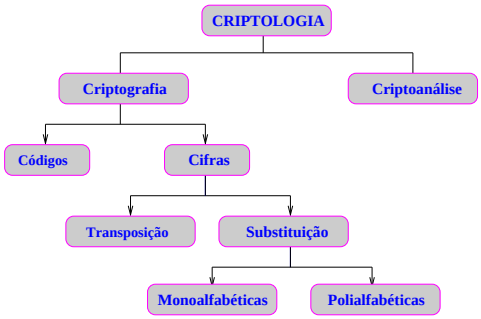

Figura 1.1: Algumas áreas de estudo da criptologia

Neste trabalho, procura-se fazer um estudo sobre alguns métodos criptográficos e criptoanalíticos das cifras clássicas. Analisar-se-á a cifra de Deslocamento Simples, Deslocamento Linear e por ultimo a de Vigenère. Com o resultado do estudo desenvolveu-se uma Página da Rede que permitirá aos seus utilizadores a compreensão dos métodos, assim como providenciando uma plataforma de experimentação ou exploração. Este trabalho está estruturado da seguinte forma, no primeiro capítulo definiu-se alguns termos que são importantes para a percepção do conteúdo do trabalho e também alguns tópicos sobre a Teoria dos Números . No segundo capítulo fez-se o estudo dos métodos criptográficos clássicos, apresentando-se alguns exemplos para ilustrar o funcionamento das cifras. Seguidamente, no terceiro capítulo estudou-se os métodos criptoanalíticos, alguns exemplos que ilustram o funcionamento dos métodos, comparação entre os méto-

dos criptoanalíticos para cada uma das cifras, e a aplicabilidade de um dos métodos criptoanalítico para a cifra de Vigenère. No quarto capítulo falou-se sobre a Página da Rede e dos códigos dos programas, e por último as conclusões.

1.1 Terminologia Básica

Antes de prosseguir com o estudo das cifras clássicas, apresentam-se algumas definições para os termos encontrados ao longo deste trabalho [3, 5, 6, 8]. Assim:

Criptologia é o estudo da criptografia e da criptoanálise.

Criptografia é o estudo das técnicas matemáticas relacionadas com os aspectos de segurança da informação tais como: confidencialidade, integridade da informação, autenticação de entidades e autenticação da origem da informação.

Criptoanálise é o estudo das técnicas matemáticas para tentar comprometer as técnicas criptográficas, e mais genericamente, os serviços de segurança da informação.

Encriptar é o processo de aplicar uma transformação a uma mensagem em texto claro para uma mensagem cifrada.

Desencriptar é o processo de aplicar uma transformação a uma mensagem cifrada. É suposto ser o processo inverso da encriptação.

Chave é um valor usado no processo de encriptação e desencriptação.

Cifras são técnicas de encriptação que são aplicadas a uma mensagem.

Cifras Simétricas são cifras em que a chave de encriptação e desencriptação é a mesma, ou nas quais uma qualquer das chaves pode ser obtida a partir da outra.

Cifra de Substituição é uma cifra em que se substitui uma letra por outra.

Cifra Monoalfabética é uma cifra em que cada letra do texto claro é substituída sempre por uma mesma letra na mensagem cifrada.

Cifra Polialfabética é uma cifra em que cada letra do texto claro é substituída por mais do que uma letra na mensagem cifrada.

Procura Exaustiva no Espaço das Chaves é o processo que consiste no testar de todas as chaves possíveis.

Análise de Frequências processo de comparar as frequências relativas das ocorrências de caracteres da língua base da mensagem à frequência observada dos caracteres na mensagem cifrada.

Digramas e Trigramas entendem-se por digramas e trigramas os conjuntos de duas e três letras seguidas respectivamente e que constituem sub-palavras, isto é, que fazem parte de uma palavra.

Palavras Curtas São palavras cujo comprimento está abaixo da média das palavras na língua de base da mensagem. No caso da Língua Portuguesa trata-se das palavras com comprimento um, dois,e três.

Letras Iniciais e Letras Finais São letras que constituem o início e o fim de uma palavra respectivamente.

1.2 Alguns Tópicos sobre Teoria dos Números

Os métodos criptográficos que vão ser estudados, baseiam-se, essencialmente em Teoria dos Números, em particular na Aritmética Modular. Portanto será feita uma revisão a algumas noções básicas desta área da Matemática. Os resultados que se seguem são bem conhecidos da literatura, por exemplo [1, 2, 6, 7].

Definição 1 (Divisibilidade nos Inteiros) *Dados inteiros a e b, com $a \neq 0$, dizemos que a divide b (ou a é um divisor de b, ou b é múltiplo de a, ou b é divisível por a) se existir um inteiro x tal que $b = ax$. A notação usada para esta relação é $a|b$.*

Definição 2 (Máximo Divisor Comum) *Sejam a e b inteiros não ambos nulos. Ao maior dos divisores comuns de a e b chama-se máximo divisor comum de a e b. A notação usada é $\mathrm{mdc}(a, b)$. Dois inteiros a e b são ditos primos relativos ou co-primos se $\mathrm{mdc}(a, b) = 1$.*

Definição 3 (Congruência) *Seja m um inteiro positivo. Dois inteiros a e b dizem-se congruentes módulo m se m divide $b - a$. A notação é $a \equiv b \pmod{m}$. Ao inteiro m chama-se o módulo da congruência.*

Dada a congruência $a \equiv b \bmod m$, se dividir a e b por m, obtém-se, um quociente e um resto inteiro, tais que os restos estão entre 0 e $m - 1$ ou seja,

$$a = q_1 m + r_1 \;\; e \;\; b = q_2 m + r_2 \qquad onde \qquad 0 \leq r_1 \leq m - 1,\; 0 \leq r_2 \leq m - 1.$$

Então não é difícil ver-se que, $a \equiv b \bmod m$ se e só se $r_1 = r_2$.

Por conseguinte $a \equiv b \bmod m$ é equivalente a $a \bmod m = b \bmod m$.

São válidas as afirmações seguintes, onde a, b, $c \in \mathbb{Z}$ e $m \in \mathbb{N}$:

1. **Reflexiva**: $a \equiv a \bmod m$, para todo o $a \in \mathbb{Z}$;

2. **Simétrica**: se $a \equiv b \bmod m$, então $b \equiv a \bmod m$;

3. **Transitiva**: se $a \equiv b \bmod m$ e $b \equiv a \bmod m$, então $a \equiv c \bmod m$.

Assim , a congruência módulo m é uma relação de equivalência sobre $\mathbb{Z}$. Deste modo, podem-se definir classes de congruência módulo m da forma

$$[a] = \{n \in \mathbb{Z} : n \equiv a \bmod m\}, \ com\ a \in \mathbb{Z}$$

ou seja $[a]$ é o conjunto de inteiros da forma $a + km$, onde k varia em $\mathbb{Z}$.

É natural a identificação do conjunto das classes de equivalência das congruências módulo m: $\mathbb{Z}_m$ com os elementos $\{0, 1, ..., m-1\}$, equipado com duas operações, $+$ e $\times$.

A adição e a multiplicação em $\mathbb{Z}_m$ funcionam, exactamente, como adição e multiplicação usuais, excepto os resultados que são reduzidos módulo m. As operações $+$ e $\times$ verificam o mesmo conjunto de propriedades:

1. a adição é fechada, isto é, para qualquer $a, b \in \mathbb{Z}_m$, tem-se que, $a + b \in \mathbb{Z}_m$;

2. a adição é comutativa, isto é, para qualquer $a, b \in \mathbb{Z}_m$, tem-se que, $a + b = b + a$;

3. a adição é associativa, isto é, para qualquer $a, b, c \in \mathbb{Z}_m$, tem-se que, $(a+b)+c = a + (b + c)$;

4. 0 é o elemento neutro da adição, isto é, para qualquer $a \in \mathbb{Z}_m$, tem-se que, $a + 0 = 0 + a = a$;

5. o inverso aditivo de qualquer $a \in \mathbb{Z}_m$ é $m - a$, isto é, $a + (m - a) = (m - a) = 0$ para qualquer $a \in \mathbb{Z}_m$;

6. a multiplicação é fechada, isto é, para qualquer $a, b \in \mathbb{Z}_m$, tem-se que, $ab \in \mathbb{Z}_m$;

7. a multiplicação é comutativa, isto é, para qualquer $a, b \in \mathbb{Z}_m$, tem-se que, $ab = ba$;

8. a multiplicação é associativa, isto é, para qualquer $a, b, c \in \mathbb{Z}_m$, tem-se que, $(ab)c = a(bc)$;

9. 1 é o elemento neutro da multiplicação, isto é, para qualquer $a \in \mathbb{Z}_m$, $a \times 1 = 1 \times a = a$;

10. a propriedade distributiva é válida, isto é, para qualquer $a, b, c \in \mathbb{Z}_m$, $(a+b)c = (ac) + (bc)$ e $a(b+c) = (ab) + (ac)$.

As propriedades $1, 3-5$ em $\mathbb{Z}_m$, formam uma estrutura algébrica chamada de *grupo*, com a respectiva operação de adição. Quando a propriedade 2 se verifica, o grupo é designado de *grupo abeliano*.

Com as propriedades $1-10$, $\mathbb{Z}_m$ é um *anel*.

Nota: Nem todos os elementos de $\mathbb{Z}_m$ têm inverso multiplicativo, mais a frente falar-se-á sobre o assunto.

Capítulo 2

Métodos Criptográficos Clássicos

A criptografia tem uma longa e fascinante história [3, 5, 8]. Pode dizer-se que o uso da criptografia é tão antiga quanto a necessidade do homem em esconder informação.

A palavra criptografia vem do Grego Kryptós, "esconder" e graphein, "escrever", é geralmente entendida como sendo o estudo das técnicas matemáticas pelas quais a informação pode ser transformada da sua forma original para outra ilegível, a menos que seja conhecida uma "chave secreta". Deste modo, só o receptor da mensagem, conhecendo a chave secreta, pode ler a informação com facilidade.

O objectivo essencial da criptografia é possibilitar a comunicação entre duas entidades, por exemplo, duas pessoas, o Miguel e o Jorge, através de um canal não necessariamente seguro, de forma a que um intruso (por exemplo o João) que interfira na comunicação não compreenda a mensagem que o Miguel transmite ao Jorge.

2.1 Criptosistema

$\mathcal{A}$ denota um conjunto finito de símbolos, designado por alfabeto de definição.

Em termos matemáticos tem-se que um criptosistema é um 5-uplo $(\mathcal{P}, \mathcal{C}, \mathcal{K}, \mathcal{E}, \mathcal{D})$, onde as seguintes condições são satisfeitas [6]:

1. $\mathcal{P}$ é o espaço das mensagem, um conjunto finito de textos em claro. $\mathcal{P}$ é composto por sequências de elementos de $\mathcal{A}$;

2. $\mathcal{C}$ é o espaço das mensagens cifradas, um conjunto finito de texto cifrado. $\mathcal{C}$ é composto por sequências de elementos de $\mathcal{A}$;

3. $\mathcal{K}$ é o espaço das chaves, um conjunto finito de chaves possíveis;

4. $\mathcal{E}$ é o conjunto das funções de encriptação;

5. $\mathcal{D}$ é o conjunto das funções de desencriptação;

6. Para cada k $\in \mathcal{K}$, tem-se a função de encriptação $e_k \in \mathcal{E}$ e a correspondente função de desencriptação $d_k \in \mathcal{D}$. Cada $e_k : \mathcal{P} \longrightarrow \mathcal{C}$ e $d_k : \mathcal{C} \longrightarrow \mathcal{P}$ são funções tais que $d_k(e_k(x)) = x$ para todo o elemento do texto original $x \in \mathcal{P}$.

A condição 6 indica que a função d_k é inversa à esquerda da função e_k para uma mesma chave k, isto é, que o criptosistema é simétrico.

O processo de comunicação entre o Miguel e o Jorge, numa cifra clássica ou simétrica, utiliza um sistema criptográfico específico, que segue os seguintes passos:

1. o Miguel e o Jorge escolhem, aleatoriamente, uma chave $k \in \mathcal{K}$. Este passo é efectuado quando eles estão juntos, sem serem observados pelo João, ou quando têm acesso a um canal seguro, neste caso, não precisam de estar no mesmo local;

2. o Miguel cifra uma mensagem e envia-a ao Jorge, por meio de um canal não necessariamente seguro. Para isso, admite-se que a mensagem é uma sequência de símbolos (cada x_i é um símbolo do texto original)

$$x = x_1 x_2 x_3 \cdots x_n$$

   para algum inteiro $n \geq 1$, onde $x_i \in \mathcal{A}$, $1 \leq i \leq n$. Cada x_i é encriptado em $y_i = e_k(x_i)$, $1 \leq i \leq n$ utilizando a função e_k, especificada para a chave $k \in \mathcal{K}$ pré-definida. Deste modo obtém-se a mensagem encriptada

$$y = y_1 y_2 y_3 \cdots y_n$$

   que é enviada através de um canal não necessariamente seguro;

3. O Jorge recebe a mensagem encriptada $y = y_1 y_2 y_3 \cdots y_n$ e desencripta-a através da função d_k, isto é, obtém-se a mensagem original, $x = x_1 x_2 x_3 \cdots x_n$, uma vez que $d_k(y) = x$.

O processo é ilustrado na figura 2.1

É imprescindível que a função e_k seja injectiva, caso contrário a decifração não poderia ser efectuada da forma inequívoca, isto é, é necessário que se possa definir a função inversa inversa $e_k^{-1} = d_k$.

Por exemplo, se uma mensagem fosse encriptada através de uma função tal que

$$y = e_k(x_i) = e_k(x_j), \; onde \; x_i \neq x_j,$$

o receptor do texto encriptado, não saberia qual das opções a tomar, isto é, se deveria desencriptar y em x_i ou em x_j.

Note-se que, para o caso em que $\mathcal{P} = \mathcal{C}$, segue-se que cada função de encriptação (e_k) é uma permutação, isto é, se o conjunto de textos em claro e de textos cifrados são iguais, então cada função e_k apenas permuta os elementos deste conjunto.

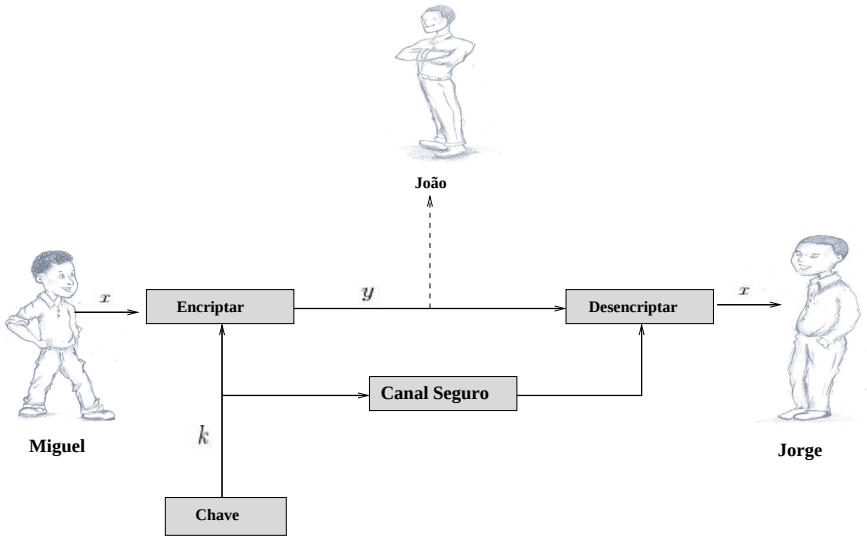

Figura 2.1: Modelo típico de comunicação

2.2 Definição do Alfabeto

O alfabeto utilizado neste trabalho é o alfabeto Português, constituído por caracteres de 'a' a 'z' e ainda por caracteres acentuados, bem como pelo 'c' cedilhado, todas minúsculas. O comprimento do alfabeto é de 43 como mostra a tabela 2.1. A tabela 2.2 mostra a codificação do alfabeto na qual se estabeleça uma correspondência biunívoca entre os caracteres do alfabeto e os inteiros entre 0 e 42. Todas as cifras serão estudadas sobre este alfabeto.

Notas:

1. Os caracteres que não fazem parte do alfabeto durante o processo de encriptar ou desencriptar uma mensagem vão manter-se inalterados;

2. A escolha do alfabeto foi baseado na codificação ISO 8859-1 (ISO Latin1 - línguas da Europa Ocidental).

3. As mensagens a ser encriptadas também podem ser escritas em maiúsculas sendo automaticamente convertidas para minúsculas.

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	à	á	â	ã	ç	è
é	ê	ì	í	ò	ó	ô	õ	ù	ú	ü					

Tabela 2.1: Definição do Alfabeto com o comprimento de 43 caracteres.

a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
q	r	s	t	u	v	w	x	y	z	à	á	â	ã	ç	è
16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
é	ê	ì	í	ò	ó	ô	õ	ù	ú	ü					
32	33	34	35	36	37	38	39	40	41	42					

Tabela 2.2: Codificação do Alfabeto.

De seguida apresentam-se os três métodos criptográficos que vão ser designadas por cifras clássicas. São cifras que, actualmente, caíram em desuso dado, como mais à frente se verá, serem pouco seguras. Estas cifras podem ser utilizadas manualmente ou através de dispositivos mecânicos simples.

2.3 Cifra de Deslocamento Simples

A cifra de deslocamento simples é provavelmente a mais bem conhecida da história das cifras [5, 8]. Trata-se de uma cifra de substituição monoalfabética em que cada letra é substituída por outra letra, n posições à frente no alfabeto. Em termos matemáticos, isso significa somar uma dada quantidade à codificação da letra, soma essa feita módulo o comprimento do alfabeto.

Criptosistema 1 (Cifra de Deslocamento Simples) *Sejam $\mathcal{P}$ e $\mathcal{C}$ sequência de símbolos de $\mathbb{Z}_{43}$, seja $\mathcal{K} = \mathbb{Z}_{43}$. Para $k \in \mathcal{K}$, define-se*

$$e_k(x) = (x + k) \bmod 43$$

e

$$d_k(y) = (y - k) \bmod 43$$

onde $x, y \in \mathbb{Z}_{43}$.

Podemos verificar que, $d_k(e_k(x)) = x$, $\forall x \in \mathbb{Z}_{43}$.

$$d_k(e_k(x)) = d_k(x + k) = x + k - k = x, \ \forall x \in \mathbb{Z}_{43}.$$

Para uma chave particular $k = 3$, a cifra é muitas vezes designada **Cifra de César**, que foi, supostamente, usada por Júlio César (Imperador Romano) durante as suas campanhas militares [6].

Para encriptar uma mensagem usando a cifra de deslocamento simples, primeiro escolhe-se uma chave, que é um inteiro $k \in \mathbb{Z}_{43}$. Usando a codificação da tabela 2.2 converter-se cada letra da mensagem para o inteiro correspondente. O próximo passo é adicionar o valor da chave a cada inteiro. E, por último, converter os inteiros obtidos para a letra correspondente no alfabeto, obtendo deste modo a mensagem cifrada.

Para desencriptar a mensagem o processo é similar. A diferença é que durante o processo de desencriptação deve-se subtrair o valor da chave em vez de adicioná-la.

De seguida apresenta-se um pequeno exemplo para ilustrar a cifra em estudo.

Exemplo 1 *Encriptar e Desencriptar a palavra "caboverde" usando a cifra de deslocamento simples com uma chave $k = 3$.*

Passo 1 Converter cada letra da palavra para o inteiro correspondente na codificação definida na tabela 2.2.

c	a	b	o	v	e	r	d	e
2	0	1	14	21	4	17	3	4

Passo 2 Adicionar, módulo 43, a chave $k = 3$ para cada um dos inteiros, reduzindo o resultado para o módulo 43 dos inteiros.

$$\begin{array}{rrrrrrrrrr} & 2 & 0 & 1 & 14 & 21 & 4 & 17 & 3 & 4 \\ + & 3 & 3 & 3 & 3 & 3 & 3 & 3 & 3 & 3 \\ \hline & 5 & 3 & 4 & 17 & 24 & 7 & 20 & 6 & 7 \end{array}$$

Passo 3 Converter cada um dos inteiros do resultado final, para a letra correspondente no alfabeto definido na tabela 2.2.

5	3	4	17	24	7	20	6	7
f	d	e	r	y	h	u	g	h

E obtém-se a palavra cifrada "fderyhugh".

Para desencriptar o processo é similar. A diferença é que durante o processo de desencriptação deve-se subtrair o valor da chave em vez de adicioná-la.

Passo 1 Converter cada letra da palavra para um inteiro correspondente no alfabeto definido na tabela 2.2.

f	d	e	r	y	h	u	g	h
5	3	4	17	24	7	20	6	7

Passo 2 Subtrair, módulo 43, a chave $k = 3$ para cada um dos inteiros.

$$\begin{array}{rrrrrrrrrr} & 5 & 3 & 4 & 17 & 24 & 7 & 20 & 6 & 7 \\ - & 3 & 3 & 3 & 3 & 3 & 3 & 3 & 3 & 3 \\ \hline & 2 & 0 & 1 & 14 & 21 & 4 & 17 & 3 & 4 \end{array}$$

Passo 3 Converter cada um dos inteiros do resultado final, para a letra correspondente no alfabeto na tabela 2.2.

2	0	1	14	21	4	17	3	4
c	a	b	o	v	e	r	d	e

E tem-se a palavra original "caboverde".

2.4 Cifra de Deslocamento Linear

A cifra de deslocamento linear é uma cifra de substituição monoalfabética. A chave da cifra é um par de inteiros (a,b), com $(a,\ b) \in \mathcal{K}$. A função de encriptação é uma função da forma:

$$e_k(x) = (ax + b) \bmod 43.$$

Estas funções são chamadas de funções lineares, daí o nome da cifra de deslocamento linear (observação: quando $a = 1$, temos a cifra de deslocamento simples).

Para que o processo de desencriptação seja possível, é necessário perguntar quando é que a função linear é injectiva. Por outras palavras, para qualquer $y \in \mathbb{Z}_{43}$, pretende-se que a congruência

$$ax + b \equiv y \pmod{43},$$

tenha uma solução única para x. Sendo esta equivalente a

$$ax \equiv y - b \pmod{43},$$

No entanto, como y varia em $\mathbb{Z}_{43}$, assim também $y - b$ varia em $\mathbb{Z}_{43}$. Daí que é suficiente o estudo da congruência $ax \equiv y \pmod{43}$.

Esta congruência tem uma solução única para todo o y se e só se $\text{mdc}(a, 43) = 1$ (onde a função mdc denota o máximo divisor comum).
Note-se o seguinte exemplo: supondo-se que $\text{mdc}(a, 43) = 1$, para algum x_1 e x_2 tais que

$$ax_1 \equiv ax_2 \pmod{43}.$$

então

$$a(x_1 - x_2) \equiv 0 \pmod{43},$$

e assim

$$43 \mid a(x_1 - x_2).$$

Fazendo agora uso da propriedade fundamental da divisão inteira: se $\mathrm{mdc}(a,b)=1$ e $a \mid bc$, então $a \mid c$. Como $43 \mid a(x_1-x_2)$ e $\mathrm{mdc}(a,43)=1$, obtém-se então

$$43 \mid (x_1-x_2),$$

isto é, $x_1 \equiv x_2 \pmod{43}$.

Até este ponto, mostrou-se que, se $\mathrm{mdc}(a,43)=1$, então a congruência da forma $ax \equiv y \pmod{43}$ tem na maior parte das vezes, uma solução em $\mathbb{Z}_{43}$.

Pelo teorema 1 conclui-se que a congruência $ax \equiv y \pmod{43}$, $y \in \mathbb{Z}_{43}$ tem solução única quando $(a,43)=1$.

Teorema 1 *A congruência $ax \equiv b \pmod{m}$ tem uma única solução $x \in \mathbb{Z}_{\mathrm{m}}$ para todo $b \in \mathbb{Z}_{\mathrm{m}}$ se e só se $\mathrm{mdc}(a,m)=1$ [1].*

Os valores de $a \in \mathbb{Z}_{43}$ tais que $\mathrm{mdc}(a,43)=1$, são os elementos $\mathbb{Z}_{43}\backslash\{0\}$. O valor de b pode ser qualquer elemento em $\mathbb{Z}_{43}$.

Definição 4 *Suponha-se que $a \geq 1$ e $m \geq 2$ são inteiros. se $\mathrm{mdc}(a,m)=1$, então diz-se que a e m são primos relativos. O número de inteiros pertencentes a $\mathbb{Z}_{\mathrm{m}}$ que são primos relativos com m é usualmente denotado por $\phi(n)$ (função phi de Euler).*

Atente-se na função de desencriptação da cifra de deslocamento linear com o módulo $m=43$. Supondo que $\mathrm{mdc}(a,43)=1$, para desencriptar, é preciso resolver a congruência $y \equiv ax+b \pmod{43}$ em ordem a x.

Considerando a congruência $y \equiv ax+b \pmod{43}$. É equivalente a

$$ax \equiv y-b \pmod{43}.$$

Definição 5 *Supondo-que $a \in \mathbb{Z}_{\mathrm{m}}$. O inverso multiplicativo de a módulo m, denotado por $a^{-1} \bmod m$, é um elemento de $a' \in \mathbb{Z}_{\mathrm{m}}$ tal que $aa' \equiv a'a \equiv 1 \pmod{m}$. Se m é fixo é usual escrever-se a^{-1} em vez de $a^{-1} \bmod m$.*

Proposição 1 *Os elementos de $\mathbb{Z}_{\mathrm{m}}$ que tem inversos multiplicativo, são aqueles primos relativos com m, isto é, os números a para os quais existe b com $ab \equiv 1 \bmod m$ são precisamente aqueles a em que $\mathrm{mdc}(a,m)=1$ [2] .*

Se existir inverso multiplicativo ele é único módulo m. Também se observa que se $b=a^{-1}$ então $a=b^{-1}$.

Usando o algoritmo para o cálculo do inverso multiplicativo em $\mathbb{Z}_m$ calculam-se os inversos multiplicativos em $\mathbb{Z}_{43}$ [6]:

$1^{-1} = 1$	$15^{-1} = 23$
$2^{-1} = 22$	$16^{-1} = 35$
$3^{-1} = 29$	$17^{-1} = 38$
$4^{-1} = 11$	$19^{-1} = 34$
$5^{-1} = 26$	$20^{-1} = 28$
$6^{-1} = 36$	$21^{-1} = 41$
$7^{-1} = 37$	$25^{-1} = 31$
$9^{-1} = 24$	$30^{-1} = 33$
$10^{-1} = 13$	$32^{-1} = 39$
$12^{-1} = 18$	$42^{-1} = 42$
$14^{-1} = 40$	

(Todos podem ser verificados facilmente. Por exemplo, $9 \times 24 = 216 \equiv 1 (\text{mod } 43)$, então $9^{-1} = 24$ e $24^{-1} = 9$). Dado que $\text{mdc}(a, 43) = 1$, a tem inverso multiplicativo módulo 43. Multiplicando ambas as partes da congruência por a^{-1}, tem-se

$$a^{-1}(ax) \equiv a^{-1}(y - b) \text{ (mod } 43).$$

Por associatividade da multiplicação módulo 43, encontra-se

$$a^{-1}(ax) \equiv (a^{-1}a)x \equiv 1x \equiv x \text{ (mod } 43).$$

Consequentemente, $x \equiv a^{-1}(y - b) \text{ mod } 43$. Esta é a formula explicita para x.

A função de desencriptação é

$$d(y) = a^{-1}(y - b) \text{ (mod } 43).$$

Então, finalmente, obtém-se a descrição completa da cifra de deslocamento linear.

Criptosistema 2 (Cifra de Deslocamento Linear) *Sejam $\mathcal{P}$ e $\mathcal{C}$ sequência de símbolos de $\mathbb{Z}_{43}$ e seja $\mathcal{K} = \{(a, b) \in \mathbb{Z}_{43} \times \mathbb{Z}_{43} : \text{mdc}(a, 43) = 1\}$. Para $k = (a, b) \in \mathcal{K}$, define-se*

$$e_k(x) = (ax + b) \text{ mod } 43$$

$$d_k(y) = a^{-1}(y - b) \text{ mod } 43$$

com $x, y \in \mathbb{Z}_{43}$.

Seguidamente, um pequeno exemplo para ilustrar a cifra em estudo.

Exemplo 2 *Encriptar e Desencriptar a palavra "coimbra" usando a cifra de deslocamento linear.*

Supondo a chave $k = (5, 3)$, a função de encriptação é

$$e_k(x) = 5x + 3,$$

o inverso multiplicativo de $5^{-1} \mod 43 = 26$, então a correspondente função de desencriptação é

$$d_k(y) = 26(y - 3) = 26y - 35,$$

onde todas as operações são efectuadas em $\mathbb{Z}_{43}$. É fácil verificar que se está perante um criptosistema, donde $d_k(e_k(x)) = x$ para todo o $x \in \mathbb{Z}_{43}$. Calculando em $\mathbb{Z}_{43}$, obtém-se

$$\begin{aligned} d_k(e_k(x)) &= d_k(5x + 3) \\ &= 26(5x + 3) - 35 \\ &= x + 35 - 35 \\ &= x. \end{aligned}$$

Passo 1 Converter cada letra da palavra para um inteiro correspondente na codificação definida na tabela 2.2 .

c	o	i	m	b	r	a
2	14	8	12	1	17	0

Passo 2 Encriptar a palavra "coimbra" usando a função de encriptação:

$$\begin{aligned} (5 \times 2 + 3) \bmod 43 &= 13 \bmod 43 &= 13 \\ (5 \times 14 + 3) \bmod 43 &= 73 \bmod 43 &= 30 \\ (5 \times 8 + 3) \bmod 43 &= 43 \bmod 43 &= 0 \\ (5 \times 12 + 3) \bmod 43 &= 63 \bmod 43 &= 20 \\ (5 \times 1 + 3) \bmod 43 &= 8 \bmod 43 &= 8 \\ (5 \times 17 + 3) \bmod 43 &= 88 \bmod 43 &= 2 \\ (5 \times 0 + 3) \bmod 43 &= 3 \bmod 43 &= 3. \end{aligned}$$

Passo 3 Converter cada um dos inteiros, para a letra correspondente usando tabela 2.2. Assim, obtém-se a palavra encriptada "nçauicd".

Para desencriptar a palavra "nçauicd"

Passo 1 Converter cada letra da palavra para um inteiro correspondente na codificação definida na tabela 2.2 .

n	ç	a	u	i	c	d
13	30	0	20	8	2	3

Passo 2 Desencriptar a palavra "nçauicd"usando a função de desencriptação.

$$\begin{aligned}
(26 \times 13 - 35) \bmod 43 &= 303 \bmod 43 &= 2\\
(26 \times 30 - 35) \bmod 43 &= 745 \bmod 43 &= 14\\
(26 \times 0 - 35) \bmod 43 &= 8 \bmod 43 &= 8\\
(26 \times 20 - 35) \bmod 43 &= 485 \bmod 43 &= 12\\
(26 \times 8 - 35) \bmod 43 &= 173 \bmod 43 &= 1\\
(26 \times 2 - 35) \bmod 43 &= 17 \bmod 43 &= 17\\
(26 \times 3 - 35) \bmod 43 &= 43 \bmod 43 &= 0.
\end{aligned}$$

Passo 3 Converter cada um dos inteiros, para a letra correspondente usando tabela 2.2 o resultado é a palavra original "coimbra".

2.5 Cifra de Vigenère

Em ambas a cifras apresentadas até o momento, uma vez escolhida a chave, cada letra do alfabeto é substituída por uma única letra do alfabeto. Por esta razão, estas cifras são chamadas de cifras de substituição monoalfabéticas. Seguidamente, apresenta-se uma cifra de substituição polialfabética em que cada letra do alfabeto é substituída por mais do que uma letra na mensagem cifrada, sendo esta a cifra de Vigenère, a qual, toma o nome do seu autor, Blaise de Vigenère que viveu no século XVI. De seguida apresenta-se o criptosistema.

Criptosistema 3 (Cifra de Vigenère) *seja m um inteiro positivo. Define-se $\mathcal{P} = \mathcal{C} = \mathcal{K} = (\mathbb{Z}_{43})^m$. Para a chave $K = (k_1, k_2, ..., k_m)$. definimos*

$$e_K(x_1, x_2, ..., x_m) = (x_1 + k_1, x_2 + k_2, ..., x_m + k_m)$$

e

$$d_K(y_1, y_2, ..., y_m) = (y_1 - k_1, y_2 - k_2, ..., y_m - k_m)$$

onde todas as operações são fechadas em $\mathbb{Z}_{43}$.

Verifiquemos que d_K é a função inversa de e_K.

De facto, $d_K(e_K(x_1, x_2, ..., x_m)) = d_K(x_1 + k_1, x_2 + k_2, ..., x_m + k_m) = (x_1 + k_1 - k_1, x_2 + k_2 - k_2, ..., x_m + k_m - k_m) = (x_1, x_2, ..., x_m)$, $\forall (x_1, x_2, ..., x_m) \in (\mathbb{Z}_{43})^m$.

A palavra-chave para a cifra de Vigenère é uma palavra constituída por letras do alfabeto. Para encriptar uma mensagem com cifra de Vigenère é necessário:

- converte-se a palavra-chave e a mensagem a ser cifrada para uma sequência de inteiros;
- dividir a mensagem em n blocos de comprimento m, onde este é o comprimento da palavra-chave;
- "adicionar" a palavra-chave a cada bloco módulo 43;
- os blocos são concatenados formando o texto cifrado.

A cifra de Vigenère usa m cifras de deslocamento simples.

Para mostrar como funciona, apresentar-se-á um pequeno exemplo.

Exemplo 3 *Encriptar e desencriptar a mensagem "matemáticacomputação", usando a palavra-chave "mestre" de comprimento* $m = 6$.

Neste sentido, converte-se a palavra chave e a mensagem para uma sequência de inteiros de acordo com a tabela 2.2. Visto que $m = 6$ divide-se a mensagem em quatro blocos e faz-se a adição módulo 43, e no final encontra-se a mensagem encriptada.

Se, caso o comprimento da mensagem não for divisível pelo comprimento da palavra chave, usa-se somente uma parte da chave para encriptar os últimos caracteres da mensagem.

Passo 1 Converter a palavra-chave e a mensagem para uma sequência de inteiros.

m	e	s	t	r	e
12	4	18	19	17	4

m	a	t	e	m	á	t	i	c	a	c	o
12	0	19	4	12	27	19	8	2	0	2	14

m	p	u	t	a	ç	ã	o
12	15	20	19	0	30	29	14

Passo 2 Aplicar a função de encriptação.

$$\begin{array}{r|rrrrrr||rrrrrr}
 & 12 & 0 & 19 & 4 & 12 & 27 & 19 & 8 & 2 & 0 & 2 & 14 \\
+ & 12 & 4 & 18 & 19 & 17 & 4 & 12 & 4 & 18 & 19 & 17 & 4 \\
\hline
 & 24 & 4 & 37 & 23 & 29 & 31 & 31 & 12 & 20 & 19 & 19 & 18
\end{array}$$

$$\begin{array}{r|rrrrrr||rr}
 & 12 & 15 & 20 & 19 & 0 & 30 & 29 & 14 \\
+ & 12 & 4 & 18 & 19 & 17 & 4 & 12 & 4 \\
\hline
 & 24 & 19 & 38 & 38 & 17 & 34 & 41 & 18
\end{array}$$

Passo 3 Converter a sequência de inteiros, para o caracteres correspondente no alfabeto definido na tabela 2.2.

24	4	37	23	29	31	31	12	20	19	19	18
y	e	ó	x	ã	è	è	m	u	t	t	s

24	19	38	38	17	34	41	18
y	t	ô	ô	r	ì	ú	s

Como resultado, obtém-se mensagem cifrada "yeóxãèèmuttsytôôrìús".

Tal como anteriormente, para desencriptar a mensagem "yeóxãèèmuttsytôôrìús", segue-se a mesma sequência de passos e com a mesma palavra-chave "mestre", mas agora subtraindo o valor da palavra-chave módulo 43.

$$\begin{array}{rcccccccccccc} & 24 & 4 & 37 & 23 & 29 & 31 & 31 & 12 & 20 & 19 & 19 & 18 \\ - & 12 & 4 & 18 & 19 & 17 & 4 & 12 & 4 & 18 & 19 & 17 & 4 \\ \hline & 12 & 0 & 19 & 4 & 12 & 27 & 19 & 8 & 2 & 0 & 2 & 14 \end{array}$$

$$\begin{array}{rcccccccc} & 24 & 19 & 38 & 38 & 17 & 34 & 41 & 18 \\ - & 12 & 4 & 18 & 19 & 17 & 4 & 12 & 4 \\ \hline & 12 & 15 & 20 & 19 & 0 & 30 & 29 & 14 \end{array}$$

Convertendo a sequência de inteiros para os caracteres correspondente na tabela 2.2 consegue-se a mensagem que foi encriptada "matemáticacomputação".

Capítulo 3

Criptoanálise das Cifras Clássicas

Neste capítulo analisar-se-ão alguns métodos criptoanalíticos para as cifras em estudo. Sendo a criptoanálise o estudo das técnicas matemáticas para tentar comprometer as técnicas criptográficas, e mais genericamente, os serviços de segurança da informação, o objectivo de um criptoanalista é descobrir a mensagem original sem ter conhecimento da chave.

Em geral, assume-se que o criptoanalista conhece o método criptográfico usado, isto é, assume-se o Princípio de Kerckhoffs, que defende que o criptoanalista tem toda a informação sobre o criptosistema usado, com excepção da chave secreta, sendo que a segurança do método se baseia totalmente no desconhecimento da chave. Decididamente, se o criptoanalista não conhecer o método criptográfico que foi usado, a sua tarefa torna-se mais difícil [3, 8].

Neste trabalho são explorados dois métodos criptoanalíticos: o método de Procura Exaustiva no Espaço das Chaves, e o método por Análise de Frequências. Na tabela 3.1 temos a frequência relativa das letras do alfabeto definido [4].

Nota: O modelo de ataque usado neste trabalho é o **ataque de texto cifrado**. Trata-se de um ataque onde o criptoanalista tenta deduzir a chave de encriptação ou o texto claro por análise unicamente do texto encriptado.

3.1 Método Criptoanalítico por Procura Exaustiva no Espaço das Chaves

O método criptoanalítico por procura exaustiva no espaço das chaves consiste em testar todas as chaves possíveis, usualmente referido como um ataque de "Força Bruta". Este é um método de ataque que só é possível realizar quando a dimensão do espaço das chaves é pequena.

3.1.1 Dicionário de Verificação

Como foi referido na secção 2.2, os caracteres que não fazem parte do alfabeto mantém-se intactos, o carácter espaço faz parte deste leque de caracteres, pois permite

letra	frequência	letra	frequência	letra	frequência
a	13,579356	p	2,429165	ç	0,444995
b	1,065210	q	1,198975	è	0,001199
c	3,282540	r	6,735379	é	0,392201
d	5,093185	s	7,877891	ê	0,120443
e	12,208640	t	4,180165	ì	0,000099
f	0,985004	u	4,411459	í	0,156633
g	1,213487	v	1,704822	ò	0,000514
h	1,489500	w	0,007734	ó	0,160987
i	5,732058	x	0,216702	ô	0,041455
j	0,369693	y	0,034694	õ	0,068893
k	0,007382	z	0,470044	ù	0,000045
l	3,092033	à	0,087362	ú	0,046214
m	4,657645	á	0,402305	ü	0,001514
n	4,915297	â	0,025176		
o	10,348816	ã	0,742794		

Tabela 3.1: Frequência de cada letra do alfabeto.

distinguir num texto uma palavra, sendo, deste modo, a mensagem mais fácil de ler.

Para o método procura exaustiva no espaço das chaves utilizou-se um dicionário com 687.842 palavras, que permite verificar se as palavras do texto decifrado existem ou não na língua em causa, neste caso a Língua Portuguesa.

Neste trabalho considerou-se que um texto decifrado com um ajuste de mais de 70% em relação às palavras existente no dicionário é uma forte candidata para ser a chave de encriptação.

De acordo com a definição do alfabeto, como mostra a tabela 2.1, temos o seguinte:

- **Cifra de Deslocamento Simples** - O número das chaves possíveis para a Cifra de Deslocamento Simples é $|\mathcal{K}| = 43$, sendo o número de possibilidades pequeno, tornando fácil a aplicação do método criptoanalítico;

- **Cifra de Deslocamento Linear** - Sendo o número das chaves para a Cifra de Deslocamento Linear $|\mathcal{K}| = \phi(43) * 43 = 42 * 43 = 1806$, onde a função $\phi()$ denota o número de inteiros que são primos relativos com 43, o número de possibilidades é já maior que a cifra anterior, mas, no entanto, ainda é fácil de quebrar se se considerar o uso de um computador;

- **Cifra de Vigenère** - O número de possibilidades das palavras chave para a Cifra de Vigenère é $|\mathcal{K}| = 43^m$, onde m é o comprimento da palavra chave. Mesmo para um valor de m pequeno a procura exaustiva requer muito tempo. Por exemplo, se $m = 4$, o número de possibilidades das palavras chave excede $3,4 \times 10^6$. Isto já é realmente um número bastante grande, mas com o uso do computador ainda é possível quebrar a cifra.

3.2 Método Criptoanalítico por Análise de Frequências

Um outro método utilizado na análise das mensagens encriptadas pelos métodos já referidos, é o método criptoanalítico por análise de frequência.

As cifras em estudo são, respectivamente, duas cifras de substituição monoalfabéticas e uma cifra de substituição polialfabética, que consistem em substituir uma letra por outra. E, com base nisso, é possível fazer a criptoanálise, fundamentando-se, no cálculo das frequências relativas das letras, digramas, trigramas, primeiras e últimas letras de uma palavra, palavras curtas e índice de coincidência, já que são características de uma dada língua.

As ocorrências das letras reflectem como um povo utiliza a sua língua, caracterizando-a de forma única. O criptoanalista pode então usar este conhecimento único para analisar as mensagem encriptadas. Por exemplo, tem-se que se um 'a' for substituído por um 'd' as características próprias da letra 'a' na Língua Portuguesa (a sua frequência relativa, etc) vão ficar inalteradas, no entanto escondidas na letra 'd'.

O estudo comparativo dos valores encontrados para a Língua Portuguesa e dos valores encontrados no texto da mensagem cifrada vai permitir um emparelhamento entre letras, e dessa forma obter a chave secreta, quebrando, assim, a cifra.

3.2.1 Criptoanálise da Cifra de Deslocamento Simples

Uma simples ilustração de como é possível aplicar o método criptoanalítico por análise frequência para a cifra de deslocamento simples usando:

- Frequência Relativa das Letras;
- Digramas;
- Trigramas;
- Palavras Curtas;
- Letras Iniciais de Palavras;

- Letras Finais de Palavras.

Os valores de referências para a Língua Portuguesa usados na aplicação do método acima, podem ser encontrados em [4].

Exemplo 4 *Mensagem encriptada, a partir da Cifra de Deslocamento Simples*

d fulswrjudild suí-frpsxwdflrqdo hud irupdgd sru xp frqmxqwr gh píwrgrv gh vxevwlwxlêér h wudqvsrvlêér grv fdudfwhuhv gh xpd phqvdjhp txh sxghvvhp vhu hàhfxwdgrv pdqxdophqwh shor hplvvru h shor ghvwlqdwçulr gd phqvdjhp. r vxujlphqwr gh pçtxlqdv hvshfldolâdgdv h, srvwhulruphqwh, grv frpsxwdgruhv rfdvlrqrx xpd vljqlilfdwlyd hyroxêér gdv wífqlfdv fulswrjuçilfdv.

Frequência Relativa das Letras

De acordo com a tabela 3.1 temos as 3 primeiras letras mais frequentes da Língua portuguesa por ordem decrescente: “a”, “e”, “o”.

As letras mais frequente na mensagem encriptada foram:

Letra	h	d	r
Frequência	34	31	31

Tem-se então como chaves candidatas $k = 7$, $k = 3$, $k = 36$:

Letra	a	e	o
Letra - h	7	3	36

Agora, a partir das três chaves candidatas, podia-se prosseguir tentando exaustivamente encontrar as chaves e depois descobrir a chave correcta.

Para melhor ilustrar este método continuar-se-á com as outras medidas já referidas.

Digramas

Os 4 digramas mais frequentes na Língua Portuguesa por ordem decrescente: “de”, “ra”, “os”, “es”.

Na mensagem encriptada existem como digramas mais frequentes os seguintes: “gh”, “rv”, “dv”, “hq”, “wr”, “fd”.

Como chaves possíveis aparecem então:

Digramas	de	ra	os	es
gh	(3,3)	(32,7)	(35,32)	(2,32)
rv	(14,17)	(0,21)	(3,3)	(13,3)
dv	(0,17)	(29,21)	(32,3)	(42,3)
hq	(4,12)	(33,16)	(36,41)	(3,41)
wr	(19,13)	(5,17)	(8,42)	(18,42)
fd	(2,42)	(31,3)	(34,28)	(1,28)

As chaves candidatas são os pares em que o primeiro elemento é igual ao segundo elemento. Assim neste caso temos como chave candidata $k = 3$. O estudo prosseguirá com os trigramas.

Trigramas

Os 4 trigramas mais frequentes na Língua Portuguesa por ordem decrescente são os seguintes: "que", "ent", "nte", "com".

Na mensagem encriptada os trigramas mais frequentes são : "phq", "grv".

Assim, temos como chaves possíveis:

Trigramas	que	ent	nte	com
phq	(42,30,12)	(11,37,40)	(2,31,12)	(13,36,4)
grv	(33,40,17)	(2,4,2)	(36,41,17)	(4,3,9)

As chaves candidatas são os ternos em que todos os elementos são iguais. Para este exemplo concreto não foi possível obter nenhuma informação acerca do estudo dos trigramas.

Palavras Curtas

São palavras cujo comprimento está abaixo do comprimento médio das palavras na Língua Portuguesa:

1. Palavras de uma letra

   As 3 palavras curtas de uma letra mais frequente na Língua Portuguesa por ordem decrescente: "a", "o", "e".

   O texto encriptado contém três palavras com uma letra: "h", "r", "d".

   Temos então como chaves possíveis:

Comprimento = 1	a	o	e
h	7	36	3

2. Palavras de duas letras

   As 6 palavras curtas de duas letras mais frequente na Língua Portuguesa por ordem decrescente: "de", "um", "do", "da", "os", "as".

   O texto encriptado contém as seguintes palavras com duas letras: "gh", "xp", "gd".

   Neste caso, há correspondência de um para outro par, essa correspondência só fará sentido se para ambos os elementos do par a chave correspondente for a mesma.

Comprimento = 2	de	um	do	da	os	as
gh	(3,3)	(29,38)	(3,36)	(3,7)	(35,32)	(6,32)
xp	(20,11)	(3,3)	(20,1)	(20,15)	(9,40)	(23,40)
gd	(3,42)	(29,34)	(3,32)	(3,3)	(35,28)	(6,28)

3. Palavras de três letras

   As seguintes são as 6 palavras curtas de três letras mais frequente na Língua Portuguesa, por ordem decrescente: "que", "com", "uma", "não", "por", "dos".

   A mensagem encriptada contém as seguintes palavras de três letras: "grv", "xpd", "sru", "suí", "txh", "vhu", "hud", "gdv".

   Como chaves possíveis temos:

Comprimento = 3	que	com	uma	não	por	dos
grv	(33,40,17)	(4,3,9)	(29,5,21)	(36,31,7)	(34,3,4)	(3,3,3)

   Os ternos válidos são os ternos em que todos os elementos são iguais.

Com o estudo das palavras curtas tem-se como chave candidata $k = 3$.

Letras Iniciais

As 3 letras iniciais mais frequentes na Língua Portuguesa por ordem decrescente são: "d", "c", "p".

As letras iniciais mais frequentes na mensagem encriptada: "g", "s".

Como chaves candidatas encontram-se as seguintes:

L. Iniciais	d	c	p
g	3	4	34
s	15	16	3

Letras Finais

As 3 letras finais mais frequentes na Língua Portuguesa são, por ordem decrescente: "a", "o", "s".

A letra final mais frequente na mensagem encriptada é: "v".

Como chaves candidatas verificam-se as seguintes:

L. Finais	a	o	s	e
v	21	7	3	17

De acordo com o estudo global feito tudo indica que, a chave usada na encriptação da mensagem foi a chave 3. Verificar-se que assim é.

Mensagem desencriptada com a chave 3

> "a criptografia pré-computacional era formada por um conjunto de métodos de substituição e transposição dos caracteres de uma mensagem que pudessem ser executados manualmente pelo emissor e pelo destinatário da mensagem. o surgimento de máquinas especializadas e, posteriormente, dos computadores ocasionou uma significativa evolução das técnicas criptográficas".

Pode-se concluir que a chave 3 é a chave correcta, porque conseguiu-se obter um texto com significado em Português.

3.2.2 Criptoanálise da Cifra de Deslocamento Linear

Será aplicado o método criptonalítico por análise de frequências para a cifra de Deslocamento Linear, fazendo um estudo da frequência relativa das letras na mensagem encriptada.

Exemplo 5 *Mensagem encriptada, a partir da Cifra de Deslocamento Linear*

j tigúsòõijìgj úiù-tòàúxsjtgòèjv ãij ìòiàjyj úòi xà tòèlxèsò yã àùsòyòn yã nxonsgsxgçzò ã sijènúòngçzò yòn tjijtsãiãn yã xàj àãènjõãà dxã úxyãnnãà nãi ãôãtxsjyòn àjèxjvàãèsã úãvò ãàgnnòi ã úãvò yãnsgèjspigò yj àãènjõãà. ò nxiõgàãèsò yã àpdxgèjn ãnúãtgjvgfjyjn ã, úònsãigòiàãèsã, yòn tòàúxsjyòiãn òtjngòèòx xàj ngõègìgtjsgâj ãâòvxçzò yjn sùtègtjn tigúsòõipìgtjn.

Na tabela 3.2 obtém-se a frequência relativa de cada letra da mensagem encriptada. As letras mais frequente são: ã (34 ocorrências), ò, j (31 ocorrências), n (28 ocorrências), g (22 ocorrências), s (19 ocorrências), à, i (18 ocorrências), x (16 ocorrências cada), è,

letra	frequência	letra	frequência
ã	34	v	6
ò	31	õ	6
j	31	ì	4
n	28	ç	3
g	22	ù	3
s	19	z	3
i	18	p	3
à	18	d	2
x	16	â	2
y	15	l	1
è	15	f	1
t	15	ô	1
ú	12	o	1

Tabela 3.2: Frequência da ocorrência das letras no texto cifrado

t, y (15 ocorrências), ú (12 ocorrências), v, õ (6 ocorrências cada), ì (4 ocorrências), ç, ù, z, p (3 ocorrências), d, â(2 ocorrências cada), l, f, ô, o (1 ocorrências cada).

Como uma primeira tentativa, pode-se supor que "ã" é a encriptação de "a" e "ò" é a encriptação de "e", visto que "a" e "e" são respectivamente as duas letras mais frequente na Língua Portuguesa.

Numericamente, expressando, tem-se que $e_k(0) = 29$ e $e_k(4) = 36$. Relembrando que $e_k(x) = ax + b$, onde a e b são duas incógnitas, é possível obter duas equações lineares com duas incógnitas:

$$\begin{aligned} 0a + b &= 29 \\ 4a + b &= 36 \end{aligned}$$

Este sistema tem uma solução única $a = 34$, $b = 29$ em $\mathbb{Z}_{43}$. O valor de a tem de ser primo relativo com 43, isto é, $\text{mdc}(a, 43) = 1$, para garantirmos que a tem inverso multiplicativo em $\mathbb{Z}_{43}$.

No próximo ensaio "ã" é a encriptação de "a" e "j" é a encriptação de "e". Seguindo o procedimento anterior, obtinha-se $a = 38$ e $b = 29$. Continuando com os testes "ã" é encriptação de "a" e "n" é encriptação de "e", até aqui tem-se mantido sempre os mesmos valores para a primeira equação, enquanto os da segunda equação foram alterando.

No, próximo teste "ò" é encriptação de "a" e "ã" é encriptação de "e", mantendo desta vez os valores da primeira equação ao mesmo tempo que se alteraram os valores da segunda equação até chegar a tentativa "ò" que é encriptação de "a" e "o" é encriptação de "e".

Fazendo sucessivamente como acima para todas as restantes letras da tabela 3.2, até chegarmos à última tentativa "o" é a encriptação de "a" e "ô" é a encriptação de "e".

O par de chaves com a forte possibilidade de ser a chave candidata é o par que consegue descobrir o maior número de palavras na língua usada, neste caso, a Língua Portuguesa.

Para o exemplo em estudo, o par de chaves usado na encriptação da mensagem foi $(a, b) = (5, 9)$, com a tentativa "j" é encriptação de "a" e "ã" é encriptação de "e".

Veja-se, em seguida, a mensagem desencriptada com o par de chave (5,9):

> "a criptografia pré-computacional era formada por um conjunto de métodos de substituição e transposição dos caracteres de uma mensagem que pudessem ser executados manualmente pelo emissor e pelo destinatário da mensagem. o surgimento de máquinas especializadas e, posteriormente, dos computadores ocasionou uma significativa evolução das técnicas criptográficas."

conclui-se, deste modo, que a chave correcta é o par (5,9).

3.2.3 Criptoanálise da Cifra de Vigenère

A cifra de Vigenère é uma cifra de substituição polialfabética, ao contrário das outras duas cifras, que são cifras de substituição monoalfabéticas, em que o método criptoanalítico se aplica calculando a frequência relativa (letras, digramas, trigramas,...), do texto encriptado, mas para a cifra de Vigenère não é possível aplicar esta técnica.

Apresenta-se, em seguida, uma das técnicas para quebrar a cifra de Vigenère com os seguintes passos:

O primeiro passo é determinar o comprimento da palavra-chave, denotado m. O comprimento da chave vai ser determinado através do índice de coincidência, tendo sido o conceito definido por William Friedman em 1920 [6].

Definição 6 ***(Índice de Coincidência)*** *Supondo que $\overline{x} = x_1x_2 \ldots x_n$ é um texto de n caracteres do alfabeto. O índice de coincidência de $\overline{x}$, denotado por $I_c(\overline{x})$, é definido como a probabilidade de dois caracteres aleatórios de $\overline{x}$ serem iguais.*

Supondo que $f_0, f_1, \cdots, f_{42}$ são as frequências dos 43 caracteres no alfabeto, para um dado texto de comprimento n. Podem ser escolhidos dois caracteres diferentes de $\overline{x}$ em $\binom{n}{2}$, para cada $0 \leq i \leq 43$, existe $\binom{f_i}{2}$ formas diferentes de escolher ambos os caracteres para i.

Daí resulta a formula:

$$I_c(\overline{x}) = \frac{\sum_{i=0}^{43}\binom{f_i}{2}}{\binom{n}{2}} = \frac{\sum_{i=0}^{43} f_i(f_i-1)}{n(n-1)}$$

Para a Língua Portuguesa o valor do índice de coincidência é de 0,072723 [4].

Esta técnica será ilustrada com o seguinte exemplo

Exemplo 6 *Mensagem encriptada, a partir da Cifra de Vigenère*

> i jrqààooâhfql pza-cwxwuáljiwyhl pya qvrulka àvr èt kzujâyào ol uaàolzz lp sâmztqçáiô ùv m àriyzpwãpçóz dwã ciâhcápyeà ke èta xlnàlneu xum wulpzsmx smâ eèpjuálkoà tavèhluputm wetz eutzswâ e àllw keàçpniçìrqz di tevãhgmx. w zuzrpmmyào ol uôxuqyhs pzpmnpattéallz m, pwãàeztvruputm, dwã cwxwuálkozpz wnhsqzuoâ ámi zioypfqnhtqéh mévlâúòo ohs çõcvtjaà jrqààooâìfqnhs.

Calculem-se os valores dos índices de coincidência:

- $m = 1$, calcule-se o índice coincidência da mensagem completa, $I_c = 0{,}035244$;
- $m = 2$, divide-se a mensagem em dois sub-textos, o primeiro sub-texto composto por caracteres começado no primeiro carácter da mensagem, somando sempre mais duas posições, até ao final da mensagem e o segundo sub-texto composto por caracteres começado no segundo carácter da mensagem, somando sempre mais duas posições, até ao final da mensagem. Assim:

  sub-texto 1 ijqàohqzwwájwhyvukvtkuààluàlzlâzqáôvmàizwpówiháyàktlànuxm wl zmmèjákàtvhuumwtuzwlwkàpiìqitvhmwzzpmàluxqhzmptélzmwàzvuumw wwákzzwhquâáizopqhqhmvâòhõvjàjqàoìqh

  sub-texto 2 ràoâflpacxuliylpaqrlaàrèzjyooaozpsmtçiùrypãçzdãcâcpeeèaxnleuups xsâepuloaèlptezetsâeàleçnçrzdeãgxurmyooôuysppnatalpãetrptdãcxulopnszo miyfntéélúoosçctaràoâfns

  1. $I_c = 0{,}052627$,
  2. $I_c = 0{,}041522$;

- $m = 3$, divide-se a mensagem em três sub-textos de forma análoga ao já descrito acima.

  sub-texto 1 iràâqpxáihprkàuyàzlszçôàywçãicyalluuwpmspáotèutwzusàwepçqdt ãmzrmolôqszntazwevpmdxáoznqoáofhémlòocjqoìn

  sub-texto 2 qohlzcwlwlyquavèkjàouozpâtáùmrzãódâáekènnmuzxmejlàahpme twelànìziehxwupyxypptlãzruwcwlzhzâmzyqthéâohçvajàofh

  sub-texto 3 jàofawujyavlrtzâolalmqivippzwchpàetxàexlsâèukvlutezâlkçirvg zmàouuhpmaélmpàtutãwukpwsuiipnqvúsõtàràâqs

  1. $I_c = 0{,}030674$;
  2. $I_c = 0{,}033327$;
  3. $I_c = 0{,}035644$;

- $m = 4$, divide-se a mensagem em quatro sub-textos de forma análoga ao já descrito acima.

  sub-texto 1 jàofawujyavlrtzâolalmqivippzwchpàetxàexlsâèukvlut ezâlkçirvgzmàouuhpmaélmpàtutãwukpwsuiipnqvúsõtàràâqs

  sub-texto 2 àâlaxlypqlàèzyoazpmçùyãzãâpèxlpxâplèpztâàççzãxry oôypntlãtpãxlpnzynééúoçtàân

  sub-texto 3 jàhwjhyvkvtuàlàzzávàzphyktlnxwzjkthuwzlkpìthzpàl xhzpézàvuwkzhuázphhvòhõjjàìh

  sub-texto 4 rofpcuilararjoostirpçdcceeaneuusseuoalteeselenrd egumouspaapertdcuosomiftloscarofs

  1. $I_c = 0{,}077256$;
  2. $I_c = 0{,}053694$;
  3. $I_c = 0{,}066316$;
  4. $I_c = 0{,}065123$;

Calculando as médias dos índices coincidência para $m = 1$ é 0,035244, $m = 2$ obtém-se 0,047075, $m = 3$ é 0,033215 e $m = 4$ obtém-se 0,065597. De acordo com as médias dos índices coincidência assume-se que o comprimento da palavra-chave é 4.

Nota:

O m válido é um m cuja a média é aproximadamente igual a 0,065. Caso não se tenha atingido esse valor escolhe-se o m com o maior valor.

Supondo que se determina correcto o valor de m, como encontrar a palavra-chave, $K = (k_1, k_2, ..., k_m)$?

Descreve-se agora um método simples e eficaz. Sejam $f_0, ..., f_{42}$ as frequências dos 43 caracteres do alfabeto no sub-texto Y_i e $n' = n/m$ o comprimento de Y_i, com $1 \leq i \leq m$.

Então a distribuição da probabilidade dos 43 caracteres em Y_i é:

$$\frac{f_0}{n'}, \cdots, \frac{f_{42}}{n'}.$$

Tendo em conta que em Y_i todos os seus elementos foram encriptados pelo mesmo elemento, k_i, da chave de encriptação, falta agora descobrir o seu valor exacto.

Para o valor exacto de k_i é de esperar que a distribuição de probabilidade com esse deslocamento

$$\frac{f_{k_i}}{n'}, \cdots, \frac{f_{42+k_i}}{n'}$$

seja aproximadamente igual a distribuição de probabilidade $p_0, ..., p_{42}$ para a língua de referência (no caso presente o Português).

Tabelando todos os valores possíveis para o deslocamento

$$Mg = \sum_{i=0}^{42} \frac{p_i f_{i+g}}{n'}, \quad com \quad 0 \leq g \leq 42$$

Tendo-se que para $g = k_i$ ter-se-á que:

$$M_g \approx \sum_{i=0}^{42} p_i^2 = 0,072$$

valor dado pelo índice de coincidência da Língua Portuguesa.

Exemplo 6 (**Continuação**) Assumindo que o comprimento da palavra-chave é 4, calcula-se os valores M_g como se descreve acima, para $1 \leq i \leq 4$. Estes valores estão na tabela 3.3, para cada i, tem-se o valor de M_g. Estes $g's$ determinam os deslocamentos de $k_1, \cdots, k_4$.

De acordo com os dados na tabela 3.3, vê-se que a chave mais provável é $K = (8, 11, 7, 0)$, à qual corresponde a palavra-chave "ilha".

Assim, obtém-se a mensagem desencriptada, com a palavra-chave "ilha".

> "a criptografia pré-computacional era formada por um conjunto de métodos de substituição e transposição dos caracteres de uma mensagem que pudessem ser executados manualmente pelo emissor e pelo destinatário da mensagem. o surgimento de máquinas especializadas e, posteriormente, dos computadores ocasionou uma significativa evolução das técnicas criptográficas."

Concluiu-se deste modo que a palavra-chave está correcta.

i	Valores de $M_g(Y_i)$										
1	.024	.020	.033	.027	.040	.027	.023	.027	$\boxed{.063}$	.024	.023
	.024	.046	.015	.022	.010	.035	.017	.025	.015	.019	.014
	.023	.004	.003	.010	.004	.002	.001	.003	.004	.003	.004
	.009	.012	.011	.013	.026	.018	.016	.020	.028	.022	
2	.021	.022	.016	.019	.016	.023	.022	.035	.020	.024	.025
	$\boxed{.051}$	.023	.023	.012	.031	.014	.015	.011	.026	.022	.023
	.016	.025	.028	.025	.007	.007	.010	.012	.009	.006	.005
	.009	.018	.012	.018	.020	.029	.022	.027	.020	.027	
3	.012	.023	.023	.038	.026	.034	.033	$\boxed{.061}$	.029	.026	.023
	.038	.017	.018	.013	.021	.016	.025	.016	.020	.019	.029
	.018	.011	.006	.020	.002	.004	.002	.005	.010	.010	.018
	.018	.016	.016	.026	.015	.020	.020	.017	.014	.018	
4	$\boxed{.069}$	.041	.039	.035	.043	.028	.030	.022	.020	.018	.024
	.025	.018	.020	.034	.022	.018	.019	.019	.009	.014	.002
	.004	.005	.007	.011	.013	.014	.016	.024	.019	.018	.016
	.023	.022	.020	.017	.023	.021	.037	.034	.032	.034	

Tabela 3.3: Valores de M_g

Aplicabilidade do Método a Casos Práticos

O método criptoanalítico por análise de frequência para a cifra de Vigenère, em termos teóricos, foi descrito acima, mas a sua aplicabilidade a casos práticos exige alguns pré-requisitos para que se possa obter um resultado concreto. O estudo é apresentado nas figuras 3.1 e 3.2.

Para o primeiro estudo em que se verificou como se reagia o método, com palavaras-chaves constituídas, nomeadamente, por caracteres diferentes, utilizaram-se 12 textos com diferentes tamanhos em palavras, o menor texto com 10 palavras e o maior texto com 10000 palavras e palavras-chaves com comprimentos diferentes entre 1 e 7.

A conclusão que se tirou deste estudo é que o método para descobrir o comprimento da palavras-chave, garante que o comprimento desta está correcto num texto de 30 palavras no mínimo, caso contrário o método pode falhar.

Em relação à palavra-chave pode-se assim concluir que quanto maior for o texto maior a probabilidade de descobrir a palavra-chave correcta. Com um texto de 30 palavras tem-se mais de 40% de hipótese de encontrar a palavra-chave correcta.

Notas:

- Há que ter em conta também o texto em estudo. Quando o texto for composto por caracteres mais frequentes da Língua Portuguesa, o método reage melhor.
- Nas palavras-chave constituidas por caracteres de a-z o método reage melhor do que quando contém caracteres acentuados ou carácter cedilhado de acordo com as implementações feitas.

Na figura 3.2 fez-se um estudo de palavras-chaves com caracteres repetido, o que se pode concluir que o método, não consegue em muitos casos descobrir correctamente o comprimento da chave, o que torna difícil descobrir a palavra chave correcta.

3.2.4 Estudo Comparativo Entre os Métodos Criptoanalíticos

O estudo que se apresenta a seguir foi efectuado sob as mesmas condições computacionais: sistema GNU/Linux 2.6.22; Intel Pentium 4 a 3,0GHZ; 2GiB RAM. Os tempos referem-se ao tempo gasto pelo processador e estão em segundos.

Cifra Deslocamento Simples

De acordo com o gráfico da figura 3.3 vê-se que a cifra é insegura, aplicando os dois métodos para descobrir a chave de encriptação, demora-se pouco. O tempo aumenta de acordo com o tamanho do texto. Dos dois métodos, o método criptoanalítico por análise de frequência é o mais rápido, conseguindo descobrir a chave mais rapidamente.

Cifra Deslocamento Linear

Analisando o gráfico da figura 3.4, a criptoanálise da cifra Deslocamento Linear já demora mais tempo em relação a cifra de Deslocamento Simples, mas mesmo assim ainda é fácil de a quebrar. Comparando os dois métodos criptoanalíticos pode-se concluir que para esta cifra o método criptoanalítico por procura exaustiva demora menos tempo, mas o tempo aumenta de acordo com o tamanho do texto.

Cifra de Vigenère

Fazendo uma comparação entre os métodos criptoanalíticos de acordo com o gráfico da figura 3.5 o método por procura exaustiva demora muito tempo para descobrir a palavra-chave e quanto maior for o comprimento da palavra-chave mais tempo ainda demora. Para esse estudo utilizou-se uma palavra-chave de comprimento 3. Enquanto

com o método por análise frequência consegue-se descobrir a palavra-chave muito mais rápido.

Todos os métodos criptográficos em estudo são inseguros, mas fazendo uma escolha, de acordo com o método criptoanalítico que o criptoanalista usar obtém-se:

- Por um lado se o criptoanalista utilizar o método por procura exaustiva, o melhor método criptográfico é a cifra de Vigenère se se optar por usar uma palavra-chave com comprimento razoável;

- Por outro se o criptoanalista utilizar o método por análise de frequência, o melhor método criptográfico é a cifra de Deslocamento Linear.

C. Chave	P. Chave	10 Palavras		20 Palavras		30 Palavras		40 Palavras		50 Palavras		100 Palavras		1000 Palavras		5000 Palavras		10000 Palavras	
		Comp.	P. chave	Comp.	P. chave	Comp.	P. chave	Comp.	P. chave	Comp.	P. chave	Comp.	P. chave	Comp.	P. chave	Comp.	P. chave	Comp.	P. chave
k =1	z	1	z	1	z	1	z	1	z	1	z	1	z	1	z	1	z	1	z
	p	1	p	1	p	1	p	1	p	1	p	1	p	1	p	1	p	1	p
	é	1	é	1	é	1	é	1	é	1	é	1	é	1	é	1	é	1	é
	ã	1	ã	1	ç	1	è	1	ã	1	í	1	ç	1	í	1	í	1	ã
k =2	b	1	b	2	bq	2	bu	2	bu	2	bu	2	bu	2	bu	2	bu	2	bu
	r	1	r	2	rí	2	rí	2	rí	2	rí	2	rí	2	rí	2	rí	2	rí
	vê	1	r	2	vê	2	vç	2	vç	2	vç	2	vê	2	vê	2	vê	2	vê
	ás	1	t	2	ás	2	ás	2	ás	2	ás	2	ás	2	ás	2	ás	2	ás
k=3	vou	1	r	3	voq	3	vou	3	vou	3	vou	3	voq	3	vou	3	vou	3	vou
	que	3	ibue	3	qqe	3	que	3	que	3	que	3	que	3	que	3	que	3	que
	vós	3	vás	3	vós	3	vçs	3	vós	3	vís	3	vçs	3	vós	3	vós	3	vós
	céu	3	béu	3	céq	3	céu	3	céu	3	céu	3	céq	3	céu	3	céu	3	céu
k=4	bola	2	bd	4	bola	4	bola	4	bola	2	ba	4	bola	4	bola	4	bola	4	bola
	zona	2	zd	4	vona	4	zora	4	zona	4	zona	4	zona	4	zona	4	zona	4	zona
	ligo	2	lr	4	ligê	4	ligo	4	ligo	4	ligs	4	liio	4	ligo	4	ligo	4	ligo
	água	2	ud	4	água	4	ãgga	4	ãgga	4	ãgua	4	ãgua	4	ãgua	4	ãgua	4	ãgua
k=5	cinto	5	cinxo	5	cinrs	5	cinto	5	ciôto	5	cinto	5	cinto	5	cinto	5	cinto	5	cinto
	turma	5	tqua	5	tqrôa	5	turma	5	tqrma	5	turôa	5	turma	5	turma	5	turma	5	turma
	êxito	5	êxixo	5	êxirs	5	êxito	5	êxito	5	êãito	5	êxito	5	êxito	5	êxito	5	êxito
	união	5	uniço	5	qniãs	5	união	5	wiio	5	qriéo	5	união	5	uniço	5	uniio	5	uniio
k=6	rufino	2	fl	6	rqfiôq	6	rqfino	6	tufirs	6	rufino	6	rufins	6	rufino	6	rufino	6	rufino
	bicos	2	ir	6	biicos	6	bicos	6	bicos	6	bicos	6	bicos	6	bicos	6	bicos	6	bicos
	ilhéus	2	ho	3	élh	6	ilhéus	6	ilhéus	6	ilhéus	6	ilhéus	6	ilhéus	6	ilhéus	6	ilhéus
	sétima	2	tê	6	sética	6	ãétia	6	sétia	6	sétima	6	sétima	6	sétima	6	sétima	6	sétima
k=7	benfica	11	acmiodcfcfue	7	benfica	7	berfica	7	berfica	7	benfica	7	benfica	7	benfica	7	benfica	7	benfica
	coimbra	11	ôcmyvdahmiik	7	coiybra	7	coiqbra	7	coimbra	7	coimbra	7	coimbra	7	coimbra	7	coimbra	7	coimbra
	hortelã	11	üzedéthôée	7	horaelé	7	hortelã	7	hovtelã	7	hortelã	7	hortelã	7	hortelã	7	hortelã	7	hortelã
	furacão	11	cqcfqinmrod	7	furacár	7	furacas	7	furaçys	7	furacis	7	furaçys	7	furacio	7	furacio	7	furacão
%		39%	11%	96%	29%	100%	43%	100%	57%	96%	36%	100%	57%	100%	71%	100%	71%	100%	79%

Figura 3.1: Testes para o Método Criptoanalítico por Análise Frequência - Cifra Vigenère

C. Chave	P. Chave	10 Palavras		20 Palavras		30 Palavras		100 Palavras		1000 Palavras		5000 Palavras		10000 Palavras	
		Comp.	P. chave	Comp.	P. chave	Comp.	P. chave	Comp.	P. chave	Comp.	P. chave	Comp.	P. chave	Comp.	P. chave
k=3	vvv	1	v	1	v	1	v	1	v	1	v	1	v	1	v
	vvo	1	k	3	vvo	3	vvo	3	vvs	3	vvo	3	vvo	3	vvo
k=4	bbbb	1	b	1	b	1	b	1	b	1	b	1	b	1	b
	bobo	1	b	2	bo	2	bo	2	bo	2	bo	2	bo	2	bo
	bboo	2	br	4	bboê	4	bboo	4	bboo	4	bboo	4	bboo	4	bboo
	bbbo	1	b	2	bb	4	bbbo	2	bb	4	bbbo	2	bb	2	bb
k=5	ççççç	1	ç	1	ç	1	ç	1	ì	1	ç	1	ç	1	ç
	çaçaç	2	íí	5	iaçaç	5	iaçaç	5	ìaìaì	5	çaçaç	5	çaçaç	5	çaçaç
	ççaç	1	i	2	iç	3	çiç	1	i	5	ççaç	5	ççaç	5	ççaç
	çaçça	2	çd	5	iaçça	5	iaçça	5	ìaììa	5	çaçça	5	çaçça	5	çaçça
k=6	ssssss	1	s	1	s	1	s	1	s	1	s	1	s	1	s
	sssssas	1	s	1	s	2	ss	2	ss	3	sas	2	ss	2	ss
k=7	kkkkkkk	1	k	1	k	1	k	1	k	1	k	1	k	1	k
	kakekik	2	kũ	7	kakekik	7	kaôekkik	7	kakekik	7	kakekik	7	kakekik	7	kakekik
	akkkkkb	1	k	2	kk	7	akôokkb	7	akkkkkb	7	akkkkkb	7	akkkkkb	7	akkkkkb
	líílílí	2	l	7	ũãlglãg	7	lãôalgã	7	lãlãlãí	7	lãlãlãã	7	lãlãlãã	7	lãlãlãã
%		0%	31%	38%	50%	50%	56%	44%	44%	56%	88%	50%	81%	50%	81%

Figura 3.2: Testes para o Método Criptoanalítico por Análise Frequência - Cifra Vigenère

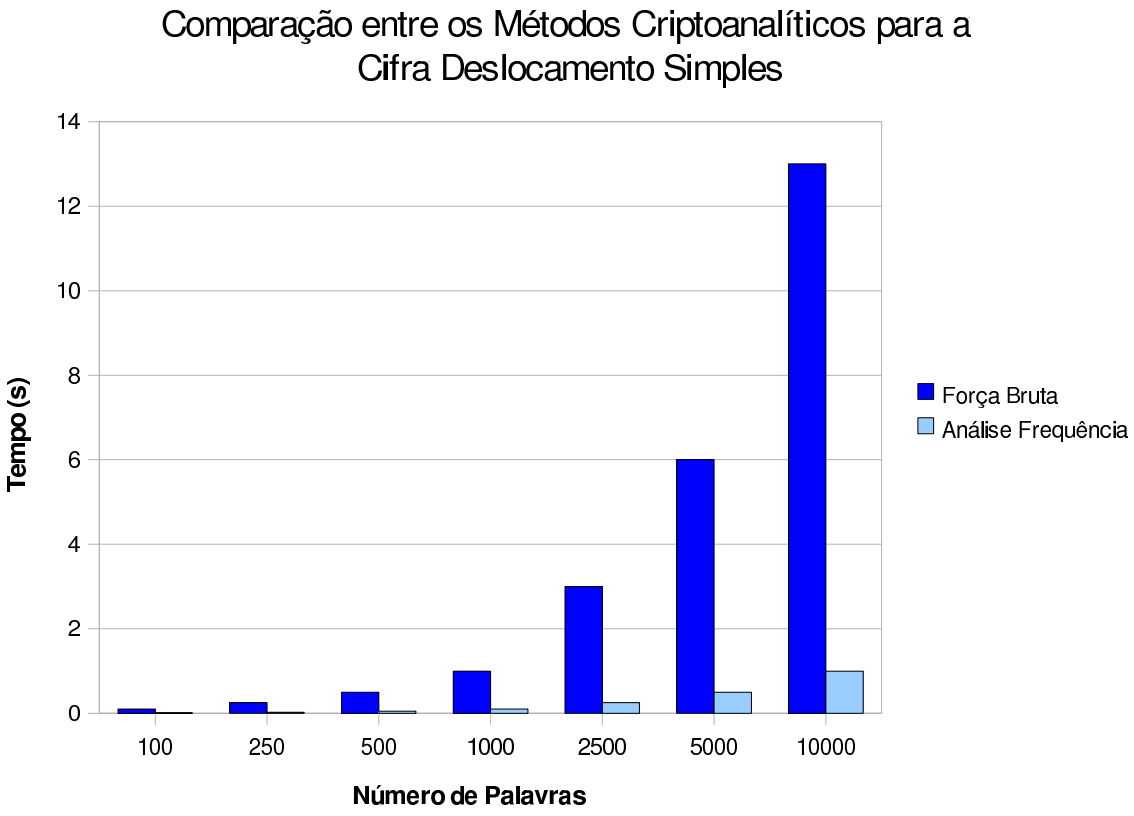

Figura 3.3: Cifra Deslocamento Simples

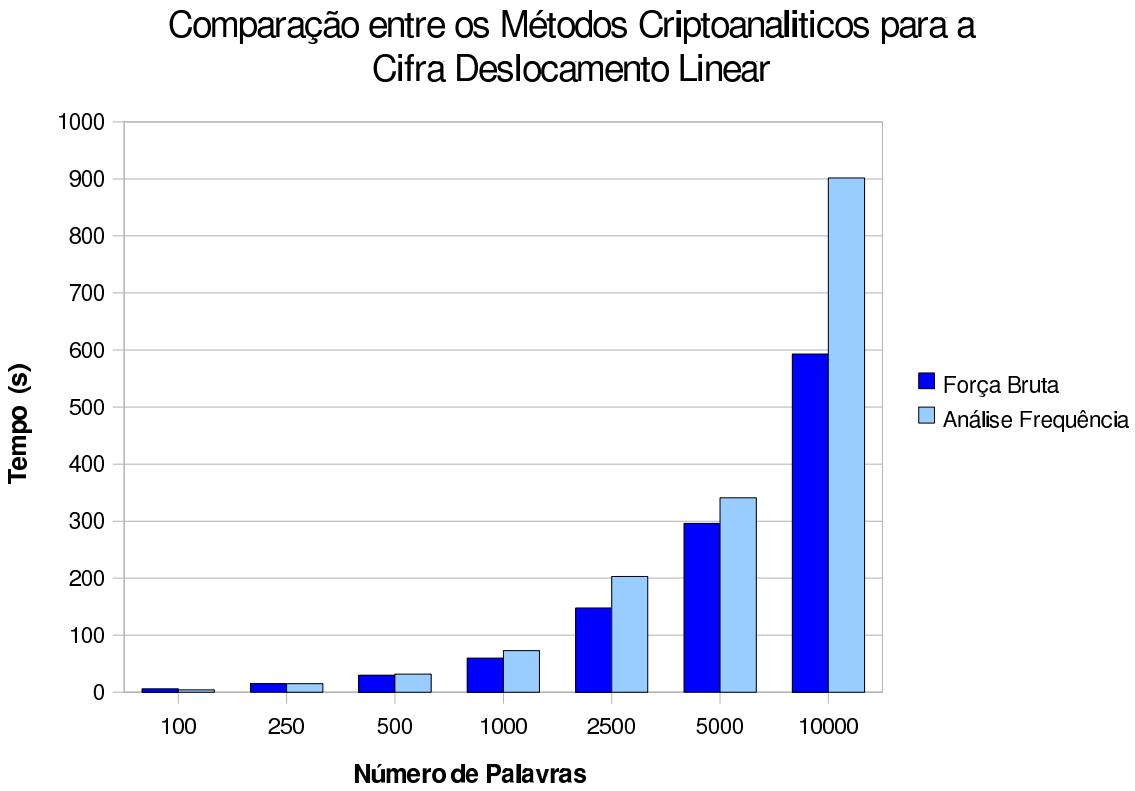

Figura 3.4: Cifra Deslocamento Linear

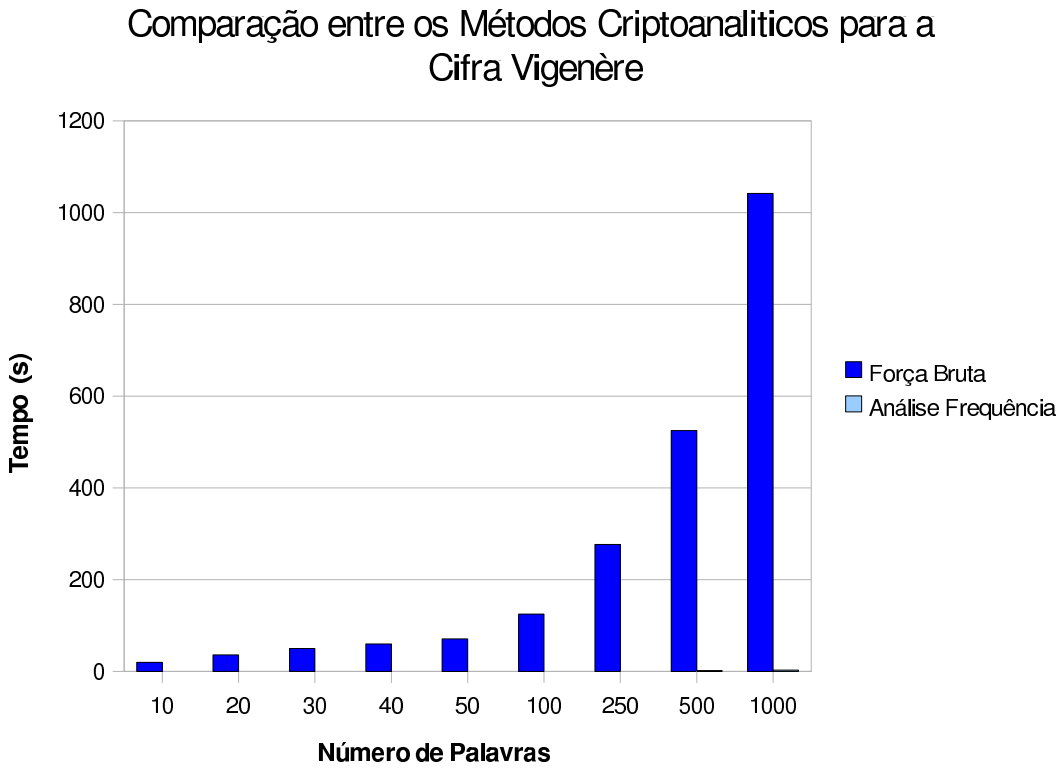

Figura 3.5: Cifra de Vigenère

Capítulo 4

A Página da Rede & Programas em C

4.1 A Página da Rede

Um dos objectivos deste trabalho para além do estudo dos métodos criptográficos e criptoanalíticos foi a implementação destes métodos em termos computacionais e a construção de uma página de rede, a qual deverá permitir aos seus utilizadores a compreensão dos métodos assim como providenciar uma plataforma de experimentação e/ou exploração dos métodos.

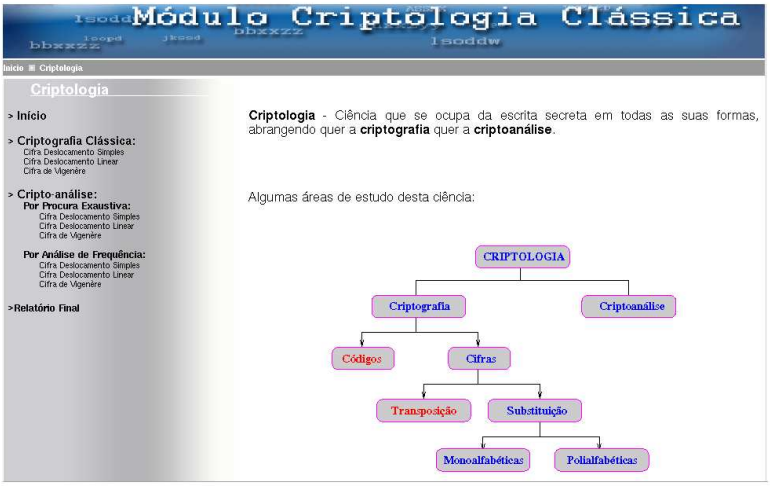

Figura 4.1: A página da Rede

A implementação dos métodos foi feita com base na linguagem de programação C e a construção da página foi feita com as linguagens PHP e HTML.

A página da Rede pode ser consultada em `http://hilbert.mat.uc.pt/~Criptologia`.

4.1.1 Estrutura da Página

Na figura 4.2 pode-ver-se a estrutura da página, na qual se pode constatar que esta tem três menus principais: Início, Criptografia Clássica e a Criptoanálise. Descrição dos menus:

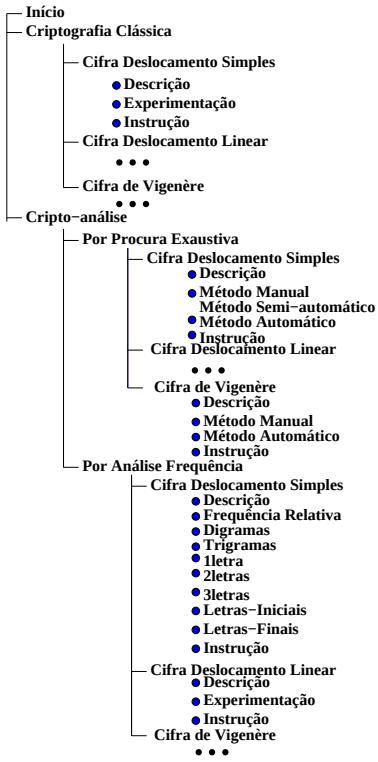

Figura 4.2: A estrutura da página

1. Início — página inicial;

2. Criptografia Clássica — página inicial para os três métodos criptográficos. Para cada um dos métodos consta:

   - Descrição — uma pequena descrição do método criptográfico;
   - Experimentação — o utilizador tem a possibilidade de experimentar/explorar o processo de encriptação e desencriptação do método;
   - Instrução — mostra ao utilizador, como encriptar/desencriptar uma mensagem de acordo com o método escolhido;

3. Criptoanálise, aqui existem como sub-menus principais os dois métodos criptoanalíticos: *Procura Exaustiva* e *Análise de Frequência*.
   Procura Exaustiva para cada um dos métodos existe:

   - Descrição — uma pequena descrição do método criptoanalítico;
   - Método Manual — depois de o utilizador encriptar a mensagem com um dos métodos criptográficos, o utilizador vai tentar no espaço das chaves possíveis descobrir a mensagem original;
   - Método Semi-automático — depois de o utilizador encriptar a mensagem com

um dos métodos criptográficos, o computador apresenta ao utilizador todas as possibilidades possíveis cabendo ao utilizador escolher a chave correcta;

- Método Automático — depois de o utilizador encriptar a mensagem com um dos métodos criptográficos, o computador apresenta ao utilizador as três melhores chaves candidatas, por ordem decrescente;
- Instrução — Mostra ao utilizador, como utilizar um dos métodos criptoanalíticos.

Nota: Para a cifra de Vigenère, não se implementou o método criptoanalítico semi-automático porque o espaço das chaves é muito grande e entendeu-se que não tinha nenhum interesse para o utilizador. Por exemplo para uma palavra-chave de comprimento 3 o número de possibilidades é de 97336.

Análise de Frequência cada método criptográfico tem o seu processo para aplicar o método criptoanalítico por análise de frequência:
Cifra Deslocamento Simples

- Descrição — uma pequena descrição do método criptoanalítico;
- Frequência Relativa — para que o utilizador depois de encriptar a mensagem com a cifra Deslocamento Simples, tente descobrir a chave de encriptação calculando a frequência relativa dos caracteres na mensagem encriptada;
- Digramas — para que a partir dos digramas existentes na mensagem encriptada, o utilizador tente descobrir a chave que foi usada no processo de encriptação;
- Trigramas — com os trigramas da mensagem encriptada, o utilizador pode tentar descobrir a chave de encriptação;
- 1 letra — a partir de palavras de comprimento um, o utilizador tenta descobrir a chave de encriptação da mensagem;
- 2 letras — a partir de palavras de comprimento dois, o utilizador tenta descobrir a chave de encriptação da mensagem;
- 3 letras — a partir de palavras de comprimento três, o utilizador tenta descobrir a chave de encriptação da mensagem;
- Letras Iniciais — a partir das letras iniciais da mensagem encriptada, o utilizador pode tentar descobrir a chave de encriptação;
- Letras Finais — a partir das letras finais da mensagem encriptada, o utilizador tenta descobrir a chave de encriptação;

- Instrução — mostra ao utilizador, como usar o método escolhido.

Cifra Deslocamento Linear

- Descrição — uma pequena descrição do método criptoanalítico;
- Experimentação — o utilizador depois de encriptar a mensagem com a cifra de Deslocamento Linear, tenta descobrir a chave de encriptação;
- Instrução — mostra ao utilizador, como usar o método.

Cifra de Vigenère

- Descrição — uma pequena descrição do método criptoanalítico;
- Experimentação — o utilizador depois de encriptar a mensagem com a cifra de Vigenère, tenta descobrir a chave de encriptação;
- Instrução — mostra ao utilizador, como usar o método.

4.2 Programas em C

Algumas das funções importantes na implementação dos métodos criptográficos e o algoritmo do Método de Procura Exaustiva para cada um dos métodos criptográficos.

- COMPALFA - o comprimento do alfabeto;
- NAOALFA - o valor das letras que não pertencem ao alfabeto.

Listing 4.1: Função de encriptação da Cifra de Deslocamento Simples

```
int cifrarDS(int m,int c){
  return((m + c) % COMPALFA);
 }
```

Listing 4.2: Função de desencriptação da Cifra de Deslocamento Simples

```
int decifrarDS(int e,int c){
  return((e - c) % COMPALFA);
}
```

Listing 4.3: Função de encriptação de um carácter com a Cifra de Deslocamento Simples

```
  -> parâmetro de entrada: o valor do carácter e a chave para
     encriptar o carácter
```

```
2   <- parâmetro de saída: o valor do carácter encriptado ou sem
        encriptar
3 int encriptarDS(int m,int chave){
4   int valornoalfabeto,encriptar,valoriso;
5   //obter o valor do carácter no Alfabeto
6   valornoalfabeto=iso8859_1paraAlfabeto(m);
7   //verificar se o carácter pertence ao Alfabeto
8   if(valornoalfabeto!= NAOALFA){
9     encriptar=cifrarDS(valornoalfabeto,chave);
10    //obter o valor do carácter no ISO8859_1
11    valoriso=AlfabetoParaISO8859_1(encriptar);
12    return valoriso;
13  }
14  else
15    //caso o carácter não pertencer ao alfabeto retorna-o sem
        encriptar
16    return m;
17 }
```

Listing 4.4: Função de desencriptação de um carácter com a Cifra de Deslocamento Simples

```
1 -> parâmetro de entrada: o valor do carácter e a chave para
      desencriptar o carácter
2 <- parâmetro de saída: o valor do carácter desencriptado ou o
      valor do carácter de entrada na função
3 int desencriptarDS(int c,int chave){
4   int valornoalfabeto,desencriptar,valoriso,diferenca;
5   valornoalfabeto=iso8859_1paraAlfabeto(c);
6   if(valornoalfabeto!=NAOALFA){
7     diferenca=decifrarDS(valornoalfabeto,chave);
8     desencriptar=mod(diferenca,COMPALFA);
9     valoriso=AlfabetoParaISO8859_1(desencriptar);
10    return valoriso;
11  }
12  else{
```

```
13      return c;
14    }
15  }
```

Listing 4.5: Função de encriptação da Cifra de Deslocamento Linear

```
1  int cifrarDL(int m,int chave1,int chave2){
2    return((chave1*m + chave2) % COMPALFA);
3   }
```

Listing 4.6: Função de desencriptação da Cifra de Deslocamento Linear

```
1  int decifrarDL(int c,int chave1,int chave2){
2    int inverso;
3    //chamar a função inversoMultiplicativo()
4     inverso=inversoMultiplicativo(chave1,COMPALFA);
5    return(inverso*(c - chave2) % COMPALFA);
6  }
```

Listing 4.7: Função de encriptação de um carácter com Cifra de Deslocamento Linear

```
1  -> parametro de entrada três inteiros:
2  - o valor do carácter e as duas chaves (chave1,chave2)
3  <- retorna um valor inteiro:
4  -o valor do carácter encriptado
5  -ou o valor do carácter sem encriptar
6  int encriptarDL(int m,int chave1,int chave2){
7    int valornoalfabeto,encriptar,valoriso;
8    //obter o valor do carácter no Alfabeto
9    valornoalfabeto=iso8859_1paraAlfabeto(m);
10   //verificar se o carácter pertence ao Alfabeto
11   if(valornoalfabeto!=NAOALFA){
12     //aplicar o algoritmo para encriptar
13     encriptar=cifrarDL(valornoalfabeto,chave1,chave2);
14     //obter o valor do carácter no ISO8859
15     //retornar o valor
16     valoriso=AlfabetoParaISO8859_1(encriptar);
17     return valoriso;
18   }
```

```
19    else
20      //caso o carácter não pertencer ao alfabeto retorna-o sem
          encripatr
21      return m;
22  }
```

Listing 4.8: Função de desencriptação de um carácter com Cifra de Deslocamento Linear

```
1  ->parâmetro de entrada três inteiros:
2  -o valor do carácter e as duas chaves (chave1,chave2)
3  <- retorna um valor inteiro:
4  -o valor do carácter desencriptado
5  -ou o valor do carácter de entrada
6  int desencriptarDL(int c,int chave1,int chave2){
7     int inverso,desencriptar,valoriso,valornoalfabeto,resultado;
8     //obter o valor do carácter no Alfabeto
9     valornoalfabeto=iso8859_1paraAlfabeto(c);
10    if(valornoalfabeto!=NAOALFA){
11      //aplicar o algoritmo para desencriptar
12      resultado=decifrarDL(valornoalfabeto,chave1,chave2);
13      desencriptar=mod(resultado,COMPALFA);
14      //obter o valor do carácter no ISO8859
15      //retornar o valor
16      valoriso=AlfabetoParaISO8859_1(desencriptar);
17      return valoriso;
18    }
19    else
20      return c;
21 }
```

Listing 4.9: função de encriptação da Cifra de Vigenère

```
1  void encriptarVig(int m[],int chave[],int c[],int tamanhochave)
     {
2     int i=0;
3     int valornoalfabeto,encriptar,valoriso;
4     while(i<=tamanhochave){
```

```
5        //obter o valor do carácter no Alfabeto
6        valornoalfabeto=iso8859_1paraAlfabeto(m[i]);
7        //verificar se o carácter pertence ao Alfabeto
8       if(valornoalfabeto!= NAOALFA){
9          encriptar=cifrarDS(valornoalfabeto,chave[i]);
10         //obter o valor do carácter no ISO8859_1
11         valoriso=AlfabetoParaISO8859_1(encriptar);
12         c[i]=valoriso;
13         i++;
14       }
15       else{
16         c[i]=m[i];
17         i++;
18       }
19     }
20 }
```

Listing 4.10: função de desencriptação da Cifra de Vigenère

```
1  void desencriptarVig(int c[],int chave[],int m[],int
     tamanhochave){
2     int i=0;
3     int valornoalfabeto,desencriptar,valoriso,diferenca;
4     while(i<=tamanhochave){
5        //obter o valor do carácter no Alfabeto
6        valornoalfabeto=iso8859_1paraAlfabeto(c[i]);
7        //verificar se o caracter pertence ao Alfabeto
8        if(valornoalfabeto!= NAOALFA){
9          diferenca=decifrarDS(valornoalfabeto,chave[i]);
10         desencriptar=mod(diferenca,COMPALFA);
11         //obter o valor do carácter no ISO8859_1
12         valoriso=AlfabetoParaISO8859_1(desencriptar);
13        m[i]=valoriso;
14         i++;
15         }
16       else {
```

```
17          m[i]=c[i];
18          i++;
19        }
20      }

22  }
```

Listing 4.11: Algoritmo do Método de Procura Exaustiva para a Cifra de Deslocamento Simples

```
1  -> Ficheiro com a mensagem encriptada
2  <- Chave
3  for(k=0; k < K; k++){
4       Desencriptar a mensagem com a chave k
5       Procurar palavras com significado em Português
6       Guardar a chave e o número de palavras encontradas
7  }
8       Apresentar a chave com o maior número de palavras com
           significado
```

Listing 4.12: Algoritmo do Método de Procura Exaustiva para a Cifra de Deslocamento Linear

```
1  -> Ficheiro com a mensagem encriptada
2  <- Chave (a,b)
3  for(a=1; a < K; a++){
4   for(b=0; b < K; b++){
5       Desencriptar a mensagem com o par chave(a,b)
6       Procurar palavras com significado em Português
7       Guardar o par chave(a,b) e o número de palavras encontradas
8   }
9  }
10      Apresentar o par chave(a,b) com o maior número de palavras
           com significado
```

Listing 4.13: Algoritmo do Método de Procura Exaustiva para a Cifra de Vigenère

```
1  -> Ficheiro com a mensagem encriptada
2  <- Palavra-chave
```

```
3  do{
4      j=0
5       do{
6          Gerar palavra-chave com comprimento j
7          Desencriptar a mensagem com a palavra-chave gerada
8          Procurar palavras com significado em Português
9          Calcular percentagem de palavras encontradas na mensagem
               desencriptada
10         if(percentagem >= 0.80){
11             Apresentar a palavra chave
12             j = COMPALFA
13         }
14        }
15        while((não gerar todas as palavras-chave de comprimento j
              ) && (j < COMPALFA))
16        j++;
17    }
18 while(j < COMPALFA)
```

Capítulo 5

Conclusões

Sendo a criptologia uma área tão vasta, este trabalho concentrou-se em três métodos criptográficos clássicos, com o intuito de mostrar o processo de encriptação, desencriptação, bem como alguns dos possíveis ataques a estes métodos.

Deste modo, foram implementadas em C as seguintes cifras bem com os seguintes métodos criptoanalíticos: Cifra de Deslocamento Simples, Cifra de Deslocamento Linear, Cifra de Vigenère, Método Criptoanalítico por Procura Exaustiva no Espaço das Chaves e o Método Criptoanalítico por Análise de Frequências, respectivamente. Assim, avaliou-se a segurança das cifras em relação aos métodos criptoanalíticos e, pode-se concluir que aquelas, actualmente, são inseguras, visto que devido ao uso das novas tecnologias é mais rápido, seguro e eficaz quebra-las do que há vários séculos atrás. Contudo, as cifras não são somente interessantes do ponto vista histórico, já que muitos conceitos e técnicas similares estão a utilizar-se em métodos criptográficos modernos. Além disso, o estudo dos métodos criptográficos clássicos ajuda na compreensão das técnicas básicas usadas para o sigilo da informação assim como os tipos de ataques que podem ser usadas para quebrar um sistema criptográfico.

Foi ainda construída uma página da Rede, na qual as cifras e os métodos atrás referidos podem ser experimentados.

Por último este é um trabalho que futuramente poderá ter continuidade, estudando outros métodos criptográficos e criptoanalíticos clássicos, assim como alguns métodos da criptografia moderna.

Bibliografia

[1] Gareth Jones and Mary Jones. *Elementary Number Theory.* Springer-Verlag, 1998.

[2] Neal Koblitz. *A course in Number Theory and Cryptography.* Springer-Verlag, 2nd ed, 1994.

[3] Alfred Menezes, Paul van Oorschot, and Scott Vanstone. *Handbook of Applied Cryptography.* CRC Press, 1996.

[4] Pedro Quaresma and Augusto Pinho. Análise de Frequências da Língua Portuguesa. *In Actas da Conferência Ibero-Americana InterTIC 2007*, 2007.

[5] Richard Spillman. *Classical and Contemporary Criptology.* Pearson Prentice-Hall, Upper Saddle, NJ 07458, 2005.

[6] Douglas Stinson. *Cryptography Theory and Practice.* Chapman & Hall/CRC, 3rd edition, 2006.

[7] James Tattersall. *Elementary Number Theory in Nine Chapters.* Cambridge University, 2005.

[8] Viktoria Tkotz. *Criptografia - Segredos Embalados para Viagem.* Novatec Editora, São Paulo, Brasil, 2005.