UNIVERSIDADE DE BRASÍLIA
DEPARTAMENTO DE CIÊNCIA DA COMPUTAÇÃO

CIC0201 - Segurança Computacional - 2023.2 - Turma 01

Trabalho de Implementação 1

Cifra de Vigenère

Rafael Rodrigues Gama D. de Aragão - 190134780

1. Introdução

Este trabalho explora a cifra de Vigenère, tendo duas partes: o cifrado/decifrador e o ataque de recuperação de senha por análise de frequência. A cifra de Vigenrère é uma cifra de substituição polialfabética em que cada símbolo é deslocado baseado em uma chave criptográfica, diferentemente de uma cifra monoalfabética, onde cada símbolo é deslocado por um valor fixo. Durante muito tempo, essa cifra foi sinônimo de confiança e chegou a ser classificada como inquebrável. Porém, em 1854, Charles Babbage provou que era possível quebrar a cifra. Ao analisar repetições no texto consegue-se achar o tamanho da chave e então, utilizando análise de frequência, descobrir a chave em si.

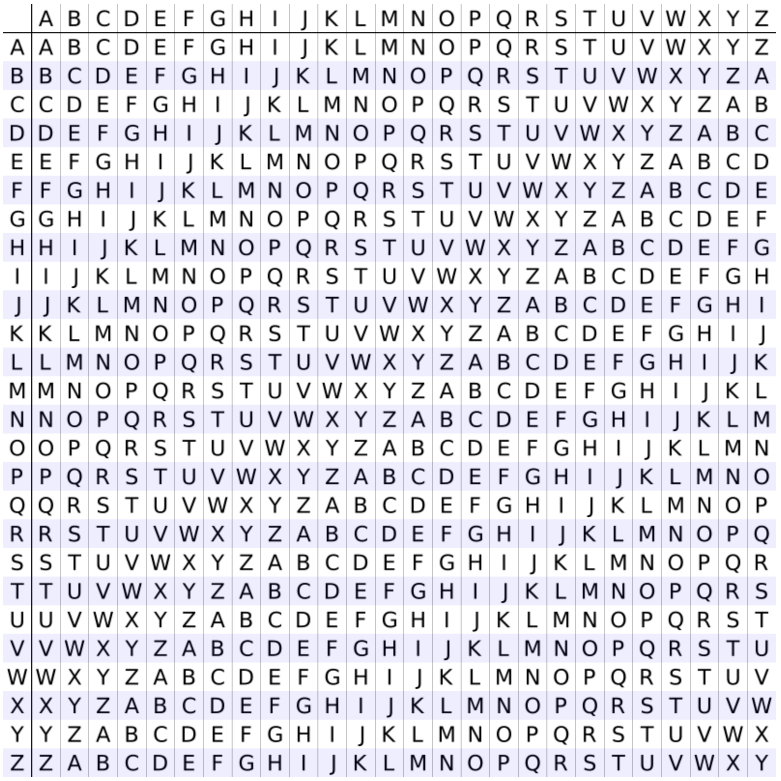
2. Fundamentação teórica

Numa cifra de César, cada letra do alfabeto é deslocada da sua posição um número fixo de lugares; por exemplo, se tiver um deslocamento de 3, "A" torna-se "D", "B" fica "E", etc. A cifra de Vigenère consiste no uso de várias cifras de César em sequência, com diferentes valores de deslocamento ditados por uma palavra-chave".

Para cifrar, é usada uma tabela de alfabetos que consiste no alfabeto escrito 26 vezes em diferentes linhas, cada um deslocado ciclicamente do anterior por uma posição. Uma palavra é escolhida como "palavra-chave" e é repetida até ter o comprimento do texto a ser cifrado. Cada letra desta palavra vai indicar a linha a ser utilizada para cifrar ou decifrar uma letra da mensagem.

Por exemplo, supondo que se quer criptografar o texto “THINKABOUT” usando a chave “VINTAGE”. Primeiramente, repetindo a palavra-chave até o tamanho do texto tem-se que a chave resultante é “VINTAGEVINTA”. A primeira letra do texto, T, é cifrada usando o alfabeto na linha V, que é a primeira letra da chave. Basta olhar para a letra na linha V e coluna T na grelha de Vigenère, e que é um O, repetindo isso para cada letra obtém-se o texto cifrado que é “OPVGKGFJCGBT”. A desencriptação é feita utilizando o processo inverso.

Figura 1 - A grade de Vigenère, conhecido também por tabula recta, usado para criptografia e descriptografia.



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

Fonte: Wikipedia (2011).

Do ponto de vista matemático, a cifra pode ser descrita pela equação

$$C_i = P_i + K_i \pmod{26}$$

em que C_i é a letra criptografada, P_i é a letra a ser criptografada, K_i é a letra da chave, mod é o resto da divisão e a-z são representados como números inteiros de 0 a 25. De forma semelhante, a desencriptação pode ser descrita como

$$P_i = C_i - K_i \pmod{26}.$$

O ponto fraco da cifra é que a chave é repetida diversas vezes, resultando em pedaços iguais no texto criptográfico. Essa repetição é mais provável que ocorra quando a mesma sequência de letras do texto simples é criptografada com a mesma parte da chave.

Para quebrar a cifra, primeiramente busca-se pedaços do texto que são repetidos várias vezes. Em seguida, calcula-se a distância entre os pedaços repetidos e os divisores de cada distância. O divisor que aparece mais vezes é provavelmente o tamanho da chave.

Figura 2 - Exemplo de repetições em texto criptografado.

Fonte: pt.frwiki.wiki (2023).

Figura 3 - Distância entre repetições e seus divisores. O número 5 é divisor de todas as distâncias sendo provavelmente o tamanho da chave criptográfica.

		Possíveis comprimentos de chave (divisores de distância)			
Seqüência repetida	Distância entre repetições	2	3	5	19
WUU	95			x	x
EEK	200	x		x	
WXIZAYG	190	x		x	x
NUOCZGM	80	x		x	
DOEOY	45		x	x	
GMU	90	x	x	x	

Fonte: pt.frwiki.wiki (2023).

Sabendo o tamanho da chave, pode-se então dividir o texto em n subtextos, em que n é o tamanho da chave, cada texto tendo sido criptografado pela mesma letra. Desse modo, o problema se torna igual ao da quebra da cifra de César e pode ser descriptografado por análise de frequência.

Cada letra do alfabeto tem uma frequência usual em que aparece nos textos de cada língua. A figura 4 mostra a frequência das letras do alfabeto para diferentes idiomas. Pode-se calcular a frequência em que cada letra aparece do subtexto criptografado e comparar com a frequência no idioma. De forma prática, isso pode ser feito utilizando a soma dos quadrados da diferença. Calcula-se esse valor para cada um dos 26 deslocamentos possíveis, o deslocamento que obtiver o menor resultado é provavelmente o

deslocamento utilizado para aquele subtexto. Realizando esse procedimento para cada um dos subtextos, descobre-se a chave criptográfica utilizada.

Figura 4 - Tabela com frequência das letras do alfabeto para diferentes idiomas..

	Alemão	Espanhol	Francês	Inglês	Italiano	Português
a	6, 51 %	12, 53 %	7, 64 %	8, 66 %	11, 74 %	14, 63 %
b	1, 89 %	1, 42 %	0, 90 %	1, 49 %	0, 92 %	1, 04 %
c	3, 06 %	4, 68 %	3, 35 %	2, 78 %	4, 50 %	3, 88 %
d	5, 08 %	5, 86 %	3, 67 %	4, 25 %	3, 73 %	4, 99 %
e	17, 40 %	13, 68 %	16, 00 %	12, 70 %	11, 79 %	12, 57 %
f	1, 66 %	0, 69 %	1, 07 %	2, 23 %	0, 95 %	1, 02 %
g	3, 01 %	1, 01 %	0, 87 %	2, 02 %	1, 64 %	1, 30 %
h	4, 76 %	0, 70 %	0, 74 %	6, 09 %	1, 54 %	1, 28 %
i	7, 55 %	6, 25 %	7, 58 %	6, 97 %	11, 28 %	6, 18 %
j	0, 27 %	0, 44 %	0, 56 %	0, 15 %	0, 00 %	0, 40 %
k	1, 21 %	0, 01 %	0, 05 %	0, 77 %	0, 00 %	0, 02 %
l	3, 44 %	4, 97 %	5, 46 %	4, 03 %	6, 51 %	2, 78 %
m	2, 53 %	3, 15 %	2, 97 %	2, 41 %	2, 51 %	4, 74 %
n	9, 78 %	6, 71 %	7, 10 %	6, 75 %	6, 88 %	5, 05 %
o	2, 51 %	8, 68 %	5, 38 %	7, 51 %	9, 83 %	10, 73 %
p	0, 79 %	2, 51 %	3, 02 %	1, 93 %	3, 05 %	2, 52 %
q	0, 02 %	0, 88 %	1, 36 %	0, 10 %	0, 51 %	1, 20 %
r	7, 00 %	6, 87 %	6, 55 %	5, 99 %	6, 37 %	6, 53 %
s	7, 27 %	7, 98 %	7, 95 %	6, 33 %	4, 98 %	7, 81 %
t	6, 15 %	4, 63 %	7, 24 %	9, 06 %	5, 62 %	4, 34 %
u	4, 35 %	3, 93 %	6, 31 %	2, 76 %	3, 01 %	4, 63 %
v	0, 67 %	0, 90 %	1, 63 %	0, 98 %	2, 10 %	1, 67 %
w	1, 89 %	0, 02 %	0, 11 %	2, 36 %	0, 00 %	0, 01 %
x	0, 03 %	0, 22 %	0, 39 %	0, 15 %	0, 00 %	0, 21 %
y	0, 04 %	0, 90 %	0, 31 %	1, 97 %	0, 00 %	0, 01 %
z	1, 13 %	0, 52 %	0, 14 %	0, 07 %	0, 49 %	0, 47 %

Fonte: https://cosmovivencias.blogspot.com/2019/07/0184-frequencia-das-letras-do-alfabeto.html (2019).

3. Metodologia

Para implementação da cifra foi utilizada a linguagem de programação Python. O código está disponível em https://github.com/rafael2903/Vigenere-cipher. O arquivo vigenere.py contém a implementação das funções de encriptação, desencriptação e quebra da cifra. O arquivo tests.py contém alguns testes unitários que testam essas funções. Para executar os testes é necessário rodar o comando `python -m unittest -v tests.py`. Também foi adicionado ao programa uma interface de linha de comando, sendo possível executar todas as funções pelo terminal passando um arquivo de entrada. A seguir são mostrados alguns exemplos de utilização dessa interface.

Encriptação

```
python vigenere.py -e -i plain_text.txt -o
cipher_text.txt -k "key"
```

Desencriptação

```
python vigenere.py -d -i cipher_text.txt -o
plain_text.txt -k "key"
```

Quebra da cifra

```
python vigenere.py -b -i cipher_text.txt -o
plain_text.txt -l PORTUGUES
```

Mais exemplos de uso estão disponíveis no GitHub citado.

4. Conclusão

Em conclusão, a implementação da cifra de Vigenère e suas funcionalidades de criptografia, descriptografia e quebra por análise de frequência ilustram a aplicação prática dos princípios criptográficos. Este trabalho não apenas aprofundou nosso entendimento da criptografia clássica, mas também fortaleceu nossas habilidades de programação e resolução de problemas, enriquecendo nosso conhecimento em segurança da informação e contribuindo significativamente para nossa formação acadêmica nessa área.

5. Referências

- Criptoanálise da cifra de Vigenère - frwiki.wiki. Disponível em: <https://pt.frwiki.wiki/wiki/Cryptanalyse_du_chiffre_de_Vigen%C3%A8re>. Acesso em: 1 out. 2023.
- Como uma cifra do século XVII se torna uma criptografia intransponível? Disponível em: <https://www.kaspersky.com.br/blog/vigenere-cipher-history/5688/>. Acesso em: 1 out. 2023.
- WIKIPEDIA CONTRIBUTORS. Vigenère cipher. Disponível em: <https://en.wikipedia.org/wiki/Vigen%C3%A8re_cipher>.

- Vigenere CIpher. YouTube, 6 jul. 2015. Disponível em: <https://www.youtube.com/watch?v=SkJcmCaHqS0>
- Cryptography - Breaking the Vigenere Cipher. Disponível em: <https://www.youtube.com/watch?v=P4z3jAOzT9I>.