Implementação e ataque da Cifra de Vigenère

Professor: João Gondim

Segurança Computacional - 2021/2

Rafael Fernandes Barbosa
17/0163857
Universidade de Brasília
rafaelfbarbosa7@gmail.com

Marina Pinho Garcia
17/0110702
Universidade de Brasília
ninapgarcia@gmail.com

1. Introdução

Criada por Leon Battista Alberti por volta de 1465, a cifra de Vigenère é uma forma de criptografia que por mais de 300 anos foi considerada inquebrável.

1.1. Cifra de Vigenère

A cifra de Vigenère é realizada utilizando uma substituição polialfaabetica, utilizando o quadro de Vigenère (figura 1).

	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

Figura 1. Quadro utilizado para a cifrar um texto utilizando a Cifra de Vigenère [1]

O quadro mostrado em 1 consiste no alfabeto escrito 26 vezes, deslocando-o (de forma cíclica) a cada linha uma vez para a esquerda.

A cifra é realizada utilizando uma palavra chave, e a depender de cada letra dessa chave, será usada uma linha diferente desse quadro para a cifração.

Para exemplificar vamos considerar um exemplo em que o texto que queremos cifrar é 'CIFRA DE VIGENERE' e a palavra chave utilizada será 'CODE'.

O primeiro passo a ser realizado é a extensão da chave para que possa ser usada em todo o texto (Figura 2)

```
> Texto:   CIFRA DE VIGENERE
> Chave:   CODE


> Texto:   CIFRA DE VIGENERE
> Chave:   CODEC OD ECODECOD
```

Figura 2. Expansão da palavra chave

Agora, temos que ir letra a letra olhando no nosso Quadro de Vigenère, a letra do texto encontraremos nas colunas, e a letra da chave nas linhas, e teremos então nossa letra crifrada. Após realizar esse processo para todas as letras, teremos nosso texto cifrado (figura 3)

```
> Texto:   CIFRA DE VIGENERE
> Chave:   CODEC OD ECODECOD
> Cifra:   EWIVC RH ZKUHRGFH
```

Figura 3. Texto cifrado

Por fim, para decifrar a mensagem, devemos apenas realizar o processo inverso, com a letra do texto e a letra da chave, deslocar o texto para a direita.

1.2. Criptoanálise da Cifra de Vigenère

Criptoanálise é a arte de tentar descobrir o texto cifrado ou a lógica por trás da sua cifração. Para quebrar a cifra de Vigenère basta adivinhar a chave com que essa foi cifrada, já que com a chave o processo de decifrar é trivial. Para encontrar a chave três passos são necessários[2]:

- É determinado o tamanho da chave.
- O texto é dividido com o tamanho da chave.

- É usada análise frequência em cada parte.

Na seção 2.2 estes passos são discutidos com maiores detalhes.

2. Implementação

2.1. Cifra

Antes de começar a cifrar o texto, foram realizados alguns pré processamentos neste:

- Transformar todas as letras para maiúsculas
- Retirar todos os acentos das palavras
- Retirar todos os caracteres que não estão presentes no alfabeto (como espaços, pontuações, ...)

Após o pré processamento do texto realizamos a expansão da palavra chave juntamente com o deslocamento do texto de acordo com o índice das letras do alfabeto e obtemos nosso texto cifrado, assim como foi explicados na seção 1.1. E para a decifração, realizamos o mesmo processo porém com o deslocamento na direção contrária.

2.2. Criptoanálise

Como dito em 1.2, o primeiro passo para se decifrar um texto criptografado por Vigenère é encontrar o tamanho da chave, para isso é necessário primeiramente contar picos de coincidências em um texto ao realizar a comparação do texto com ele mesmo deslocado. Como é apresentado na figura 4.

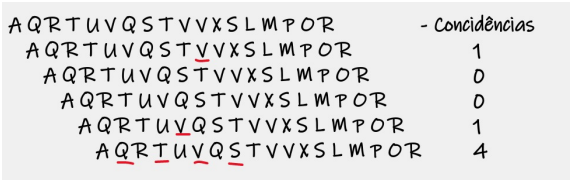

Figura 4. Esquemático de comparação do texto deslocado

Fazendo este deslocamento ao longo de todo o texto e armazenando coincidências em um vetor podemos plotar um gráfico com a forma apresentada na figura 5 neste é fácil perceber que em intervalos de tamanho 5 os shifts apresentam uma maior coincidências, isto indica uma chave de tamanho 5 (de fato a chave tem este tamanho). Identificar tais picos, não é uma tarefa tão fácil de se observar computacionalmente, principalmente quando se trata de chaves maiores como é apresentado na figura 6. Por este motivo, como trata-se de análises de frequências foi feita a Transformada Rápida de Fourier para melhor compreender o funcionamento dos sinais (quantidade de coincidências por shifts), o mesmo sinal da figura 6 é visto em Fourier na figura 7 onde é muito mais fácil encontrar os intervalos de repetição, na análise de Fourier deve-se calcular a distância entre os intervalos e inverter este valor: na figura o tamanho da chave seria 1/(0.199-0.132) ≈ 15.

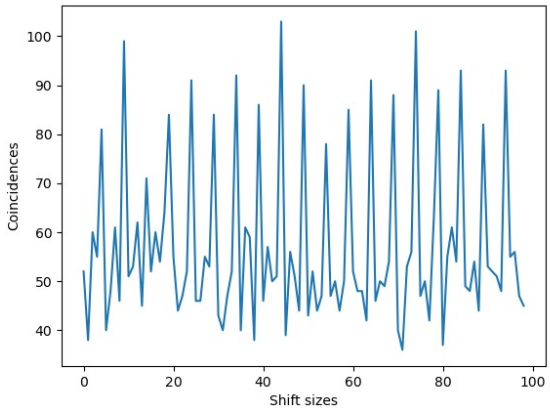

Figura 5. Presença de Coincidências Palavra de tamanho 5.

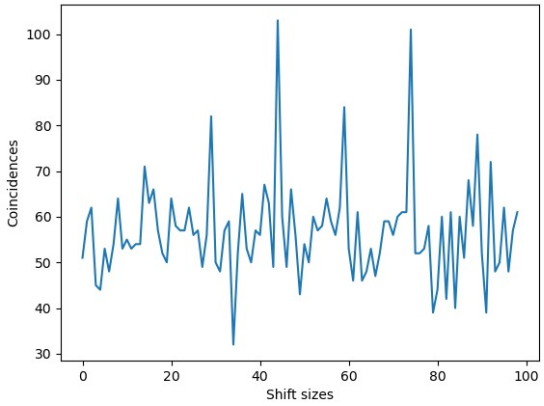

Figura 6. Presença de Coincidências Palavra de tamanho 15.

Para fazer encontrar o valor dessa chave automaticamente, dois algoritmos foram utilizados: O primeiro calcula a moda da diferença entre os intervalos dos 10 maiores picos, sendo o inverso desse valor de moda considerada a estimativa do tamanho da chave; O segundo estabelece um threshold normalizado (limiar para considerar um valor de pico) sendo este limiar uma porcentagem (45% obtido de forma empírica) e utiliza o primeiro pico encontrado acima deste valor como o que tem a frequência fundamental do sinal(inverso do tamanho da chave).

Com o tamanho da chave é feita a separação do texto com o tamanho da chave, isto é se uma chave tem 3 letras, é uma mensagem cifrada é ”ABCDEFGHIJKLMNO”, temos 3 intervalos a serem analisados ”ADGJM”, ”BEHKN”e ”CFILO”para cada um deste intervalo é feito um shift até

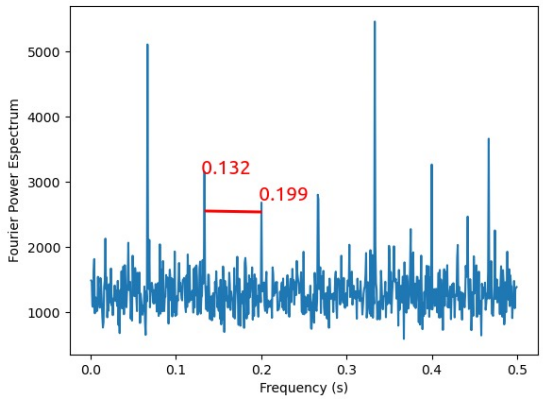

Figura 7. Presença de Coincidncias no Espectro de Fourier

que a distribuição das letras corresponda à distribuição da língua analisada como pode ser observado nas figuras 8, 9, 10 para a palavra chave TRE (nesse caso para um texto cifrado em inglês). Assim, com a iteração do melhor encaixe para cada letra é possível adivinhar a chave, e com esta obter a mensagem original é uma tarefa trivial com tratado anteriormente.

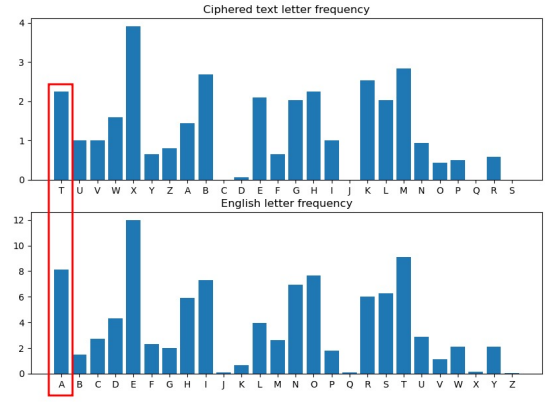

Figura 8. Melhor encaixe T.

3. Resultados

3.1. Cifração e Decifração

Nesta parte os resultados obtidos foram completamente satisfatórios, ao informarmos a mensagem desejada e a palavra chave que deve ser usada, conseguimos obter o texto cifrado e em seguida decifrá-lo sem problemas.

3.2. Ataque à Cifra

Para testar a o ataque à cifra foi feito um código que gerava senhas aleatórias com tamanho de 2 a 15 caracteres e

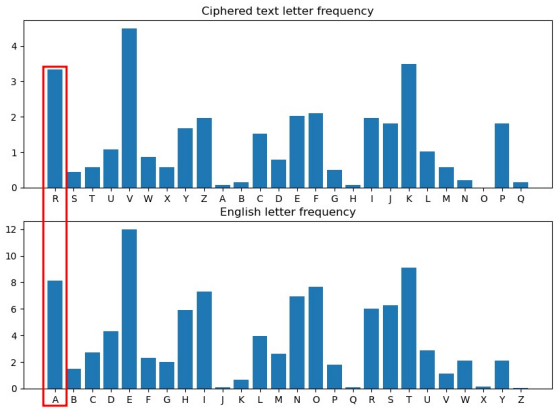

Figura 9. Melhor encaixe R.

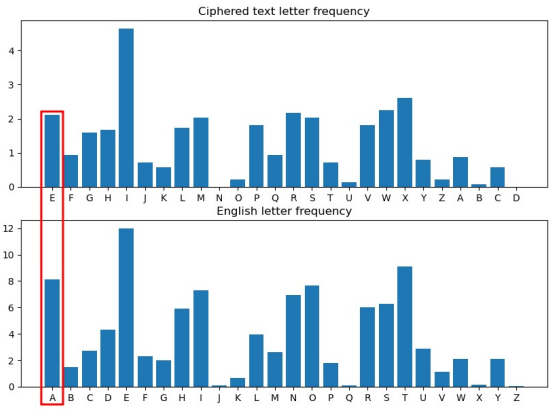

Figura 10. Melhor encaixe E.

era então feita a cifra de um texto. Neste texto é aplicada a detecção automática de chave apresentada e comparado este resultado com a chave aleatória gerada, se estas forem iguais é aumentado um contador e no final é obtida uma métrica do desempenho do algoritmo em porcentagem. Na tabela 1 é possível observar que o algoritmo que faz uso do threshold normalizado tem o melhor desempenho e por isso utilizamos ele na detecção automática da senha e além disso pode-se observar que para um tamanho de chave conhecido o algoritmo sempre acerta a chave, sendo assim a análise de frequência se comportou perfeitamente dada as limitações impostas (texto com mais de 1000 caracteres e chave de 2 à 15 caracteres). Aos erros de detecção de chaves, o principal motivo é provavelmente a presença de letras repetidas nas palavras sendo que este fator pode torna a análise de frequência mais complexa causando confusão no algoritmo de detecção de tamanhos.

Calculo da Moda	Threshold	Tamanho conhecido
89,47%	93,21%	100%

Tabela 1. Acurácia em porcentagem para os dois algoritmos de detecção de tamanho de chave e para uma chave de tamanho conhecido.

4. Conclusão

Apesar de existirem algumas limitações na implementação do projeto, consideramos o resultado obtido bastante satisfatório, com 93,21% de acurácia no nosso ataque à cifra de Vigenère (dentro das restrições apresentadas).

Por fim, por meio deste trabalho foi possível aprofundar e entender na prática os conhecimentos obtidos em sala de aula sobre como as cifras funcionam.

Referências

[1] Cifra de vigenère. https://pt.wikipedia.org/wiki/Cifra_de_Vigenère - Acessado em 13 de março de 2022.

[2] Cryptography - breaking the vigenere cipher. https://www.youtube.com/watch?v=P4z3jAOzT9I - Acessado em 13 de março de 2022.