# Segurança Computacional Trabalho 1 - Cifra de Vigenère

# Mateus Freitas Cavalcanti 16/0137519

mat.fcavalcanti@gmail.com

# Mariana Mendanha da Cruz 16/0136784

mariana.mendanha.mm@gmail.com

#### **Abstract**

Este trabalho explora a cifra de Vigenère, tendo duas partes: o cifrador/decifrador e o ataque de recuperação de senha por análise de frequência. Para a parte de cifração e decifração, foi mostrado a lógica por trás da codificação de Vigenère, tal como seu processo reverso. Por fim, para a quebra de cifra, foi mostrado o modelo implementado, baseado na busca pelo melhor candidato de tamanho de chave e cálculo de X² para cada caracter da possível chave.

# 1. Introdução

Cifra, também chamada de algoritmo de criptografia são sistemas para criptografar e descriptografar dados. As cifras convertem mensagens, chamadas de plaintext, em textos cifrados, isto é, textos que se pretendem ilegíveis para quem não possui uma informação mantida secreta que chamamos de chave.

Cifras geralmente são categorizadas de acordo com o seu funcionamento e com a utilização de suas chaves. Estas podem ser simples ou complexas, a criptografia moderna usa técnicas cada vez mais sofisticadas e desempenha papel crucial em várias atividades do nosso dia a dia a começar por proteger dados confidenciais e pagamentos assegurando principalmente a integridade dos dados.

A cifra de Vigenère é uma técnica de criptografia por substituição polialfabética que usa uma série de diferentes cifras de César baseadas nas letras de uma chave.

A cifra de César consiste em um deslocamento de letras por uma posição fixa, por exemplo, com o deslocamento de 2 a letra "A" se torna "C". Sabendo disso a cifra de Vigenère faz o mesmo em sequência, com valores diferentes de deslocamento que dependem de uma chave dada.

Para cifrar, usamos a tabela que pode ser vista na Figura 1 onde o alfabeto é deslocado em cada linha até completar 26 deslocamentos, aonde se encerra o ciclo. Desta forma, a chave indica cada linha que iremos utilizar para cifrar enquanto cada letra da mensagem indica a coluna.

Portanto, se temos uma chave "AB" e mensagem "VIDA" a mensagem criptografada será "VJDB" visto que repetimos a chave para caber na mensagem, ou seja, "ABAB" observamos a linha "A" e coluna "V" e obtemos "V" e seguimos com o mesmo processo até termos a mensagem cifrada.



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

#### 2. Materiais e Métodos

- · Linguagem Python
- VS Code
- Cifra de Vigenère
- Ataque por Análise de frequência

#### 2.1. Parte I

#### 2.1.1 Modelo

O cifrador recebe uma senha e uma mensagem que é cifrada segundo a cifra de Vigenère, gerando um criptograma, enquanto o decifrador recebe uma senha e um criptograma que é decifrado segundo a cifra de Vigenère, recuperando uma mensagem.

#### 2.1.2 Cifrador

A função Cifrador() recebe um plaintext e uma chave, com isso, para cada letra na mensagem faz-se a operação correspondente a chave atribuída com o código unicode e um offset para o alfabeto ascii, após a operação o código para a letra volta a ser um caractere e então é retornada a mensagem criptografada.

$$E_i = (P_i + K_i) \mod 26$$

Figure 2. Codificação de Vigenère

#### 2.1.3 Decifrador

A função Decifrador() faz a operação inversa, recebendo o criptograma e a chave, e retornando a mensagem decifrada.

$$D_i = (E_i - K_i + 26) \mod 26$$

Figure 3. Decodificação de Vigenère

#### 2.2. Parte II

#### 2.2.1 Modelo

Foram fornecidas duas mensagens cifradas (uma em português e outra em inglês) com senhas diferentes. Cada uma das mensagens deve ser utilizada para recuperar a senha geradora do keystream usado na cifração e então decifradas. Sendo assim, o algoritmo de quebra de cifra foi separado em duas etapas: Determinação de melhor candidato para tamanho da chave desconhecida, e cálculo da melhor chave.

#### 2.2.2 Tamanho da chave

Para determinar o tamanho da possível chave geradora do keystream, foi utilizado o método baseado no índice de coincidência. Essa abordagem consiste em efetuar o cálculo de IC para diferentes tamanhos de chaves e, considerando as frequências das letras do idioma em questão, é escolhido como melhor candidato aquele tamanho de chave que apresenta o índice de coincidência associado mais próximo do índice das frequências do idioma. Sendo assim, tendo em vista um intervalo fixo de tamanhos de chaves entre 2 e 20,

é possível determinar o melhor tamanho para a chave desconhecida.

$$\mathbf{IC} = rac{\displaystyle\sum_{i=1}^{c} n_i (n_i - 1)}{N(N-1)/c}$$

Figure 4. Equação do índice de coincidência de uma subamostra

Para o cálculo do IC de chaves com tamanhos maiores ou iguais a 2, tem-se que o índice de coincidência associado é a média dos índices das N subamostras da mensagem cifrada, com N igual ao tamanho da chave.

#### 2.2.3 Melhor chave

Tendo como base o tamanho da chave encontrado (N), é utilizado o método X² para se determinar a possível chave do criptograma. Esse método consiste em separar o texto cifrado em N subamostras, onde todas as letras de uma determinada subamostra são cifradas pela mesma letra da chave. Sendo assim, é possível decifrar cada subamostra considerando uma cifra de césar, onde cada letra do texto em claro é cifrado pela mesma letra. Calculando o X² para cada possível valor de 'letra chave', tem-se que a letra com menor X² associado é adotada como melhor candidato. Por fim, para cada N subamostra, pode-se obter as N possíveis letras da chave.

$$\chi^{2}(C, E) = \sum_{i=A}^{i=Z} \frac{(C_{i} - E_{i})^{2}}{E_{i}}$$

Figure 5. Cálculo do chi square (X2) de uma subamostra

#### 3. Resultados

# 3.1. Cifração

De maneira inicial, o algoritmo de cifração proposto foi testado para diversas mensagens. Considerando a cifração da mensagem referente à enunciação da **primeira lei de Newton**, é possível verificar a mensagem encriptada, considerando a chave **newton** (figura 6).

Figure 6. Teste de cifração

# 3.2. Decifração

Considerando o resultado da cifração da sessão anterior, é possível utilizar a mesma chave, '**newton**', para efetuar o processo reverso e encontrar a mensagem em claro. O resultado deste processo é visto na figura 7.

Figure 7. Teste de decifração

### 3.3. Quebra de cifra

Para o teste da quebra de cifra, foram testadas mensagens cifradas em português e inglês. Sendo assim, foram obtidos os seguintes resultados:

Figure 8. Melhor candidato para tamanho de chave (tamanho 7).

De acordo com a figura 8, o índice de coincidência calculado para o tamanho de chave igual a 7 retorna o valor mais próximo ao índice da frequência de letras da língua portuguesa. Sendo assim, o tamanho de chave 7 é tido como melhor candidato. São calculados os valores de X² para todos os *shifts* das 7 subamostras da mensagem:

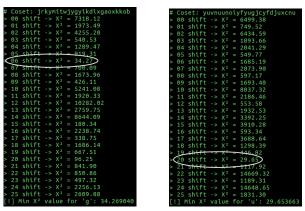


Figure 9. Cálculo X<sup>2</sup> para 1° e 2° caracteres da chave.

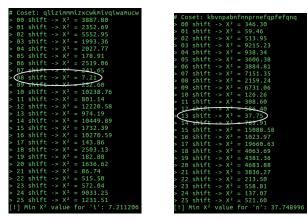


Figure 10. Cálculo X<sup>2</sup> para 3° e 4° caracteres da chave.

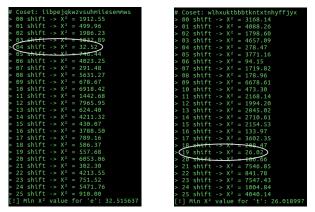


Figure 11. Cálculo X<sup>2</sup> para 5° e 6° caracteres da chave.



Figure 12. Cálculo X<sup>2</sup> para 7° caracter da chave.

Sendo assim, considerando os valores dos menores  $X^2$  para cada caracter da possível chave e seu respectivo *shift*, é possível definir o melhor candidato para a chave:

- 1° subamostra  $\rightarrow$  X° mínimo para **06** shifts  $\rightarrow$  g
- $2^a$  subamostra  $\rightarrow X^2$  mínimo para **20 shifts**  $\rightarrow u$
- $3^{\text{a}}$  subamostra  $\rightarrow$   $X^{\text{2}}$  mínimo para **08 shifts**  $\rightarrow$  i
- $4^{a}$  subamostra  $\rightarrow X^{2}$  mínimo para 13 shifts  $\rightarrow n$
- $5^{a}$  subamostra  $\rightarrow X^{2}$  mínimo para **04 shifts**  $\rightarrow e$
- $6^{\text{a}}$  subamostra  $\rightarrow$   $X^{\text{2}}$  mínimo para 19 shifts  $\rightarrow$  t
- $7^{\text{a}}$  subamostra  $\rightarrow$   $X^{\text{2}}$  mínimo para 14 shifts  $\rightarrow$  o

Por fim, a chave é obtida: **guineto**. A decodificação da mensagem encriptada é verificada na figura 13.

```
Found possible key: guineto

[...] Decoding 'jyqkiwsrutbllgkvivbhoynznpxfmuipeusiumajkstnmbqtzwomnk
bbjiifwbayyznzbhgfxnvtcyycpskoiuwrunskgknhtsdjmemxbicifitdxyvqinagfqpe
hsadifsyfojweefoxuafmfakxmqmjikcunwyconcqsxabuw' with key 'guineto'

> Decrypted text: deixedeladoessebaixoastralergaacabecaenfrenteomalque
agindoassimseravitalparaoseucoracaoequeemcadaexperienciaseaprendeumali
caoeujasofriporamarassimmedediqueimasfoitudoemvao
```

Figure 13. Resultado da decodificação da mensagem com a chave 'guineto'.

Aplicando os devidos espaçamentos, a mensagem em claro é dada por:

# Conselho

Deixe de lado esse baixo astral Erga a cabeça enfrente o mal Que agindo assim será vital Para o seu coração É que em cada experiência Se aprende uma lição Eu já sofri por amar assim Me dediquei, mas foi tudo em vão

Almir Guineto

De maneira análoga para a língua inglesa, considerando uma mensagem cifrada 'fhvzpiehmmnpdlzvdt uoeazncdseaxmdwsoodedqaahpahaztljyhbgtyulrbuifyy igmfxfreimoiywedakqzvvqsamvnzzzeyunamnbrghgmfc cweuazdfutargwaruahpvfhpoebwwbneheeawqtktxysm wgryanlyggbxadexwtytwogsviaifvasefnwwehmupicxg bxnqeubzjrvudekppzrlsztrvjbulnsedjlzedseieosqqgezie bsmzldplywgbratmlcbsmyrrgmzxdczjezeiiewevoztymt vghrzekbpvqwodegmlbcuepewqymqfhgnbalaahcqsjicg zdkunxbsqfwhqfzzdbguuqgvvpznwodoebsmvqtqremeq gxsqsrltkglozaigznbyedlrbtvjrrpstwxjvqepwzbsiudnpf ltznzrdgljmybrgcqskzfkgxrqskwrmahrmtytzzrpibsluh pvfhxofsdzrdsanrjwmgkeseemcighdxoimxqcvuyijbsme hfarviwenbsrrvmqzbprqpvbtbvrnunamnbrghgmfccwe qozcyicipwedijbtkjrrpsvbn' em inglês, a quebra da cifra pode ser vista na figura 14.

Figure 14. Teste de quebra de cifra

É possível verificar que, de maneira inicial, o algoritmo determinou o melhor candidato para tamanho da chave como sendo 9. Sendo assim, foi calculada a possível chave de tamanho 9 e foi obtido 'marilinda'. Por fim, aplicando a decodificação da mensagem encriptada com a chave obtida, tem-se a seguinte mensagem:

'therearemanyvariationsofpassagesofloremipsumav ailablebuthmajorityhavesufferedalterationinsomeform byinjectedhumourorrandomisedwordswhichdontlooke venslightlybelievableifyouaregoingtouseapassageoflore mipsumyouneedtobesurethereisntanythingembarrassin ghiddeninthemiddleoftextalltheloremipsumgeneratorso ntheinternettendtorepeatpredefinedchunksasnecessary makingthisthefirsttruegeneratorontheinternetitusesadi ctionaryofoverlatinwordscombinedwithahandfulofmod elsentencestructurestogenerateloremipsumwhichlooksr easonablethegeneratedloremipsumisthereforealwaysfre efromrepetitioninjectedhumourornoncharacteristicwordsetc'

Aplicando os devidos espaçamentos na mensagem decodificada, é possível recuperar a mensagem original.

#### 4. Discussão e Conclusões

A implementação do algoritmo foi bem sucedida, foi possível cifrar e decifrar mensagens em inglês e português com e sem o uso da palavra-chave a partir dos métodos de frequência e índice de coincidência apresentados anteriormente.

A abordagem escolhida se mostrou muito eficaz para os casos em que as mensagens possuem frequência parecida com seu respectivo idioma, o que geralmente acontece em textos longos e sem muitas repetições de palavras, mas este resultado pode variar caso a mensagem escolhida seja uma tentativa forçada de burlar essa frequência com muitas repetições de palavras ou a escolha de idiomas diferentes do que está sendo testado, por exemplo.

Comparando os resultados previstos pela teoria com os obtidos, chegamos à conclusão de que todos os testes feitos alcançaram resultados positivos, isto pode ser confirmado anteriormente em "resultados", onde todos os criptogramas testados foram decifrados com sucesso, seja com ou sem o uso da chave.

Com a conclusão deste projeto os conceitos passados em sala de aula como técnicas de segurança computacional foram reforçados, abordando a cifra de Vigenère enquanto que novos conceitos foram introduzidos e causaram questionamento, acarretando em muita pesquisa e aprendizado.

# 5. Referências

- 1. Slides de Aula
- 2. YOUTUBE: Frequency analysis to crack Vigenere 1
- 3. YOUTUBE: Breaking the Vigenere using frequency analisis
- 4. YOUTUBE: Brute Force Attack Vigenère Cipher
- 5. YOUTUBE: Vigenere Cipher Decryption (Known Key)
- 6. YOUTUBE: Vigenere Cipher Decryption (Unknown Key)
- 7. YOUTUBE: Using the Index of Coincidence to Determine Vigenère Keyword Length
- 8. YOUTUBE: Index of Coincidence
- 9. YOUTUBE: Cryptanalysis of Vigenere cipher: not just how, but why it works