Criptoanálise - Cifra de Vigènere

Larissa Fiorini Martins

Escola Politécnica – Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS)
Porto Alegre – RS – Brasil

larissa.martins@edu.pucrs.br

Abstract. *This article describes an implementation of a program that finds clear text from a ciphertext encrypted with Vigenère Cipher, the key used for the encryption is unknown. The program behavior will be detailed, with a description of each decryption step.*

Resumo. *Este artigo descreve uma implementação de um programa que encontra o texto claro a partir de um texto cifrado com a Cifra de Vigenère, sendo a chave utilizada na cifragem desconhecida. Será detalhado o funcionamento do programa, com uma descrição de cada etapa de decifragem.*

1. Introdução

A cifra de Vigenère é uma cifra de substituição polialfabética que consiste em uma evolução da cifra de César. A cifra de César criptografa deslocando cada letra no texto simples um determinado número de vezes. Na cifra de Vigènere, ao invés de ser executado um único deslocamento em todo o texto simples, se usa uma chave para determinar valores de deslocamento diferentes em toda a mensagem, assim um caractere não vai ser cifrado sempre por outro mesmo caractere, sendo essa a força da cifra (Michigan Technological University, 2019).

Neste artigo será apresentada uma implementação em *Java* para a decifragem da cifra de Vigenère com chave desconhecida. Será mostrada uma descrição de cada etapa realizada, detalhando o funcionamento do código. A solução implementada espera receber um texto cifrado desconhecido, e, ao final, mostra ao usuário o texto claro em português ou inglês decifrado.

2. Descobrindo tamanho da chave

A primeira etapa da decifragem consiste na descoberta da chave utilizada para cifrar o texto, visto que apenas o texto cifrado foi fornecido. Para descobrir a chave, é necessário primeiramente descobrir seu tamanho, ou seja, quantos caracteres possui. Neste trabalho, a busca pelo tamanho da chave foi implementada utilizando o Teste de Kasiski. O Teste de Kasiski se baseia na fraqueza da cifra de Vigènere que é a repetição da chave, sendo possível procurar segmentos idênticos no texto que podem ter sido cifrados com o mesmo caractere da chave (Michigan Technological University, 2019).

O funcionamento do Teste de Kasiski implementado inicia pela busca por conjuntos repetidos de letras no texto, após, é realizada uma contagem do espaçamento entre as sequências. Por fim, são obtidos os fatores de espaçamento utilizando o algoritmo euclidiano, que permite encontrar o maior divisor comum de todas as distâncias, que será então o tamanho da chave. Para os exemplos de textos cifrados fornecidos (*"texto_cifrado.txt"*, *"20192-teste2.txt"*), as chaves encontradas pelo Teste de Kasiski foram de tamanho 7. Para o

exemplo de texto cifrado *"20192-teste1.txt",* que está em inglês, o Kasiski se mostrou muito lento devido ao tamanho do texto, por isso, o programa espera que o tamanho da chave seja fornecido para prosseguir com sua execução.

3. Busca de caracteres da chave

Após encontrar o tamanho da chave, a próxima etapa da decifragem consiste na busca dos caracteres que compõem a chave. Para esta busca, o texto cifrado foi dividido em uma matriz com *n* colunas, sendo *n* o tamanho da chave. Cada coluna corresponde à uma sequência de letras que foram cifradas com a mesma letra da chave. O número de linhas é descoberto pela divisão do tamanho do texto pelo tamanho da chave.

Após, cada coluna será analisada separadamente para descobrir cada caractere da chave, pois agora cada coluna pode ser tratada como uma cifra de substituição monoalfabética. Para cada coluna será então realizada uma análise de frequências, buscando primeiro quantas vezes cada letra do alfabeto aparece na coluna de texto cifrado. A ideia inicial dessa implementação era que a letra que mais aparece no texto cifrado fosse correspondente à letra mais frequente no idioma português ou inglês, e assim sucessivamente, porém, os valores podem ser próximos e não solucionarem o problema corretamente. Então, para uma maior precisão, foi utilizado o teste estatístico Chi-Square, que permite verificar qual letra do alfabeto resulta na frequência mais próxima à frequência da língua portuguesa/inglesa (Data Science Central, 2019). Na figura 2 (abaixo), "*O"* é a frequência observada no texto cifrado e *"E"* é a frequência esperada pelo alfabeto do português ou do inglês.

$$\chi_c^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

Figura 2. Chi-Square.

Para cada coluna foi calculado o Chi-Square de cada encriptação de cada letra do alfabeto, pois essa fórmula permite aproximar as frequências e verificar qual a mais provável de ser a letra correta. Será então escolhido o valor mínimo como letra da chave, visto que um valor pequeno resultante deste teste significa que é observada uma relação entre os dados. Esse processo se repetirá até que todos caracteres da chave sejam encontrados.

Conforme a escolha inicial pelo idioma do texto em português ou inglês, o programa irá utilizar as frequências de cada letra correspondente ao idioma para a análise de frequências. As chaves encontradas para os exemplos de textos cifrados fornecidos (*"texto_cifrado.txt"*, *"20192-teste1.txt" e "20192-teste2.txt"*) foram, respectivamente, "avelino", "meunome" e "meunome".

4. Decifrando texto

Agora com a chave conhecida, a próxima etapa consiste em decifrar todo o texto cifrado pelas Cifras de César simples agora conhecidas. A descriptografia é o mesmo processo da cifragem só que invertido, subtraindo a chave em vez de adicionar, para retornar ao valor original em texto. Então, basta utilizar o caractere da chave para descobrir o quanto o caractere cifrado foi deslocado, retornando o caractere para sua posição original. Repetindo

esse processo por todo texto cifrado irá resultar no texto claro até então desconhecido. Exemplo de parte do texto cifrado fornecido e parte do texto claro encontrado:

Arquivo *"texto_cifrado.txt"*:

Texto cifrado: qpixpnqiigzmahavrywfhiqidarocjvloradztfjywcvvfuywvmsnwzc…

Texto claro: quemhacincoentaannostivesseacoragemdepublicarumlivrocomo...

Arquivo *"20192-teste1.txt"*:

Texto cifrado: flcfsnsaocftavflyhgqsrehlczimrsjvqvqenacosexuarimfluyaawf...

Texto claro: thisebookisfortheuseofanyoneanywhereatnocostandwithalmost…

Arquivo *"20192-teste2.txt"*:

Texto cifrado: cyyzvmgurwbszxmehacexuzyfgqeoslnuqqpijhpxmoelhaxmhvipcys...

Texto claro: quemhacincoentaannostivesseacoragemdepublicarumlivrocomo…

5. Conclusão

O desenvolvimento desse trabalho me permitiu um maior entendimento do funcionamento da cifra de Vigenère e de qual abordagem deve ser feita no planejamento de algoritmos para decifragem de um texto em que se desconhece informações da chave utilizada na cifragem. Conclui-se que a implementação atingiu o objetivo e conseguiu corretamente determinar a chave e descobrir o texto claro.

Referências

Michigan Technological University (2019) "Kasiski's Method", https://pages.mtu.edu/~shene/NSF-4/Tutorial/VIG/Vig-Kasiski.html.

Learn Cryptography (2019) "Vigenère Cipher", https://learncryptography.com/classical-encryption/vigenere-cipher.

Michigan Technological University (2019) "The Vigenère Cipher Encryption and Decryption", https://pages.mtu.edu/~shene/NSF-4/Tutorial/VIG/Vig-Base.html

Data Science Central (2019) "Chi-Square Statistic", https://www.statisticshowto.datasciencecentral.com/probability-and-statistics/chi-square.