Trabalho 1 de Segurança Computacional

Estudante: Luís Henrique Araújo Martins
Matrícula: 221002058

Introdução

A cifra de Viginère é uma forma de criptografia baseada no deslocamento das letras das mensagens a partir de uma chave dada. Com isso, cada letra do texto é deslocada x vezes na sequência do alfabeto, em que x é o valor numérico da letra correspondente da chave. Para decifrar uma mensagem encriptada a partir da sequência de Viginère, basta fazer o procedimento inverso, ou seja, a partir da chave realizar o deslocamento na direção reversa.

É também possível decifrar uma mensagem sem que sua chave esteja disponível através da análise de frequência. De acordo com a distância entre letras repetidas que mais aparece em uma cifra, é possível decifrar qual é o tamanho da chave. A partir daí, pode-se dividir a mensagem em grupos e em cada um deles é feita uma análise de frequência observando as letras que mais se repetem na parte cifrada e no alfabeto de uma determinada língua para descobrir a chave. Em posse da chave, pode-se facilmente decifrar a mensagem.

Neste trabalho, foram implementadas funções que cifram e decifram textos a partir de uma chave e também decifram textos sem chave através da análise de frequência. Foi utilizada a linguagem *Python*.

Estrutura do código

Main

O arquivo *main.py* é o que deve ser rodado para executar o programa principal. A partir dele, será primeiro pedido o arquivo que contém a mensagem ou cifra que será a entrada do programa. Então, haverá uma opção para cada uma das funções presentes. Se for escolhido cifrar ou decifrar mensagem com chave conhecida, será pedido que seja digitada a chave. As funções do programa em si estarão no arquivo *viginere.py*

Cifra de mensagens

A função *cipher* é responsável por cifrar a mensagem a partir de uma chave dada. Tanto a cifração quanto a decifração das mensagens são feitas mantendo as letras maiúsculas como maiúsculas e minúsculas como minúsculas, além de não cifrar caracteres que não façam parte do alfabeto e os que têm acento.

A função consiste em, para todas as letras que fazem parte do alfabeto sem acento, somar o valor numérico de cada letra da mensagem à letra da chave na posição correspondente, de forma a posição da letra na chave é dada pelo resto da posição da letra na mensagem pelo tamanho da chave. Então, esse valor é diminuído em duas vezes pelo

valor numérico da letra a e feita a operação de resto por 26, para que seja um número entre 0 e 25 que será somado novamente à letra a para dar uma letra do alfabeto.

Decifração de mensagem

A função *cipher* tem funcionamento semelhante ao da função *cipher*, porém dessa vez é feita uma subtração do valor numérico da letra da mensagem pelo valor numérica da letra da chave. Então, caso seja um valor negativo, é subtraído da letra z e caso seja positivo é somado à letra a. Novamente, é preservado o *case* das letras e são mantidos os caracteres não pertencentes ao alfabeto.

Análise de frequência

Tamanho da chave

Para descobrir o tamanho da chave, foram coletadas todas as sequências de três letras das mensagens e foi analisada a distância entre cada uma das suas repetições. Para cada distância, para cada um dos divisores desse valor menores ou iguais a 20, foi aumentado em um a quantidade de ocorrências.

Então, no fim foi observado quais os números com maior valor de quantidade de ocorrências. Nem sempre o maior valor é o tamanho da chave, pois, por exemplo, para uma chave tamanho 8, haverá mais ocorrências de 2 e 4, já que são divisores de 8. Porém, essa diferença será baixa. Por isso, será mostrado ao usuário a quantidade de ocorrências para cada um dos valores e ele decidirá para qual tamanho de chave o ataque será executado. Caso a mensagem decifrada não seja o que o usuário desejava, ele poderá tentar um outro tamanho.

Ataque de recuperação da chave

De posse do tamanho da chave, deverá agora ser feita a análise de frequência propriamente dita. Para isso, foram coletadas as frequências de cada letra na língua portuguesa e na língua inglesa. Em seguida, foram retirados do texto os caracteres não pertencentes ao alfabeto. Depois, a mensagem foi dividida pelo tamanho da chave, em que cada letra ia para a parte dada pela operação resto entre a posição da letra da mensagem pelo tamanho da chave.

Então, para cada uma das partes, foi calculada a frequência para cada uma das letras. Após isso, foi multiplicada a frequência de cada letra na sua respectiva língua e na dada parte do texto, sendo somado tudo no final. Depois, com *i* indo até o 25 foi multiplicada a frequência de cada letra na língua utilizada pela letra deslocada *i* vezes na parte do texto determinada. O valor *i* que produzir o maior produto interno será então somado ao valor numérico da letra a e será uma letra da chave. É feito esse procedimento para cada uma das partes do texto separadas.

Por fim, é utilizada a função *decipher* passando a chave descoberta como argumento.

Arquivos

- en.txt: texto em inglês para ser codificado com chave
- de.txt: texto do arquivo en.txt codificado com a chave fishes

- attack_en.txt: texto em inglês para ser decodificado sem chave
- attack_pt.txt: texto em português para ser decodificado sem chave

Referências

- KATZ, Jonathan; LINDELL, Yehuda. Introduction to modern cryptography. CRC press, 2014.
- The Vigenère Cipher: Frequency Analysis. Michigan Technological University - Information Technology, 2023. Disponível em <https://pages.mtu.edu/~shene/NSF-4/Tutorial/VIG/Vig-Frequency-Analysis.html>. Acesso em 01 de outubro de 2023.
- Vigenère Cipher. Geeks for Geeks, 2023. Disponível em <https://www.geeksforgeeks.org/vigenere-cipher/>. Acesso em 01 de outubro de 2023.
- Frequência de letras. Wikipédia, a enciclopédia livre, 2022. Disponível em <https://pt.wikipedia.org/wiki/Frequ%C3%AAncia_de_letras>. Acesso em 01 de outubro de 2023.