Trabalho de Implementação 1

Segurança Computacional - Turma B

Leonardo Rodrigues de Souza - 17/0060543
Lucas Dalle Rocha - 17/0016641

Universidade de Brasília - UnB {170060543,170016641}@aluno.unb.br

Resumo Esse trabalho objetifica-se na descrição do primeiro trabalho da disciplina de Segurança Computacional da UnB, que consistiu na construção de um algoritmo para cifragem de texto, decifragem e ataque com a cifra de *Vigenère*.

Keywords: Criptografia · Encrypt · Decrypt · Vigenère · Segurança.

1 Introdução

O primeiro trabalho da disciplina de Segurança Computacional, da Universidade de Brasília, consistiu na construção de um algoritmo para cifragem e decifragem de texto, além de implementar um ataque à cifra de *Vigenère*. A cifra de *Vigenère* é um método de criptografia de texto, baseada na *Cifra de César*, em que cada letra do alfabeto é deslocada em um número fixo de posições. Por sua vez, a cifra de *Vigenère* implementa diferentes séries da cifra de César, baseadas em letras de uma senha, que é replicada ao tamanho da mensagem a ser cifrada. Dessa forma, é considerada uma cifra de substituição polialfabética, com decodificação mais complexa do que a *Cifra de César*, porém possível.

1.1 Cifração

Para cifrar um texto usando a cifra *Vigenère* é usado uma tabela de alfabetos, denominada *tabula recta*, que consiste em todas as combinações de símbolos do alfabeto e depende da quantidade de símbolos do alfabeto a ser utilizado. Cada linha é o próprio alfabeto deslocado em uma posição e a cifragem é feita pela busca da posição na matriz em que a letra da mensagem é a linha da matriz e a letra da chave é a coluna da matriz.

1.2 Decifração

Para decifrar, utiliza-se o método inverso, isto é, lê-se na coluna a respectiva letra da chave, e ocorre a busca por essa coluna da letra que representa o criptograma. Ao encontrar, a letra respectiva à mensagem é a letra da linha da matriz. Por possuir função inversa, é facilmente descrita, algebricamente.

2 Implementação

2.1 Cifrador

O processo de cifragem inicia-se com a passagem de uma mensagem e uma senha escolhida pelo usuário, a fim de gerar um criptograma que respeita as seguintes primitivas: símbolos que não estão contidos no alfabeto da cifra não são rotacionados do texto limpo, ou seja, são apenas copiados em suas respectivas posições para o texto cifrado obtido; símbolos pertencentes ao alfabeto, isto é, [A-Z] são rotacionados com base na chave passada pelo usuário, que em conjunto retorna uma posição da tabela de alfabetos que corresponde ao símbolo do texto cifrado.

2.2 Decifrador

Já o algoritmo de decifração, recebe um criptograma e também recebe uma senha escolhida pelo usuário, a fim de decifrar uma mensagem em que já se sabe a senha. Assim como no algoritmo de cifragem, símbolos não contidos no alfabeto são apenas repassados, em suas respectivas posições, para o texto limpo, enquanto símbolos que pertencem ao alfabeto são deslocados, sem necessitar da averiguação na tabela de símbolos, e recuperam a mensagem decifrada, dada a função algébrica [1] :

$$M_i = (C_i - K_i + 26) mod 26$$

Obs.: i representa o índice da mensagem limpa, do criptograma, e da chave, que é replicada ao tamanho do criptograma.

3 Ataque de Recuperação de Senha por Análise de Frequência

3.1 Determinando o tamanho da chave

Inicialmente, foi aplicado o método Kasiski [2] para termos uma noção de possíveis tamanhos da chave, e os mais prováveis. O método consiste em buscar por conjuntos de letras de tamanho n (buscamos por conjuntos de tamanho três em nossa implementação) que se repetem ao longo do criptograma, com o intuito de destacar a distância entre essas repetições, visto que pode representar uma repetição da chave durante o processo de cifragem. Dessa forma, busca-se encontrar divisores da distância entre as repetições, uma vez que a chave provavelmente tem tamanho respectivo a algum dos divisores.

Não obstante, o algoritmo para definição do tamanho da chave também calcula o índice de coincidência [3] do criptograma, isto é, divide-se o criptograma em l criptogramas e a caractere da posição n do criptograma original pertencerá ao novo criptograma $n\%l$, seja l o tamanho estimado da chave. Dito isso, busca-se fazer isso para todos os possíveis tamanhos da chave, ou seja, [2..20], e o índice de coincidência das letras de cada subconjunto do criptograma é calculado:

$$\mathrm{IC} = \frac{1}{N(N-1)} \sum_{i=1}^{n} F_i(F_i - 1)$$

Figura 1. Cálculo do índice de coincidência para estimar tamanho da chave de um criptograma.

A heurística utilizada para automatização desse processo foi, meramente, o índice de coincidência. A análise dos trigramas é disposta ao usuário apenas quando o criptograma é pequeno o suficiente para não popular, de modo significativo, a estrutura utilizada para verificar a frequência das letras, de modo que caberá ao usuário escolher o tamanho da chave.

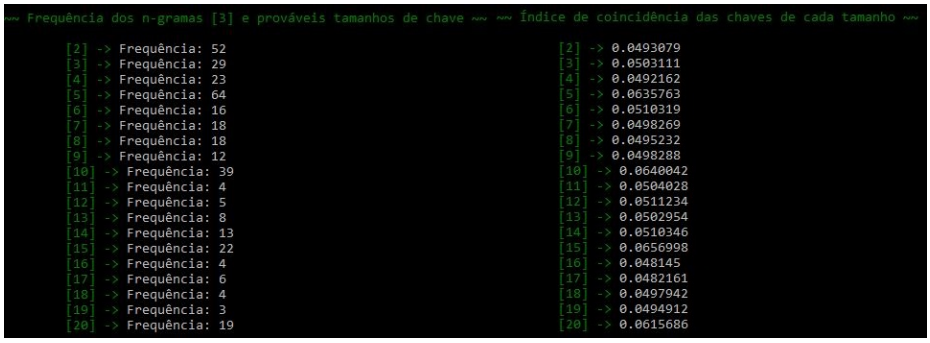

Figura 2. Valores de frequência dos trigramas para chaves de tamanho [2-20] à esquerda. Índice de coincidência, dada análise de frequência das letras no criptograma, para chaves de tamanho [2-20] à direita.

3.2 Descobrindo a chave

Por fim, quando se sabe o tamanho da chave, é possível recuperá-la com análise de frequência das letras, baseado no alfabeto da mensagem original. Para isso, utilizamos o mesmo processo de separação do criptograma em l criptogramas, com $n\%l$, assim como na determinação do tamanho da chave.

$$\chi^2 = \sum_{i=1}^{n} \frac{(f_i - F_i)^2}{F_i}$$

Figura 3. Cálculo do χ^2.

Destarte, calcula-se o χ^2 de cada subconjunto do criptograma com base no linguagem que se espera que a mensagem tenha sido escrita, e ocorre deslocamento à esquerda para todos os subconjuntos do criptograma, e para todas as possíveis combinações. Assim, ao final do processo, o deslocamento do subconjunto que obtiver o menor χ^2 representa a maior proximidade com a frequência do alfabeto escolhido inicialmente, e basta somar ao índice para obter a letra correspondente à chave.

Outro heurística foi usada na descoberta da chave, que consiste em calcular χ^2 baseado nas frequências da língua inglesa e da língua portuguesa, de modo que no final, tira-se a média χ^2 e a menor média corresponde à linguagem da mensagem original, sem que haja necessidade do usuário saber em qual linguagem o texto foi cifrado.

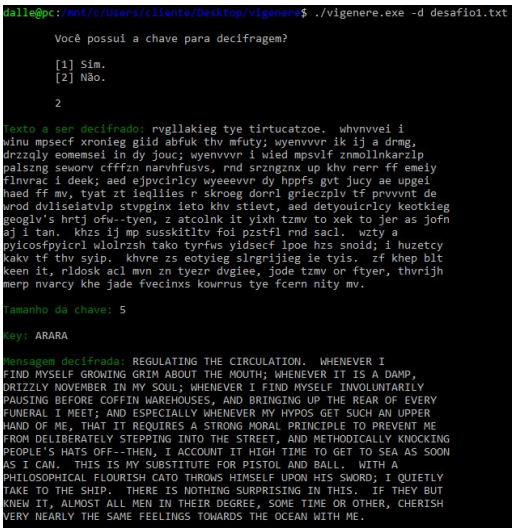

Figura 4. Execução da quebra da cifra de Vigenere.

Referências

1. Wikipedia contributors, Wikipedia, The Free Encyclopedia., Vigenère cipher, https://en.wikipedia.org/wiki/Vigen%C3%A8re_cipher, Last accessed 10 Aug 2021
2. Friedrich W. Kasiski, Die Geheimschriften und die Dechiffrirkunst, Mittler und Sohn, Berlin, 1863. (An unabridged facsimile version was published by Adamant Media Corporation in 2006.)
3. William F. Friedman, The Index of Coincidence and Its Applications in Cryptanalysis, Aegean Park Press, 1996. (This book was originally published in 1922 as Riverbank Publication No. 22, Riverbank Laboratories, Geneva, Illinois.) https://pages.mtu.edu/ shene/NSF-4/Tutorial/VIG/Vig-IOC.html, Last accessed 10 Aug 2021