Criptoanálise com Cifra de Vigenère

Leonardo A. Pasqualotto

Escola Politécnica – Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS)
Caixa Postal 1.429 – 90.619-900 – Porto Alegre – RS – Brasil

Leonardo.pasqualotto@edu.pucrs.br

Resumo. *Este artigo descreve os procedimentos utilizados na criptoanálise do primeiro trabalho da disciplina de Segurança de Sistemas do curso de Engenharia de Software da PUCRS – 2022/2. O trabalho consiste na criação de um programa que dado um texto cifrado encontre o texto claro utilizando o método de Vigenère.*

1. Contextualização e Etapas

O programa foi desenvolvido utilizando a linguagem de programação Java. A execução, restrições, inputs e outputs estão descritos no arquivo README.md no repositório https://github.com/leopasquargenton/Seguran-a-de-Sistemas.

Dada a leitura de um arquivo que contém um texto cifrado pelo método de Vigenère, o programa retorna o texto claro, juntamente com sua chave. O código é executado em 4 etapas:

- Cálculo do Tamanho da Chave pelo Índice de coincidência
- Identificação da Língua
- Descoberta da Chave
- Descriptografia

2. Cálculo do Tamanho da Chave pelo Índice de Coincidência

Assumindo que o texto claro está escrito em um idioma existente (inglês ou português), e que o método de Vigenère é baseado em uma cifra de substituição polialfabética, podemos utilizar o índice de coincidência como técnica de criptoanálise para descobrir o tamanho da chave. Ocorre que, ao contrário das cifras de substituição mais simples como a cifra de César, cada caractere é cifrado com um deslocamento diferente conforme o tamanho da chave, fazendo com que seja inviável calcular o índice de coincidência para o texto completo.

A solução é "quebrar" o texto utilizando somente os caracteres múltiplos de "n", sendo "n" o tamanho da chave. Dessa forma, o texto quebrado em blocos com o tamanho certo se comporta como uma cifra de César clássica, com deslocamento igual para todos os caracteres e o índice de coincidência pode finalmente ser aplicado.

Como não sabemos o tamanho da chave, nessa etapa o programa realiza o cálculo do índice de coincidência para textos quebrados com diferentes tamanhos de chaves. A repetição começa testando o tamanho da $chave = 1$ e aumenta incrementalmente de um em um. O critério de parada ocorre quando o cálculo retorna um valor próximo aos índices de coincidência dos idiomas existentes ($0{,}6 < ic < 0{,}8$).

3. Identificação da Língua

Assim que descobrimos o tamanho da chave e consequentemente o índice de coincidência do texto quebrado, podemos utilizar a tabela abaixo para descobrir a língua em que o texto foi escrito:

Língua	Índice de coincidência[1]
Alemão	2,05 (0,078)
Espanhol	1,94 (0,074)
Francês	2,02 (0,077)
Inglês	1,73 (0,066)
Italiano	1,94 (0,074)
Português	1,94 (0,074)
Aleatório	1,00 (0,0385)

Figura 1: índices de Coincidência

É importante lembrar que o texto, mesmo após quebrado, deve ter tamanho mínimo de caracteres suficiente para que o cálculo do índice coincidência não seja prejudicado.

No programa desenvolvido, considerou-se que os índices refletem os idiomas conhecidos da seguinte forma:

Inglês → $0{,}6 < ic < 0{,}7$

Português → $0{,}7 < ic < 0{,}8$

Texto Aleatório → $ic < 0{,}6$

4. Descoberta da Chave

Assim como na cifra de César, é necessário saber o deslocamento dos caracteres do texto quebrado. Conhecendo as letras mais repetidas em cada idioma através do https://en.wikipedia.org/wiki/Letter_frequency é possível fazer uma comparação com a letra mais frequente do idioma encontrado nos "n" blocos de texto quebrados. Dessa forma é possível aplicar o deslocamento para cada caractere da chave e finalmente descobrir o texto da chave em si.

5. Descriptografia

O texto claro é finalmente encontrado após a substituição de cada caractere cifrado pelo seu correspondente utilizando o deslocamento encontrado anteriormente. Cada bloco é juntado novamente e impresso na tela letra por letra.

References

Mička, Pavel. "Letter frequency (English)". Algoritmy.net. Archived from the original on 4 March 2021. Retrieved 14 June 2022. Source is Leland, Robert. Cryptological mathematics. [s.l.] : The Mathematical Association of America, 2000. 199 p. ISBN 0-88385-719-7Knuth, D. E. (1984), The TeXbook, Addison Wesley, 15th edition.