TRABALHO 1 DE SEGURANÇA DE SISTEMAS – VIGÈNERE[1]

Gabriel Ferreira Kurtz[2] <gabriel.kurtz@acad.pucrs.br>
Prof. Avelino Zorzo[3] – Orientador

Pontifícia Universidade Católica do Rio Grande do Sul – Faculdade de Informática – Curso de Ciência da Computação
Av. Ipiranga, 6681 Prédio 32 Sala 505 – Bairro Partenon – CEP 90619-900 – Porto Alegre – RS

23 de setembro de 2021

RESUMO

Este é o primeiro trabalho da disciplina de Segurança de Sistemas do curso de Engenharia de Software da PUCRS, realizado durante o segundo semestre letivo de 2021. O objetivo do exercício é criar um algoritmo para decifrar textos que utilizam a criptografia Vigènere através da aplicação de conceitos como Índice de Coincidência e Análise de Frequência.

Palavras-chave: Segurança de Sistemas; Criptografia; Vigénere; Índice de Coincidência; Análise de Frequência.

ABSTRACT

Title: *"Systems Security Task 1 – Vigènere"*

This paper is an academic work in paper format for the Systems Security discipline of the Software Engineering course at Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS – Pontifical Catholic University of Rio Grande do Sul). The exercise is to create an algorithm to decipher texts that use the Vigènere cryptography though the application of concepts such as Index of Coincidence and Frequency Analysis.

Key-words: *Systems Security; Cryptography; Vigènere; Index of Coincidence; Frequency Analysis.*

INTRODUÇÃO

O trabalho consiste na implementação de um algoritmo capaz de decifrar textos encriptados com Vigènere. Este foi um método de criptografia bastante utilizado alguns séculos atrás e, em determinado momento, considerado indecifrável. Com o aumento do conhecimento a respeito de assuntos como criptografia e linguística ao longo dos anos, foi possível encontrar padrões que possibilitaram a quebra desta criptografia.

A dificuldade em decifrá-lo consiste em que não é utilizada uma chave comum para todo o texto (como na criptografia César, por exemplo). É aplicada uma chave de múltiplos caracteres, cujo tamanho é variável, onde cada letra da chave é aplicada a um caractere do texto sucessivamente e repetidamente, até que aplique-se a chave ao texto inteiro.

Sendo assim, para decifrá-lo são necessárias algumas etapas: descobrir o tamanho da chave, localizar a chave correta a partir do tamanho, e finalmente decifrar o texto aplicando a chave descoberta.

TAMANHO DA CHAVE

Para descobrir o tamanho da chave, foi utilizado o Índice de Coincidência. Para cada linguagem, é possível saber uma média aproximada de, ao selecionar dois caracteres de um texto, que eles sejam o mesmo caractere do alfabeto. Utilizando esta média, que pode ser encontrada na Wikipedia, podemos comparar os Índices de Coincidência do texto cifrado. O segredo aqui consiste em separar o texto cifrado em vetores de texto, utilizando incrementos crescentes entre a distancia entre os caracteres simulando um tamanho de chave Vigènere (por exemplo, para simular uma chave

1 Trabalho realizado para a disciplina de Segurança de Sistemas do curso de Engenharia de Software da PUCRS.

2 Aluno do curso de Engenharia de Software da PUCRS.

3 Professor da disciplina de Segurança de Sistemas do curso de Engenharia de Software da PUCRS.

de tamanho 2, um vetor teria os elementos 0, 2, 4, ..., enquanto o outro teria 1, 3, 5, ...) e assim vamos separando em incrementos sucessivamente maiores, até que algum tamanho de incremento produza um Índice de Coincidência similar à média da língua. Este é o provável tamanho da chave.

Descobrindo a Chave

Sabendo o tamanho da chave, podemos utilizar a Análise de Frequência para ajudar a encontrar a chave. Cada língua possui letras (normalmente) mais frequentes, uma informação que também pode ser obtida na Wikipedia. Sabemos que a aplicação da chave na cifra Vigènere nada mais é do que o tamanho de um deslocamento de cada caractere no alfabeto. Ou seja, o caractere C, por exemplo, ao ser aplicada a chave B, sofreria um deslocamento tornando-se D (supondo que a chave A seja a ausência de deslocamento).

Assim, uma forma bastante simples de encontrar a chave é utilizar novamente os vetores que foram separados anteriormente que indicam que todos caracteres daquele vetor foram codificados com o mesmo caractere de chave, e deslocar todos eles até que a letra mais frequente entre eles seja a letra mais frequente esperada na língua (para o inglês, a letra E).

Poderia-se desenvolver um algoritmo um pouco mais complexo nesta etapa que considerasse, por exemplo, algumas das letras mais e menos frequentes do alfabeto ou mesmo todas as letras, fazendo algum tipo de soma das semelhanças. No entanto, o algoritmo que considera a primeira letra produziu um resultado correto e decidi mantê-lo.

Decifrando o Texto

Conhecendo a chave exata, basta fazer o processo oposto da criptografia Vigènere. Varremos o texto criptografado e, a cada caractere, desfazemos o deslocamento que foi usado para criptografá-lo de acordo com o deslocamento indicado na chave.

Resultado

Foi possível decifrar com sucesso o texto em inglês “20201-teste1.txt” bem como os textos bíblicos em português na pasta zipada, utilizando a letra E como a mais frequente. Para o texto “20201-teste2.txt”, em português, o algoritmo não funcionou em com as configurações em português nem em inglês. Supondo que não tenha sido um erro na implementação, creio que seria necessário implementar um pouco mais de inteligência ao algoritmo na fase de Análise de Frequência.

Texto em inglês cifrado:

“flcfsnsaocftavflyhgqsrehlczimrsjvqvqenacosexuarimfluyaawfriesexdmwgwareabnhesqzyemayyespcbcuxavjqmfeqnmavdiofsuxgrxrfflqxyeaesrxbrddsviwgugxqrvrfspugyagqmzgfhrqhimnuhtmeivbcwsdshywzim”

Texto decifrado:

“thisebookisfortheuseofanyoneanywhereatnocostandwithalmostnorestrictionswhatsoeveryoumaycopyitgiveitawayorreuseitunderthetermsoftheprojectgutenberglicenseincludedwiththisebookoronlinea”

Vídeo de Apresentação e Repositório

- Vídeo: https://youtu.be/9wP7pvhh6h8
- Repositório: https://github.com/gabrielkurtz/ss1-vigenere

Conclusão

Com o trabalho pudemos perceber a importância, ao criptografar, de não utilizar padrões identificáveis. Ou seja, não havendo um padrão aleatório, ocorrem pontos que podem ser atacados e utilizados para decifrar o texto como ocorreu eventualmente com o padrão Vigènere.