T2 Confiabilidade e Segurança de Software

1st Tomas Haddad Caldas
Escola Politécnica
Pontifícia Universidade Católica do Rio Grande do Sul
Porto Alegre, Brasil
tomas.caldas@edu.pucrs.br

I. Introdução

A cifra de Vigenere é um método de criptografia bem conhecido. Ela utiliza uma chave para realizar um deslocamento das letras no alfabeto, codificando o texto original. Diferente da cifra de César, que usa um único deslocamento para todo o texto, a cifra de Vigenère é polialfabética, pois aplica deslocamentos variados ao longo do texto, tornando a criptografia mais resistente a ataques de frequência.

O trabalho se baseia em conseguir quebrar a criptografia de uma cifra de Vigenere através de um ataque de força bruta utilizando o método de análise de frequência, para assim, poder decifrar um texto encriptado. Porém, como a cifra de Vigenere é resistente a esse tipo de ataque, é preciso antes encontrar o tamanho da palavra-chave usada para a cifra do texto. Isso é feito utilizando duas abordagens: o teste de Kasiski e o índice de coincidência. Com o tamanho da palavra encontrado, é possível aplicar a análise de frequência nos segmentos do texto que foram codificados com o mesmo deslocamento. Sabendo que o texto encriptado está em portugês, usamos a distribuição de frequência de caracteres da língua.

II. Desenvolvimento

Para quebrar a cifra de Vigenere foram utilizados ambos os métodos de Kasiski e do Índice de coincidência.

A. Método de Kasiski

O teste de Kasiski faz múltiplas comparações no texto cifrado. O método aproveita o fato de que quando sequências idênticas no texto plano são cifradas com a mesma parte da chave, estas produzem sequências cifradas idênticas. Assim, o programa varre o texto cifrado procurando sequências de 3 ou mais caracteres repetidas. Quando acha um par de sequências, mede a distância de caracteres no texto entre o começo de cada sequência. Quando temos todas as distâncias entre possíveis sequências encontradas, todas as distâncias têm o maior divisor comum destas calculado. Este resultado é o provável tamanho da cifra.

B. Índice de coincidência

O teste com índice de coincidência é uma análise estatística que utiliza do valor da probabilidade de dois caracteres aleatórios serem iguais em um texto. Este índice é específico para cada língua (Ic = 0,07797, no caso da língua portuguesa). O método se inicia dividindo o texto encriptado em L (tamanho da chave hipótese) grupos com cada grupo contendo caracteres cifrados com a mesma letra da chave. Para cada grupo se calcula o índice de coincidência médio local. O tamanho L que tiver o índice de coincidência médio local mais próximo do Ic da língua alvo é dado como o tamanho provável da cifra. Um exemplo posicional para os grupos com L = 3:

grupo 1 = (0,3,6,9,12,....)
grupo 2 = (1,4,7,10,13,....)
grupo 3 = (2,5,8,11,14,....)

graphicx

A implementação de código feita para quebrar a cifra adotou os dois métodos explicados acima, usando o índice de coincidência com valores L de 1 a 20. Após o texto encriptado e as frequências de cada caractere serem carregados, os métodos são aplicados para encontrar o tamanho da cifra. Como foram usados os dois métodos, Kasiski foi executado primeiro, seguido do teste de índice de coincidência, no caso de Kasinski não encontrar resultados satisfatórios o suficiente.

Com o tamanho da cifra encontrado, cada posição da cifra passa pelo processo de encontrar o melhor caractere a partir de decifrações parciais no texto encriptado. Para cada posição, o caractere que bater a melhor pontuação entre os 26 possíveis para cada posição da cifra, é o selecionado. Tendo o tamanho e caracteres da cifra encontrados, é executado o algoritmo de decriptação da cifra de Vigenere. Ao final da execução, o programa imprime a cifra encontrada com seu tamanho e as 200 primeiras caracteres do texto, agora decriptado.

III. Conclusão

Por fim, foi encontrado um tamanho de 18 caracteres para a cifra: 'teimarteimarteimar' e texto é do livro "Quincas Borba" de Machado de Assis. Os desafios enfrentados no desenvolvimento do trabalho estavam relacionados a encontrar parâmetros de aceitação para os resultados de execução do método de Kasiski e do teste do índice de coincidência.

References

[1] DOS, C. tipo simples de sistema de criptografia polialfabético. Disponível em: ¡https://pt.wikipedia.org/wiki/Cifra$_d e_V igen$
[2] DOS, C. Método Kasiski. Disponível em: ¡https://pt.wikipedia.org/wiki/M
[3] Index of coincidence. Disponível em: ¡https://en.wikipedia.org/wiki/Index$_o f_c oincidence >$
.