A quebra da cifra de Vigenère

João Marcelo Fantin Lerina

[1]Escola Politécnica – Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS)
Caixa Postal 1429 – 90.619-900 – Porto Alegre – RS – Brasil

joao.fantin@acad.pucrs.br

Abstract. *This article presents the solution to breaking Vigenère ciphers through frequency analysis. Cracking a cipher means to access its clear text without having the encryption key at hand. To solve the problem, techniques such as Friedman's index of coincidence and alphabetic substitution algorithms were needed. This task was given as the first assignment for the Systems Security class of the Software Engineering course, taught by Prof. Avelino Zorzo.*

Resumo. *Este artigo apresenta a solução para quebrar cifras de Vigenère através de análise de frequências. Quebrar uma cifra significa obter seu texto claro sem dispor da chave correspondente. Para resolver o problema, técnicas como o índice de coincidência de Friedman e algoritmos de substituições alfabéticas se fizeram necessários. Essa tarefa foi proposta como primeira avaliação da disciplina de Segurança de Sistemas do curso de Engenharia de Software, ministrada pelo Prof. Avelino Zorzo.*

1. Contexto

A cifra de Vigenère [Lamagna] é um exemplo de cifra de substituição polialfabética. Uma cifra de substituição polialfabética é semelhante a uma substituição monoalfabética, exceto que o alfabeto cifrado é alterado periodicamente durante a codificação da mensagem. Blaise de Vigenère desenvolveu o que hoje é chamado de cifra de Vigenère em 1585. Ele usou uma mesa conhecida como quadrado de Vigenère para cifrar mensagens.

Na prática, a cifra é composta por um conjunto variado de chaves de César. No entanto, a chave da cifra também exerce a função de alterar a forma dos caracteres originais baseado no deslocamento das posições entre cada caractere da chave com os caracteres pareados a ela no texto claro, continuamente. Para descobrir essa chave, é possível utilizar os métodos de Kasiski ou de Friedman, e o segundo foi implementado para solucionar o problema.

2. Problema

O problema envolve a quebra de um texto cifrado para revelar o texto claro criptografado com Vigenère. Isso não é possível antes de descobrir o tamanho da chave, que por sua vez se aproveita do estudo de frequência de caracteres entre diferentes idiomas, a falha fundamental dessa técnica de criptografia. A ação foi realizada através do cálculo de índice de coincidência para, em meio às permutas de caracteres da cifra, chegar ao índice da língua portuguesa, que tem tudo a ver com o valor estatístico das repetições de caracteres, e não necessariamente com o valor dos caracteres em si, mascarados uns pelos outros.

3. Solução

O efeito da chave na cifra é o de deslocar corretamente os caracteres para os valores originais, explorando a distância entre esses valores e os da própria chave. O primeiro passo é apontar o tamanho da chave, que pode ser revelado conforme proximidade com o de índice de coincidência, compreendido pela seguinte fórmula:

$$\frac{\sum_{i=0}^{25} fi(fi-1)}{n(n-1)}$$

Para se aproximar do possível tamanho de chave, utiliza-se uma tabela para armazenar diferentes índices de coincidência entre os tamanhos de chave maiores que 1 verificados. O tamanho 1 não é considerado, pois tornaria a cifra simples como a de César. O tamanho de chave então é aplicado na cifra através da fragmentação da mesma em *chunks* da mesma proporção, entendidos aqui como conjuntos de caracteres. Para cada um desses conjuntos é calculado o índice de coincidência médio, e quando esses cálculos tendem ao índice de coincidência do próprio idioma, nesse caso o Português, é encontrado um candidato para possível tamanho de chave.

Com o tamanho de chave em mãos, o texto é dividido de forma a coincidir com esse tamanho, e para cada um desses blocos assumimos que o *n-ésimo* caractere foi cifrado com o *n-ésimo* caractere da chave. O teste de hipóteses é realizado para procurar convergências entre os resultados de duas distribuições, em busca dos valores dos caracteres da chave.

Primeiramente, cada caractere de um bloco se desloca até a posição 25, da última letra do alfabeto, e utiliza no processo a letra mais frequente do alfabeto (para o Português, a letra A). Realizadas as flutuações, confere-se a ocorrência de cada caractere sobre o respectivo bloco, e para cada uma dessas ocorrências, ocorre o cálculo de distribuição sobre o número total de ocorrências. Por fim, a distribuição incrementa ao somatório e o resultado fica armazenado em lista, e o índice com menor valor na distribuição se torna candidato a deslocamento.

Após essas etapas, a letra mais frequente do alfabeto deve revelar o caractere da chave via soma dos índices.

4. Resultados

Foram disponibilizados 31 cifras para quebrar. Com base nos testes realizados, é seguro afirmar que a maioria (se não a totalidade) possui texto claro em Português. O programa lê o arquivo cifrado, realiza as análises de frequência, forja a chave e salva em um arquivo de saída o texto decifrado. Estes são alguns dos resultados:

arquivo	cifra	chave	texto claro
1	dzjdbisniiivtbrnflkshronqw...	cristian	bibliasagradatraducaojoaof...
2	eiwtldsvouddvbuddpkdrjjiri...	david	bibliasagradatraducaojoaof...
4	flvlzdgejlaudhvdxutdcnruow...	eduardo	bibliasagradatraducaojoaof...
8	imsnqlaonvrfiezokytcwuwovj...	hercilio	bibliasagradatraducaojoaof...
16	tkiftejsiyuoekjckonefbqhiq...	schuler	bibliasagradatraducaojoaof...

4.1. Conclusão

Este trabalho será inesquecível, tanto pela luta que passei em fazer sentido da teoria, quanto pelo *insight* que me fez gerar a respeito de *exploits* e ataques de força bruta contra sistemas de segurança digital. Obviamente, a cifra de Vigenère está completamente obsoleta e não diz respeito ao estado da arte em criptografia - mesmo assim, considero que essa tenha sido uma ótima primeira experiência para entender o cuidado necessário para proteger sistemas de forma consistente contra maneiras tão criativas de quebrá-los.

References

Lamagna, E. Classical cryptography: Vigenère cipher. `https://www.cs.uri.edu/cryptography/classicalvigenere.htm`. Accessed 24 Sep. 2021.