Cifra de Vigenère

Ingrid Carolina Maciel da Nóbrega

I. Introdução

A Cifra de Vigenère, durente muitos anos, foi considerada uma cifra inquebrável pelo fato de que ela não é vulnerável à análise de frequência. Como a cifra usa diferentes deslocamentos ao longo do texto, a mesma letra do texto original não será sempre convertida na mesma letra do texto cifrado. Por exemplo, se "G" fosse a letra mais comum no texto cifrado, poderia-se supor que representasse "A". Contudo, na Cifra de Vigenère, essa correspondência fixa não ocorre. Para decodificar essa cifra, utilizou-se o método de Kasiski, que realiza uma análise de frequência em trechos para identificar a chave usada na codificação de textos criptografados.

II. Metodologia

A principal fraqueza da Cifra de Vigenère está na repetição da chave. Se a palavra-chave "cifra" for utilizada, o fluxo de chave será "cifracifracifra...", fazendo com que a cada terceira letra do texto original seja criptografada com o mesmo deslocamento. Na prática, isso equivale a ter três Cifras de César intercaladas, onde cada uma pode ser quebrada separadamente por análise de frequência. A maior dificuldade, portanto, é descobrir o comprimento da palavra-chave.

Para decodificar a cifra, utilizou-se a análise de Kasiski, que identifica padrões repetidos no texto criptografado e analisa as distâncias entre essas repetições. Essas distâncias são então utilizadas para calcular os divisores, que determinarão o tamanho da chave empregada na cifra.

A. Comprimento da chave

Para determinar o comprimento da chave, primeiramente busca-se por padrões repetidos ao longo do texto (escolhidos arbitrariamente como uma sequência de 4 caracteres). Uma vez encontrados, as posições de cada padrão são armazenadas em um dicionário. Em seguida, para os padrões que ocorrem mais de uma vez, calculam-se as distâncias entre suas ocorrências.

Com as distâncias calculadas, obtêm-se os divisores de cada uma e organiza-se esses divisores pela frequência. Posteriormente, observou-se em todos os casos de exemplo que o divisor ideal é o quarto mais frequente, pois os primeiros geralmente correspondem a valores como 1, 2, 3 ou 4. Por fim, com o número de divisores, determina-se o comprimento da chave, que define a quantidade de substrings que serão geradas para definir a chave.

B. Obtenção da Chave

Dado o comprimento "n" da chave, em cada substring, realiza-se uma análise de frequência de cada letra e se compara a letra mais frequente dessa substring com a letra mais comum do idioma, que, no caso do português, é "A". Por exemplo, se a letra mais frequente em uma substring for "D", significa que há um deslocamento de 3 posições de "A" até "D". Fazendo isso para cada substring, a chave será revelada.

C. Decodificação

Para decifrar o texto, considerando que a chave se repete ao longo de todo o conteúdo, usa-se o caractere correspondente da chave para cada caractere cifrado. Esse caractere da chave determina o deslocamento necessário para identificar o caractere decifrado. A fórmula utilizada para a decodificação é:

$$decrypt=(cipherChar - keyChar+26) mod 26$$

Onde *cipherChar* é o caractere cifrado e *keyChar* é o caractere da chave. A adição de 26 assegura que o número seja positivo, enquanto o uso do módulo garante que os índices permaneçam dentro do intervalo do alfabeto. No exemplo abaixo, considerando que a chave seja "RASGOU" e uma frase cifrada, os seguintes passos seriam seguidos:

IEDOEOZAK...
RASGOURAS...

1) 1º caractere: I (texto cifrado) e R (chave)
   I → posição 8, R → posição 17
   (8 - 17 + 26) % 26 = 17 → **R**

2) 2º caractere: E (texto cifrado) e A (chave)
   E → posição 4, A → posição 0
   (4 - 0 + 26) % 26 = 4 → **E**

3) 3º caractere: D (texto cifrado) e S (chave)
   D → posição 3, S → posição 18
   (3 - 18 + 26) % 26 = 11 → **L**

4) 4º caractere: O (texto cifrado) e G (chave)
   O → posição 14, G → posição 6
   (14 - 6 + 26) % 26 = 8 → **I**

5) 5º caractere: E (texto cifrado) e O (chave)
   E → posição 4, O → posição 14
   (4 - 14 + 26) % 26 = 16 → **Q**

6) 6º caractere: O (texto cifrado) e U (chave)
   O → posição 14, U → posição 20
   (14 - 20 + 26) % 26 = 20 → **U**

7) 7º caractere: Z (texto cifrado) e R (chave)
   Z → posição 25, R → posição 17
   (25 - 17 + 26) % 26 = 8 → **I**

8) 7º caractere: A (texto cifrado) e A (chave)
   Z → posição 25, R → posição 17
   (0 - 0 + 26) % 26 = 0 → **A**

9) 7º caractere: K (texto cifrado) e S (chave)
   Z → posição 25, R → posição 17
   (10 - 18 + 26) % 26 = 18 → **S**

Por fim, a string "IEDOEOZAK", ao ser cifrada com a chave "RASGOU", resulta em "RELIQUIAS". Esse processo é repetido ao longo de todo o texto até que ele seja completamente decifrado.

III. Resultados

Para o texto cifrado "42294.txt", foi utilizada a chave "RASGOU" de comprimento 6, o que permitiu revelar a obra Relíquias de Casa Velha, de Machado de Assis, em cerca de 6 segundos. Desvendar essa cifra, considerada por muito tempo inquebrável, apresentou alguns desafios, especialmente na criação de um código dinâmico capaz de determinar o tamanho da sequência de letras do padrão a ser encontrado e escolher o divisor correto.

Embora o algoritmo decodifique o texto corretamente e em pouco tempo, algumas melhorias ainda são possíveis, como a adaptação automática da letra mais frequente de cada idioma para permitir a codificação em várias línguas. Por exemplo, em português, a letra mais comum é "A", enquanto em inglês e espanhol é "E". Atualmente, essa alteração precisa ser feita manualmente no código para decifrar o texto em outros idiomas.