Cifra de Vigenère

Segurança da Informação

Ualiton Ventura da Silva — 202033580

Maio 2023

1 Introdução

A cifra de Vigenère refere-se a um método de encriptação ao qual dada uma mensagem M e uma chave K, gera-se uma mensagem criptografada C, podendo ser descrito da seguinte forma:

Considere incialmente que cada letra do alfabeto pode ser mapeada para um número, onde "a"= 0, "b"= 1, "c"= 2, e assim sucessivamente.

Dada a mensagem **Atacar base inimiga** e a chave **guerra**, então encriptá-la através de Vigenere será:

1. Definir o deslocamento de cada letra da mensagem, relativo a chave:
   Atacar base inimiga
   guerra guer raguerr

2. Deslocar cada uma das letras utilizando a letra da chave associada.
   a + g = g
   t + u = n
   a + e = e
   c + r = t
   a + r = r
   r + a = r
   ...

Assim obtém-se a mensagem **"gnetrr huwv znogmxr"**, sendo que cada letra "remapeada"não possuirá o mesmo valor, em "atacar", o primeiro "a"torna-se "g", já o último torna-se "r". De maneira generalizada pode-se descrever através de:

Encriptação: $Ci \equiv (Mi + Ki)(mod 26)$
Decriptação: $Mi \equiv (Ci - Ki)(mod 26)$

2 Análise da Cifra de César

Para análise de Vigenère vamos inicialmente observar a cifra de César e como realiza-se a sua quebra.

A princípio a cifra de César pode ser descrita como uma forma simplificada de Vigenère, onde o deslocamento de todas as letras de uma dada mensagem será o mesmo.

Considere a mesma mensagem anterior, porém cifrada através de César. Escolhendo um deslocamento de 3(equivalente a "d"), tem-se:

a + d = d
t + d = w
a + d = d
c + d = f
a + d = d
r + d = u
...

Observe que para obter a mensagem inicial, é suficiente que subtraia-se 3, e uma propriedade a se notar trata-se de que uma mesma letra sempre será codificada para o mesmo símbolo.

Para quebrá-la é possível utilizar a análise de frequência das letras para a linguagem utilizada. Na língua portuguesa a letra mais utilizada é "a", ao obter o texto cifrado, bastaria verificar qual

símbolo é o mais recorrente, e existe grande possibilidade que seja a letra "a", por exemplo, caso "k"fosse a letra mais frequente, poderia-se inferir que $k = a$, portanto o deslocamento realizado fora de $k - a$.

A análise realizada acima não é suficiente, pois, podem existir casos onde a recorrência das letras são muito próximas, neste caso, deve-se analisar a distribuição do alfabeto como um todo e a distribuição das letras no texto criptografado, e então verificar qual deslocamento médio irá satisfazer as frequências do alfabeto.

3 Análise da Cifra de Vigenère

Conhecendo a cifra de César e como realiza-se a sua quebra, é possível deduzir o que poderia ser suficiente para a quebra da cifra de Vigenere. A princípio deve-se perguntar se existe a possibilidade de reduzi-la a uma cifra de César, e neste caso, a resposta é afirmativa.

Considere novamente a mensagem "Atacar base inimiga"e a chave "guerra", observe que algumas das letras da frase estão cifradas pela mesma letra, ou seja:

- **g** cifra: **a, b, i**
- **u** cifra: **t, a, m**
- **e** cifra: **a, s, i**
- **r** cifra: **c, e, g**
- **r** cifra: **a, i**
- **a** cifra: **r, n**

Analisando o primeiro grupo, temos que a, b, i estão todos deslocados por "g", assim, o problema pode ser reduzido a uma cifra de César para descobrir o primeiro elemento da chave, sendo que as frequências analisadas são relativas a estes elementos, contudo, esta redução não poderá ser feita de maneira imediata.

Utilizando esta análise inicial é possível estabelecer os passos para a sua quebra:

1. Determinar o tamanho da chave
2. Obter o agrupamento de letras de cada elemento da chave
3. Para cada agrupamento, realizar análise semelhante à utilizada para César através de suas frequências.

4 Determinando Tamanho da Chave

Para determinar o tamanho da chave utilizada pode ser executado o seguinte método:

1. Escolher o possivel tamanho k
2. Deslocar a mensagem criptografada por k unidades
3. Comparar as coincidências desta nova mensagem com a anterior.

Exemplo:
Para **"abcavcakcaucabc"**, pode-se inicialmente definir $k = 3$, então utilizando o método apresentado acima, realiza-se o deslocamento e então são feitas comparações, ou seja:

```
......abcavcakcauc
abcavcakcaucabc
......*...**...**...**...*
```

Sendo que $*$ indicam as coincidências. Basta realizar o mesmo processo para diferentes tamanhos e buscar aquele que possui o maior número de coincidências, apesar de não parecer uma análise válida a princípio, pois letras iguais não necessariamente possuem o mesmo valor, mais adiante será descrita as probabilidades de encontro entre letras.

5 Agrupamento de letras

Conhecendo o tamanho da chave(considere k), devem ser formados os grupos de letras, podendo ser descrito como:

Dado um elemento na posição l de um criptograma C, este elemento irá pertencer a um conjunto i se $l \equiv i(\mod k)$.

Ao realizar esta análise deve-se notar que um agrumento de letras poderá possuir elementos repetidos, e pelo fato de serem criptografados pela mesma letra, então são iguais. Portanto, considere que obtenha-se $k_1 = \{l_1, l_4, l_1\}$, e o criptograma possua tamanho n, assim, a frequência para a letra l_1 ao analisar-se k_1 será $\frac{2l_1}{n}$.

6 Análise de Frequências

Como descrito anteriormente, utilizar somente o termo mais recorrente para a quebra da cifra de César pode não ser suficiente, observa-se então as frequências como um todo.

Considere um texto cifrado C, com chave $k = k_1k_2...k_n$. Então, para determinar o valor de k_i, será:

- Determinar o agrumento de letras para i
- Determinar as frequências dos elementos em relação ao texto cifrado
- Buscar o produto escalar de frequências que maximiza o resultado.

Através do item anterior é possível estabelecer as frequências do agrupamento, sendo que os elementos não inclusos possuem frequência 0.

Para determinar o deslocamento utilizado é necessário definir se existe algum dentre eles que maximiza o resultado probabilistico. Exemplo:

Considere um agrupamento de g, que gera frequências $f = (f_a, f_b, ..., f_z)$ para $a, b, ..., z$ respectivamente(observe que são os elementos criptografados).

Para a língua portuguesa tem-se que o alfabeto de "a"a "z"possui as frequências $f' = (f'_a, f'_b, ..., f'_z)$, caso seja feita o produto escalar $f' \cdot f'$, e comparar-se este resultado com os produtos escalares de $f' \cdot f_s$, onde f_s refere-se a um novo vetor após a realização de deslocamento; para os 26 possíveis tipos de deslocamentos, $f' \cdot f'$ será na maioria dos casos o maior produto, o mesmo aplica-se ao inglês.

Utilizando o resultado anterior, o deslocamento poderá ser buscado através da maior frequência de $f \cdot f'$, onde f sofrerá 26 tipos de deslocamentos possíveis, bastando definir então o maior dentre eles.

7 Conclusão

O método apresentado apesar de ser uma dentre muitas abordagens para a resolução do problema, possui boa precisão para textos longos e chaves curtas, onde seja possível definir com facilidade frequências. Contudo, possui limitações para casos onde estas condições não são satisfeitas, pois torna-se um desafio estabelecer diferenças entre baixas frequências, e possívelmente sejam necessárias informações de contexto que favoreçam a análise de mensagem.

Referências

[1] Introduction to Cryptography With Coding Theory - Washington, Lawrence C