Vigenère

Maria Eduarda Casanova Nascimento[1]

[1]Escola Politécnica – Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS)

maria.casanova@edu.pucrs.br

Resumo. *Este relatório descreve a alternativa dada ao problema proposto na disciplina Segurança de Sistemas. A questão consiste em: dado um texto em português cifrado, encontrar o texto claro utilizando um dos métodos da cifra de Vigenère, Teste de Kasiski ou Índice de Coincidência. É apresentada uma solução utiliza o método de Índice de Coincidência, e a sua análise de complexidade. Por fim são apresentados os resultados encontrados para os casos de teste fornecidos e seus tempos de execução.*

1. Introdução

A cifra de Vigenère é uma técnica de criptografia por substituição polialfabética que utiliza uma série de cifras de César diferentes, baseadas nas letras de uma palavra-chave.

Para criptografar um texto usando cifra de Vigenère é necessária uma palavra-chave e a grade de Vigenère,conhecida por *tabula recta*, que consiste no alfabeto escrito 26 vezes em diferentes linhas, cada um deslocado ciclicamente do anterior por uma posição. As 26 linhas correspondem às 26 possíveis cifras de César. A palavra escolhida como palavra-chave contém o valor utilizada para deslocar os caracteres do texto, para cifrar ou decifrar a mensagem.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
B	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A
C	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B
D	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C
E	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D
F	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E
G	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F
H	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G
I	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H
J	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I
K	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J
L	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K
M	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L
N	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M
O	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N
P	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
Q	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
R	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
S	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
T	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
U	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
V	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
W	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
X	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
Y	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
Z	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y

Figure 1. tabula recta

Exemplo

Supondo que se quer criptografar o texto:

tutetornaseternamenteresponsavelporaquiloquecativas

com a chave:

maria

A primeira letra do texto, T, é cifrada usando o alfabeto na linha L, que é a primeira letra da chave. Basta olhar para a letra na linha M e coluna T na grade de Vigenère, e que é um F. Para a segunda letra do texto, ver a segunda letra da chave: linha A e coluna U , que é U, continuando sempre até obter:

fukmtareisqtvznmmvvtqrvapanjivqlgwrmqlqlaqlmcmtzdae

A cifra de Vigenère pode ser vista algebricamente. Utilizando as letras A–Z mapeadas para 0–25, e a adição módulo 26 for aplicada.

Codificação

$$C_i \equiv P_i + K_i (mod 26)$$

Decodificação

$$P_i \equiv C_i - K_i + 26 (mod 26)$$

Sendo C_i a letra na posição i do texto cifrado, P_i a letra na posição i do texto claro e K_i a letra da chave que corresponde a posição i.

O problema proposto apresenta um texto criptografado utilizando Vigenère com uma chave desconhecida. O objetivo então é descobrir essa chave e decifrar o texto

2. Solução

A implementação descrita nesse relatório utiliza o método de Índice de Coincidência proposto por William F. Friedman [Friedman 1922].

A solução para este problema pode ser divida em três partes, sendo elas: (1) encontrar o tamanho da chave, (2) encontrar a chave e (3) decifrar o texto.

Encontrar o tamanho da chave

Para encontrar o tamanho da chave utilizando o método de Índice de Coincidência, é necessário testar para n valores, assumindo que o tamanho certo esta dentro desse intervalo, criando assim n grupos de cifras $C = C_1, C_2, C_2...C_n$. Para cada das n possibilidades é criado um grupo de subsbtrings formados da seguinte forma:

$$C_1 = c_1 c_{n+1} c_{n+1} c_{n+1} ...$$

$$C_2 = c_2 c_{n+2} c_{n+2} c_{n+2} ...$$

$$C_n = c_n c_{2n} c_{3n} c_{4n} ...$$

Construindo dessa forma todas as substrings, sendo n o tamanho da chave, podemos calcular para cada substring gerada, seu índice de coincidência. O grupo com o

índice de coincidência mais próximo ao índice de coincidência da língua do texto, é o grupo com o tamanho correto.

O Índice de Coincidência é calculado com a seguinte fórmula:

$$IC(x) = \frac{\sum_{i=0}^{25} f_i(f_1 - 1)}{n(n-1)}$$

Sendo x uma sequência de n letras e f_i uma função que computa a frequência da letra no texto.

Encontrar a chave

Sabendo o tamanho da chave, podemos procurar qual a chave utilizada. Sendo n o tamanho da chave, se divide o texto em n blocos, cada bloco contém os caracteres que foram cifrados pelo mesmo caracter da chave, isso é, o primeiro bloco pegará todos caracteres de n em n a partir da primeira posição, o segundo bloco pegará todos caracteres de n em n a partir da segunda posição e assim consecutivamente.

Para cada bloco gerado se a letra mais frequente e assim se forma a chave. A solução implementada utilizou a primeira e a segunda letra mais frequente de cada bloco, e gerou todas as combinações de chaves a partir disso.

Decifrar o texto

Sabendo a chave correta, se aplica a fórmula de decodificação apresentada anteriormente.

3. Resultados

A solução apresentada encontra a chave correta na maioria dos casos e em alguns consegue decifrar parcialmente os textos. Na tabela estão alguns dos casos e seus respectivos resultados.

Arquivo	Tempo para encontrar a chave	Chave	Tempo para decifrar
cipher1	22.041706085205078s	cristian	1.9600341320037842s
cipher2	12.070812225341797s	david	1.9369750022888184s
cipher3	27.121154069900513s	diego	2.054938077926635s
cipher4	11.5815908908844s	eduardo	1.9772229194641113s
cipher5	–	–	–
cipher6	14.404611110687256s	girotto	2.0288729667663574
cipher7	23.378509998321533s	gregory	1.981942892074585s
cipher8	–	–	–
cipher9	–	–	–
cipher10	–	–	–
cipher1	22.041706085205078s	cristian	1.9600341320037842s
cipher2	12.070812225341797s	david	1.9369750022888184s
cipher3	27.121154069900513s	diego	2.054938077926635s
cipher4	11.5815908908844s	eduardo	1.9772229194641113s

Table 1. Resultados

4. Considerações Finais

Com esse método de cifra clássica é possível decifrar o texto sem nenhuma informação adicional de maneira muito rápida, sendo assim não é um método seguro de criptografia.

Na implementação notei algumas coisas que considerei interessante. Ao fazer as tentativas para encontrar o tamanho da chave, fazendo 20 tentativas, os possíveis tamanhos sempre eram múltiplos da primeira opção. Sendo assim consegui fazer uma pequena otimização no método, retornando sempre o primeiro valor com um índice de coincidência uma diferença menor a 0.01 em relação ao índice de coincidência da língua portuguesa.

Para a implementação do algoritmo apresentado acima foi utilizado Python. Para utilizar o programa gerado é necessário inserir o caminho do arquivo que se deseja testar no terminal, será gerado um arquivo de saída com as possíveis chaves. Para decifrar o código é necessário inserir a chave no terminal e o resultado será escrito no mesmo arquivo

References

Friedman, W. F. (1922). The index of coincidence and its applications in cryptanalysis.