Vigenere Solver (https://github.com/gmenti/vigenere-solver)

Giuseppe Ferreira Menti

O trabalho neste artigo tem como objetivo exercitar conceitos relativos à criptografia de textos utilizando vigenere, entender como funciona a criptografia e decriptografia dos dados e explorar suas fraquezas, descobrindo o tamanho da chave, a chave de segurança e após realizar a decriptografia dos dados de forma automatizada.

Introdução

O método de criptografia Vigenere utiliza cifras baseadas em letras de uma chave de segurança. É uma cifra fácil de entender e pôr em prática, parecendo inquebrável, porém facilmente quebrada por pessoas com mais conhecimento no assunto.

Utiliza-se das cifras de César para criptografar os dados, onde a letra é deslocada por um número fixo de posições, gerando uma nova letra, o método vigenere basicamente define esse valor de descolamento a partir da chave de segurança fornecida, fazendo com que a mesma letra possa se tornar diferentes letras, tornando o dado criptografado mais indefinido e aleatório, mas ainda sim explorável por criptoanalistas.

Criptografia / Decriptografia

Para realizar a criptografia, é utilizado uma matriz com as letras do alfabeto (26 letras), onde a combinação da letra com a letra da chave irá transformar em outra letra, realizando a cifra para todos os caracteres do texto desejado.

Exemplo, digamos que se quer criptografar o texto “bibliasagradatraducaojoao” utilizando a chave “jerusalem” e a repetindo até o comprimento do texto, então teremos:

Texto: bibliasagradatraducaojoao

Chave: jerusalemjerusalemjerusal

Texto criptografado: kmsfaadesaeuulrlhglefdgaz

A cifra de Vigenere também pode ser convertida para a álgebra, transformando as letras A-Z para números inteiros 0-25 e aplicando o mód26, teremos a seguinte fórmula:

Criptografia:

$$C_i \equiv P_i + K_i \pmod{26},$$

Decriptografia:

$$P_i \equiv C_i - K_i + 26 \pmod{26}.$$

	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

Figura 1 Grade de vigenere, utilizado durante a criptografia e decriptografia dos dados.

Descoberta da palavra-chave

Para descobrir a palavra-chave utilizada em um texto cifrado pelo método Vigenere, precisamos primeiro encontrar o tamanho da palavra-chave, para isso iremos precisar medir a frequência de petição de cada palavra no texto e depois calcular o índice de coincidência deste texto.

$$I_c(x) = \frac{\sum_{i=0}^{25} \binom{f_i}{2}}{\binom{n}{2}} = \frac{\sum_{i=0}^{25} f_i(f_i-1)}{n(n-1)} \approx \sum_{i=0}^{25} {p_i}^2$$

Figura 2 Fórmula para calcular índice de coincidência

Sabendo que cada linguagem (português, inglês) possui um índice de coincidência médio fixo, saberemos quando encontrarmos o tamanho da palavra-chave correto, pois será muito próximo ao índice da linguagem. Neste caso, foi obtido os valores abaixo a partir do cálculo do índice de coincidência de textos genéricos encontrados na internet, para descobrir o índice de cada linguagem.

Índice de coincidência inglês: 0.067

Índice de coincidência português: 0.083

Para cada linguagem também foi mapeado uma lista com as letras das linguagens ordenadas pela ordem de maior frequência, estes valores também foram descobertos analisando textos encontrados na internet para cada linguagem.

Inglês: e t a o i n s h r d l c u m w f g y p b v k j x q z

Português: e a o s r i d u m n t c l p v h q f b g j k z x w y

Com a análise de coincidência e frequência do texto cifrado já realizadas, o índice de coincidência de cada linguagem e as letras mais frequentes e o tamanho da palavra-chave já descobertos, podemos iniciar a decriptografia. Para isso, é necessário quebrar o texto cifrado com a distância do tamanho da chave e montar diferentes textos, para separar os textos por caráter de cifragem da palavra-chave, o caractere que mais se repete nestes blocos se corresponde ao caractere mais utilizado na linguagem, e sabendo disso teremos a distância de deslocamento do caractere do texto cifrado, e sabendo disto podemos fazer o mesmo processo para todos os outros blocos para obter o restante da palavra-chave.

Uso da ferramenta

Foi desenvolvida uma ferramenta em Node.js opensource [1] para realizar a automatização dessa tarefa de decifragem, para utilizar é necessário somente a dependência do node instalado na máquina e executar o arquivo “index.js” da raiz do projeto. Ele irá realizar a decifragem de todos os arquivos cifrados que estão na pasta “assets”, imprimir os resultados e salva-los em um arquivo JSON na raiz do projeto chamado “output.json”.

Para realizar a decifragem mais rápido, foi adicionado uma opção (flag –preview) ao executar a ferramenta para cifrar o texto a partir dos primeiros 10000 caracteres e não do texto inteiro, aumentando a performance de decifragem como um todo, porém imprimindo somente o início do texto decifrado ao invés de todo.

Decifragem completa do texto: node ./index.js

Decifragem otimizada de prévia: node ./index --preview

Resultados

Na tabela abaixo podemos visualizar os resultados gerados pela aplicação desenvolvida em Node.js [1] para todos os textos cifrados [2] disponibilizados pelo Professor Avelino Francisco Zorzo.

Arquivo	Palavra-chave	IC	Início do texto decifrado
Cypher0.txt	plato	0.066	neithermustweforgetthatthe...
Cypher1.txt	cristian	0.082	bibliasagradatraducaojoao...
Cypher2.txt	david	0.082	bibliasagradatraducaojoao...
Cypher3.txt	diego	0.082	bibliasagradatraducaojoao...
Cypher4.txt	eduardo	0.082	bibliasagradatraducaojoao...
Cypher5.txt	felipe	0.082	bibliasagradatraducaojoao...
Cypher6.txt	girotto	0.082	bibliasagradatraducaojoao...
Cypher7.txt	gregory	0.082	bibliasagradatraducaojoao...
Cypher8.txt	hercilio	0.082	bibliasagradatraducaojoao...
Cypher9.txt	maurer	0.082	bibliasagradatraducaojoao...
Cypher10.txt	rangel	0.082	bibliasagradatraducaojoao...
Cypher11.txt	jerusalem	0.082	bibliasagradatraducaojoao...
Cypher12.txt	software	0.082	bibliasagradatraducaojoao...
Cypher13.txt	igor	0.082	bibliasagradatraducaojoao...
Cypher14.txt	joaopedro	0.082	bibliasagradatraducaojoao...
Cypher15.txt	stein	0.082	bibliasagradatraducaojoao...
Cypher16.txt	schuler	0.082	bibliasagradatraducaojoao...
Cypher17.txt	marcelo	0.082	bibliasagradatraducaojoao...
Cypher18.txt	mateus	0.082	bibliasagradatraducaojoao...
Cypher19.txt	matheus	0.082	bibliasagradatraducaojoao...
Cypher20.txt	mathias	0.082	bibliasagradatraducaojoao...
Cypher21.txt	paulo	0.082	bibliasagradatraducaojoao...
Cypher22.txt	ritter	0.082	bibliasagradatraducaojoao...
Cypher23.txt	companhoni	0.082	bibliasagradatraducaojoao...
Cypher24.txt	cadaval	0.082	bibliasagradatraducaojoao...
Cypher25.txt	renata	0.082	bibliasagradatraducaojoao...
Cypher26.txt	ricardo	0.082	bibliasagradatraducaojoao...
Cypher27.txt	rodrigo	0.082	bibliasagradatraducaojoao...
Cypher28.txt	branco	0.082	bibliasagradatraducaojoao...
Cypher29.txt	kroth	0.082	bibliasagradatraducaojoao...
Cypher30.txt	virgilius	0.082	bibliasagradatraducaojoao...
Cypher31.txt	vitor	0.082	bibliasagradatraducaojoao...

Conclusão

Vigenere pode parecer muito bom e indecifrável no início, porém com um pouco de técnica e análise do texto podemos decifrar todo o texto. Por não gerar de forma aleatória e sim repetir de acordo com o tamanho da chave, deixa uma enorme fraqueza, pois tem um padrão identificável que poderá ser usado

por criptoanalistas para quebrar essa cifra, então sempre devemos gerar textos aleatórios sem padrões para manter a segurança dos dados.

Referências

1. Repositório do projeto no github, https://github.com/gmenti/vigenere-solver

2. Arquivos cifrados utilizados nos testes,

3. Vigenere Cypher, https://en.wikipedia.org/wiki/Vigen%C3%A8re_cipher

4. Vigenere Solver Online, https://www.guballa.de/vigenere-s