Lista 01

Vinícius de Oliveira Peixoto Rodrigues (245294)

Agosto de 2022

Nota

Não consegui subir os scripts que eu escrevi para esta atividade no Moodle, de modo que eu os coloquei em um repositório no Github (link).

Questão 1

No script `playfair.py` em anexo se encontra uma implementação de decodificação do Playfair. A mensagem resultante é `"ODCAEHPARTEDAFEXEC"`, que é claramente `O DCA É PARTE DA FEEC` (o `X` apareceu porque há dois `E` repetidos).

Questão 2

Item (a)

- Cada rotor tem 10 configurações iniciais (obtidas por meio de rotação), de modo que há $10^N = 10^4$ posições relativas entre os rotores (e, consequentemente, alfabetos distintos)

Item (b)

- Agora há 4! permutações entre os rotores, de modo que há agora $4! \cdot 10^4$ alfabetos

Item (c)

- Como não pode haver rotores iguais, o número máximo é de 10 rotores, de modo que resultam $10! \cdot 10^{10}$ alfabetos

Questão 3

O efeito avalanche é a propriedade de uma ferramenta criptográfica de produzir mudanças drásticas no saída mediante pequenas mudanças na entrada.

Existe também um critério de avalanche mais formal e probabilístico, que é satisfeito quando a mudança de um bit na entrada faz com que cada bits da saída tenha 50% de chance de trocar.

Esse efeito é definido (e desejável) tanto em relação à chave quanto ao texto de entrada, visto que se tanto um quanto o outro não exibissem esse efeito, seria possível se aproveitar da correlação entrada-saída como via de ataque estatístico.

Questão 4

Suponha que um atacante consiga acesso ao *comprimento* da chave. Usaremos como exemplo a chave `carta`, de comprimento 5, que será usada para cifrar um fragmento ($\approx$ 4.5k palavras) do primeiro capítulo de *"A Hora da Estrela"*, de Clarice Lispector. Todos os resultados apresentados aqui foram gerados a partir do script `vigenere.py`, que se encontra em anexo.

Conhecendo-se o k da chave, sabe-se imediatamente que todos os caracteres a uma distância múltipla de k uns dos outros foram cifrados com a mesma cifra de César (visto que cada linha da tabela de Vigenere é uma cifra de César com offset igual à distância de um caractere da cifra até o `'a'`). Desse modo, é possível estudar a distribuição de frequências para esses grupos. A imagem abaixo mostra a distribuição para o grupo com os caracteres na posição 1, 6, 11, 16, ...

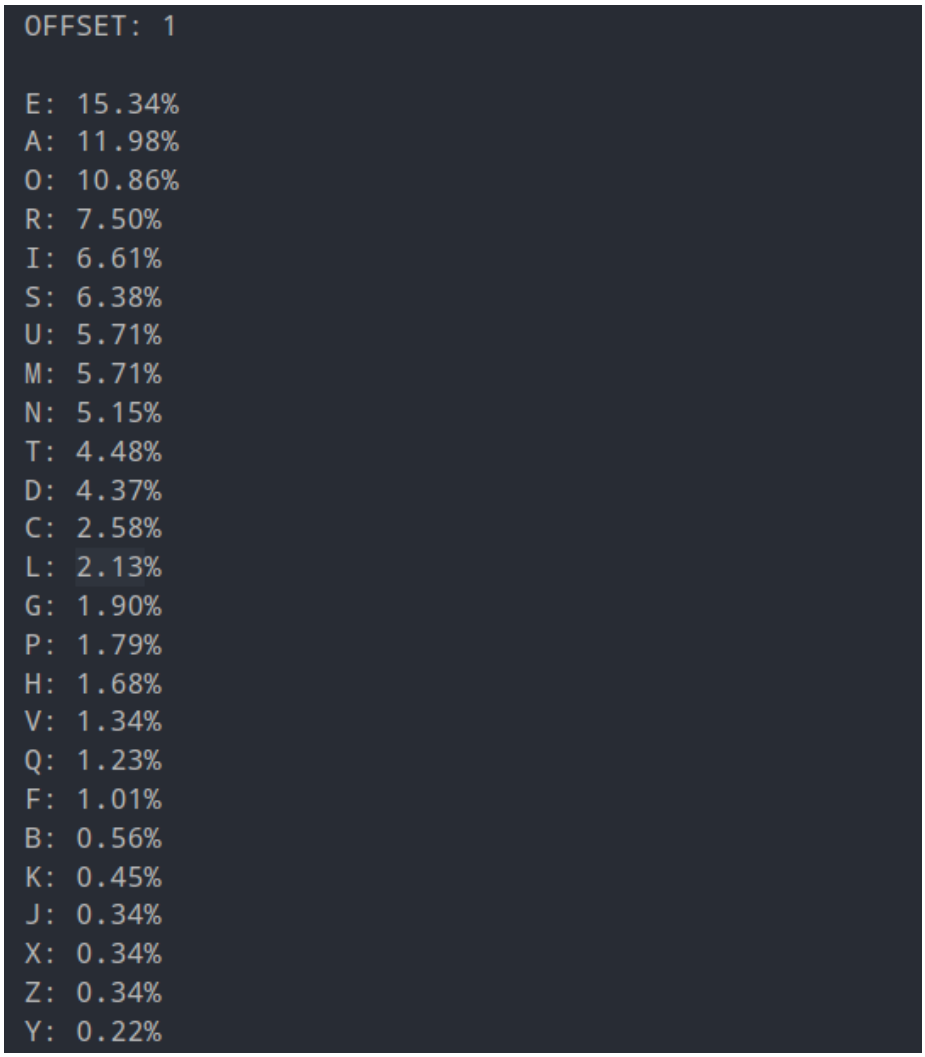

Figura 1: Distribuição de frequência dos caracteres em posições $\equiv 1 \mod 5$

Da tabela, é possível perceber que há um grupo de três caracteres mais frequentes que os demais. Os dados para todos os grupos apresentam essa regularidade (e estão em anexo no arquivo `vigenere_statistics.txt`).

Na Wikipedia, é possível encontrar uma tabela de frequência de letras no português:

Letra	Frequência
a	14.63%
e	12.57%
o	10.73%
s	7.81%
r	6.53%
i	6.18%
n	5.05%
d	4.99%

Figura 2: Frequência de letras no português

A partir desses dados, que nos informa que o grupo das três letras com frequência mais alta são `a, e, o`, é possível descobrir a chave.

Por exemplo, seguindo os dados da Figura 1, vemos imediatamente que o grupo `e, a, o` tem a frequência mais alta, de modo que o caractere na posição 1 da chave deve ser `_a___`.

```
OFFSET: 2

V: 13.33%
R: 11.09%
F: 10.19%
Z: 7.73%
J: 7.50%
```

Figura 3: Frequência de letras em posições $\equiv 2 \mod 5$

Do grupo acima, vemos que `v, r, f` deve corresponder a alguma permutação de `a, o, e`. Analisando com cuidado:

- `'o' - 'a' = 14`
- `'o' - 'e' = 10`
- `'e' - 'a' = 4`

Comparando com grupo cifrado:

- `'v' - 'r' = 4` $\Rightarrow$ `'a' -> 'r'`, `'e' -> 'v'`

De modo que se encontra mais um caractere da chave: `_ar___`
Por meio desse processo é possível quebrar a chave inteira.

Questão 5

O one-time pad é uma extensão natural da cifra de Vigenere, visto que a vulnerabilidade descrita na questão anterior advém do fato de a chave ter tamanho menor que o texto e portanto ter que ser concatenada, fazendo com que porções igualmente espaçadas do texto sejam cifradas "juntas". Essa regularidade é eliminada no one-time pad, onde se use uma chave aleatória do tamanho do texto (de modo que a ausência de um padrão estatístico na chave implica na ausência de um padrão estatístico no texto cifrado).

Questão 6

Os alemães, durante a Segunda Guerra, usavam 5 aspectos do *Enigma* como chaves:

- A ordem dos rotores
- A posição do anel ajustável do alfabeto em relação à fiação cada rotor
- As conexões no *plugboard* da máquina
- A configuração do refletor reconfigurável
- A posição inicial dos rotores

Todos esses fatores juntos funcionam como a "chave"da cifra.

Questão 7

Difusão e confusão são dois conceito relacionados que têm origem nos trabalhos de Claude Shannon, pai da Teoria da Informação.

- **Difusão** se refere ao obscurecimento de traços estatísticos do texto em claro no texto cifrado. Isso normalmente é alcançado por meio de várias iterações seguidas de operações de "embaralhamento", como substituição e permutação (por exemplo, na permutação das S-boxes do DES e no ShiftRows/MixColumns do AES). Serve para prevenir ataques estatísticos.
- **Confusão** se refere a tornar a relação entre a chave e o texto cifrado o mais complexa e imprevisível possível. Isso é importante para garantir que mesmo com um número muito grande de pares P-C, ainda seja muito difícil obter informação sobre a chave.

Questão 8

Ataques estatísticos tomam vantagem de deficiências de difusão para encontrar informação sobre chaves (ou até sobre o próprio texto em claro) a partir da análise estatística do texto cifrado. Um exemplo é a sequência de passos apresentada na Questão 4 para quebrar a cifra de Vigenere.

Como mencionado na Questão 7, o uso de algoritmos criptográficos com alta difusão torna difícil o ataque estatístico.

Questão 9

Dados:

Meu nome: `VINICIUSDEOLIVEIRAPEIXOTORODRIGUES` (comprimento 34)
Meu RA: `245294` → `k1 = 4`
`k2 = 7 2 1 8 3 0 5 6 4`

Cifrar

Inicialmente:

0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
v	i	n	i	c	i	u	s	d
e	o	l	i	v	e	i	r	a
p	e	i	x	o	t	o	r	o
d	r	i	g	u	e	s	0	0

Em seguida, concatenamos as colunas na ordem da chave:
Resultado: `"SRRONLIIIOERDAOOIIXGVEPDIETEUIOSCVOU"`

s	n	i	d	i	v	i	u	c
r	l	o	a	i	e	e	i	v
r	i	e	o	x	p	t	o	o
0	i	r	0	g	d	e	s	u

Decifrar

Para decifrar, calculamos o número de colunas dividindo o tamanho da cifra pela chave `k1`: `36/4 = 9`, de modo que temos os grupos

7	**2**	**1**	**8**	**3**	**0**	**5**	**6**	**4**
SRR0	NLII	IOER	DAO0	IIXG	VEPD	IETE	UIOS	CVOU

Reorganizando novamente em colunas de acordo com as posições na chave:

v	i	n	i	c	i	u	s	d
e	o	l	i	v	e	i	r	a
p	e	i	x	o	t	o	r	o
d	r	i	g	u	e	s	0	0

De onde obtemos de volta o texto em claro
`"VINICIUSDEOLIVEIRAPEIXOTORODRIGUES"`.

Questão 10

Partindo-se do pressuposto que o algoritmo criptográfico usado é conhecido pelo atacante (princípio conhecido como máxima de Shannon), é possível delinear algumas categorias:

- Texto cifrado conhecido, quando o atacante só tem acesso a um conjunto (potencialmente grande) de texto encriptado
- Texto em claro conhecido, quando o atacante tem acesso a um conjunto de pares P-C
- *Chosen-plaintext/chosen-ciphertext*, quando o atacante consegue obter texto cifrado a partir de texto em claro conhecido ou vice-versa
- *Adaptive chosen-plaintext/chosen-ciphertext*, quando o atacante escolhe premeditadamente textos em claro baseado em informações obtidas de ciframentos anteriores