

Criptanálise

Definição (Criptoanálise)

Criptanálise é o estudo dos procedimentos necessários para tentar comprometer as técnicas criptográficas, e mais genericamente, os serviços de segurança da informação.

- Stinson, Douglas, *Cryptography: Theory and Practice*, CRC, 2006.
- Pedro Quaresma, Augusto Pinho, Análise de Frequências da Língua Portuguesa, Livro de Actas InterTIC 2007, 3 a 5 de Dezembro de 2007, Porto, Portugal, pags 267-272, IASK, 2007.

Criptanálise - Objectivos do Adversário

Cifra (parcialmente) Quebrada O objectivo principal de um adversário que queira atacar uma cifra é o de recuperar, de forma sistemática, texto claro a partir de texto cifrado. Se este objectivo for atingido diz-se, que a cifra foi parcialmente quebrada.

Cifra (formalmente) Quebrada Um objectivo mais ambicioso é o de obter a chave privada de uma dada entidade, nesse caso a cifra é completamente, e formalmente, quebrada.

Classes de Ataques

Definição (Ataque Passivo)

Um ataque passivo é um ataque em que o adversário só monitoriza o canal de comunicação. Um atacante passivo só ameaça a confidencialidade da informação.

Definição (Ataque Activo)

Um ataque activo é um ataque em que o adversário tenta apagar, acrescentar, ou de alguma forma modificar a informação. Um atacante activo ameaça a integridade da informação assim como a confidencialidade.

Ataques no Esquema de Encriptação

O objectivo dos ataques que a seguir se descrevem têm como objectivo a obtenção do texto claro a partir do texto cifrado, ou mesmo a dedução da chave de decifração.

- 1 **ataque de texto cifrado**, é um ataque aonde o adversário (o criptoanalista) tenta deduzir a chave de decifração ou o texto claro por observação unicamente do texto cifrado. Um esquema de encriptação que seja vulnerável a este tipo de ataque considera-se **completamente inseguro**.
- 2 **ataque de texto claro conhecido**, é um ataque aonde o adversário consegue obter um excerto de texto claro e o correspondente texto cifrado. Este tipo de ataque é tipicamente só marginalmente mais difícil a montar.
- 3 **ataque de texto claro escolhido**, é um ataque aonde o adversário escolhe o texto em claro obtendo de seguida o correspondente texto cifrado. Toda a informação daí deduzida é posteriormente usada em outros textos cifrados.

Espaço das Chaves

A dimensão do espaço da chaves é dado pelo número de pares encriptação/desencriptação que está disponível no sistema de cifração.

A dimensão do espaço das chaves é importante dado que está directamente relacionada com a possibilidade, ou não, de um ataque directo por procura exaustiva, usualmente designado por *método da força-bruta*.

Facto

Uma condição necessária, mas de uma forma geral não suficiente, para um esquema de encriptação ser seguro é que o espaço das chaves seja suficientemente grande para evitar procuras exaustivas.

Por exemplo, para uma cifra de chaves simétrica em que o espaço das chaves seja da ordem de 3×10^{22} , a procura exaustiva já se torna impraticável, no entanto é um sistema com um grau de segurança baixo, dado ser susceptível a outro tipo de ataques.

Ataque por Procura Exaustiva

Um dos ataques possíveis, para uma qualquer cifra, é o **ataque por procura exaustiva no espaço das chaves**. Este ataque é também designado por **ataque de força-bruta**.

Ou seja tenta-se, de forma exaustiva, todas as chaves possíveis.

Uma cifra que seja susceptível a este tipo de ataque é considerada uma **cifra fraca**.

Todas as cifras clássicas são cifras fracas (ou completamente inseguras).

- Deslocamento Simples - $|\mathcal{A}|$ - extremamente fraca
- Deslocamento Linear - $\varphi(|\mathcal{A}|) \times |\mathcal{A}| + |\mathcal{A}|$ - muito fraca
- Cifra de Vigenère - $|\mathcal{A}|^{\frac{|\mathcal{A}|(2+|\mathcal{A}|)}{2}}$ - fraca

Exercício Prático 2

Implemente o método de ataque directo para a cifra de deslocamento simples.

- Alfabeto Português incompleto $\mathcal{A} = \{a-z\}$.
- Cifração carácter a carácter.
- Utilize uma lista de palavras Portuguesas para, de forma automática, avaliar que a cifra foi, ou não, quebrada.
- Interface Entrada/Saída: ficheiros e linha de comando.
- Implementação em C (ou C++).

Análise de frequências

Os métodos de substituição, tanto mono-alfabéticos, como poli-alfabéticos, preservam todas as características de uma dada linguagem.

Por exemplo se todas as ocorrências da letra *a* são substituídas pela letra *x*, uma mensagem cifrada contendo numerosas instâncias da letra *x*, iria sugerir ao criptoanalista, que a letra *x* representa a letra *a* (em Português).

- Frequência relativa das letras.
- Digramas, Trigramas.
- Letras iniciais e finais das palavras.
- Palavras pequenas: palavras de uma; duas; três letras.
- Índice de Coincidência; Índice de Coincidência Mútuo.

Análise de Frequência — Um exemplo

Considere-se a seguinte mensagem, resultado da aplicação de uma cifra de deslocamento simples, a um dado texto.

*mcolr fívdv r pdlv lpsruwdqwh lpshudgru urpdqr
ghvorfdyd dv ohwudv gd phqvdjhp ruljlqdo wuîv
svlêahv sdud hylwdu txh r lqlpljr ôî-vh rv vhxv
sodqrv.*

- Tendo o texto(s) encriptado(s) faz-se o estudo de frequências para esse(s) texto(s).
- Comparam-se os valores encontrados com um dado estudo de referência (para a língua de base do texto).
- Testam-se os valores encontrados para as chaves de forma exhaustiva. Este conjunto de valores será muito menor que o dado pela dimensão do espaço das chaves.

53 / 246

Análise para a Língua Portuguesa

Para fazer um estudo estatístico de uma dada língua é necessário escolher um conjunto amplo e significativo de textos, os quais devem poder representar fielmente a língua em estudo.

O conjunto escolhido deve respeitar dois critérios principais:

- a dimensão, isto é, o número de caracteres contados;
- o tipo de textos escolhidos. Os textos devem ser de diferentes tipos, cobrindo muitos autores, contextos históricos, e tipos literários.

Para o Português moderno, escolheram-se autores “recentes”. No total analisaram-se 141 textos, de 47 autores, totalizando 11.133.372 caracteres e 2.400.295 palavras.

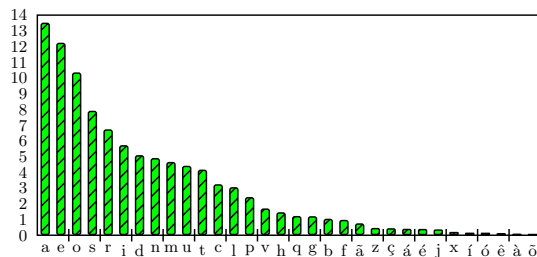
<http://www.mat.uc.pt/~pedro/cientificos/Cripto/>

54 / 246

Frequência Relativa das Letras

As letras mais frequente na mensagem cifrada foram:

Letra	d	v	r
Frequência	16	14	13



Temos então como possíveis chaves:

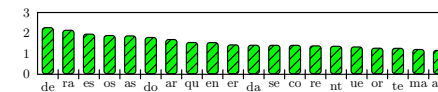
Letra	a	e	o
d	3	42	32
v	21	17	7
r	17	13	3

55 / 246

Digramas & Trigramas

Os digramas e os trigramas são conjuntos de duas ou três letras seguidas e que constituem sub-palavras.

Eles vão dar uma ideia das “vizinhanças” que se encontram numa dada língua.



Na mensagem temos como digramas mais frequentes: “dq”, “du”, “lp”, “ru”, “rv”, “ud”.

	de	ra	es	os
dq	(0,12)	(14,16)	(1,2)	(11,2)
du	(0,16)	(14,20)	(1,41)	(11,41)
lp	(8,11)	(22,15)	(36,3)	(3,3)
ru	(14,16)	(28,20)	(30,41)	(40,2)
rv	(14,17)	(28,21)	(30,40)	(40,40)
ud	(17,42)	(3,3)	(27,15)	(37,15)

56 / 246

Conclusões

+ letras iniciais e finais das palavras + palavras pequenas

No fim ter-se-ia uma ou duas chaves prováveis, no caso presente a chave mais provável é dada por $c = 3$.

$$d_c(y) = (y - c) \bmod 43$$

Obter-se-ia:

*júlio César o mais importante imperador romano
deslocava as letras da mensagem original três posições
para evitar que o inimigo lê-se os seus planos.*

Os dados completos deste estudo estão disponíveis em:

<http://www.mat.uc.pt/~pedro/cientificos/Cripto/>.

Exercício Prático 3

Implemente o método de análise de frequências para a cifra de deslocamento simples.

- Alfabeto Português completo
 $\mathcal{A} = \{a-z, à, á, â, ã, ç, è, é, ê, ì, í, ò, ó, ô, õ, ù, ú, ü\}$.
- Cifração carácter a carácter.
- Através da análise de frequências das letras do texto cifrado encontre os 5 valores mais prováveis para a chave.
- Teste os valores encontrados por comparação com uma lista de palavras Portuguesas.
- Interface Entrada/Saída: ficheiros e linha de comando.
- Implementação em C (ou C++).

Criptanálise de Cifras Poli-alfabéticas

A vantagem das cifras poli-alfabéticas em relação às mono-alfabéticas advém do facto de a frequência das letras não ser, pelo menos de forma directa, preservada.

No entanto assim que o comprimento do bloco é descoberto podemos dividir as letras da mensagem cifrada em grupos, tantos quanto o comprimento do bloco, e fazer o estudo de frequências dentro de cada grupo.

- 1 Descobrir o comprimento da chave - Índice de Coincidência.
- 2 Descobrir a chave - Índice de Coincidência Mútuo.

Índice de Coincidência

No caso da cifra de Vigenère com uma chave de comprimento m , k_1, k_2, \dots, k_m , temos que a mesma sub-chave k_i é usada de m em m caracteres. Podemos usar essa característica para quebrar a cifra.

Definição (Índice de Coincidência)

Suponhamos que $x = x_1x_2 \dots x_n$ é um texto de n caracteres do alfabeto. O índice de coincidência de x , designado por $I_c(x)$, é definido como a probabilidade de dois caracteres aleatórios de x serem iguais.

$$I_c(x) = \frac{\sum_{i=1}^{|\mathcal{A}|} (f_i)^2}{\binom{n}{2}} = \frac{\sum_{i=1}^{|\mathcal{A}|} f_i(f_i - 1)}{n(n - 1)}$$

com f_i a frequência do carácter de codificação i na mensagem.

Índice de Coincidência

Calculando o índice de coincidência para a língua Portuguesa é de esperar que um texto em Português se aproxime desse valor.

$$I_c(x) \approx I_c(Pt) = 0,072723$$

Temos então:

$$i \leftarrow 2$$

enquanto ($i \leq |\mathcal{A}|$) faz

divide o texto em i sub-textos, (de i em i caracteres)

calcula o índice de coincidência para cada um dos sub-textos

faz-se a média dos textos com chave de comprimento i

Dado que só os sub-textos obtidos a partir de uma sub-divisão com um comprimento igual ao da chave é que vão corresponder a sub-textos de um texto em Português (a menos da encriptação), só esses é que terão um valor de índice de coincidência próximo do índice de coincidência do Português.

Índice de Coincidência Mútuo

Agora resta-nos calcular cada uma das sub-chaves $K = k_1, \dots, k_l$, (com l o comprimento da chave).

Não podemos fazer o estudo das frequências das letras directamente, tal como fizemos para a cifra de deslocamento simples, dado que não temos um texto, mas sim sub-textos.

Definição (Índice de Coincidência Mútuo)

Suponhamos que temos dois textos $x = x_1x_2 \dots x_m$ e $y = y_1y_2 \dots y_n$. O índice de coincidência mútuo de x e y , designado por $I_{cm}(x, y)$, é definido como a probabilidade de um carácter aleatório de x ser igual a um carácter aleatório de y .

$$I_{cm}(x, y) = \frac{\sum_{i=1}^{|\mathcal{A}|} f_i^x f_i^y}{mn}$$

com f_i^x, f_i^y as frequências do carácter de codificação i em x e y respectivamente.

Usando as frequências da língua Portuguesa como referência, podemos obter desta forma as diferentes sub-chaves.

Índice de Coincidência Mútuo

Sabendo já de antemão o comprimento correcto da chave, $K = k_1, k_2, \dots, k_m$, podemos usar o índice de coincidência mútuo, com os valores de referência da língua Portuguesa, como forma de obter as diferentes sub-chaves.

$$I_{g_j} = \sum_{i=1}^{|\mathcal{A}|} \frac{p_i f_{i+g_j}}{n'} \quad j = 1, \dots, m$$

com n' o comprimento do sub-texto j ; p_i os valores de referência para a língua Portuguesa; os índices $i + g_j$ calculados módulo $|\mathcal{A}|$.

Caso $g_j = k_j$ teremos $I_{g_j} \approx I_c(Pt)$.

Exercício Prático 4

Dado o seguinte texto, encriptado através da cifra de Vigenère.

i jrçààooâhfql pza-cwxuáljiwyhl pya qvrulka àvr èt
kzujâyào ol uaàolzz lp sãmztqçáiðvñ m àriyzpwãpçóz dwã
ciãhcápyeà ke èta xlnâlneu xum wulpzsmx smã eèpjuálkoà
tavèhlpupm wetz eutzswã e àllw keàçpniçì rçz di
tevãhgmx. w zuzrçpmyào ol uðxuqyhs pçpmpat- téallz m,
pwãæztvruputm, dwã cwxuálkoczpz wnhsqzuoã ámi
ziopyfqñhtqéh mévlãúodo ohs çöçvtjaà jrçààooâi fqñhs.

Descubra o texto original.