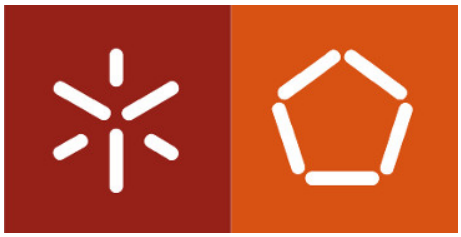
Tecnologia Criptográfica

6 de Dezembro de 2020

Trabalho Prático 2

a83899	André Morais
a84485	Tiago Magalhães

Análise de Criptogramas gerados pelo OTP



Mestrado Integrado em Engenharia Informática
Universidade do Minho

Conteúdo

1 Introdução

No âmbito da Unidade Curricular de Tecnologia Criptográfica, foi nos proposto para analisar os criptogramas fornecidos pelo professor, e descobrir qual desses textos cifrados tinham usado a mesma chave. Estes foram cifrados com a cifra *One-time-pad*(OTP).

OTP é uma técnica de chave única, ou seja, esta chave só pode ser usada uma única vez, tendo esta que ter o mesmo ou maior tamanho que o texto limpo.

2 Estratégia de resolução

Sabendo que para cifrar mensagens no OTP, para cada letra da mensagem temos de fazer adição modular com a letra da chave, isto é $c(i) = m(i) + k(i) \mod 26$, sendo c o texto cifrado, k a chave, 26 o número de letras do alfabeto e i a posição da letra da mensagem que estamos a cifrar.

No enunciado é dito que dois criptogramas foram obtidos com a mesma chave, a reutilização da chave pode revelar informação acerca dos textos limpos, uma vez que sabendo que temos 2 mensagens diferentes (m_1 e m_2), estas foram cifradas da seguinte maneira:

$(c_1 = m_1 + k \mod 26)$ e $(c_2 = m_2 + k \mod 26)$

Daqui podemos retirar que:

$c_2 - c_1 = (m_2 + k) \mod 26 - (m_1 + k) \mod 26$
$\equiv c_2 - c_1 = ((m_2 \mod 26 + k \mod 26) \mod 26 - (m_1 \mod 26 + k \mod 26)) \mod 26$
$\equiv c_2 - c_1 \mod 26 = ((m_2 \mod 26 + k \mod 26) \mod 26 - (m_1 \mod 26 + k \mod 26)) \mod 26) \mod 26$
$\equiv c_2 - c_1 \mod 26 = ((m_2 \mod 26 + k \mod 26) - (m_1 \mod 26 + k \mod 26)) \mod 26$
$\equiv c_2 - c_1 \mod 26 = (m_2 \mod 26 + k \mod 26 - m1 \mod 26 - k \mod 26) \mod 26$
$\equiv c_2 - c_1 \mod 26 = (m_2 \mod 26 - m_1 \mod 26) \mod 26$
$\equiv c_2 - c_1 \mod 26 = (m2 - m1) \mod 26$

Podemos concluir que a diferença de dois criptogramas cifrados com a mesma chaves nos irá dar a diferença entre as letras nas mensagens. Como normalmente nos vários idiomas existem letras que são mais frequentes é de esperar que nas diferenças dos criptogramas cifrados com a mesma chave, no caso de texto em inglês, ocorram coincidentemente as diferenças entre **E - T**(=4 - 19 = -15 $\mod 26 = 11$) **= L**, **T-E**=15**=P**, **E-A**=4**=E**, **A-E**=-4**=W**,**E-E=0=A**, uma vez que estas são a letras mais frequentes.

Assim para detetarmos quais do criptogramas foram cifrados com a mesma chave, fizemos para cada combinação de criptogramas o texto resultante da sua diferença, isto é, aplicamos $(c_2 - c_1 \mod 26)$. Após isto para cada combinação gerada de diferenças entre criptogramas aplicamos uma análise de frequências de letras para ver quais as diferenças mais esperadas.

Para a escolha do texto final que identificaria quais criptogramas foram cifrados com a mesma chave, procuramos pelo texto com maior percentagem existente da letra 'A' uma vez que revelaria uma maior coincidência, isto é, letras iguais nas mesmas posições em ambas as mensagens, o que nos levou a um texto com 6,965% de ocorrência desta letra, o que é próximo ao índice de coincidência de um texto em inglês(6,67%)[1].

Em baixo, está representado a frequência das letras do texto obtido com maior frequência da letra 'A', sendo este obtido pela diferença dos criptograma 6 e 14.

Em relação aos outros textos podemos constatar que a análise de frequência revelou textos sem padrões onde as letras apareciam com frequências todas próximas uma das outras, enquanto que neste texto existe uma maior disparidade. Também nos outros textos a letra 'A' apresenta frequências na ordem dos 3,8% o que se encontra próximo do índice de coincidência de um texto aleatório (3,85%)[1], ou seja o uso de chaves diferentes não permite obter qualquer informação acerca dos textos limpos.

```
    Frequência de Letras
([('A', 371, 6.965828013518588),
('P', 254, 4.769057453999249),
('E', 240, 4.50619601952685),
('L', 240, 4.50619601952685),
('N', 231, 4.337213668794593),
('W', 231, 4.337213668794593),
('K', 218, 4.093128051070222),
('M', 217, 4.074352234322193),
('O', 215, 4.036800600826136),
('Z', 215, 4.036800600826136),
('R', 203, 3.8114907998497936),
('T', 202, 3.792714983101765),
('Q', 199, 3.7363875328576794),
('V', 196, 3.680060082613594),
('B', 190, 3.5674051821254222),
('G', 190, 3.5674051821254222),
('H', 185, 3.4735260983852796),
('F', 183, 3.435974464889223),
('Y', 183, 3.435974464889223),
('J', 176, 3.304543747653023),
('I', 175, 3.2857679309049947),
('U', 170, 3.1918888471648517),
('C', 166, 3.1167855801727375),
('D', 164, 3.0792339466766805),
('S', 163, 3.060458129928652),
('X', 149, 2.7975966954562526),
6,14]  --> Criptograma 6 e 14
```

Como esperado, as letras 'A' , 'L', 'P', 'E' e 'W' são as que apresentam maior percentagem.

A partir destes resultados seria possível obter a chave através de palpites tais como utilizar a palavra *the* que é das mais usadas em inglês e poderá ocorrer nas mensagens.

Assim os criptogramas cifrados com a utilização da mesma chave foram o **6** e o **14**.

3 Conclusão

Dado como terminado este trabalho, este permitiu-nos aprofundar os conhecimentos obtidos nas aulas teóricas, percebendo melhor na prática o *One-time-pad* e que para obter a sua chave é realmente mais complicado do que nas cifras clássicas.

Apesar de ser considerada das cifras mais seguras, uma vez se a chave for *pseudo-aleatória* os criptogramas não irão revelar informação acerca do texto limpo. A reutilização da chave pode comprometer a sua segurança, assim, teremos de gerar sempre chaves diferentes, devido ao facto de que uma mensagem cifrada com a mesma chave irá dar sempre o mesmo texto cifrado e mensagens diferentes cifradas com a mesma chave irão ter o problema abordado neste relatório, além de que a chave tem de ser do tamanho da mensagem, ou seja terá de ser grande e sempre diferente, o que na prática leva a que seja pouco eficiente.

Referências

[1] Index of Coincidence, disponível em:
`https://alexbarter.com/statistics/index-of-coincidence/`