Henrique Tsuyoshi Yara
11796083

Lista 1

MAC0336 - Routo Terada
1 de outubro de 2021

Exercício 1

Uma fonte de informação $\mathcal{X}$ gera saídas $\{x_1, \ldots, x_n\}$ com as probabilidades $p(x_1)$, ..., e $p(x_n)$. Lembre que a entropia de Shannon é definida como:

$$H(\mathcal{X}) = \sum_{i=1}^{n} p(x_i) \log_2 \left(\frac{1}{p(x_i)} \right).$$

1.1

Escrever todos os passos do cálculo da entropia de X para as seguintes probabilidades:

- $p(x_1) = 1/4$,
- $p(x_2) = 1/16$,
- $p(x_3) = 1/16$,
- $p(x_4) = 1/16$,
- $p(x_5) = 1/4$,
- $p(x_6) = 1/16$,
- $p(x_7) = 1/4$.

Resolução.
Usando a definição de entropia e aplicando os valores das probabilidades acima, temos ...

1. $E(X) = p(x_1)log_2[\frac{1}{p(x_1)}]+p(x_2)log_2[\frac{1}{p(x2)}]+p(x_3)log_2[\frac{1}{p(x_3)}]+p(x_4)log_2[\frac{1}{p(x_4)}]+p(x_5)log_2[\frac{1}{p(x_5)}]+ p(x_6)log_2[\frac{1}{p(x_6)}] + p(x_7)log_2[\frac{1}{p(x_7)}]$
2. $E(X) = \frac{1}{4}log_2[\frac{1}{(\frac{1}{4})}] + \frac{1}{16}log_2[\frac{1}{(\frac{1}{16})}] + \frac{1}{16}log_2[\frac{1}{(\frac{1}{16})}] + \frac{1}{16}log_2[\frac{1}{(\frac{1}{16})}] + \frac{1}{4}log_2[\frac{1}{(\frac{1}{4})}] + \frac{1}{16}log_2[\frac{1}{(\frac{1}{16})}] + \frac{1}{4}log_2[\frac{1}{(\frac{1}{4})}]$
3. $E(X) = \frac{1}{4}log_2[4] + \frac{1}{16}log_2[16] + \frac{1}{16}log_2[16] + \frac{1}{16}log_2[16] + \frac{1}{4}log_2[4] + \frac{1}{16}log_2[16] + \frac{1}{4}log_2[4]$
4. $E(X) = \frac{1}{4} * 2 + \frac{1}{16} * 4 + \frac{1}{16} * 4 + \frac{1}{16} * 4 + \frac{1}{4} * 2 + \frac{1}{16} * 4 + \frac{1}{4} * 2$
5. $E(X) = \frac{1}{2} * 3 + 4 * \frac{1}{4}$
6. $E(X) = \frac{3}{2} + 1$
7. $E(X) = \frac{5}{2}$
8. $E(X) = 2,5$

□

Henrique Tsuyoshi Yara
11796083

Lista 1

MAC0336 - Routo Terada
1 de outubro de 2021

1.2

Para $j = 1, 2, ...n$, seja $max_j\lceil log_2[\frac{1}{p(x_j)}]\rceil = C$. Demostrar (i.e., provar matematicamente) que esse valor maximo C e o numero de bits dos x_j : j = 1,2,...n.

Resolução.

Supondo que temos n indices na tabela o valor do indice precisa ser pelo menos $\lceil log_2(n)\rceil$.

Para provar $C = max_j\lceil log_2[\frac{1}{p(x_j)}]\rceil \geq \lceil log_2(n)\rceil$.

Sabendo que o menor max_j possivel ocorre quando $p(x_i) = \frac{1}{n}$, pois para algum $p(x_i) < \frac{1}{n}$ vai existir $p(x_j) > \frac{1}{n}$ ja que $\sum_{i=1}^{n} p(x_i) = 1$ e o max_j aumentaria.

Portanto o menor $max_j\lceil p(x_i)\rceil$ possivel ocorre quando:

$p(x_i) = \frac{1}{n}$

1. $max_j\lceil log_2[\frac{1}{p(x_j)}] \geq \lceil log_2(\frac{1}{p(x_j)})\rceil$
2. $max_j\lceil log_2[\frac{1}{p(x_j)}] \geq \lceil log_2(\frac{1}{n^{-1}})\rceil$
3. $max_j\lceil log_2[\frac{1}{p(x_j)}] \geq \lceil log_2(n)\rceil$

□

1.3

Demonstrar que $log_2 n$ é a entropia **máxima** de qualquer $X = \{x_1, x_2, ...x_n\}$.
Supor dado o Lema: "A função $log_2()$ é estritamente côncava". E aplicar o Teorema de Jensen: "Se f: $\mathbb{R} \to \mathbb{R}$ é uma função contínua estritamente côncava no intervalo I, então $\sum_{i=1}^{n} a_i f(x_i) \leq f(\sum_{i=1}^{n} a_i x_i)$
Resolução.

1. Sendo a funçao $log_2()$ côncava, usando o teorema de jensen obtemos:
   $\sum_{i=1}^{n} a_i log_2(x_i) \leq \sum_{i=1}^{n} a_i x_i$
2. Relacionando com a formula da entropia:
   $a_i = p(x_i); x_i = p(x_i)^{-1}$

3. $\sum_{i=1}^{n} p(x_i) log_2(p(x_i)^{-1}) \leq log_2(\sum_{i=1}^{n} p(x_i) p(x_i)^{-1})$
4. $E(X) \leq log_2(\sum_{i=1}^{n} p(x_i) p(x_i)^{-1})$
5. $E(X) \leq log_2(\sum_{i=1}^{n} 1)$
6. $E(X) \leq log_2(n)$

□

1.4

Para qual conjunto X essa entropia **máxima** ocorre? Demonstrar matematicamente esse fato.

Resolução.
Supondo um conjunto $X = \{x_1, x_2, ...x_n\}$, sendo $\frac{1}{n} = p(x_1) = p(x_2) = ... = p(x_n)$, $\sum_{i=1}^{n} p(x_i) = 1$ e $E(X) \leq log_2(n)$. Usando a formula da entropia:

1. $E(X) = \sum_{i=1}^{n} \frac{1}{n} log_2(n)$
2. $E(X) = \frac{n}{n} log_2(n)$
3. $E(X) = log_2(n)$

Portanto o conjunto X tem a entropia máxima.

□

Exercício 2

2.1.1

Listar sua data de nascimento
Resolução.
29/07/01

2.1.2

Listar o valor de (E_0, D_0) em **hexadecimal**, como definimos anteriormente, para seu NUSP e sua data de nascimento.

Resolução.
$(E_0, D_0) : (0x11796022; 0x11796022)$
$K : 0x0209000700010000$

2.1.3

Aceitar como entrada (E_0, D_0) e a subchave K_1, e calcular e listar em **hexadecimal** a saída da primeira iteração (round 1), $(E1, D1)$. conforme o desenho dado de 1 iteração (round). A subchave deve ser gerada com a chave K definida com os seus dados.
Resolução.
$(E_1, D_1) : (0x00ee2288; 0x32a6f73e)$
$K_1 : 0x000000010009$

2.1.4

Complementar apenas o bit mais à esquerda de E0 e calcular e listar em **hexadecimal** a saída da primeira iteração (round 1), (E_1^c, D_1^c)
Resolução.
$(E_0^c, D_0) : (0x91796022; 0x11796022)$
$(E_1, D_1) : (0x01ee2288; 0x76aff62e)$
$K_1 : 0x000000010009$

2.1.5

Calcular e listar o número de bits diferentes entre (E_1, D_1) e (E_1^c, D_1^c)
Resolução.

A diferença entre (E_1, D_1) e (E_1^c, D_1^c) é de 7 bits

2.2.1

Efetuar os mesmo passos (2) a (5) para cada iteração $j = 2, 3, 4, ..16$, ou seja, calcular e listar o número de bits diferentes entre E_j, D_je E_j^c, D_j^c

Resolução.

$Round_j$	subch. K_j	E_j	E_j^c	D_j	D_j^c	bits diff
0	0x000000000000	0x11796022	0x91796022	0x11796022	0x11796022	1
1	0x000000010009	0x00ee2288	0x01ee2288	0x32a6f73e	0x76aff62e	7
2	0x000000200680	0x32a6f73e	0x76aff62e	0x63c30f2a	0xbd7b26f0	24
3	0x000000180003	0x63c30f2a	0xbd7b26f0	0xcc519dad	0x286c8ce8	32
4	0x000000064000	0xcc519dad	0x286c8ce8	0xde091d5e	0x2470e4f7	35
5	0x000000002140	0xde091d5e	0x2470e4f7	0x8902388e	0xffbb0a71	42
6	0x000000a08000	0x8902388e	0xffbb0a71	0xb3fc4207	0x22ac1cb8	38
7	0x000000400602	0xb3fc4207	0x22ac1cb8	0x3a35e3ac	0x7b588e0a	33
8	0x0000001c0008	0x3a35e3ac	0x7b588e0a	0x3bd3c313	0xcdb83d43	36
9	0x000000000424	0x3bd3c313	0xcdb83d43	0x284bd431	0xee94f446	38
10	0x000000480880	0x284bd431	0xee94f446	0x34799fd6	0xaad372a1	39
11	0x000000004019	0x34799fd6	0xaad372a1	0x4357cdb7	0xed95b64d	41
12	0x000000031000	0x4357cdb7	0xed95b64d	0x61f21e01	0x06ed9081	35
13	0x000000800120	0x61f21e01	0x06ed9081	0xde7ca3e3	0xc3da3f00	32
14	0x000000000a04	0xde7ca3e3	0xc3da3f00	0x0bd007e4	0xc9985cba	32
15	0x000000500090	0x0bd007e4	0xc9985cba	0xbe348e0f	0x4d2d2eee	30
16	0x00000080a004	0xbe348e0f	0x4d2d2eee	0x41078409	0xdd39c753	31

2.3.1

Listar K em **hexadecimal**
Resolução.

$K = 0x0209000700010000$

2.3.2

Aceitar como entrada (E_0, D_0), como definido anteriormente, listar esses valores, e calcular a saída da primreira iteração $(round1), (E_1, D_1)$

Resolução.

$(E_0, D_0) : (0x11796022; 0x11796022)$
$K_1 : 0x000000010009$
$(E_1, D_1) : (0x00ee2288; 0x32a6f73e)$
$K_1 : 0x000000010009$

2.3.3

Listar os valores da subchave $K_1, (E_1, D_1)$ em **hexadecimal**
Resolução.

$(E_1, D_1) : (0x00ee2288; 0x32a6f73e)$
$K_1 : 0x000000010009$

2.3.4

Complementar apenas o bit mais à esquerda da **chave** K (Sem alterar a entrada), e listar esse valor em **hexadecimal**
Resolução.

$K = 0x8209000700010000$

2.3.5

Calcular a subchave K_1^c e a saída da primeira iteração (round 1), (E_1^c, D_1^c)

Resolução.

$(E_1^c, D_1^c) : (0x00ee2288; 0xb2a6e77e)$
$K_1^c : 0x000010010009$

2.3.6

Listar os valores $K_1^c, (E_1^c, D_1^c)$ em **hexadecimal**
Resolução.

$(E_1^c, D_1^c) : (0x00ee2288; 0xb2a6e77e)$
$K_1^c : 0x000010010009$

2.3.6

Listar os valores $K_1^c, (E_1^c, D_1^c)$ em **hexadecimal**
Resolução.

A diferença entre (E_1, D_1) e (E_1^c, D_1^c) é de 3 bits

2.4.1

Efetuar os mesmos passos (2) a (7) para cada iteração $j = 2, 3, 4, ..16$, ou seja, calcular e listar K_j em hexadecimal e o número de bits diferentes entre $(E_j, D_j) e (E_j^c, D_j^c)$
Resolução.

R_j	K_j	K_j^c	E_j	E_j^c	D_j	D_j^c	bits diff
1	0x000000010009	0x000010010009	0x00ee2288	0x00ee2288	0x32a6f73e	0xb2a6e77e	3
2	0x000000200680	0x004000200680	0x32a6f73e	0xb2a6e77e	0x63c30f2a	0x3065e338	18
3	0x000000180003	0x000100180003	0x63c30f2a	0x3065e338	0xcc519dad	0x0ac7d791	30
4	0x000000064000	0x000001064000	0xcc519dad	0x0ac7d791	0xde091d5e	0x2eb32c77	30
5	0x000000002140	0x010000002140	0xde091d5e	0x2eb32c77	0x8902388e	0x0d68f79c	29
6	0x000000a08000	0x000080a08000	0x8902388e	0x0d68f79c	0xb3fc4207	0x10a4345b	30
7	0x000000400602	0x100000400602	0xb3fc4207	0x10a4345b	0x3a35e3ac	0x26c7b20c	29
8	0x0000001c0008	0x0000001c0008	0x3a35e3ac	0x26c7b20c	0x3bd3c313	0x8e5a95d4	30
9	0x000000000424	0x002000000424	0x3bd3c313	0x8e5a95d4	0x284bd431	0xc4488c24	30
10	0x000000480880	0x000400480880	0x284bd431	0xc4488c24	0x34799fd6	0xf569044f	26
11	0x000000004019	0x400000004019	0x34799fd6	0xf569044f	0x4357cdb7	0xc4bc8f1e	29
12	0x000000031000	0x008000031000	0x4357cdb7	0xc4bc8f1e	0x61f21e01	0x874ed128	35
13	0x000000800120	0x000002800120	0x61f21e01	0x874ed128	0xde7ca3e3	0x7c8f81e0	32
14	0x000000000a04	0x200000000a04	0xde7ca3e3	0x7c8f81e0	0x0bd007e4	0x3c6a71e5	29
15	0x000000500090	0x000000500090	0x0bd007e4	0x3c6a71e5	0xbe348e0f	0x7ca6e7a4	31
16	0x00000080a004	0x00004080a004	0xbe348e0f	0x7ca6e7a4	0x41078409	0x30797def	36

2.5.1

Supondo a chave K permaneça fixa, o resultado numérico que V obteve indica algum nível de dificuldade de um mal-intencionado calcular a entrada (E_0, D_0) corresponde a uma dada saída de 64 bits? Por que?

Resolução.
Supondo que a chave K permaneça fixa V vai obter vários resultados numéricos criptografados por uma mesma chave, o problema que isso causa é que se V testar a mesma chave em várias mensagens ele vai conseguir decriptografar todas.

2.5.2

A mesma pergunta, se fosse só uma iteração?

Resolução.
Sendo apenas uma iteração, a mensagem criptografada vai estar mais vulnerável ainda. O nível de entropia vai estar bem baixa, consequentemente a difusão e a confusão também, então duas mensagens criptografada com a mesma chave vão estar bem parecidas.

2.5.3

Supondo que a entrada (E_0, D_0) permaneça fixa, o resultado numérico que V obteve indica algum nível de dificuldade de um mal-intencionado calcular a chave K correspondente a uma dada saída de 64 bits? Por quê?

Resolução.
Supondo que a entrada (E_0, D_0) permaneça fixa V vai obter vários resultados numéricos criptografados por chaves diferentes, o problema que isso causa é que várias chaves vão conseguir decriptografar a mensagem.

2.5.4

A mesma pergunta, se fosse só uma iteração?

Resolução.
Sendo apenas uma iteração, a mensagem criptografada vai estar mais vulnerável já que nível de entropia vai estar bem baixa então uma mensagem criptografada com chaves diferentes vão estar bem parecidas.

2.5.5

Qual a relação do resultado numérico que V obteve com o conceito de Entropia de Informação segundo Shannon?

Resolução.
A relação que foi obtida é que a entropia (confusão e difusão) tem uma grande importância em proteger as informações das mensagens e das chaves