Examen 17 junio 2013, preguntas y respuestas

Informació i Seguretat (Universitat Autònoma de Barcelona)

INFORMACIÓ I SEGURETAT
17 de juny de 2013

Nom i cognoms: _____ Grup: _____

- **Cal que justifiqueu convenientment totes les respostes**
- Valoració dels exercicis: 1) 1,25+1,25 punts; 2) 1+1+0,5 punts; 3) 0,5+1+1; 4) 1+0,5+1 punts
- $\log 3 = 1.58$, $\log 5 = 2.32$, $\log 7 = 2.8$

1. Sigui $\{A_1, A_2\}$ l'alfabet d'entrada i $\{B_1, B_2, B_3, B_4\}$ el de sortida d'un canal discret sense memòria amb matriu de probabilitats condicionades:

$$\Pi = \begin{pmatrix} 1/6 & 1/3 & 1/6 & 1/3 \\ 1/3 & 1/6 & 1/3 & 1/6 \end{pmatrix}$$

   (a) Calculeu la capacitat del canal.

   (b) Doneu la funció de descodificació a *màxima versemblança* (MV) per a aquest canal i digueu quan valdria la probabilitat mitjana d'error en la descodificació.

   Solució:

   (a) Es tracta d'un canal simètric, per tant, $C = \log m - H$, on m és el nombre de sortides i H és l'entropia d'una fila de la matriu:

$$\begin{aligned} C &= \log 4 - H(1/6, 1/3, 1/6, 1/3) = 2 - (\frac{2}{3}\log 3 + \frac{2}{6}\log 6) \\ &= 2 - (\log 3 + \frac{1}{3}\log 2) = 2 - \frac{1}{3} - \log 3 \approx 1.67 - 1.58 = 0.09 \text{ bits/entrada.} \end{aligned}$$

   (b) Agafant els valors màxims a cada columna de Π tenim la funció de descodificació a màxima versemblança:

$$f(B_1) = f(B_3) = A_2; \quad f(B_2) = f(B_4) = A_1.$$

   La probabilitat mitjana d'error serà:

$$\begin{aligned} \overline{P}_e &= 1 - \left(\frac{1}{3}P(A_2) + \frac{1}{3}P(A_1) + \frac{1}{3}P(A_2) + \frac{1}{3}P(A_1)\right) \\ &= 1 - \frac{1}{3}\left(2P(A_1) + 2P(A_2)\right) = 1 - \frac{2}{3} = \frac{1}{3}. \end{aligned}$$

2. Una font F pot emetre 7 missatges diferents amb probabilitats 1/3, 1/3, 1/9, 1/9, 1/27, 1/27, 1/27. La font pot emetre els missatges utilitzant un alfabet binari o un alfabet ternari. Sigui C_1 un codi ternari de longitud constant per la font F, C_2 un codi binari òptim associat a la font F i $C_3 = \{0, 1, 20, 21, 220, 221, 222\}$ un codi ternari associat també a F.

   (a) Doneu la longitud mitjana de cada un dels tres codis.

   (b) Quin dels tres codis és més eficient?

   (c) Per aquesta font, existeix algun codi binari amb eficiència igual a 1?

Solució:

(a) Per codificar 7 missatges en ternari amb un codi de longitud constant, cal que $L \geq \frac{\log 7}{\log 3} = 1.77$. Hauriem d'agafar $L = 2$ i la longitud mitjana serà també $\bar{L}_1 = 2$ digits ternaris/símbol.

Si C_2 és òptim aleshores és el resultat d'aplicar l'algorisme de Huffman a la font F. El codi seria, $C_2 = \{0, 10, 110, 1110, 11110, 111110, 111111\}$ amb longitud mitjana $\bar{L}_2 = \frac{1}{3} + \frac{2}{3} + \frac{3}{9} + \frac{4}{9} + \frac{5}{27} + \frac{6}{27} + \frac{6}{27} = 2.4$ bits/símbol.
La longitud mitjana de C_3 seria, $\bar{L}_3 = \frac{1}{3} + \frac{1}{3} + \frac{2}{9} + \frac{2}{9} + \frac{3}{27} + \frac{3}{27} + \frac{3}{27} = 1.44$ digits ternaris/símbol.

(b) Si volem calcular l'eficiència, primer calculem l'entropia de la font:
$H(F) = \sum_{i=1}^{7} p_i \log p_i = -\frac{2}{3} \log \frac{1}{3} - \frac{2}{9} \log \frac{1}{9} - \frac{3}{27} \log \frac{1}{27} = 2.28$ bits/símbol.
Calculem l'eficiència dels tres codis,
$\eta_1 = \frac{2.28}{2 \cdot \log 3} = 0.72$.
$\eta_2 = \frac{2.28}{2.4} = 0.95$.
$\eta_3 = \frac{2.28}{1.44 \cdot \log 3} = 1$.
Per tant, el més eficient és el codi C_3.
De fet, no calia calcular l'entropia. Com que

$$\eta = \frac{H(F)}{\bar{L} \log D},$$

només cal veure en quin cas el denominador és més petit. Per al codi C_1, el denominador és $2 \cdot \log 3 = 3.16$; per al codi C_2 és 2.4 i per al codi C_3 és $1.44 \cdot \log 3 = 2.28$. El més eficient és doncs C_3.

(c) No existeix ja que el codi C_2 és binari òptim i la seva eficiència és menor que 1. També es pot veure que no existeix perquè les probabilitats dels missatges no són potències de dos (condició necessària segons el primer teorema de Shannon).

3. Considereu el codi C que té matriu de control

$$H = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 1 \end{pmatrix}.$$

(a) Determineu els paràmetres n (longitud), k (dimensió), d (distància mínima), δ (capacitat detectora), t (capacitat correctora) i R_T (taxa de transmissió). Quantes paraules-codi té C? Justifiqueu els valors de n, k i d.

(b) Utilitzant l'alfabet següent:

A	B	C	D	E	F	G	H
000	001	010	011	100	101	110	111

Codifiqueu el missatge "CAFE".

(c) A la sortida del canal rebem "101011 100101 100001". Utilitzant la descodificació via síndrome calculeu la síndrome de cada vector i digueu si s'han produït errors durant la transmissió. En cas positiu digueu quin és el vector d'error, quants errors hi ha i si es poden corregir o no.

Solució:

(a) La matriu generadora és

$$G = \begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 \end{pmatrix}.$$

La longitud és el nombre de columnes de G, per tant, $n = 6$, la dimensió és el nombre de files de G $k = 3$, la taxa de transmissió és $R_T = \frac{k}{n} = \frac{1}{2}$. Mirant la matriu de control, podem veure que tota parella de columnes és linealment independent i que les columnes $1, 2$ i 5 són linealment dependents; per tant, $d = 3$, $\delta = d - 1 = 2$ i $t = \lfloor \frac{3-1}{2} \rfloor = 1$. El codi té $2^k = 8$ paraules-codi.

(b) Volem codificar "CAFE"; és a dir 010, 000, 101, 100:

$$(0,1,0) \cdot G = (0,1,0,1,1,1),$$
$$(0,0,0) \cdot G = (0,0,0,0,0,0),$$
$$(1,0,1) \cdot G = (1,0,1,1,1,0),$$
$$(1,0,0) \cdot G = (1,0,0,1,0,1).$$

(c) El codi és 1-corrector; per tant, la taula de síndrome seria:

e	H(e)
000000	000
100000	101
010000	111
001000	011
000100	100
000010	010
000001	001

Calculem la síndrome dels vectors rebuts (1,0,1,0,1,1), (1,0,0,1,0,1), (1,0,0,0,0,1):

- $H((1,0,1,0,1,1)) = (1,0,1) = H((1,0,0,0,0,0))$; s'ha produït un error a la coordenada 1. Aleshores, el vector original és: $(0,0,1,0,1,1)$. El vector d'error és: 100000
- $H((1,0,0,1,0,1)) = (0,0,0)$; no s'ha produït cap error. El vector d'error és: 000000
- $H((1,0,0,0,0,1)) = (1,0,0) = H((0,0,0,1,0,0))$; s'ha produït un error a la coordenada 4. Aleshores, el vector original és: $(1,0,0,1,0,1)$. El vector d'error és: 000100

  Per tant el vector d'error és: 100000 000000 000100

4. Dos usuaris d'una xarxa volen intercanviar informació de manera privada. Les dades públiques dels dos usuaris és la següent:

	RSA	Funció Hash
A	(e_A, n_A)	H
B	(e_B, n_B)	

I aquesta és la informació privada de cada usuari:

	AES128	RSA
A	k	d_A
B		d_B

(a) Què és el sobre digital i quin problema vol solucionar? Si A vol enviar a B el missatge m fent servir el sobre digital, què ha d'enviar (doneu les equacions)?

(b) Què és una funció Hash? Per què les fem servir en les signatures digitals?

(c) A envia a B (c_m, s_r), on c_m és el missatge m xifrat amb RSA i s_r és un resum de m signat. Determineu quins passos ha de seguir B per tal d'obtenir el missatge i verificar la signatura de A (doneu les equacions).

Solució:

(a) El sobre digital és un mecanisme criptogràfic que utilitza una combinació de clau pública i clau compartida; es fa servir la clau pública per enviar xifrada la clau compartida. D'aquesta manera, enviem la clau compartida de forma segura i podem fer servir un criptosistema de clau compartida per xifrar el missatge que és molt més ràpid que si ho el xifrem fent servir un criptosistema de clau compartida.
L'usuari A ha d'enviar $(k^{e_B} \bmod (n_B), AES128_k(m))$.

(b) Una funció Hash és una funció que assigna a cada missatge m un resum $H(m)$ de mida fixa. Com signar fent servir criptosistema de clau pública és lent, en comptes de signar tot el missatge, es signa només un resum d'aquest.

(c) El missatge m s'obté desxifrant c_m: $m = c_m^{d_B} \bmod (n_B)$. L'usuari B calcula el hash del missatge, $H(m)$ i verifica la signatura de A calculant $r_m^{e_A} \bmod (n_A)$. Si aquest darrer valor coincideix amb $H(m)$, aleshores podem garantir que A ha signat el resum del missatge.