Codage de l'information

Devoir surveillé n° 2

Mardi 11 janvier 2011 - Durée 2h - Documents et calculatrices autorisés.

Veuillez indiquer le numéro de votre groupe de TD sur la copie.
Ce sujet contient trois exercices indépendants.

Exercice 1-1 *Code*

On considère le langage L défini sur l'alphabet $\mathcal{A} = \{0, 1\}$, n et m étant deux entiers naturels, par

$$L = \{0, 1^n0, 1^m\}$$

où la notation 1^n désigne le mot constitué de n fois la lettre 1.

Question 1 À quelle condition portant sur les entiers n et m le langage L est–il un code ?

Question 2 Dans cette question, on pose $n = 3$ et $m = 2$. On note A le sous-ensemble de $\mathbb{N}$ dont l'écriture en base 2 est un mot binaire qui se décompose en mots de L.

Q 2–1 Parmi les entiers de 0 à 10, quels sont ceux qui appartiennent à A ?

Q 2–2 Y a–t–il des puissances de 2 dans A ?

Q 2–3 Y a–t–il des nombres de la forme $2^k - 1$ dans A ? si oui pour quels entiers k ?

Exercice 1-2 *Hexadécimal*

Un éditeur hexadécimal montre le (début du) contenu d'un fichier

```
00000000    45 58 45 52 43 49 43 45 20 44 45 20 43 4F 44 41
00000010    47 45
```

Question 1 En sachant qu'il s'agit d'un fichier texte codé en ASCII, décodez-le. (Rappel le caractère `A` a 65 pour code décimal ASCII, et le caractère espace 32.

Raymond Calbuth a examiné attentivement un fichier avec un éditeur hexadécimal et il a noté la fréquence des symboles hexadécimaux dans la table 1.

s	0	1	2	3	4	5	6	7
Pr(s)	24,0%	10,3%	10,3%	4,4%	10,3%	4,4%	4,4%	1,9%
s	8	9	A	B	C	D	E	F
Pr(s)	10,3%	4,4%	4,4%	1,9%	4,4%	1,9%	1,9%	0,8%

TABLE 1 – Fréquences des symboles hexadécimaux

Question 2 Calculez la quantité d'information de chacun des seize symboles hexadécimaux, puis l'entropie de cette source de symboles.

Question 3 En utilisant le théorème du codage sans bruit, donnez un encadrement de la longueur moyenne d'un codage binaire optimal de cette source.

Question 4 Construisez un codage optimal pour cette source, et calculez sa longueur moyenne.

Raymond Calbuth construit un générateur aléatoire de chiffres hexadécimaux. Ce générateur utilise une pièce de monnaie qui présente PILE avec la probabilité $p = 0,3$ et FACE avec la probabilité complémentaire $q = 0,7$. Pour produire un chiffre hexadécimal, le générateur lance consécutivement quatre fois la pièce et, en traduisant 1 lorque PILE apparaît et 0 lorsque c'est FACE, il obtient l'écriture binaire sur quatre bits d'un chiffre hexadécimal. Ainsi si la séquence obtenue après les quatre lancers est FACE, PILE, PILE, PILE, le chiffre obtenu est en binaire `0111`, autrement dit c'est le chiffre `7`.

Question 5 Établissez la distribution de probabilité pour cette source de chiffres.

Raymond propose à sa femme Monique de trouver le symbole que son générateur vient de produire. Il accepte de répondre à toutes les questions que Monique lui posera et auxquelles il peut répondre par oui ou non. Les réponses qu'il donnera seront toujours correctes (aucun mensonge).

Question 6 Quelle stratégie conseillez-vous à Monique de sorte que le nombre moyen de questions lui permettant de trouver le symbole généré soit minimal ?

Exercice 1-3 *Codes correcteurs*

En 1972, la sonde spatiale Mariner 9 a transmis des photographies de la planète Mars en utilisant un code correcteur de Reed-Muller d'ordre 1 et de longueur 32.

Les codes de Reed-Muller d'ordre 1 forment une famille de codes qui dépend d'un paramètre entier $m \geq 1$, et sont notés $\mathcal{R}(m)$. Ils sont définis par récurrence sur m par

$$\mathcal{R}(1) = \{\mathtt{00}, \mathtt{01}, \mathtt{10}, \mathtt{11}\}$$

$$\text{et pour } m \geq 1,\ \mathcal{R}(m+1) = \{\mathbf{uu} \mid \mathbf{u} \in \mathcal{R}(m)\} \cup \{\mathbf{u\overline{u}} \mid \mathbf{u} \in \mathcal{R}(m)\},$$

où la notation $\overline{\mathbf{u}}$ désigne le mot complémentaire du mot $\mathbf{u}$, c'est–à–dire le mot obtenu en changeant tous les `1` en `0` et vice versa.

Par exemple, le code $\mathcal{R}(2)$ est

$$\mathcal{R}(2) = \{\mathtt{0000}, \mathtt{0101}, \mathtt{1010}, \mathtt{1111}\} \cup \{\mathtt{0011}, \mathtt{0110}, \mathtt{1001}, \mathtt{1100}\}.$$

Question 1 Déterminez les capacités détectrices et correctrices d'erreurs des codes $\mathcal{R}(1)$ et $\mathcal{R}(2)$.

Question 2 Donnez la liste des mots de $\mathcal{R}(3)$. Quelles sont les capacités détectrices et correctrices d'erreurs de ce code ?

Question 3 Dans cette question, C désigne un code binaire quelconque de longueur n et de distance minimale d.

Q 3–1 Quelle est la distance minimale du code C_1 de longueur $2n$ dont les mots sont les mots de C doublés ?

$$C_1 = \{\mathbf{uu} \mid \mathbf{u} \in C\}.$$

Q 3–2 Quelle est la distance minimale du code C_2 de longueur $2n$ dont les mots sont les mots de C concaténés à leur complémentaire ?

$$C_2 = \{\mathbf{u\overline{u}} \mid \mathbf{u} \in C\}.$$

Question 4 Déterminez en fonction de m les caractéristiques (longueur, nombre de mots, distance minimale) du code $\mathcal{R}(m)$.

Question 5 Quelles sont les capacités détectrices et correctrices d'erreurs du code utilisé par la sonde Mariner 9 ?