UPMC- Master Informatique- CCC **Mai 2008**

Examen, durée 2h

Exercice 1 : Compression statistique.

1. Donner un codage de compression optimal pour l'ensemble des fréquences suivantes, correspondant aux 8 premiers nombres de Fibonacci : `a :1 ; b :1, c :2 ; d :3 ; e :5 ; f :8 ; g :13 ; h :21`
   Généraliser la réponse pour proposer un codage optimal lorsque les fréquences sont les n premiers nombres de Fibonacci ($F_n = F_{n-1} + F_{n-2}$).
2. Soit T un texte contenant 256 caractères, et tel que la fréquence d'apparition est a peu près la même pour chaque caractère : la fréquence maximale vaut moins de deux fois la fréquence minimale. Montrer que dans ce cas, le codage de Huffmann n'est pas plus efficace qu'un codage de longueur fixe sur 8 bits.

Exercice 2 : Compression par tranches.
La méthode suivante compresse un texte T en le coupant en tranches successives. Le découpage se fait à partir d'un dictionnaire évolutif D, qui est un ensemble préfixiel. à chaque étape de compression, on dispose d'un numéro d'ordre n, du dictionnaire courant D et du texte t restant à compresser (à la première étape, $n = 0, D =$ et $t = T$). Tant que t est non vide, le découpage d'une tranche consiste à chercher le plus court préfixe t_1 de t qui n'est pas dans D. Si t_1 existe, on l'ajoute au dictionnaire avec le numéro d'ordre n, puis on retire la tranche t_1 à t et on incrémente n. Si t_1 n'existe pas, alors, par construction, D contient t, et en retirant t on le texte restant à compresser est vide. Lorsque, à l'étape n, l'on retire la tranche $s = pb$, p est un mot du dictionnaire courant D et $b \in \{0, 1\}$ est le bit de prolongement ; on émet alors la paire (k, b), composée du numéro d'ordre de p et du bit de prolongement.
Exemple 1 : le texte 0100110111100 est découpé en tranches $0, 1, 00, 11, 01, 111, 00$ et le texte compressé est (0,0), (0,1), (1,0), (2,1), (1,1), (4,1), (1,0).

1. Expliciter le découpage en tranches et le texte compressé obtenu à partir du texte original 0100110111100011011.
2. Donner l'ordre de grandeur de la taille du texte compressé –c'est-à-dire du nombre de paires émises– dans le cas particulier oİ le texte d'origine est une suite de n bits identiques.
3. Décrire une structure de donnée arborescente (arbre digital avec information d'ordre dans les noeuds), associant leur numéro d'ordre aux tranches, pour représenter le dictionnaire.
   Dessiner l'arbre associé au dictionnaire de l'exemple 1.
4. Écrire l'algorithme de compression en utilisant cette structuration des données. Quel est l'ordre de grandeur de la complexité en temps de la compression d'un texte composé de n bits.
5. Expliquer la méthode de décompression et écrire l'algorithme de décompression.
6. Comparer la compression par tranches à la méthode LZW vue en cours.

Exercice 3 : Le schèma de Feistel

Le schéma de Feistel fonctionne de la manière suivante. Etant donnés une fonction F de n bits vers n bits et un message $x = x_1||x_2$ de $2n$ bits ($x_1,\ x_2\ \in \{0,1\}^n$), on définit :

$$G(x) = y,$$

où $y = y_1 \| y_2$, avec $y_1,\ y_2\ \in \{0,1\}^n$ tels que

$$\begin{cases} y_1 &= x_2 \\ y_2 &= x_1 \oplus F(x_2) \end{cases}$$

1. Représenter schématiquement un tel schéma.
2. Comment retrouve-t-on x_1, x_2 à partir de y_1, y_2 ? Représenter schématiquement le fonctionnement du déchiffrement.

Exercice 4 : Une autre vue des codes de Reed-Solomon

Nous allons définir un code sur un corps $\mathbb{F}$. Pour cela, on interprète les mots de $\mathbb{F}^k$ comme des polynômes dans $\mathbb{F}[X]$, le mot $(a_0, \ldots, a_{k-1})$ correspondant au polynôme $a_0 + a_1 X + \cdots + a_{k-1}X^{k-1}$.

On notera $\mathcal{C}^k_{[\alpha_0,\ldots,\alpha_{n-1}]}$ le code de dimension k dont les mots sont les évaluations aux points $\alpha_0, \ldots, \alpha_{n-1}$ des polynômes correspondant aux mots de longueur k.

Ainsi le code $C^3_{[1,2,3,4]}$ est l'image de $\mathbb{F}^3$ par l'application

$$\begin{array}{rcl} \mathbb{F}^3 & \mapsto & \mathbb{F}^4 \\ (a_0, a_1, a_2) & \to & (P(1), P(2), P(3), P(4)) \text{ où } P(X) = a_0 + a_1 X + a_2 X^2. \end{array}$$

On considère le code $\mathcal{C}^4_{[1,2,3,4,5,6]}$ défini sur le corps $\mathbb{F} = \mathbb{Z}/7\mathbb{Z}$.

1. Quelles sont la longueur et la dimension de ce code ?
2. Donner une matrice génératrice du code.
3. Mettre cette matrice sous forme systématique.
4. En déduire une matrice de contrôle.
5. Les mots $(4, 4, 5, 4, 1, 0)$ et $(5, 6, 1, 0, 6, 1)$ sont-ils dans le code ?