EPFL - Information, Calcul et Communication - Partie théorique

Série 12

1 Distance minimale d'un code binaire

Considérons tout d'abord le code correcteur d'erreurs suivant (pouvant être utilisé par exemple pour encoder les quatre points cardinaux N,S,E,W):

$$\mathcal{C} = \{111100, 110011, 001111, 000000\}$$

a) Combien d'effacements un tel code peut-il corriger?

b) Combien d'erreurs (i.e., d'inversions $0 \leftrightarrow 1$) peut-il corriger?

Définitions:

- La *distance de Hamming* $d(\mathbf{c}, \mathbf{c}')$ entre deux mots de code binaires $\mathbf{c}$ et $\mathbf{c}'$ (de même longueur) est donnée par le nombre de bits qui diffèrent entre ces deux mots (par exemple, si $\mathbf{c}$ =101 et $\mathbf{c}'$ =011, alors $d(\mathbf{c}, \mathbf{c}') = 2$).

- La distance minimale d'un code binaire $\mathcal{C} = \{\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_M\}$ est définie ainsi:

$$d = \min_{i \neq j} d(\mathbf{c}_i, \mathbf{c}_j)$$

c) Calculer la distance minimale d du code $\mathcal{C}$ défini ci-dessus.

d) En général, combien d'effacements un code binaire $\mathcal{C}$ avec distance minimale d peut-il corriger? Et combien d'erreurs? (en appliquant la règle de la majorité vue au cours)

2 Code de Hamming

L'encodage de Hamming est défini ainsi: pour envoyer 4 bits, disons x_1, x_2, x_3, x_4, on ajoute à ceux-ci 3 bits de parité définis ainsi:

$$x_5 = x_1 \oplus x_2 \oplus x_3, \quad x_6 = x_1 \oplus x_2 \oplus x_4, \quad x_7 = x_1 \oplus x_3 \oplus x_4$$

a) Vérifier que:

- Si deux séquences de 4 bits x_1, x_2, x_3, x_4 et y_1, y_2, y_3, y_4 diffèrent en 1 bit seulement (p.ex., 0011 et 0001), alors les séquences correspondantes des bits de parité x_5, x_6, x_7 et y_5, y_6, y_7 diffèrent en 2 bits au moins.

- Si deux séquences de 4 bits x_1, x_2, x_3, x_4 et y_1, y_2, y_3, y_4 diffèrent en 2 bits (p.ex., 0011 et 0000), alors les séquences x_5, x_6, x_7 et y_5, y_6, y_7 diffèrent en 1 bit au moins.

b) Déduire de a) que la distance minimale d du code de Hamming est plus grande ou égale à 3.

c) Montrer que cette distance est en fait égale à 3 en exhibant deux mots de code $\mathbf{x} = (x_1, x_2, x_3, x_4, x_5, x_6, x_7)$ et $\mathbf{y} = (y_1, y_2, y_3, y_4, y_5, y_6, y_7)$ tels que que $d(\mathbf{x}, \mathbf{y}) = 3$.

d) Combien d'effacements / d'erreurs ce code peut-il corriger?

Remarque: Cette technique d'encodage se généralise à des messages longs de n bits auxquels on rajoute de l'ordre de $\log_2(n)$ bits de parité (rappelez-vous les rats et les bouteilles). Ainsi, on obtient un moyen très efficace de protéger l'information transmise, avec proportionnellement très peu de redondance.

e) Supposons maintenant que vous ayez le choix entre utiliser le code de Hamming ou le code de l'exercice 1 pour envoyer des informations. Lequel préféreriez-vous utiliser?

3 Pour le plaisir: codes-barres*

3.1 Codes-barres unidimensionnels

Les codes-barres qu'on trouve sur les articles de supermarché sont généralement basés sur le code binaire "2 parmi 5" (ou une variante de celui-ci) qui permet d'encoder les chiffres décimaux de 0 à 9 sous forme binaire, en associant à chaque chiffre une suite de 5 bits dont 2 prennent la valeur 1 seulement (on obtient ainsi "2 parmi 5" possibiltés de mots de code, c'est-à-dire $\frac{5(5-1)}{2} = 10$ possibilités). Une version de ce code est la suivante:

0	1	2	3	4	5	6	7	8	9
00110	10001	01001	11000	00101	10100	01100	00011	10010	01010

Ensuite, les 0 sont représentés par des barres fines, les 1 par des barres épaisses, et chaque chiffre du code entier est donc représenté par 5 barres; 2 épaisses et 3 fines. Notez cependant que dans le code-barres illustré ci-dessus, un symbole de départ est ajouté à gauche, ainsi qu'un symbole de fin à droite, et les barres représentant les bits sont entrelacées et sont alternativement noires ou blanches (d'un chiffre à l'autre), ce qui complique sérieusement la lecture du code! Par exemple, les deux premiers chiffres 19 sont représentés ainsi: 1 est représenté en noir par la suite 10001, tandis que 9 est représenté en blanc par la suite 01010, donc on obtient: .

a) Pour quelle raison à votre avis entrelace-t-on des barres noires et blanches ici (plutôt que de n'utiliser que des barres noires)?

Revenons maintenant au code "2 parmi 5".

b) Combien d'erreurs/d'effacements un tel code peut-il corriger? Quelle est la distance minimale de ce code?

c) A vrai dire, la fonction principale de ce code est surtout de *détecter* des erreurs, non de les corriger. En particulier, supposons que les seules erreurs possibles soient des inversions du type $0 \longrightarrow 1$, mais pas $1 \longrightarrow 0$. Combien d'erreurs de ce type est-il alors possible de détecter avec un tel code?

Note: Pour renforcer la capacité de correction d'un code-barres, on peut aussi ajouter un *chiffre de parité* à la fin de celui-ci, qui n'est rien d'autre que la somme modulo 10 des autres chiffres (et ce dernier chiffre est lui-même encodé au format "2 parmi 5").

3.2 Codes-barres bidimensionnels (QR-codes)

Les QR-codes (QR pour "Quick Response") ci-dessus, utilisés généralement pour encoder l'adresse d'un site web, sont des codes-barres à 2 dimensions. Les codes correcteurs d'erreurs utilisés ici sont les codes de Reed-Solomon, qui permettent de corriger jusqu'à 30% d'erreurs, comme vous pouvez le voir en essayant de scanner le code de droite dont une bonne partie a été noircie (sans pour autant toucher les 3 balises carrées...).

d) Quel avantage y a-t-il à utiliser un code-barres bidimensionnel plutôt qu'unidimensionnel?