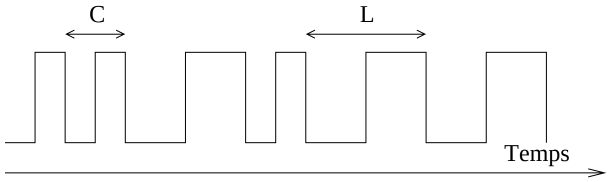
Examen 2003/2004 Compression

Le 19/12/2003, Tout document autorisé

Compression temporelle

Vous parlez Morse ?

Pour envoyer un message on ne dispose que de deux signaux différents. Qui sont le signal L et le signal C. Un message est une suite de L et de C. Ces deux durées sont constantes et sont nommées par abus de langage L et C. Le message affiché dans la figure est $CCLCLL$ et sa durée est $3L + 3C$.



En compression classique on s'intéresse à réduire le nombre de bits émis, mais dans le cadre de cet énoncé on désire minimiser la durée du message.

1. Par rapport à ce que l'on a vu en cours, quels algorithmes vont devoir être modifiés ?
2. Soient les événements X et Y avec respectivement 2 et 3 occurrences. Donnez les arbres de codage possibles ainsi que l'arbre optimum.
3. Soient les événements X, Y et Z avec respectivement 1, 2 et 3 occurrences. Donnez TOUS les arbres de codage possibles.
4. Pour chaque arbre, évaluez la durée d'émission du message en fonction de L et C.
5. Pour $L = 2, C = 1$ donnez le/les arbres optimum pour le codage.
6. Pour $L = 3, C = 1$ donnez le/les arbres optimum pour le codage.
7. Les méthodes de Shannon Fano et Huffmann sont-elles réutilisables ? Pourquoi ?

Interface utilisateur sans clavier/souris

Un utilisateur ne peut interagir avec la machine qu'en cliquant, il n'a pas de clavier ni de souris. On désire lui permettre de taper du texte. Pour ce faire un programme lui propose des choix successifs à interval régulier t, l'utilisateur doit cliquer quand le choix montré est celui qu'il désire. Une fois arrivé à la fin des choix on recommence au début, en effet l'utilisateur a pu oublié de cliquer. Pour simplifier on considérera tous les choix équiprobables. On supposera qu'il y a n choix. On supposera que l'utilisateur clique juste à la fin de l'interval t

8. Si les choix sont proposés les uns après les autres, quel est le temps moyen pour choisir ?
9. Une approche récursive est faite dans laquelle on propose à chaque niveau la moitié des choix restant. quel est le temps moyen pour choisir ?
10. Donnez la formule générale en fonction de N le nombre de paquets fait à chaque niveau.
11. Application numérique : quel est le meilleur découpage ?
12. Ce meilleur découpage n'est pas optimum. Pourquoi ?
13. Si l'on a 12 événements, quel est le temps moyen de choix pour le découpage suivant : $((ABC)DEF)(GHI)JKL$
14. Donner la version de ce découpage pour 24 événements ainsi que le temps moyen de choix.
15. Prouvez que l'on ne peut pas faire mieux que ce découpage.

Traitement du signal

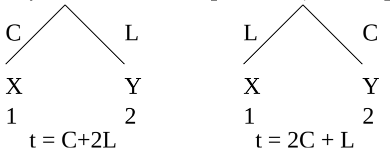
On veut transmettre en continu la hauteur et la vitesse de l'eau dans une rivière.

16. Ce sont des signaux ou des données ?
17. Proposez une méthode simple et générale de compression.
18. Proposez des améliorations successives de votre méthode.

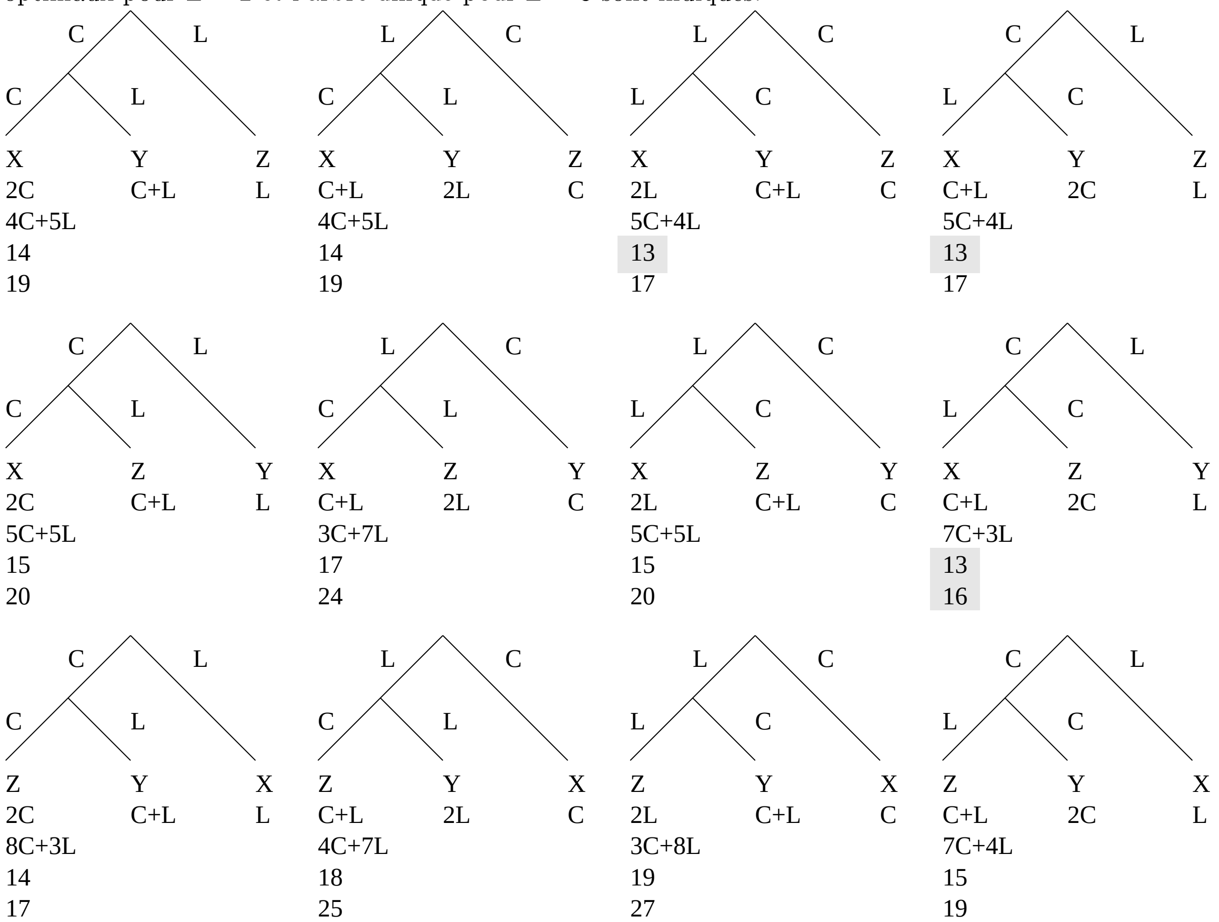
Une correction

1. Les seuls algorithmes à modifier sont ceux égalisant les probabilités. En effet, avant ceux-ci on traite des événements et non des suites de bits. Les algos d'égalisation ont pour but de sortir autant de 0 que de 1 afin de maximiser la quantité d'information transmise. Malheureusement avec des L et des C ce n'est plus le cas, en effet s'il y a le même nombre de L et de C il ne sont pas équiprobables, en effet si vous regardez le message à un instant donné la probabilité qu'il soit en train de transmettre un L est plus grande.

2. Il y a deux arbres possible, celui qui transmettra le message le plus rapidement sera celui de droite car $2C+L < C+2L$.



4-5-6 La liste des arbres avec au dessous : la durée d'envoi d'un événement, la durée totale du message en considérant les nombres d'occurrences et finalement l'application numérique demandée pour les deux valeurs de L. Les trois arbres optimaux pour $L = 2$ et l'arbre unique pour $L = 3$ sont indiqués.

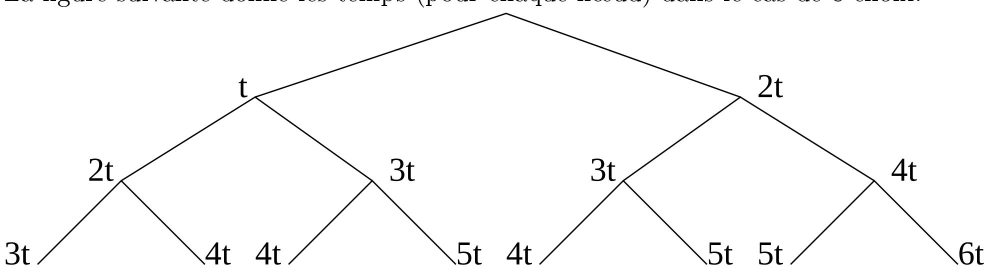


7. Non car elle ne peuvent pas donner de bon résultats. En effet, elles ne peuvent pas coder un événement courant avec plus de bits qu'un événement rare. Or la question 6 montre que cela peut être nécessaire. Peut être peut-on se débrouiller avec SHANNON-FANO en ne coupant pas l'arbre en deux (en nombre d'occurrences) mais en mettant plus d'occurrences du coté du C.

8. Le temps pour choisir le premier événement est t, le dernier nt. La réponse est donc la moyenne de la somme de ces temps de choix :

$$\frac{\sum_{i=1}^{n} it}{n} = \frac{n(n+1)}{2n}t = \frac{n+1}{2}t$$

9. La figure suivante donne les temps (pour chaque nœud) dans le cas de 8 choix.



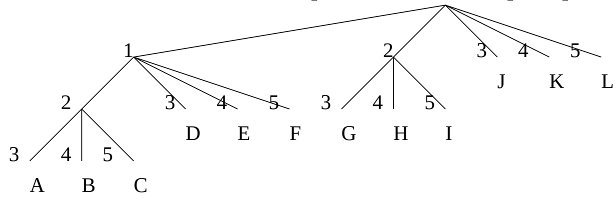
L'évaluation du temps moyen est triviale, en effet comme la probabilité d'allé à gauche ou à droite à chaque nœud est de 0.5 il suffit de faire somme des temps moyens pour chacun des niveaux de l'arbre. A chaque niveau on prend un temps $\frac{2+1}{2}t$ (application de la formule de la question précédente). Le temps moyen est donc $\frac{3}{2}t\ log_2(n)$

10. Pour N branches sur chaque nœuds, le temps moyen est $\frac{N+1}{2}t\ log_N(n)$

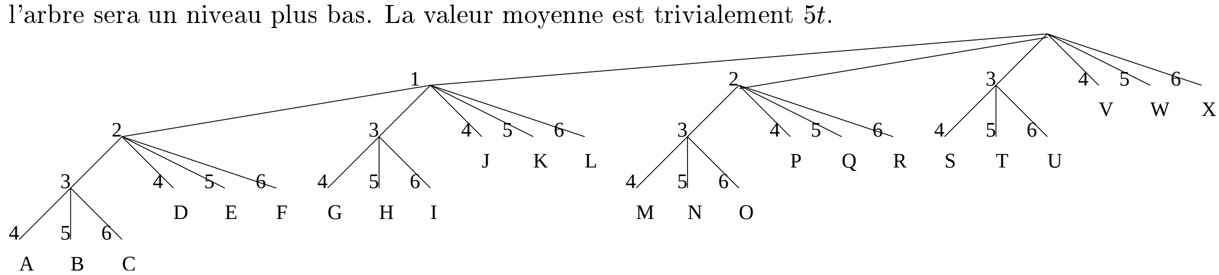
11. L'application numérique donne comme optimum $N = 4$.

12. La figure de la question 9 montre que l'on prend plus de temps pour les choix de droite que pour ceux de gauche. C'est encore plus vrai pour $N = 4$. Or, comme les choix sont équiprobables, le découpage optimum devrait permettre de choisir n'importe quel événement avec le même temps.

13. Les nombres indiquent le nombre d'intervals t pour choisir le nœud. La valeur moyenne est trivialement $4t$. Cet arbre est intéressant car les temps de choix sont presque les même pour tous les choix.



14. Pour doubler le nombre de feuille et garder la même propriété, il faut ajouter une branche à la racine pour incrémenter de 1 tous les temps de l'ancien arbre. Et ajouter l'ancien arbre à gauche, les temps seront là aussi incrémenté de 1 car l'arbre sera un niveau plus bas. La valeur moyenne est trivialement $5t$.



15. La formule générale pour le temps moyen en nombre de t est $log_2 \frac{4}{3} n = log_2(\frac{4}{3}) + log_2(n)$ Avec ce découpage si l'on double le nombre de choix, le temps de choix est seulement augmenté de t (ce qui est l'optimal). Ceci n'était pas le cas dans la question 11 ou si l'on quadruple n, l'incrément de temps est $\frac{5}{2}t\ (log_4(4n) - log_4(n)) = \frac{5}{2}t$ et ceci est supérieur à 2t.

    Si l'on veux faire mieux, il faut que tous les choix soient fait en un temps x ou $x+1$. Il vous suffit d'enlever le troisième choix dans chaque triplet final de l'arbre. La formule générale pour le temps moyen est $0.5 + log_2(n)$ Cette forme arbre vérifie aussi le fait que quand on double n le temps augmente de t. Par contre $0.5 > log_2 \frac{4}{3}$ donc cet arbre donne de moins bons résultats.

    J'avoue que ceci n'est pas une preuve rigoureuse.

16. C'est un signal car il varie continûment (si l'interval d'échantillonnage est suffisant) et qu'une variation infime de la mesure ne perturbe en rien ce que l'on peut en faire (contrairement à une donnée ou toute variation de la donnée entraîne l'invalidation de l'ensemble). Attention : un capteur analogique peut parfaitement retourner une valeur qui est une donnée et non un signal (par exemple un compteur de particules).

17. On envoit les différences par rapport au valeur précédentes.

18. On peut éliminer le bruit par différentes méthodes car il y en a certainement à cause des remous et des vagues.

    Si les variations sont faibles les flux de données contiendront de très nombreux 0, il faut donc faire un RLE.

    Si les variations sont vraiment très faible, on peut stocker les différences de longueurs d'une bande de 0 à l'autre. Ceci revient à faire une prédiction sur la dérivée du signal.

    On peut transformer les 2 signaux afin d'éliminer la redondance. Par exemple en transférant la hauteur et la différence de la vitesse réelle par rapport à une vitesse prédite en fonction de la hauteur.