Club Math

Centre informatique
pédagogique (CIP)
Rue Théodore-de-Bèze 2
Case Postale 3144
1211 GENÈVE 3
Tél: (022) 318.05.30
Fax: (022) 318.05.35
Directeur:
Raymond Morel

Lettre nº 44

Lundi 4 décembre 1995 à 17 h

Communiquer

Bernard Vuilleumier

Communiquer, c'est transmettre des messages, et pour élaborer des messages, nous recourons à des codes

Communiquer, c'est transmettre des messages. Pour élaborer un message, nous recourons à des systèmes de signes et à divers langages. Le langage naturel fait appel à des phrases construites de mots, eux-mêmes constitués de lettres. Les symboles utilisés peuvent être des formes sonores émises par l'appareil vocal, des marques sur une feuille de papier ou des «bits» informatiques. De tels systèmes de signes permettant de représenter ou de transmettre de l'information sont appelés *codes*.

La plupart des codes que nous utilisons – langages naturels, notations musicales, pictogrammes de signalisation routière, etc. – sont le fruit d'une longue évolution. D'autres codes en revanche sont conçus d'emblée à des fins précises. Ainsi la cryptographie cherche à élaborer des codes inviolables afin d'assurer un caractère confidentiel aux messages entre correspondants (voir lettre du Club Math nº 25). La révolution numérique qui affecte de nos jours l'informatique et les télécommunications – réseaux mondiaux comme Internet, transmissions par satellites, réseaux numériques à intégration de services – oblige à concevoir des codes capables de transmettre de grandes quantités d'information avec un faible taux d'erreur. C'est le rôle dévolu aux *codes correcteurs d'erreurs*.

La révolution numérique qui affecte l'informatique et les télécommunications oblige à concevoir des codes capables de transmettre beaucoup d'information avec un faible taux d'erreur

Pour coder un message, il faut un alphabet, c'est-à-dire un ensemble fini de symboles. Le message à coder est converti en une suite de «mots», chaque mot étant une séquence de symboles de longueur fixe. Un ensemble de mots définit un «vocabulaire». Le plus simple des alphabets utilisés pour coder un message est sans doute l'alphabet $A_2=\{0, 1\}$. Un élément de A_2 est un symbole binaire appelé «bit». Cet alphabet est celui de la technologie informatique. Il permet de construire des mots de différentes longueurs: les mots de 4 bits (nombres hexadécimaux), les mots de 8 bits (octets), etc. Dans cette technologie, les messages à communiquer sont mis sous la forme d'une suite de mots constitués de k symboles pris dans un alphabet A. Les mots passent ensuite par un canal de communication qui peut être affecté par diverses sources de bruit. Il y a donc toujours un risque que certains symboles du message soient omis ou transmis de façon erronée. Et dans le cas de l'alphabet $A_2=\{0, 1\}$ par exemple, la réception d'un «1» au lieu d'un «0» peut être catastrophique pour l'interprétation du message. Ces erreurs, qui sont quasiment inévitables, rendent la transmission sans code correcteur très vulnérable.

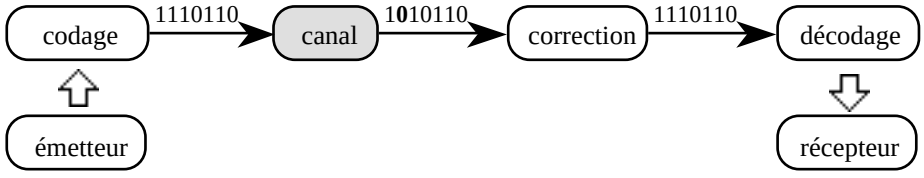

Fig. 1: La transmission d'un message fait intervenir un codage. En passant dans le canal de communication les messages codés subissent fréquemment des modifications indésirables. Il faut utiliser un code correcteur d'erreurs pour détecter et corriger ces erreurs.

En ajoutant des symboles de contrôle aux mots à transmettre, les codes correcteurs permettent non seulement de détecter les erreurs, mais de les corriger

Les codes correcteurs utilisent des *symboles de contrôle*. Au lieu d'envoyer des mots de k symboles dans le canal, ils allongent ces mots en ajoutant à chacun d'eux un certain nombre de symboles de contrôle. Dans le cas d'un alphabet binaire par exemple, il est possible de rajouter un «bit de parité» de telle façon que le nombre de «1» figurant dans chaque mot soit pair. Ainsi le mot 0100 sera allongé en 01001, tandis que 0110 donnera 01100. Si un bit du mot est modifié durant la transmission, la parité change. Ce procédé permet de détecter une erreur, mais pas de la corriger. En faisant dépendre linéairement les symboles de contrôle des symboles d'information, on obtient des *codes linéaires* capables de corriger les erreurs.

Travaux pratiques

Pour préciser quelques notions à propos de codes

A propos de codes

• Les mathématiciens appellent généralement codage ou code, une application d'un vocabulaire *U* vers un vocabulaire *V*, c'est-à-dire une application qui fait correspondre à tout mot de *U* un unique mot de *V*. De plus, pour que le décodage soit possible, deux mots différents de *U* doivent être associés à deux mots différents de *V*. C'est donc par abus de langage que les vocabulaires *U* et *V* dont les alphabets et la longueur des mots ne sont pas nécessairement identiques, sont appelés codes.

• Un code correcteur d'erreurs peut être caractérisé par les trois paramètres[*n, k, d*] suivants:

- la longueur *n* de chaque mot (symboles d'information + symboles de contrôle);
- le nombre *k* des symboles d'information;
- le nombre minimal *d* de symboles différents d'un mot à l'autre du code.

Le rapport *k*/*n* mesure le taux de transmission et le rapport *d*/*n* la fiabilité de la transmission. Si un symbole de contrôle permet de détecter une erreur – c'est le cas du bit de parité – il en faut au moins trois pour la corriger. Dans le cas du codage à triple répétition qui transforme 0 en 000 et 1 en 111, on décode en supposant que le symbole émis est celui qui se répète. Si l'on reçoit par exemple le mot 101, on en déduit que c'est 111 qui a été envoyé. Ce code est caractérisé par les paramètres [3, 1, 2].

• Pour en savoir plus, voir: Gilles Lachaud et Serge Vladut, Les codes correcteurs d'erreurs, *La Recherche* n° 278, juillet-août 1995, pp. 778-782.

Pour illustrer les possibilités d'un alphabet à deux symboles

Exercice 1

Combien de mots peut-on obtenir à partir de l'alphabet $A_2=\{0, 1\}$ lorsque les mots comportent 8, 16, 32, 64 symboles?

Pour se familiariser avec deux applications correspondant à des codages fréquemment utilisés

Exercice 2

a) Formez tous les mots de quatre symboles à partir de l'alphabet $A_2=\{0, 1\}$.

b) Etablissez une correspondance entre le vocabulaire des mots de quatre symboles obtenu à partir de l'alphabet $A_2=\{0, 1\}$ et les chiffres du système de numération hexadécimal. Le code ASCII (American Standard Code for Information Interchange) utilisé par les ordinateurs pour transcrire les fichiers de type «TEXT» représente les caractères typographiques par des mots de 8 symboles (octets) de l'alphabet $A_2=\{0, 1\}$.

c) Etablissez la table de correspondance entre le vocabulaire *U* de l'alphabet $A_2=\{0, 1\}$ et le vocabulaire *V* des signes typographiques.

Pour mettre en œuvre un code correcteur d'erreurs

Exercice 3

Avant d'envoyer des mots de 4 bits $u_1u_2u_3u_4$ dans un canal de transmission on leur ajoute 3 bits de contrôle $u_5=u_1+u_2+u_3$, $u_6=u_2+u_3+u_4$, $u_7=u_1+u_2+u_4$, avec les règles arithmétiques suivantes: $0+0=0$, $0+1=1$ et $1+1=0$.

a) Etablissez les vocabulaires des mots de 4 et de 7 bits de ce code.

b) Donnez les trois paramètres [*n, k, d*] de ce code.

c) Calculez le taux de transmission et la fiabilité de transmission de ce code.

d) Simulez la transmission et la correction de mots comportant une erreur sur un des 7 bits.

Prochaine réunion: lundi 8 janvier 1996 à 17 h.