Codes correcteurs d'erreurs

Comment communiquer dans un monde non fiable

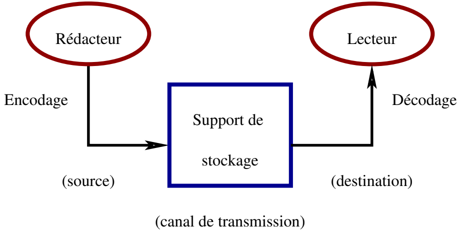
Jean-Marc.Vincent@imag.fr

Université de Grenoble-Alpes, UFR IM^{2}AG
DU Informatique et Sciences du Numérique : Information

Novembre 2016

Codes correcteurs d'erreur

Transmission de l'information



Critères de qualité d'un code :

- **Intégrité de l'information :** tolérance aux fautes (détection/correction des erreurs)
- **Sécurité de l'information :** authentification (cryptage)
- **Efficacité de la transmission :** compression des données

Donnée (message) : séquence finie de bits, éventuellement structurée

Exemples de taux d'erreur

Types d'erreurs

- altération : modification d'un ou plusieurs bits
- insertion / suppression d'un ou plusieurs bits

Ordres de grandeur de taux d'erreurs

ligne	taux d'erreur
Disquette	10^{-9} : à 5 Mo/s, 3 bits erronés par minute
CD-ROM optique	10^{-5} : 7ko erronés sur un CD de 700 Mo
DAT audio	10^{-5} : à 48 kHz, deux erreurs par seconde
Mémoires à semi-conducteurs	$< 10^{-9}$
Liaison téléphonique	entre 10^{-4} et 10^{-7}
Télécommande infrarouge	10^{-12}
Communication par fibre optique	10^{-9}
Satellite	10^{-6} (Voyager), 10^{-11} (TDMA)
ADSL	10^{-3} à 10^{-9}
Réseau informatique	10^{-12}

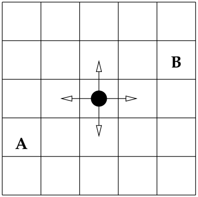
TAB. 4.1: Ordre de grandeur du taux d'erreurs.

CODES CORRECTEURS D'ERREUR

Informatique débranchée

Activité à 3 : un petit robot, un pilote et l'environnement, pour l'instant non malicieux
L'objectif est de diriger le petit robot sur une grille pour aller d'un point A à un point B.
Pour cela le pilote ne peut qu'envoyer des jetons au robot (une séquence de jetons permettant d'aller de A à B).

Le terrain



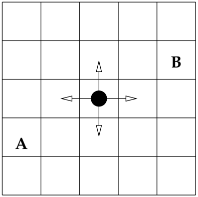
La commande

Proposer un codage binaire permettant de piloter le robot sur la grille.

INFORMATIQUE DÉBRANCHÉE

Activité à 3 : un petit robot, un pilote et l'environnement, pour l'instant non malicieux
L'objectif est de diriger le petit robot sur une grille pour aller d'un point A à un point B.
Pour cela le pilote ne peut qu'envoyer des jetons au robot (une séquence de jetons permettant d'aller de A à B).

Le terrain



La commande

Proposer un codage binaire permettant de piloter le robot sur la grille.

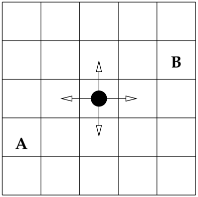
L'environnement

L'environnement n'est pas fiable, certains bits peuvent être changés (l'environnement "retourne" un jeton lors de la transmission).
Proposer un codage qui détecte une erreur de transmission, deux erreurs, 3 erreurs, ...

Informatique débranchée

Activité à 3 : un petit robot, un pilote et l'environnement, pour l'instant non malicieux
L'objectif est de diriger le petit robot sur une grille pour aller d'un point A à un point B.
Pour cela le pilote ne peut qu'envoyer des jetons au robot (une séquence de jetons permettant d'aller de A à B).

Le terrain



La commande

Proposer un codage binaire permettant de piloter le robot sur la grille.

L'environnement

L'environnement n'est pas fiable, certains bits peuvent être changés (l'environnement "retourne" un jeton lors de la transmission).
Proposer un codage qui détecte une erreur de transmission, deux erreurs, 3 erreurs, ...

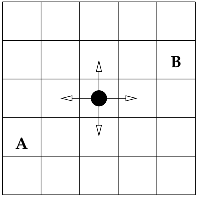
Correction des erreurs

Proposer un algorithme qui corrige une erreur, s'il y en a une.

INFORMATIQUE DÉBRANCHÉE

Activité à 3 : un petit robot, un pilote et l'environnement, pour l'instant non malicieux
L'objectif est de diriger le petit robot sur une grille pour aller d'un point A à un point B.
Pour cela le pilote ne peut qu'envoyer des jetons au robot (une séquence de jetons permettant d'aller de A à B).

Le terrain



La commande

Proposer un codage binaire permettant de piloter le robot sur la grille.

L'environnement

L'environnement n'est pas fiable, certains bits peuvent être changés (l'environnement "retourne" un jeton lors de la transmission).
Proposer un codage qui détecte une erreur de transmission, deux erreurs, 3 erreurs, ...

Correction des erreurs

Proposer un algorithme qui corrige une erreur, s'il y en a une.

Pour chaque code donner ses qualités (taux de détection/correction) et son efficacité (nombre de bits utiles/bits transmis)

Hypercubes

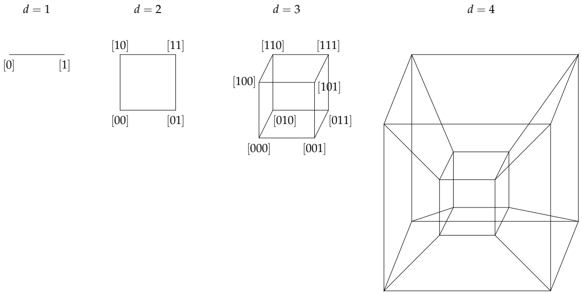
Espace des vecteurs de bits

Exercice : dessiner les hypercubes de dimension $d = 1, 2, 3, 4, \ldots$

HYPERCUBES

Espace des vecteurs de bits

Exercice : dessiner les hypercubes de dimension $d = 1, 2, 3, 4, \ldots$



Distance de Hamming

- Nombre de bits différents entre 2 vecteurs de bits. Calculer la taille de la boule à distance $1, 2, 3, \ldots$ d'un vecteur x de n bits.
- Exercice : calculer la **borne de Hamming**, capacité d'un code de longueur n à corriger 1 bit erroné.

Analogie

Un groupe de bits dans un ordinateur est un mot, chaque bit est considéré comme étant une lettre.

Analogie avec la langue naturelle

- Toutes les combinaisons possibles de l'alphabet ne sont pas des mots de la langue. Les seuls mots autorisés sont ceux énumérés dans un dictionnaire.
- Des erreurs qui se produisent en transmettant ou en stockant des mots français peuvent être détectées en déterminant si le mot reçu est dans le dictionnaire.
- S'il ne l'est pas, des erreurs peuvent être corrigées en déterminant quel mot français existant est le plus proche du mot reçu.

Idée pour la correction d'erreurs :

- Ajouter des lettres supplémentaires (redondantes).
- Ces lettres supplémentaires donnent une structure à chaque mot.
- Si cette structure est changée par des erreurs, les changements peuvent être détectés et corrigés.

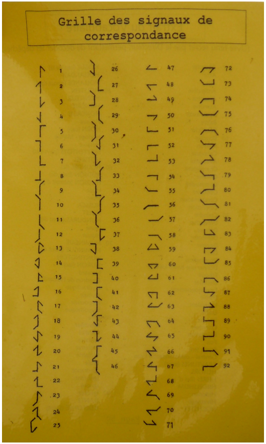
CODES CORRECTEURS D'ERREUR

Code de Chappe

Une station de communication

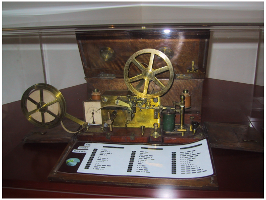
Station du Haut-Barr en Alsace

Table de codes

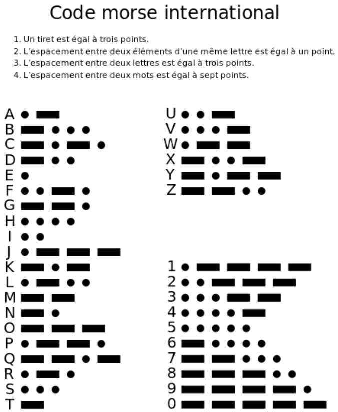


Code Morse

Samuel Morse



Source Wikipedia



CODES CORRECTEURS D'ERREUR

Contrôle de parité

Une technique de base pour construire un code détecteur

1. Découper le message en mots de 7 bits $m = [x_0, \ldots, x_6]$
2. Ajouter aux mots leur parité : $f(m) = [x_0, \ldots, x_6, p]$

Le nombre de 1 dans le mot est soit pair ($p = 0$) soit impair ($p = 1$)

$$p \stackrel{def}{=} \sum_{i=0}^{6} x_i \;\; modulo \;\; 2.$$

Standard n5 du Comité Consultatif International Télégraphique et Téléphonique (CCITT 5)

le plus populaire et celui utilisé par exemple aux USA.

Lettre	Codage de base sur 7 bits	Mot de code avec bit de parité
a	1000 001	1000 0010
e	1010 001	1010 0011
u	0110 101	0110 1010

Détecte un nombre impair d'erreurs

Codes actuels

Numéro de Sécurité Sociale

Un numéro de sécurité sociale est un nombre de $n = 15$ chiffres : un numéro d'identification K sur $k = 13$ chiffres suivi de la clé C de $r = 2$ chiffres calculée pour que $K + C$ soit un multiple de 97.

1. Quelle est la clé du numéro de sécurité sociale 2.63.05.38.516.305 ?
2. Quel est le rendement de ce code ?
3. Combien d'erreurs de chiffres, la clé du numéro de sécurité sociale permet-elle de détecter ?

Formule de Luhn pour les cartes bancaires

Carte bancaire : 4 nombres de 4 chiffres

- Pour les chiffres de rang pair (le premier chiffre est de rang 0) on double ce chiffre modulo 9
- On additionne ces chiffres aux chiffre de rang impair.

Le résultat doit être divisible par 10
Exercice : vérifier sur votre carte bancaire, comment calculer la clé (dernier chiffre).

International Standard Book Number (ISBN)

Exercice : chercher la clé.

Parité longitudinale et transversale

0	0	0	0	0	0	0	
0	1	1	1	1	1	0	
0	1	1	1	1	1	0	
0	0	0	1	0	0	0	
0	0	1	0	1	0	0	
0	1	0	1	0	1	0	
1	0	1	0	1	0	1	

Exercice : Calculer le rendement de ce type de code, calculer sa capacité de détection, de correction, calculer la probabilité d'avoir une erreur non détectée.

Parité longitudinale et transversale

0	0	0	0	0	0	0	**0**
0	1	1	1	1	1	0	**1**
0	1	1	1	1	1	0	**1**
0	0	0	1	0	0	0	**1**
0	0	1	0	1	0	0	**0**
0	1	0	1	0	1	0	**1**
1	0	1	0	1	0	1	**0**

Exercice : Calculer le rendement de ce type de code, calculer sa capacité de détection, de correction, calculer la probabilité d'avoir une erreur non détectée.

PARITÉ LONGITUDINALE ET TRANSVERSALE

0	0	0	0	0	0	0	**0**
0	1	1	1	1	1	0	**1**
0	1	1	1	1	1	0	**1**
0	0	0	1	0	0	0	**1**
0	0	1	0	1	0	0	**0**
0	1	0	1	0	1	0	**1**
1	0	1	0	1	0	1	**0**
1	**1**	**0**	**0**	**0**	**1**	**1**	

Exercice : Calculer le rendement de ce type de code, calculer sa capacité de détection, de correction, calculer la probabilité d'avoir une erreur non détectée.

Parité longitudinale et transversale

0	0	0	0	0	0	0	**0**
0	1	1	1	1	1	0	**1**
0	1	1	1	1	1	0	**1**
0	0	0	1	0	0	0	**1**
0	0	1	0	1	0	0	**0**
0	1	0	1	0	1	0	**1**
1	0	1	0	1	0	1	**0**
1	**1**	**0**	**0**	**0**	**1**	**1**	**0**

Exercice : Calculer le rendement de ce type de code, calculer sa capacité de détection, de correction, calculer la probabilité d'avoir une erreur non détectée.

CODES CORRECTEURS D'ERREUR

UNE FORME STANDARD : CODE LINÉAIRE

Un code correcteur (n, k)

$$\begin{array}{llcl} f: & V^k & \longrightarrow & V^n \\ & m & \longmapsto & f(m) \\ & [x_0, \cdots, x_{k-1}] & & [y_0, \cdots, y_{n-1}] \end{array}$$

Idée : si f est linéaire, alors les opérations de codage/décodage peuvent se faire en temps linéaire/taille de message

- Rapide (proportionnel à la taille du message)
- Il faut : V^k, V^n espaces vectoriels donc V un corps
- Opérations modulo 2 (ex : parité) : $V = \mathbb{Z}/2\mathbb{Z}$ est un corps !
- On travaille en général avec V à 2, 2^8 ou 2^{256} éléments

Alors f linéaire correspond à une matrice G (génératrice) et $f(m) = mG$:

$$[y_0, y_1, \cdots, y_{n-1}] = [x_0, \cdots, x_{k-1}] \begin{bmatrix} g_{0,0} & g_{0,1} & \cdots & g_{0,n-1} \\ \vdots & \vdots & \vdots & \vdots \\ g_{k-1,0} & g_{k-1,1} & \cdots & g_{k-1,n-1} \end{bmatrix}$$

PROPRIÉTÉS D'UN CODE LINÉAIRE

Exercice : Calculer la matrice associée au code de parité sur $k = 3$ bits, au code de parité longitudinale/transversale sur $k = 4$ bits.

Temps de calcul

Pour tout code linéaire $C(n, k)$, il existe une matrice normalisée $G' = [Id_k|T]$ qui engendre le même code

$$[y_0, y_1, \dots, y_{n-1}] = [x_0, x_1, \dots, x_{k-1}, b_k, \dots, b_{n-1}]$$

Codage : $y = xG'$ (temps quadratique)

Décodage : $x =$ premiers bits de y (immédiat)

Détection : $H = [T^t| - Id_{n-k}]$, alors z est erroné ssi $Hz \neq 0$!!! (quadratique)

Correction : table précalculée des e de poids min. tels que $He = Hz \neq 0$ $y = z - e$ est le mot correct le plus proche de z (temps constant)

Code de Hamming

Code 1-correcteur à nombre de bits ajoutés minimal, $\delta = 3$

Idée : ajouter un contrôle de parité pour chaque puissance de 2 : $b_1, b_2, b_4, \cdots$
$\Rightarrow$ localisation de l'erreur
Rendement pour un code $C(n, n - \log_2(n) - 1)$

$$\simeq 1 - \frac{\log_2(n)}{n}.$$

Matrice pour $n = 6, k = 3$

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 \end{bmatrix}$$

Exemples classiques

Minitel(136,120) = Hamming(128,120)

1-correcteur avec un ajout de 8 bits toujours à 0, pour les grosses erreurs.
$\Rightarrow$ taux d'erreur $= 10^{-4}$; rendement $\simeq 88\%$; bits ajoutés $= 16$

Consultative Committee for Space Data Standard : échange de données spatiales avec RS(255,223)

$\Rightarrow$ 32-détecteur et 16-correcteur ; rendement $\simeq 87,5\%$; bits ajoutés $= 256$

CD audio : CIRC(32,28) (Cross Interleaved RS Code)

1. RS(255,251) distance 5
2. On ne prend que les mots de code commençant par un nombre donné de 0, puis on enlève les 0 (32,28) = RS raccourci.
3. La distance est conservée, donc aussi 4-détecteur et 2-correcteur
   $\Rightarrow$ taux d'erreur $= 10^{-5}$; rendement $\simeq 87,5\%$; bits ajoutés $= 32$

CODES CORRECTEURS D'ERREUR

Références

- Exposé de Jean-Guillaume Dumas à MidiSciences Codes détecteurs et correcteurs d'erreurs
- **Introduction aux sciences de l'information**, Jean-Yves Le Boudec, Patrick Thiran et Rüdiger Urbanke, Presses polytechniques et universitaires romandes, 2015
- **Théorie des codes**, J-G. Dumas, J-L. Roch, E. Tannier et S. Varrette, Dunod 2007, Site
- **Introduction aux codes correcteurs** Pierre Csillag, Ellipses 1990
- **L'information : L'histoire - La théorie - Le déluge** James Gleick, Cassini 2015

UNIVERSITÉ Grenoble Alpes