Sécurité des Réseaux

Jean-Yves Antoine

LI - Université François Rabelais de Tours

Jean-Yves.Antoine AT univ-tours.fr

UNIVERSITÉ FRANÇOIS - RABELAIS TOURS

Sécurité des réseaux

Codage : codes linéaires

UNIVERSITÉ FRANÇOIS - RABELAIS TOURS

Codes linéaires: motivations

- Concevoir de codes ayant la plus grande distance minimale possible (bonne capacité de détection et de correction)
- Concevoir des codes ayant un rendement optimal (transmission rapide)
- Concevoir des codes faciles à implémenter (structures particulières)

Espace vectoriel (rappels)

Définition

Soit K un corps muni de deux loi de composition internes **+** et **.**

On dit que E est un espace vectoriel sur un corps K ssi pour tous éléments u, v et w de E, on a :

- $(u+v) + w = u + (v+w)$
- $\exists 0 : u + 0 = 0 + u = u$
- $\forall u\ \exists(-u) : u + (-u) = 0$
- $u + v = v + u$
- $\forall c \in K,\ c . (u+v) = c.u + c.v$
- $\forall a,b \in K,\ (a+b).u = a.u + b.v$
- $\forall a,b \in K\ (a.b).u = a. (b.u)$
- $\exists 1 : 1 . u = u$

Base canonique

Soit $B = \{b_1, b_2, \ldots, b_n\}$ un ensemble d'éléments de l'espace vectoriel E

- B est dite famille génératrice de E si elle engendre E par combinaison linéaire: $\forall e \in E, \exists\ a_1, a_2, \ldots, a_n \in K$ tq $e = a_1. b_1 + a_2. b_2 + \ldots a_n. b_n$
- B est dite libre si aucun de ses éléments n'est une combinaison linéaire des autres
- B est une base de E ssi elle est libre et génératrice

Espace vectoriel $\{0,1\}^n$

- On note B le corps {0,1}. Pour n entier naturel fixé, il est possible de munir B^n d 'une structure d'espace vectoriel avec les lois de composition internes suivantes :
- Addition **+** addition booléenne bit à bit (ou exclusif logique)
- Multiplication **.** multiplication booléenne bit à bit (et logique)

Base canonique de $\{0,1\}^n$

- L'espace vectoriel B^n admet comme base (dite canonique) l'ensemble des vecteurs $\{e_1, e_2, \ldots, e_n\}$ défini comme suit :

$e_1=(1\ 0\ 0\ ..0)$ $e_2=(0\ 1\ 0\ \ldots 0)\ldots$ $e_n=(0\ 0..0\ 1)$

Poids d'un mot binaire

Définition

Soit un mot a élément de B^n. On appelle poids du mot a, noté w(a), sa distance de Hamming avec l'élément neutre $0_B{}^n = (0, \ldots, 0)$.

Donc : $w(a) = d_H(a, 0_B{}^n)$ = nombre de bits non nuls de a

Propriétés

Soient a,b,c trois éléments de l'espace vectoriel B^n, alors

- ✓ $a + a = 0_B{}^n$
- ✓ $d_H(a,b) = w(a+b)$
- ✓ $d_H(a,b) = d_H(a+c,b+c)$
- ✓ $d_H(a,b) = d_H(c,a+b+c)$
- ✓ L'équation $a+x = b$, d'inconnue x admet une unique solution $x = a+b$

Application linéaire

Soit deux espaces vectoriels E et F construit sur le corps B. On appelle application linéaire (interne) toute application f de E dans F qui vérifient

- $\forall x,y \in E,\ f(x+y) = f(x) + f(y)$
- $\forall \lambda \in B, \forall x \in E, f(\lambda.x) = \lambda. f(x)$

Définition

On appelle codage linéaire toute application linéaire injective de B^m dans B^n (m et n étant deux entiers tels que n>m)

Exemple

$$\begin{array}{c} 00 \\ 01 \\ 10 \\ 11 \end{array} \xrightarrow{\quad f \quad} \begin{array}{c} 000 \\ 100 \\ 011 \\ 111 \end{array}$$

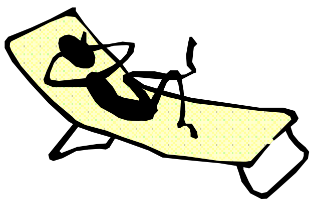
Matrice génératrice d'un code linéaire

Soit f un code linéaire de B^m dans B^n avec $n > m$. On note $(e_1..e_m)$ la base canonique de B^m .

On appelle matrice génératrice de f la matrice à deux dimensions (m lignes, n colonnes) dont la $i^{ème}$ ligne est f(ei).

Exemple

$$\begin{array}{c} 00 \\ 01 \\ 10 \\ 11 \end{array} \xrightarrow{f} \begin{array}{c} 000 \\ 100 \\ 011 \\ 111 \end{array} \qquad G = \begin{pmatrix} 011 \\ 100 \end{pmatrix}$$



Codage

- ✓ Codage facile et exhaustif de tous les éléments de $\{0,1\}^m$ à partir de la matrice génératrice du code :
- ✓ La matrice génératrice suffit pour effectuer le codage

Codes linéaires systématiques

Code systématique

Soit f un code de B^m dans B^n avec $n > m$. On dit que f est un code systématique si, $\forall x \in B^m$, le vecteur formé des m premiers bits de f(x) est égal à x.

Intérêt

La matrice génératrice d'un code linéaire systématique a une forme caractéristique (facilité d'utilisation) : matrice sous **forme normale**

Exemple : codage par bit de parité

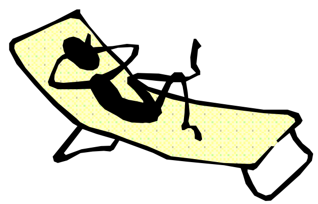
x	$\xrightarrow{f}$	$f(x)$
000		000**0**
001		001**1**
010		010**1**
011		011**0**
100		100**1**
101		101**0**
110		110**0**
111		111**1**

matrice identité

$$G = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix}$$

symboles d'information

symboles de redondance



Mise sous forme normale

Deux matrices G et G' de dimension (m x n) engendrent des codes linéaires équivalents si on peut obtenir G à partir de G' par une suite quelconque des opérations suivantes :

- permutation des lignes
- addition de deux lignes
- permutation de colonnes

Application : mise sous forme normale de la matrice, donc transformation en code systématique

Exemple

$$\begin{matrix}00\\01\\10\\11\end{matrix} \xrightarrow{f} \begin{matrix}000\\001\\110\\111\end{matrix} \qquad G = \begin{pmatrix}110\\001\end{pmatrix}$$

$$G' = \begin{pmatrix}101\\010\end{pmatrix} \qquad \begin{matrix}00\\01\\10\\11\end{matrix} \xrightarrow{f'} \begin{matrix}\mathbf{00\ 0}\\\mathbf{01\ 0}\\\mathbf{10\ 1}\\\mathbf{11\ 1}\end{matrix}$$

Propriétés

- $(0..0)_n$ est toujours mode de code d'un codage linéaire de B^m dans B^n
- Pour un code linéaire, la somme de deux mots de code linéaire est toujours un mot de code.

Distance minimale

- La distance minimale d'un code linéaire est égale à son poids minimal (i.e. poids minimal d'un mot de code non nul)

Exemple

$$G = \begin{pmatrix} 101010 \\ 110001 \\ 000111 \end{pmatrix}$$

- mots du code
- distance minimale
- capacité de correction
- code systématique ?

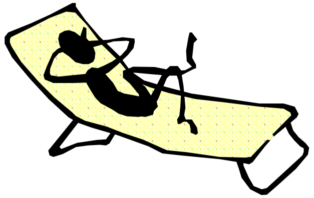
Vecteur d'erreur

Soit f un code linéaire de B^m dans B^n. On note MC le message correct émis et R le message reçu.

- ✓ Si la transmission est sans erreur alors R=MC
- ✓ Sinon, on appelle **vecteur d'erreur** le vecteur de B^n , noté e tel que R = e +MC

Le **poids de e** est le **nombre d'erreurs de transmission**

Propriété



On a **e = R + MC** ⇨ calcul direct de l'erreur !

On a **MC = R + e** ⇨ calcul direct du mot émis si erreur connue !

Correction : cas linéaire général

1. Réception message R : si R est un mot de code, alors R=C et on décode son antécédent E dans B^m
2. Si R n'appartient pas au code, on fait la liste de tous les vecteur d'erreurs envisageables e de $\{0,1\}^n$ tq e = R + C, avec C mot de code quelconque.
3. Correction par maximum de vraisemblance : le vecteur d'erreur retenu e_{max_vrais} est celui qui est de poids minimal
4. Alors on corrige le mot R dans le mot de code $C = R + e_{max\_vrais}$
5. On décode alors l'antécédent de C dans B^m

Exemple

$$G = \begin{pmatrix} 100011 \\ 010101 \\ 001110 \end{pmatrix}$$

- mots de code correspondant aux vecteurs de la base :

- codage de (1 1 0) :

- décodage de (0 0 1 1 1 0) :

- décodage de (1 1 1 0 0 1) :

Décodage: tableau standard

Relation d’équivalence (rappel)

On appelle relation d’équivalence sur un ensemble E toute relation *R* qui est réflexive, symétrique et transitive

Classe d’équivalence (rappel)

Soit *R* une relation d’équivalence sur un ensemble E. On appelle classe d’équivalence C de la relation *R* tout sous-ensemble d’élément de E en relation : $\forall e_1, e_2 \in C$, $e_1 \, R \, e_2$

Théorème : soit R une relation d’équivalence définie sur un ensemble E. Alors les classes d’équivalences de R partitionnent E

Codes linénaires : classes latérales

Soit C un code linéaire de B^m dans B^n. On définit la relation *R* tq $\forall e1,e2 \in B^n$, $e_1 \, R \, e_2$ si e_1+e_2 est un mot de code de C

Les classes d’équivalence de la relation *R* sont appelées classes latérales du code C. Elle partionnent B^n

Décodage: tableau standard

Classes latérales: propriétés

- Les mots de code forment une classe latérale de C
- Pour chaque classe latérale C_i, il existe un vecteur e, appelé *tête de liste*, tel que tous les éléments de C_i s'écrivent comme la somme d'un mode de code et de e : $C_i = \{ e + MC \mid MC \in C \}$

Tableau standard

Code linéaire de B^m dans B^n : tableau **2^{n-m} lignes x 2^m colonnes**

- **1ère ligne** : classe des mots de code
- **1ère colonne** : vecteurs têtes de liste (remplissage du tableau de haut en bas avec des têtes de listes de poids minimal)

Décodage : têtes de listes de poids minimal $\Rightarrow$ le mot de code de correction est celui qui est dans la colonne du mot reçu

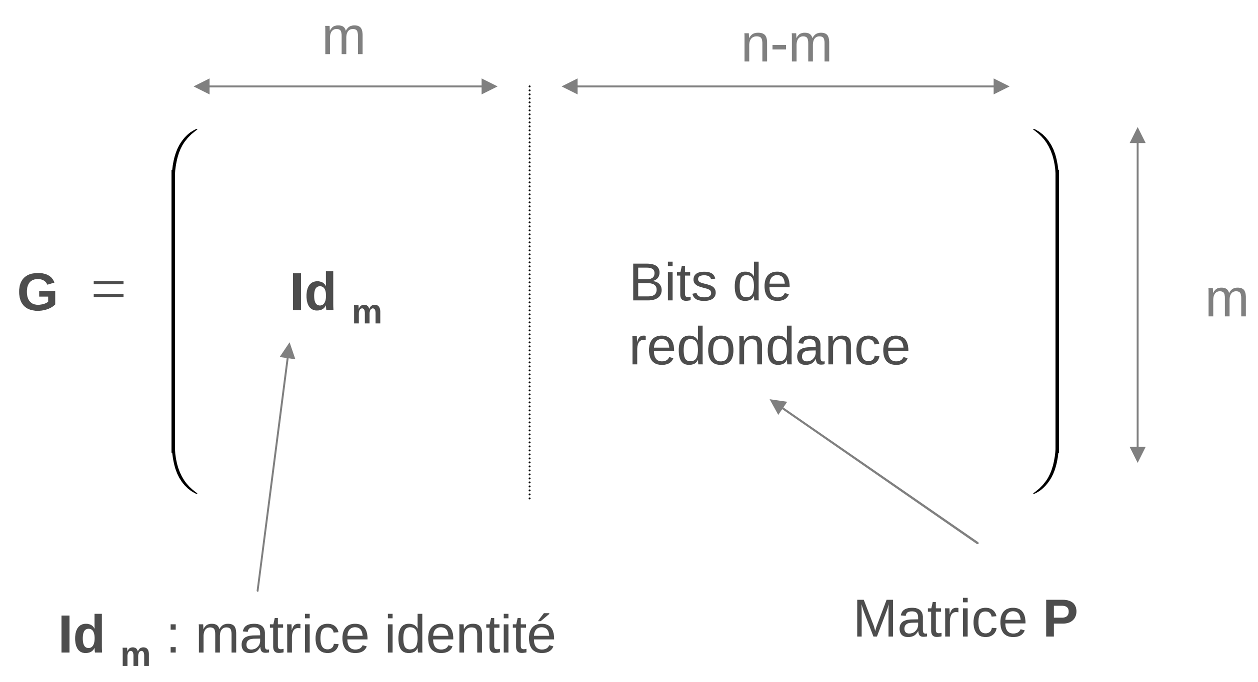
Exemples $G1 = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 \end{pmatrix}$ $G2 = \begin{pmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}$

Décodage: codes systématiques

Matrice génératrice sous forme normale (rappel)

Soit C : $B^m \rightarrow B^n$, un **code linéaire systématique** de matrice génératrice G. On peut alors écrire G sous la forme normale :

$G = (Id_m\ P)$ avec P matrice quelconque m x (n-m)



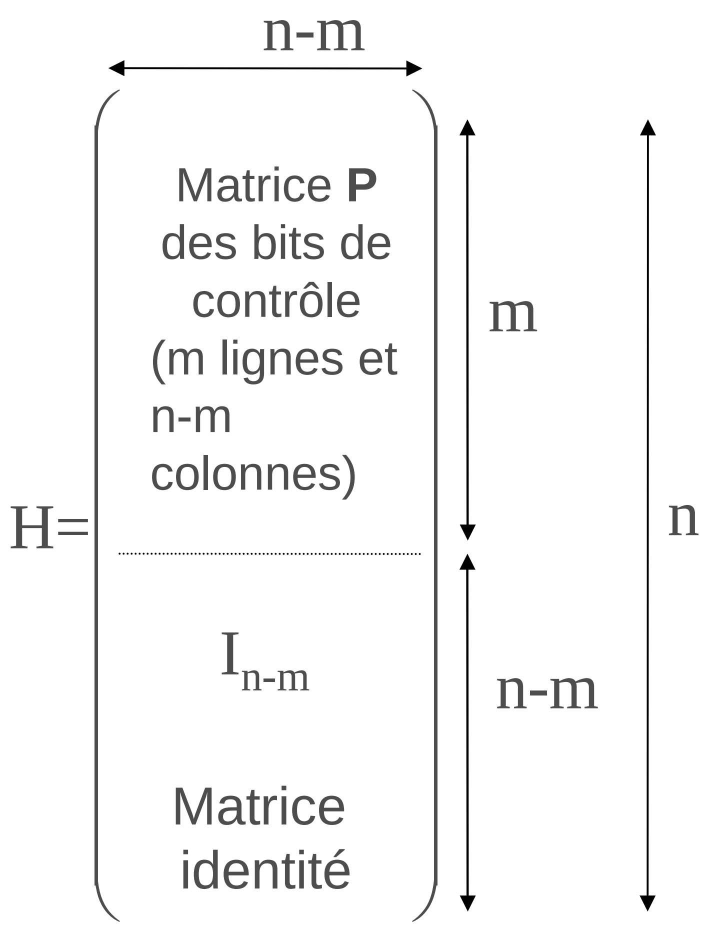
Décodage par syndrome

Matrice de contrôle

Soit C : $B^m \rightarrow B^n$, un **code linéaire systématique** de matrice génératrice $\mathbf{G = (Id_m\ P)}$

On appelle **matrice de contrôle** du code C la matrice H tq:

$$\mathbf{H} = \begin{pmatrix} \mathbf{P} \\ \mathbf{Id_{n-m}} \end{pmatrix}$$



Syndrome

Soit M un mot quelconque de B^n susceptible d'être reçu. On appelle **syndrome de M** le vecteur à n-m colonnes noté s(M) tq :

$$\mathbf{S(M) = M.H}$$

Décodage par syndrome

Syndrome : propriétés

Soit un **code linéaire systématique** noté f: $B^m \to B^n$. Soit M_i un message reçu (un élément quelconque de B^n). Alors

1. $\forall$ M $\in$ B^n, M est un mot de code si et seulement si S(M) = 0
2. Deux mots M_1 et M_2 ont même syndrome ssi ils sont dans la même classe latérale

Décodage par syndrome

Bijection entre syndromes et classes latérales:

1. Au lieu de mémoriser tout le tableau standard, on se limite aux syndromes des têtes de listes (**table des syndromes**)
2. A la réception d'un mot M, on calcul son syndrome pour connaître sa classe d'équivalence et sa tête de liste e : **erreur connue**
3. On corrige par **MC = M + e**

Décodage par syndrome

Exemple

$$G1 = \begin{pmatrix} 1\ 0\ 1\ 0\ 1\ 0 \\ 0\ 1\ 0\ 1\ 0\ 1 \end{pmatrix}$$

- mots de code :
- distance :
- tableau standard (têtes de liste) :
- décodage de (1 1 1 1 1 1) :
- décodage de (1 1 0 1 1 1) :

Exemple de code systématique

Code de Golay étendu (exemple : G_{24})

- code des sondes Voyager : longueur n = 24, rendement r = 50%
- matrice génératrice 12 x 24

$$G = (Id_{12}\ B) \text{ avec B tq } = \begin{pmatrix} 1&1&0&1&1&1&0&0&0&1&0&1 \\ 1&0&1&1&1&0&0&0&1&0&1&1 \\ 0&1&1&1&0&0&0&1&0&1&1&1 \\ 1&1&1&0&0&0&1&0&1&1&0&1 \\ 1&1&0&0&0&1&0&1&1&0&1&1 \\ 1&0&0&0&1&0&1&1&0&1&1&1 \\ 0&0&0&1&0&1&1&0&1&1&1&1 \\ 0&0&1&0&1&1&0&1&1&1&0&1 \\ 0&1&0&1&1&0&1&1&1&0&0&1 \\ 1&0&1&1&0&1&1&1&0&0&0&1 \\ 0&1&1&0&1&1&1&0&0&0&1&1 \\ 1&1&1&1&1&1&1&1&1&1&1&0 \end{pmatrix}$$

code cyclique

(chapitre 3)

- distance d = 8
- corrige 3 erreurs sur 24 bits transmis (soit 12,5%)

Définition

Codes linéaires définis à partir de leur matrice de contrôle.

On appelle **code de Hamming de redondance r**, noté Ham(r), tout code de matrice de contrôle r colonnes x (2^r-1) lignes dont les lignes correspondent à tous les vecteurs non nuls de B^r

Propriétés

- Longueur du code $n = 2^r - 1$
- Dimension $m = 2^r - r - 1$

- Équivalence par permutation des vecteurs de la matrice de contrôle : code **systématique**
- Distance du code ($r \geq 2$) $d = 3$
- Codes parfaits

B

$$H_{Ham(3)} = \begin{pmatrix} 111 \\ 110 \\ 101 \\ 011 \\ 100 \\ 010 \\ 001 \end{pmatrix}$$

Décodage

• distance d = 3 ⇨ code 1-correcteur

• syndrome : têtes de listes de poids 1 (outre $0_B{}^r$) ⇨base de B^r

Codes de Hamming éténdu

Codes obtenus à partir d'un code de Hamming Ham(r) en ajoutant d'autres bits de contrôle : $\mathbf{G = (G_{Ham(r)}\ G_{ext})}$

Extension par ajout d'un bit de parité

- Longueur du code n = 2^r
- Distance du code d = 4

Codes de Reed-Muller

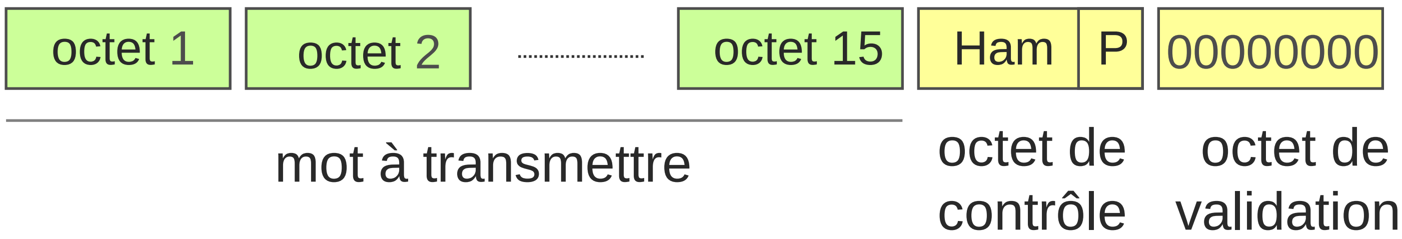
Exemple : code du Minitel

Transmission

- réseau téléphonique
- par paquets de 15 octets (120 bits)
- taux d’erreur relativement faible : 1 à 2 caractère par page

Codage par code de Hamming étendu

- code Ham(7) : longueur 127 dont 7 bits de contrôle
- extension : bit de parité
- dernier octet $0_B{}^8$ de validation (détection perturbations importantes)



Codes de Reed-Muller

Définition

Code de Reed-Muller d'ordre r et de longueur 2^n, noté RM(r,n) : code de matrice génératrice G(r,n) définie comme suit :

- $G(0,n) = (1\ 1\ 1\ \ldots\ldots\ 1) = \{\ 1\ \}^{2^n}$
- $G(n,n) = \begin{pmatrix} G(n-1,n) \\ 0 \ldots\ldots\ldots 0\ 1 \end{pmatrix}$
- $G(r,n) = \begin{pmatrix} G(r,n-1) & G(r,n-1) \\ 0 & G(r-1,n-1) \end{pmatrix}$ pour $0 < r < n$

Propriétés

- codes de Hamming étendus
- longueur $n = 2^n$
- dimension $m = \sum_{i=0}^{i=r} C_n^i$
- distance $d = 2^{n-r}$

Sonde Mariner-9

Code RM(1,5)
6 bits codés sur 32 bits
Corrige jusqu'à 7 erreurs

Comparaison de quelques codes linéaires

	Codage	rend^{mt}	distance	correction
Ham(3) + parité	4bits → 7bits	57%	4 bits	1 bit (6%)
Ham(7) + parité *Minitel*	120 bits → 128 bits	94%	4 bits	1 bit (1%)
Golay étendu G_{24} *Voyager*	12 bits → 24 bits	50%	8 bits	3 bits (12,5%)
Reed-Muller RM(1,5) : *Mariner*	6 bits → 32 bits	19%	15 bits	7 bits (22%)