Introduction
à la
Théorie de l'Information

Définition de l ’Information
Correction des erreurs

- Une information est un couple constitué:

  d'une représentation matérielle, qui en constitue le formant

  et d'un ensemble d'interprétations, qui en constitue le formé

  dont la nature, événementielle, consiste en un changement d'état qui, par l'occurence de cette représentation matérielle, provoque l'activation du champ interprétatif correspondant, selon les régles fixées par un code préétabli.

  Georges Ifrah - Histoire universelle des chiffres

  - Seule la composante matérielle (formant) d'une information fait l'objet d'une communication: ce n'est pas le sens (formé) que l'on transmet
  - L'information est la troisième dimension universelle après la matière et l'énergie. L'information n'est autre que la néguentropie (structure ordonnée)
  - Étymologie: informare= donner une forme....

- Quantité d'information

  Mesure quantitative de l'incertitude d'un message en fonction du degré de probabilité de chaque signal composant ce message

- **Information (suite...)**
  - Séquence de signaux, correspondant à des règles de combinaisons précises, transmise entre une source et un collecteur par l'intermédiaire d'un canal
  - Ecrit, fait, notion ou instruction susceptible d'être traitée en tout ou partie par des moyens automatiques.
  - Renseignements obtenus de quelqu'un ou sur quelqu'un ou quelque chose, en particulier nouvelle communiquée par la presse, la radio,...
- **Message**
  - Lot d'information formant un tout intelligible ou exploitable et transmis en une seule fois
  - Séquence de signaux qui correspondent à des règles de combinaisons précises et qu'une source transmet à un collecteur par l'intermédiaire d'un canal
- **Signal**
  - Phénomène physique porteur d'une information et pouvant représenter des données
  - Variation d'une grandeur de nature quelconque grâce à laquelle, dans un équipement, un élément en influence un autre
  - Signe convenu pour avertir, annoncer, donner un ordre.

Notations

- X Symbole émis par une source
- Y Symbole reçu par l'observateur au collecteur
- $P(x_k)$ Probabilité que $X = x_k$
- $P(x_k / y_j)$ Probabilité d'avoir émis $X = x_k$ si on a reçu $Y = y_j$
- $P(x_k ; y_j)$ Probabilité émettre $X = x_k$ et de recevoir $Y = y_j$

Définitions

$$I(x_k) = -\log_2 P(x_k) \text{ bits}$$

- Information fournie par la source

$$I(x_k ; y_j) = -\log_2 \{P(x_k / y_j) / P(x_k ; y_j)\} \text{bits}$$

- Information mutuelle
- Entropie de la source : H(X) espérance mathématique de $I(x_k)$

$$H(X/Y) = -\sum\sum P(x_k ; y_j) \log_2 \{P(x_k / y_j)\}$$

- Equivocation: Indetermination sur X quand on a reçu Y

$$I(X,Y) = H(X) - H(X/Y)$$

- Information transmise: I(X,Y)
- Capacité d'un canal discret

$$C \text{ bits} = MAX\{I(X,Y)\} \text{ pour toutes les lois } P(X)$$

- Codage binaire : erreur si 0 ➔ 1 ou 1 ➔ 0
- Correction d'une erreur: il suffit de connaître son emplacement
  puis de transformer 1 ➔ 0 ou 0 ➔ 1
- Syndrome d'erreurs; exemple
  - Chaîne à transmettre: 1 0 1 1 0 0 0 1 1 1 0 1 0 1 0 1 1 1 0 0 0 0 1 0 1 0 1 0
  - Chaîne reçue: 1 1 1 1 0 0 0 1 1 1 0 1 0 1 0 1 1 0 0 1 0 0 1 1 1 0 1 0
  - syndrome d'erreurs: 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 1 1 0 0 1 0 0 0 0
  - erreurs isolées: x
  - paquet d'erreurs de taille 7: x x x x x x x
- **Exemples d'erreurs observées**
  - pour des blocs de 100 octets sur liaison téléphonique
  - erreurs isolées 36 % des cas
  - paquets de taille > 10 : 15 % des cas

- **Jeu : soit à coder les valeurs : 1 2 3, on code 2**
  - **premier choix:** C E F, soit E
    - **ces lettres sont trop semblables et une altération peut transformer E en C ou F**
  - **second choix:** **A E J**
    - **si on reçoit C (ou F) une erreur est détectée à coup sûr.**
    - **Ce symbole étant très proche de E , il est TRES PROBABLE que ce soit un E bruité et on le corrigera**
- **Exemples de codes simples**

code 1: distance1	Code 2: distance 2	Code 3: distance 3
0000	0000	0000
0001	0011	1011
0010	0101	
0011	0110	corrige 1 erreur
0100	1001	si on reçoit 1001 on en déduit
.....	1010	que 1011 a été émis
.....	1100	ou détecte 2 erreurs
.....	1111	on peut recevoir 1000

- caractérise la faculté de séparer les symboles d'un code
- DONC sa puissance de détection ou de correction
- Pour détecter p erreurs
  - d ³ p+1
- pour corriger q erreurs
  - d ³ 2q+1
- pour détecter p erreurs ET en corriger q (autres)
  - d ³ p + q + 1
- exemples
  - si la distance de Hamming est 3 : détection de 2 erreurs OU correction 1 erreur
  - si la distance de Hamming est 4 : correction des erreurs simples
    ET détection des erreurs doubles

octet							
1	1	0	0	1	0	0	bits 1
0	1	1	1	0	0	0	
1	1	1	0	0	0	0	
0	0	0	0	1	0	0	bits j
1	1	1	1	0	1	0	
0	0	1	1	0	0	1	
0	0	0	1	0	1	1	
							Parité Croisée
0	1	1	1	1	1	?	en pratique 1

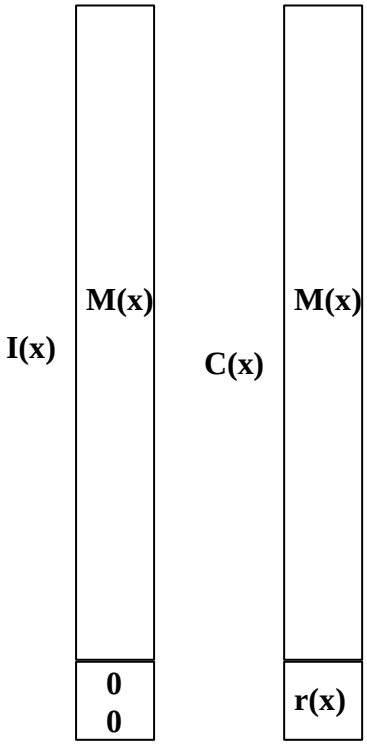
- **Technique**
  - simple
  - courante
  - moyennement performante
- **parité sur caractère**
  - parité verticale
  - toujours sur coupleur
- **parité sur bloc**
  - ième bit des caractères
  - "check sum"
- **parité**
  - paire (even): mode arythmique
  - impaire (odd): mode synchrone

- **Codage binaire: Coefficients d'un polynôme en x (à valeur dans GC2)**
  - **exemple: 1101011 est représenté par** $1*x^6 + 1*x^5 + 0*x^4 + 1*x^3 + 0*x^2 + 1*x^1 + 1*x^0$

    soit $x^6 + x^5 + x^3 + x + 1$
  - **Opérateur dans GC2: + est le "ou exclusif" (1+1=0)**
- **Code cyclique:**
  - **Tout code dont la représentation M(x) est multiple d'un polynôme GENERATEUR G(x)**
  - **G(x) doit être un polynôme irréductible**
    - **les polynômes irréductibles sont donnés par la table de Peterson dans "Error Correcting Codes" de WW Peterson et EJ Weldon**
    - **exemples: $x+1$, x^2+x+1, x^3+x+1, x^3+x^2+1**
    - **$x^{15}+x+1$ et $x+1$ donnent $x^{16}+x^{15}+x^2+1$**
    - **$x^{15}+x^{14}+x^{13}+x^{12}+x^4+x^3+x^2+x+1$ et $x+1$ donnent $X^{16}+x^{12}+x^5+1$**
    - **$x+1$ génère une parité paire**
    - **$x^{16}+x^{12}+x^5+1$ est le code à 16 bits CCITT**
    - **$x^{16}+x^{15}+x^2+1$ est le code CRC16**

- On choisit un polynôme générateur G(x)
- Soit à transmettre M(x)
- Mot du code émis: $C(x) = M(x) * G(x)$
- Transmission: syndrome d'erreurs E(x)
- On reçoit $C^*(x) = C(x) + E(x)$
- Décodage
  - $Q(x) = C^*(x) / G(x) + R(x)$
  - si $R(x) = 0$ alors $M(x) = Q(x)$
  - sinon erreur détectée
- correction
  - demander la répétition de C(x)

  ou
  - calculer $E(R(x))$ et corriger par $C(x) = C^*(x) + E(x)$



- On choisit un polynôme générateur G(x) de degré k
- Soit à transmettre M(x)
- $I(x) = x^k * M(x)$: décalage de k bits
- r(x) = I(x) - G(x) * Q(x), reste de la division de I(x) par G(x)
- Mot du code émis: C(x) = I(x) + r(x)
- Transmission: syndrome d'erreurs E(x)
- On reçoit $C^*(x) = C(x) + E(x)$
- Décodage
  - $Q(x) = C^*(x) / G(x) + R(x)$
  - si R(x) = 0 alors $M(x) = C^*(x) / x^k$: Troncature
  - sinon erreur détectée
- correction
  - demander la répétition de C(x)
  - ou calculer E(R(x)),corriger par $C(x) = C^*(x) + E(x)$ et troncature