Début du cours de Théorie de l'information

Ecrit par Marion Candau

Enseignant : M. Gilles Zémor

Master 1 Cryptologie et Sécurité Informatique
Université Bordeaux 1

2009 - 2010

Table des matières

Chapitre 1

Rappels de probabilités

1.1 Définitions

Définition : Probabilité
Une probabilité est une application dans un espace probabilisé (Ω, P) qui est définie comme suit : $P : \mathcal{P}(\Omega) \to [0, 1]$ et qui vérifie les propriétés $P(\emptyset) = 0$, $P(\Omega) = 1$ et $P(A \cap B) = P(A) + P(B)$ si $A \cap B = \emptyset$.

Définition : Variable aléatoire
Une variable aléatoire est une application $X : \Omega \to \mathbb{R}$. La loi de X est donnée par les probabilités $P(X = x) = P(X^{-1}(x)) = P(\{\omega, X(\omega) = x\})$

Définition : Espérance de X
L'espérance d'une variable aléatoire X est définie comme suit :

$$E[X] = \sum_{\omega \in \Omega} X(\omega)P(\omega) = \sum_{x \in Im(X)} x \times P(X = x)$$

Théorème
$E[X + Y] = E[X] + E[Y]$

Définition : Fonctions indicatrices
C'est une fonction définie comme suit :

$$A \in \mathcal{P}(\Omega), 1_A = \begin{cases} 1 & \text{si } \omega \in A \\ 0 & \text{sinon.} \end{cases}$$

On a la propriété suivante : $P(A) = E[1_A]$

1.2 Probabilités conditionnelles, indépendance

Définition

$P(A|B) = \dfrac{P(A \cap B)}{P(B)}$ si $P(B) \neq 0$

$P(X = x|Y = y) = \dfrac{P(X = x, Y = y)}{P(Y = y)}$ si $P(Y = y) \neq 0$

$A, B \subset \Omega$ sont indépendants si $P(A \cap B) = P(A) \times P(B)$

Théorème

$E[XY] = E[X] \times E[Y]$ si X et Y sont indépendantes.

Théorème : Inégalité de Markov

Pour $X \geqslant 0$ on a :

$$P(X \geqslant k \times E[X]) \leqslant \frac{1}{k}$$

Loi des grands nombres

Soient $X_1, X_2, \ldots, X_n$ n variables indépendantes et de même loi, on a donc $X = \sum_{i=1}^{n} X_i$ et on a la formule suivante :

$$P\left(\left|\frac{X}{n} - E[X_1]\right| > \epsilon\right) \underset{n\to\infty}{\longrightarrow} 0$$

Définition : Variance

$Var(X) = E\left[(X - E[X])^2\right]$

L'écart type est quant lui : $\sigma(X) = \sqrt{Var(X)}$

Théorème de Tchebichev

$$P\left(|X - E[X]| > k \times \sigma(X)\right) \leqslant \frac{1}{k^2}$$

Théorème

Si X et Y sont deux variables aléatoires indépendantes alors

$$Var(X + Y) = Var(X) + Var(Y)$$

et si $X_1, \ldots, X_n$ sont indépendantes et de même loi alors

$$Var(X_1 + \ldots + X_n) = n \times Var(X_1)$$

1.3 Mesure de l'information

Définition : Entropie

Si X est une variable de Bernoulli alors $P(X = 1) = p$. On définit l'entropie de X notée $H(X)$ par :

$$H(X) = p \log \left(\frac{1}{p}\right) + (1 - p) \log \left(\frac{1}{1 - p}\right)$$

Définition

Si X est une variable aléatoire ayant pour loi $p = \{p_1, p_2, \ldots, p_n\}$ alors

$$H(X) = p_1 \log_2 \left(\frac{1}{p_1}\right) + p_2 \log_2 \left(\frac{1}{p_2}\right) + \ldots + p_n \log_2 \left(\frac{1}{p_n}\right)$$

Cette valeur se mesure en bits ou shannons.

Chapitre 2

Grandeurs Informationelles

Rappels

L'entropie H(X) est la grandeur définie comme suit :

$$H(X) = \sum_x P(X = x) \times \log \frac{1}{P(X = x)}$$

$$H(X, Y) = \sum_{x,y} P(X = x, Y = y) \times \log \frac{1}{P(X = x, Y = y)}$$

Définition : Distance de Kullback

Soient $p = \{p_1, p_2, \ldots, p_n\}$ et $q = \{q_1, q_2, \ldots, q_n\}$, on a :

$$D(p||q) = \sum_{i=1}^{n} p_i \times \log \frac{p_i}{q_i}$$

Proposition

$$\forall p, q, D(p||q) \geqslant 0$$

Proposition

Si X prend m valeurs avec une loi $p = \{p_1, p_2, \ldots, p_n\}$ on a :

$$H(X) \leqslant \log_2 m$$

L'galit est atteinte si et seulement si la loi de X est uniforme c'est-à-dire $\forall i, p_i = \frac{1}{m}$.

Digression

Soit X valeurs dans $\mathbb{N}$ avec pour loi $p = \{p_1, p_2, \ldots, p_n, \ldots\}$ et $\sum_{i=1}^{\infty} p_i = 1$.

Soit $E[X] = m$ fixée. Quel est le maximum de H(X) ?
Ce maximum est donnée par la loi géométrique $p_i = \gamma^i \times (1 - \gamma)$
Cherchons γ :

$$E(X) = \sum_i i \times p_i = (1-\gamma) \sum_i i\gamma^i = (1-\gamma) \times \gamma \times \sum_i i\gamma^{i-1} = \frac{\gamma}{1-\gamma} = m$$

D'où $\gamma = \frac{m}{m+1}$.

2.1 Entropie Conditionnelle

Définition

$$H(X|Y) = \sum_{x,y} P(X = x, Y = y) \times \log_2 \frac{1}{P(X = x|Y = y)}$$

$$= \sum_y P(Y = y) \sum_x P(X = x|Y = y) \times \log_2 \frac{1}{P(X = x|Y = y)}$$

Proposition

$$H(X, Y) = H(Y) + H(X|Y)$$

Définition : Information mutuelle
$I(X, Y)$ est l'information mutuelle de X et Y.

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

$$= H(X) - H(X|Y)$$

$$= H(Y) - H(Y|X)$$

Proposition

$$I(X, Y) \geqslant 0$$

2.2 Application la cryptologie

Soit un système de chiffrement avec M le message en clair, K la clé et $C = f(M, K)$ le message chiffré. Le système est dit parfait si $H(M|C) = H(M)$.

Théorème
Si un système est parfait alors $H(K) \geqslant H(M)$.

2.3 Distance d'unicité

Définition
C'est d le plus petit m tel que $H(K|C_1, \ldots, C_m) = 0$

$$d \geqslant \frac{H(K)}{\log(\#\mathcal{C}) - h}$$

avec $h = \dfrac{H(M_1 \ldots M_m)}{m}$

Chapitre 3

Codage de source (compressif)

Soit une variable aléatoire X qui prend ses valeurs dans $\mathfrak{X}$.
codage : $c : \mathfrak{X} \longrightarrow \{0,1\}^*$.
c^* est obtenu partir de c par concaténation. $C = c(\mathfrak{X})$ est appel code.

Définition
On dit que C est uniquement déchiffrable si :

$$\forall m \in C^*, \exists! c_1, \ldots, c_k \in C, m = c_1 \ldots c_k$$

Cas particulier : Code préfixe
C est préfixe si $\forall c, c' \in C$, c n'est pas préfixe de c'.

Proposition
Un préfixe est un code uniquement déchiffrable.

Remarque
Mais un code uniquement déchiffrable n'est pas nécessairement un code préfixe.

Définition : longueur de X

$$\bar{l}(c(X)) = \sum_{x \in \mathfrak{X}} P(X = x) l(c(X))$$

où $l(m_1 \ldots m_i) = i$ (nombre de bits de $C(x_i)$).

Remarque
Un code préfixe se représente par un arbre.

Proposition
Soient $C = \{c_1, \ldots, c_m\}$ et $l_i = l(c_i)$. Si C est préfixe alors :

$$\sum_{i=1}^{m} \frac{1}{2^{l_i}} \leqslant 1$$

C'est l'inégalité de Kraft.

Théorème de Kraft
Soient $l_1, \ldots, l_m \in \mathbb{N}$ et $l_i = l(c_i)$,

$$\exists C = \{c_1, \ldots, c_m\} \text{ code préfixe} \Longleftrightarrow \sum_{i=1}^{m} 2^{-l_i} \leqslant 1$$

Théorème de McMillan
Soient $l_1, \ldots, l_m \in \mathbb{N}$ et $l_i = l(c_i)$,

$$\exists C = \{c_1, \ldots, c_m\} \text{ uniquement déchiffrable} \Longleftrightarrow \sum_{i=1}^{m} 2^{-l_i} \leqslant 1$$

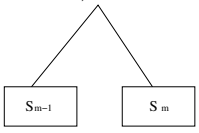
Proposition
Soit X valeurs dans $\mathfrak{X} = \{x_1, \ldots, x_n\}$, de loi $p = \{p_1, \ldots, p_n\}$. Alors pour tout codage c uniquement déchiffrable de $\mathfrak{X}$, on a :

$$\bar{l}(c) = \sum_{x} P(X = x) l(c(X)) \geqslant H(X)$$

Proposition
Il existe toujours un codage c préfixe tel que $\bar{l}(c) \geqslant H(X) + 1$.

Algorithme de Huffman
Un exemple vaut toujours mieux qu'un long discours ;) ==> 3.1.

Lemme
Soient $p = \{p_1, \ldots, p_m\}$ avec $p_1 \geqslant p_2 \geqslant \ldots \geqslant p_m$. Parmi les codes optimaux, il existe un code préfixe tel que :



et $l(s_m) = l(s_{m-1})$ maximum.

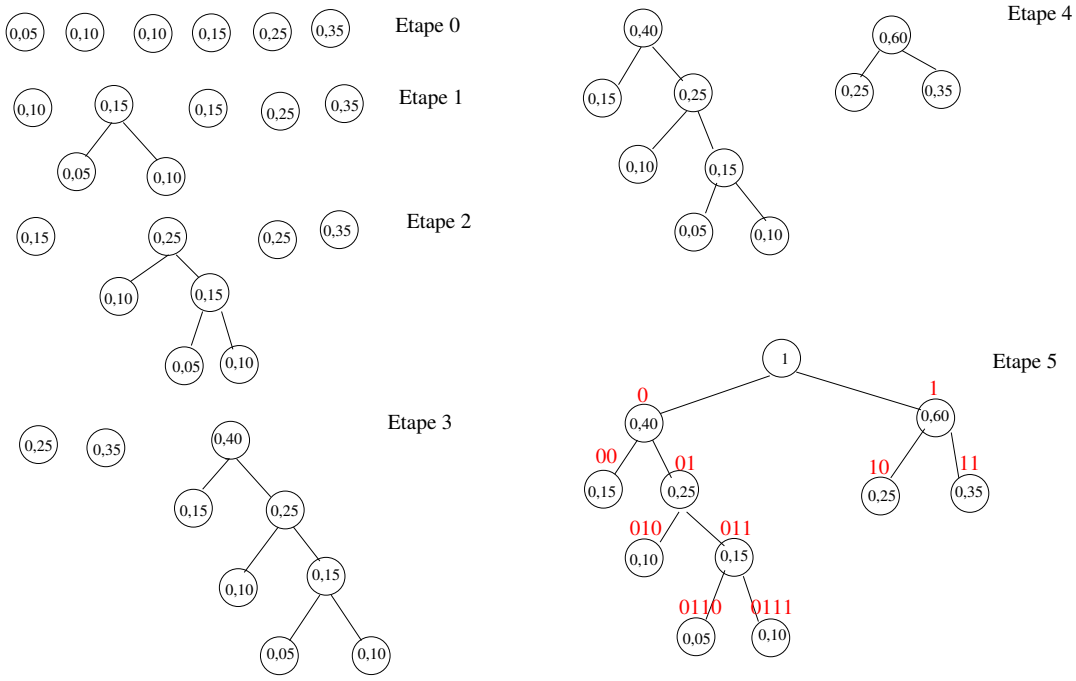
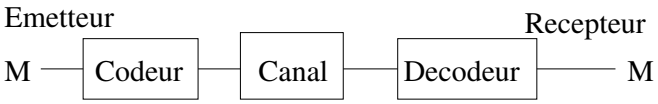


FIG. 3.1 – Exemple de l'algorithme de Huffman

Chapitre 4

Codage de canal



4.1 Canal binaire symétrique

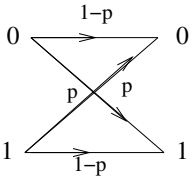
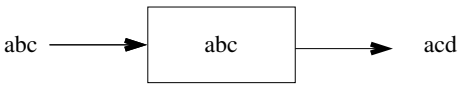


FIG. 4.1 – Exemple de canal binaire symétrique

Plus généralement on définit un canal discret sans mémoire.
Le canal a pour entrée un alphabet noté x et pour sortie un autre alphabet



noté y, il est représenté également par des probabilités $P(Y = y|X = x) = p_{xy}$. Pour tout x, $\sum_y p_{xy} = 1$.

Analyse "nave"

Soit un canal binaire qui a des probabilités p. La stratégie pour communiquer est d'utiliser un code $C \subset \{0,1\}^n$. L'ensemble des messages possibles est donc : $|C| = M$. On définit le rendement du code R tel que $|C| = 2^{Rn}$ avec $0 \leqslant R \leqslant 1$.

Quel est le rendement maximum ?

La distance de Hamming dans $\{0,1\}^n$ est définie comme suit :

$$\forall x, y \in \{0,1\}^n, d_H(x,y) = \#\{i, x_i \neq y_i\} = \text{poids}(x+y)$$

Soient $x = (x_1, \ldots, x_n)$ un mot émis et $y = (y_1, \ldots, y_n)$ un mot reçu. On a :

$$d_H(x,y) = pn + o(\sqrt{n})$$

$P(X = x|Y = y)$ ne dpend que de $d_H(x,y)$. Pour que C soit fiable ; il faudrait que les boules de rayon pn ayant pour centre un mot de code émis soient disjointes. Soit $S(x,pn) = \{y, d(x,y) = pn\}$, on a $|C| \times |S| \leqslant 2^n$ et $|s| = \begin{pmatrix} n \\ pn \end{pmatrix}$ D'où :

$$2^{Rn} = |C| \leqslant \frac{2^n}{\begin{pmatrix} n \\ pn \end{pmatrix}} \approx 2^{n-nh(p)}$$

$$R \leqslant 1 - h(p)$$

Définition

On appelle capacité du canal $C = \max_{\text{loi de X}} I(X,Y)$
avec $I(X,Y) = H(X) - H(X|Y) = H(Y) - H(Y|X)$.

Exemple pour le canal binaire symétrique

$$I(X,Y) = H(Y) - h(p)$$

Si X uniforme, $I(X,Y) = 1 - h(p) \Longrightarrow C = 1 - h(p)$.

4.2 Canal binaire à effacements

$I(X,Y) = H(X) - H(X|Y) = H(X) \times (1-p)$. D'où $C = 1 - p$.
Soit $x = (x_1, \ldots, x_n) \in C \subset \{0,1\}^n$ un mot reçu, et $y = (x_1, x_2, ?, x_4, ?, \ldots) =$

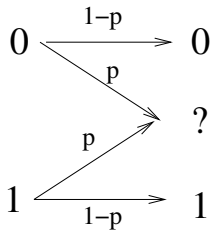


FIG. 4.2 – Canal binaire effacements

$[x_i|?\ldots?]$. Le nombre de ? est de pn donc le nombre de bits corrects est de $(1-p)n$. Pour retrouver les bits effacés, il faut donc que $|C| \gg 2^{n(1-p)}$.

4.3 Canal en "Z"

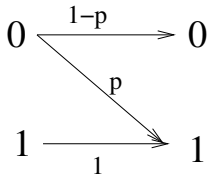


FIG. 4.3 – Canal en "Z"

On a : $I(X,Y) = H(X) - H(X|Y)$ mais $H(X|Y)$ ne s'écrit pas en fonction de $H(X)$. L'exercice 2 du TD5 illustre cet exemple.

4.4 Clavier bruité

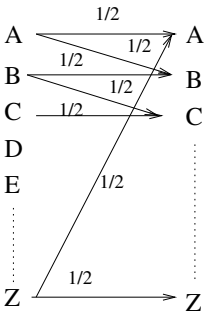


FIG. 4.4 – Clavier bruité

$I(X,Y) = H(Y) - H(Y|X) = H(Y) - 1$ d'où $C = \log_2 26 - 1 = \log_2 13$

Chapitre 5

Codes correcteurs d'erreurs et d'effacements

Code : $C \subset \{0,1\}^n$
Encodage : $\{0,1\}^k \to C \subset \{0,1\}^n$
Paramètre utile pour la correction : distance de Hamming entre n-uples :

$$\forall x = (x_1, \ldots, x_n), y = (y_1, \ldots, y_n), d_H(x,y) = \#\{i, x_i \neq y_i\}$$

Définition : distance minimale de C

$$d = d(C) = \min_{c,c' \in C, c \neq c'} d_H(c,c')$$

Correction d'effacements

Proposition
Si # effacements $< d(C)$ alors on peut retrouver le mot émis $c = (c_1, \ldots, c_n)$. c est le seul mot de C vérifiant $x_i \in \{0,1\} \Rightarrow x_i = c_i$.

Proposition
Si #erreurs$< \frac{d}{2}$, alors le mot le plus proche est le mot émis. Plus généralement, le mot de C le plus proche est le plus probable (si loi de C uniforme).

Définition : dcodage au maximum de vraisemblance
Prendre le (un) mot de code le plus proche.

5.1 Codes linéaires

Définition

$C \subset \{0,1\}^n$ est linéaire si $x, y \in C \Rightarrow x + y \in C$. C est un espace vectoriel sur $\mathbb{F}_2 = \{0,1\}$. C admet des bases.

$$g_1, \ldots, g_k \in \{0,1\}^N$$

$$\forall c \in C, \exists! I \subset \{1,2,\ldots,k\}, c = \sum_{i \in I} g_i$$

$$k = \dim C, |C| = 2^k$$

Définition

On appelle matrice génératrice de C, la matrice de taille (k, n) :

$$G = \begin{bmatrix} g_1 \\ \vdots \\ g_k \end{bmatrix}$$

où $g_1, \ldots, g_k$ est une base de C.
Encodage : $\{0,1\}^k \to C$

$$(x_1, \ldots, x_k) \to xG = \sum_{i, x_i = 1} g_i$$

Définition

On dit que G (matrice génératrice de C) est sous forme systématique si :

$$G = [I_k A]$$

Définition

Paramtres de C : [n,k,d]=[longueur,dimension,distance minimale]

Proposition

Pour tout C linéaire, il existe une permutation des coordonnées pour laquelle C admet une matrice génératrice systématique.

Autre définition d'un code : $H = [h_1 \ldots h_n]$ est la matrice de parité (controle) de C.

$$C = \left\{ (c_1, \ldots, c_n), c_1 h_1 + \ldots + c_n h_n = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} \right\}$$

Passage de G à H ou de H à G

Définition : produit scalaire dans $\mathbb{F}_2^n$
Soient $x = (x_1, \ldots, x_n)$ et $y = (y_1, \ldots, y_n)$.

$$x.y = x_1y_1 + \ldots + x_ny_n \in \mathbb{F}_2$$

Pour x donné, si $\forall y \in \mathbb{F}_2^n$, $x.y = 0$, alors $x = 0$.

Définition
C code linéaire dans $\mathbb{F}_2^n$
$C^\perp$ code orthogonal (ou dual)
$C^\perp = \{x \in \mathbb{F}_2^n, \forall c \in C, c.x = 0\}$

Définition
On dit que $x \perp y$ si $x.y = 0$.

Définition
On appelle matrice de parité (controle) de C une matrice génératrice de $C^\perp$

Dimensions :

Proposition
$\dim C + \dim C^\perp = n \Rightarrow \dim C^\perp = n - k$

Définition
H définit une fonction syndrome :

$$\sigma : \{0,1\}^n \to \{0,1\}^r$$

$$x \longmapsto H.{}^t x$$

$$(x_1, \ldots, x_n) \longmapsto \sum_{i=1}^{n} x_i h_i$$

avec $H = [h_1, \ldots, h_n]$

Proposition

$$C = \{x, \sigma(x) = 0\}$$

Proposition

$$d_{\min} = \min\{|I|, I \subset \{1, \ldots, n\}, \sum_{i \in I} h_i = 0\}$$

Proposition
Si $G = [I_k A]$ génératrice de C, alors $H = [{}^tA : I_{n-k}]$ est une matrice de parité de C.

Proposition
Si les colonnes h_i d'une matrice H sont non nulles et différentes alors $d \geqslant 3$.

Codes parfaits de $\mathbb{F}_2^n$
Un code parfait est un code tel qu'il n'y a aucun n-uples en dehors de boules de rayon t et $d \geqslant 2t + 1$.

- code à répétition $G = [1 \ldots 1]$, n impair
- Hamming $[n = 2^r - 1, k = 2^r - 1 - r, d = 3]$
- Golay $[n = 23, k = 12, d = 7]$

Distance 4

Remarque
Si $[1 \ldots 1] \in C^\perp$ alors tous les poids de C sont pairs.

Distance 5
H a la propriété supplémentaire :

$$\forall i, j, i', j' \in \{1, \ldots, n\}, h_i + h_j \neq h_{i'} + h_{j'}$$

Codes de grandes distances ?
On a un algorithme glouton qui sans la linéarité marche avec une hypothèse pessimiste : les boules sont disjointes. Chaque boule interdit $|B_{d-1}|$ nouveaux éléments.

Borne de Gilbert-Varshamov

$$|C| \geqslant \frac{2^n}{|B_{d-1}|} = \frac{2^n}{1 + \binom{n}{1} + \binom{n}{2} + \ldots + \binom{n}{d-1}}$$

Existence de bons codes linéaires

On fixe $r = n - k$, $H = [h_1 \dots h_n]$, $h_i \in \{0,1\}^r$. Supposons pour tout $I \subset \{1, \dots, n\}$:

$$1 \leqslant |I| \leqslant d-1$$

$$\sum_{i \in I} h_i \neq 0$$

- $|I| = 1 \Rightarrow n$ colonnes
- $|I| = 2 \Rightarrow \begin{pmatrix} n \\ 2 \end{pmatrix}$ colonnes
- $|I| = 3 \Rightarrow \begin{pmatrix} n \\ 3 \end{pmatrix}$ colonnes
- $\vdots$
- $|I| = d-2 \Rightarrow \begin{pmatrix} n \\ d-2 \end{pmatrix}$ colonnes

On choisit $h_{n+1} \neq \begin{cases} h_i \\ h_i + h_j \\ \vdots \\ h_{i_1} + \ldots + h_{i_{d-2}} \end{cases}$

C'est toujours possible si $1 + n + \begin{pmatrix} n \\ 2 \end{pmatrix} + \ldots + \begin{pmatrix} n \\ d-2 \end{pmatrix} < 2^r$.

Donc le "meilleur" code linéaire (le plus long) vérifie :

$$1 + n + \begin{pmatrix} n \\ 2 \end{pmatrix} + \ldots + \begin{pmatrix} n \\ d-2 \end{pmatrix} \geqslant 2^{n-k}$$

$$\iff |B_{d-2}| \geqslant \frac{2^n}{|C|}$$

$$\iff |C| \geqslant \frac{2^n}{|B_{d-2}|}$$

Remarque

$$\frac{\begin{pmatrix} n \\ d-1 \end{pmatrix}}{|B_{d-1}|} = \epsilon$$

avec ϵ petit.

Remarque

Soient $\frac{k}{n} = R$ (rendement de C) et $\frac{d}{n} = \delta$. Alors :

$$|B_{d-1}| \approx 2^{nh(\delta)}$$

avec $\log_2 |B_{d-1}| = h(\delta)$ D'où :

$$2^{Rn} \geqslant \frac{2^n}{2^{nh(\delta)}}$$

$$Rn \geqslant n - nh(\delta)$$

$$R \geqslant 1 - h(\delta)$$

Borne supérieure de Hamming

Si $t < \frac{d}{2}$ alors $B_r(C) \cap B_r(C') = \emptyset$. D'où :

$$|C||B_r| \leqslant 2^n$$

$$|C| \leqslant \frac{2^n}{|B_r|} \leqslant \frac{2^n}{|B_{\lfloor \frac{d-1}{2} \rfloor}|}$$

$$2^{Rn} \leqslant \frac{2^n}{2^{nh(\frac{\delta}{2})}}$$

$$R \leqslant 1 - h\left(\frac{\delta}{2}\right)$$

avec $\delta \leqslant \frac{1}{2}$.

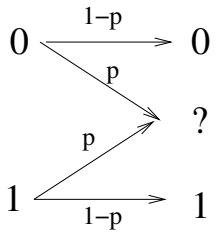


FIG. 5.1 – Canal binaire effacements

Correction d'effacements et théorème de Shannon

Soient $E \subset \{1, \ldots, n\}$ l'ensemble des positions effacées, $c \in C$ le mot de code émis, $x \in \{0, 1, ?\}^n$ reçu.

$$\forall i \notin E, x_i = c_i$$

Dcodage

Trouver $z \in C$ tel que $\forall i \notin E, z_i = x_i$. Si z et z' vérifient $\forall i \notin E, z_i = x_i = z'_i$ alors supp $(z + z') \subset E$, $z + z' \in C$, $\forall i \notin E, z_i + z'_i = 0$

Proposition

$E \subset \{1, \ldots, n\}$ est incorrigible si et seulement si E contient le support d'un mot non nul de C. Notons $\mathcal{E}$ la deuxième partie de cette proposition. $\mathcal{E} = \bigcup_{z \in C, z \neq 0} \mathcal{E}_z$ avec $\mathcal{E}_z = \{\text{supp } z \subset E\}$.

$$P(\mathcal{E}) = P\left(\bigcup_z \mathcal{E}_z\right)$$

$$P(\mathcal{E}) \leqslant \sum_z P(\mathcal{E}_z) = \sum_z p^{|z|} = \sum_{1 \leqslant i \leqslant n} A_i p^i$$

avec $|z|$ = poids de z et $A_i = \#\{z \in C, |z| = i\}$.

On choisit H aléatoire uniforme c'est-à-dire la probabilité d'avoir 1 est $\frac{1}{2}$.

Calcul de $\bar{A}_i = E[A_i]$:

$$A_i = \sum_{\substack{x \in \{0,1\}^n \\ |x| = i}} X_x$$

avec $X_x = \begin{cases} 1 & \text{si } x \in C \\ 0 & \text{sinon} \end{cases}$ et $P(X_x = 1) = P(\sigma(x) = 0) = \frac{1}{2^r}$. D'où :

$$E(A_i) = \frac{1}{2^r} \begin{pmatrix} n \\ i \end{pmatrix}$$

$$i = \lambda n \Longrightarrow \bar{A}_i = \frac{2^{nh(\lambda)}}{2^r} = 2^{n(h(\lambda) - (1-R))}$$

Le théorème de Markov avec une probabilité $\leqslant \frac{1}{n^2} \Longrightarrow A_i \geqslant n^2 \bar{A}_i$.

Proposition

Avec une probabilité $\geqslant 1 - \frac{1}{n}$, on a $A_i \leqslant n^2 \bar{A}_i \, \forall i$

Posons E l'ensemble des positions effacées choisi aléatoire, uniforme parmi les parties de $\{1, \ldots, n\}$ w éléments, $w = \omega n$.

$$P(\mathcal{E} \mid |E| = w) = \frac{\#\{E \supset \text{supp}(z),\ z \neq 0,\ z \in C\}}{\begin{pmatrix} n \\ w \end{pmatrix}}$$

On a :

$$\begin{aligned}
\#\{E \supset \mathrm{supp}(z),\, z \neq 0,\, z \in C\} &\leqslant \sum_{1 \leqslant i \leqslant w} A_i \left(\begin{array}{c} n-i \\ w-i \end{array} \right) \\
&\leqslant n^2 2^{-r} \sum_{1 \leqslant i \leqslant w} \left(\begin{array}{c} n \\ i \end{array} \right) \left(\begin{array}{c} n-i \\ w-i \end{array} \right) \\
&\leqslant n^2 2^{-r} \left(\begin{array}{c} n \\ w \end{array} \right) \sum_{1 \leqslant i \leqslant w} \left(\begin{array}{c} w \\ i \end{array} \right) \\
&\leqslant n^2 2^{-r} \left(\begin{array}{c} n \\ w \end{array} \right) 2^w
\end{aligned}$$

D'où :

$$\begin{aligned}
P(\mathcal{E} \mid |E| = w) &\leqslant n^2 2^{w-r} \\
&\leqslant n^2 2^{n(\omega-(1-R))}
\end{aligned}$$

Or $\omega = p$, d'où :

$$R < 1 - p - \epsilon$$

$$p - (1 - R) < -\epsilon$$

Théorème de Shannon

Soient p et $\epsilon > 0$.
Alors il existe une famille de codes (C_i) telle que $\dim C_i \geqslant (1-p-\epsilon)n_i$ avec :

$$P(\mathcal{E}) \underset{i \to +\infty}{\to} 0$$

Remarque

Si R proche de $1 - p$ alors $P(\mathcal{E}) \lesssim 2^{-nD(1-R||p)}$

5.2 Canaux wire-tap

A envoie à B un mot de code sur un canal sans bruit, mais Oscar écoute le canal. Seulement le canal d'Oscar est lui bruité. Comment peut faire A pour envoyer un secret B sans qu'Oscar ne découvre le secret ?

Comment transmettre un bit ?

$$0 \longmapsto (x_1, \ldots, x_n) \text{ le nombre de 1 est pair}$$

$$1 \longmapsto (x_1, \ldots, x_n) \text{ le nombre de 1 est impair}$$

O obtient, par symbole transmis :

– si c'est un canal effacement : $1-p$
– si c'est un canal binaire symétrique : $1-h(p)$

$$\text{Espoir : } \frac{\#\text{bits de secret}}{\#\text{symboles transmis}} \leqslant \begin{cases} p & \text{(effacements)} \\ h(p) & \text{(erreurs)} \end{cases}$$

A envoie $(x_1, \ldots, x_n) \in \{0,1\}^n$ et il code le secret par : $s = \sum_{i=1}^{n} x_i \times 1$ avec $s \in \{0,1\}$

Codage par syndrome

On peut également choisir $s \in \{0,1\}^r$. Pour cela :

$$s = \sum_{i=1}^{n} x_i h_i$$

avec h_i la i-ème colonne de la matrice de parité H du code de taille $n \times r$. Donc :

$$s = \sigma(x)$$

Donc pour transmettre le secret, on choisit$(x_1, \ldots, x_n)$ aléatoire uniforme parmi tous les x tels que $\sigma(x) = s$.

Cas des effacements

Soit

$$H = \begin{bmatrix} & \\ & H_E \\ & \end{bmatrix}$$

avec H_E matrice carre de taille pn car $r \approx pn$. Ainsi :

$$s = \sum_{i \in E} x_i h_i + \sum_{i \notin E} x_i h_i$$

On cherche $\sum_{i \in E} x_i h_i$.

Oscar reçoit $x + e$ avec $e \in \{0,1\}^n$ et soit w le poids de e, $w(e) \approx pn$.

$$\sigma(x+e) = s + \sigma(e)$$

avec $\sigma(e)$ aléatoire uniforme.

Ensemble des vecteurs e

Il est de cardinal : $\begin{pmatrix} n \\ pn \end{pmatrix} \approx 2^{nh(p)}$.

Nouveau modle de canal wire tap

Maintenant on suppose que le canal de transmission entre A et B est également bruité et on note p_B la probabilité d'erreurs sur ce canal. O écoute toujours et son canal est bruité avec une probabilité p_O.
Il faut placer le secret dans $h(p_O) - h(p_B)$. On a la matrice H suivante de taille $nh(p_O) \times n$:

$$\begin{array}{cc} H_1 & \left.\begin{array}{|c|}\hline \\ \hline\end{array}\right\} \; r_1 \\ H_2 & \left.\begin{array}{|c|}\hline \\ \hline\end{array}\right\} \; r_2 \end{array}$$

A choisit x aléatoire parmi ceux tels que $\sigma(x) = H.{}^t x = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \vdots \\ s \\ \vdots \end{bmatrix} + \sigma(e_O)$. B trouve x puis trouve $\sigma(x)$.

Cas adversaire (Wire tap de type 2)

L'observateur O intercepte s positions exactement choisies par lui. On a la matrice H de taille $r \times n$:

$$H = \left[\begin{array}{c|c} H_N & H_I \end{array}\right]$$

avec I positions interceptes et N positions non interceptes.
O reçoit le vecteur $x = [x_N | x_I]$ et il veut rang $H_N < r$.
Donc il existe $J \subset \{1, 2 \ldots, r\}$ et $\sum_{i \in J} l_i = [0, \ldots, 0 \underset{N}{} \; \underset{I}{*}]$ avec :

$$H = \begin{bmatrix} l_1 \\ l_2 \\ \vdots \\ l_n \end{bmatrix}$$

O veut le support d'un mot du dual.

Proposition

Soit $d^\perp$ distance minimale de $C^\perp$ et si $s < d^\perp$ alors O a 0 bits d'information sur s.

Définition : tableau orthogonal de force T

On dit que $T = (T_{ij})$ a force t si $\forall t$ colonnes,

$$T_j = (T_{ij})_{j \in J},\ |J| = t$$

Chaque $v \in \{0,1\}^t$ apparait le même nombre de fois comme ligne de T_j.

Ex : $\begin{matrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \\ 0 & 0 \end{matrix}$ est de force 2.

Théorème

Si (T_{ij}) a pour lignes l'ensemble d'un code linéaire de distance $d^\perp$ alors T a force $d^\perp - 1$. Et si on a G matrice génératrice de C et que le rang de $G_J = t$ alors on a la propriété.

Diffusion d'aléa

Soit

$$G = \begin{pmatrix} l_1 \\ l_2 \\ \vdots \\ l_k \end{pmatrix}$$

et $a \in \{0,1\}^k$ k bits d'aléa "pur". On a :

$$x = \sum_{i=1}^{k} a_i l_i$$

Pour $t < d^\perp$, si on a $(x_i)_{i \in J}$ tels que $|J| = t$ alors les x_i sont indépendants.