Théorie de l'information

Cours 3 - Systèmes de communication

Laurent Oudre
laurent.oudre@univ-paris13.fr

Université Paris 13, Institut Galilée
Master Ingénierie et Innovations en Images et Réseaux - 1ère année
2017-2018

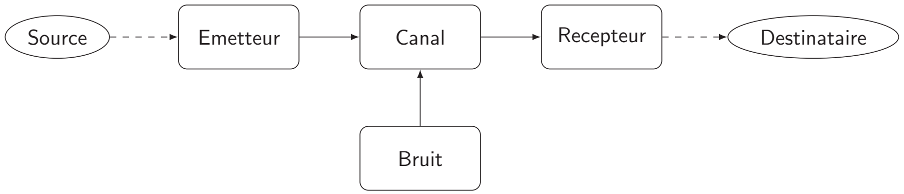
Sommaire

Sommaire

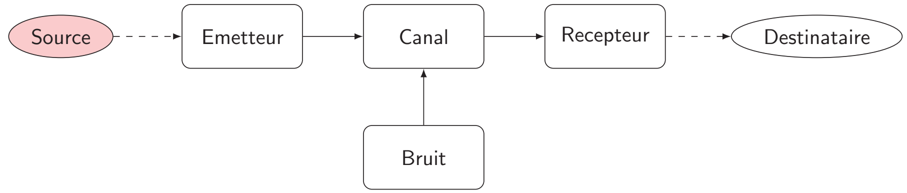
Systèmes de communication

- Une des contributions de Shannon dans son article de 1948 fut un schéma qui modélisait de façon formelle la communication entre deux machines.
- Ce schéma est désormais très largement utilisé en télécommunications, mais également parfois pour modéliser la communication humaine.
- Il met en évidence 3 acteurs indispensables à la transmission de l'information : la source (celui qui envoie l'information), le canal (qui transmet l'information) et le destinataire (qui reçoit l'information).

Schéma général



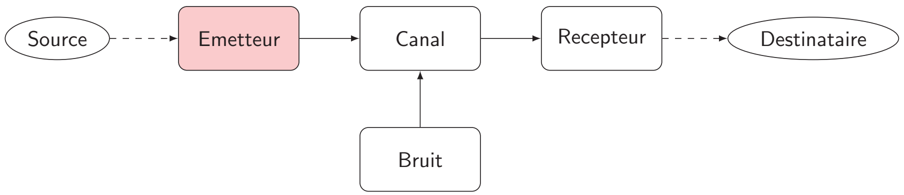
Source



La **source** envoie un message, souvent constitué d'une série de symboles pris dans un alphabet donné. On parle dans ce cas de message numérique.

- Données discrètes : texte, numéros...
- Données analogiques numérisées : image, voix, vidéo...

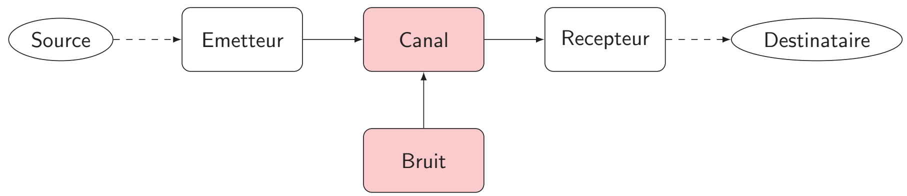
Emetteur



L'**émetteur** prend ce message numérique et réalise les étapes suivantes :

- **Codage source :** compression des données pour qu'elles prennent le moins de place possible. Cela revient à remplacer le message à envoyer par un message le plus court possible, souvent représenté sous forme d'une série de 0 et de 1.
- **Codage canal :** rajout de bits d'information supplémentaires dans le message pour permettre de corriger les éventuelles erreurs de transmission
- Transformer le message numérique en un signal physique (onde électromagnétique, signal électrique, etc...) qui puisse être transmis sur le canal de transmission

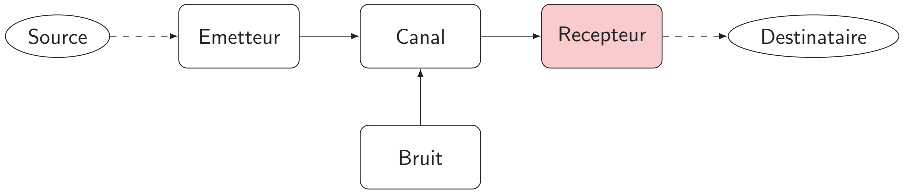
Canal



Le **canal** achemine le signal physique d'un point à l'autre

- Le canal peut être de différents types : câbles coaxiaux, paires torsadées, réseau hertzien, infrarouge, fibres optiques,....
- Généralement perturbé par un bruit qui dépendra de l'environnement et de la nature du canal : perturbations électriques, réflexion d'ondes, détérioration du câble, etc...
- Ce bruit a pour conséquence une dégradation du signal voire la perte de parties du signal

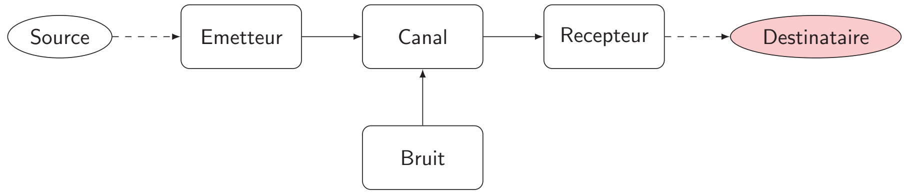
Récepteur



Le **récepteur** prend le signal physique et réalise les étapes suivantes :

- Transformer le signal physique en un message numérique
- Inverser les étapes de codage canal et de canal source pour reconstituer le message envoyé par la source

Destinataire



Le destinataire reçoit le message

Sommaire

Source discrète

Pour modéliser un système de communication on supposera dans le cadre de ce cours, que

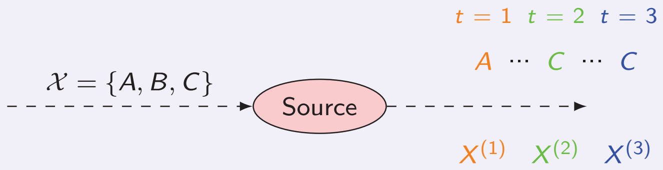
- La source est discrète et prend ses valeurs dans un ensemble $\mathcal{X} = \{x_1, \ldots, x_M\}$, souvent appelé **alphabet** ou **alphabet de source**.
- On la modélise souvent comme une variable aléatoire discrète X à valeurs dans $\mathcal{X}$.
- Un **message** de longueur n est une suite de n **symboles**

  $$x^{(1)}x^{(2)}x^{(3)}x^{(4)}\ldots x^{(n)}$$

  issus de $\mathcal{X}$
- L'émission d'un message de longueur n fait donc intervenir n variables aléatoires qu'on pourra noter $X^{(1)}, X^{(2)}, X^{(3)}, X^{(4)}, \ldots X^{(n)}$.
- $X^{(t)}$ est la variable aléatoire correspondant au symbole émis par la source à l'instant t

Source discrète

Illustration



- $\mathcal{X}$: ensemble des symboles possibles (alphabet)
- $x^{(1)}x^{(2)}x^{(3)} = ACC$: message envoyé
- $X^{(1)}, X^{(2)}$ et $X^{(3)}$: variables aléatoires associées respectivement au premier, deuxième et troisième symbole envoyé.

Source discrète sans mémoire

- On dit que la source est **sans mémoire** si et seulement si toutes les variables aléatoires $X^{(1)}, X^{(2)}, X^{(3)}, X^{(4)} \dots X^{(n)}$ sont indépendantes.
- On a donc :

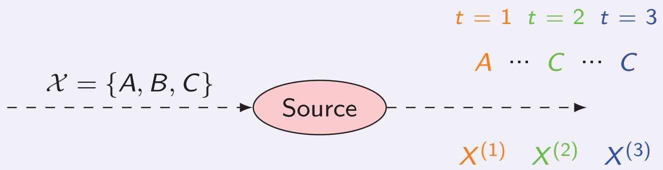
$$P_X\left(X^{(1)}X^{(2)}X^{(3)}X^{(4)}\dots X^{(n)} = x^{(1)}x^{(2)}x^{(3)}x^{(4)}\dots x^{(n)}\right) = \prod_{t=1}^{n} P_X\left(X^{(t)} = x^{(t)}\right)$$

- Si on suppose en plus que toutes les variables aléatoires $X^{(1)}, X^{(2)}, X^{(3)}, X^{(4)} \dots X^{(n)}$ sont **identiquement distribuées**, l'étude de l'émission d'un message est ramenée à celle d'une seule variable aléatoire X correspondant à l'émission d'un seul symbole.

$$P_X\left(X^{(1)}X^{(2)}X^{(3)}X^{(4)}\dots X^{(n)} = x^{(1)}x^{(2)}x^{(3)}x^{(4)}\dots x^{(n)}\right) = \prod_{t=1}^{n} P_X\left(X = x^{(t)}\right)$$

Source discrète sans mémoire

Illustration



- **Sans mémoire** : les variables aléatoires $X^{(1)}, X^{(2)}$ et $X^{(3)}$ sont toutes indépendantes
- **Identiquement distribuées** : $X^{(1)} \sim X$, $X^{(2)} \sim X$ et $X^{(3)} \sim X$

  Par exemple, $P_{X^{(1)}}(X^{(1)} = A) = P_{X^{(2)}}(X^{(2)} = A) = P_{X^{(3)}}(X^{(3)} = A) = P_X(X = A)$

Source discrète sans mémoire

- Tous les symboles émis sont indépendants les uns des autres
- Les symboles émis à chaque instant ne dépendent pas de ceux qui ont été envoyés précédemment, d'où le terme *sans mémoire*. C'est comme si à chaque instant, la source oubliait le symbole qu'elle avait envoyé avant.
- Dans le vocabulaire courant, lorsque l'on parle d'une source sans mémoire, on suppose que les symboles envoyés successivement sont indépendants, mais également identifiquement distribués.

Source discrète sans mémoire

Exemple : Source binaire

- Dans le cas d'une source binaire, on aura $X \in \{0, 1\}$.
- Si on suppose que la source est sans mémoire, c'est que par exemple dans le message 001100, tous les bits sont envoyés de façon indépendante, et on a donc :

$$p_X(001100) = p_X(0)^4 p_X(1)^2$$

Exemple : Langue française

- Dans le cas d'un mot en langue française, on a $X \in \{A, B, C, D, \ldots, X, Y, Z\}$.
- Bien souvent dans ce cas, on ne peut pas supposer que la source est sans mémoire. Par exemple si le message émis pour le moment est AX, on sait que la prochaine lettre envoyée ne peut pas être un Z. Les symboles envoyés ne sont donc pas indépendants.

Canal discret

Pour modéliser un système de communication on supposera dans le cadre de ce cours, que

- Le canal est discret et transforme le message envoyé par la source en un autre message, qui sera reçu par le destinataire
- Les symboles reçus par le destinataire seront également modélisés par une variable aléatoire discrète Y à valeurs $\mathcal{Y} = \{y_1, \ldots, y_L\}$.
- Comme le récepteur ne fait que recevoir les symboles émis par la source, on pourrait penser que $\mathcal{X} = \mathcal{Y}$, mais à cause des perturbations sur le canal, on peut créer des symboles qui n'existaient pas à l'entrée.
- Un message émis par la source

$$x^{(1)}x^{(2)}x^{(3)}x^{(4)}\ldots x^{(n)} \in \mathcal{X}^n$$

sera transformé en message

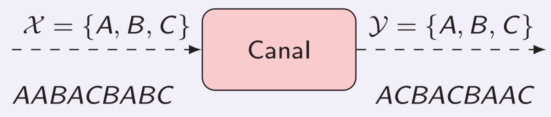
$$y^{(1)}y^{(2)}y^{(3)}y^{(4)}\ldots y^{(n)} \in \mathcal{Y}^n$$

Canal discret sans mémoire

- Un canal est dit **sans mémoire** si la valeur du symbole $y^{(t)}$ ne dépend que de la valeur du symbole $x^{(t)}$ émis au même instant.
- Si l'on suppose en plus que la source est également sans mémoire, il suffit d'étudier l'émission, la transmission et la réception d'un symbole unique.
- Dans ce cas là, la loi $p_{Y|X}(y|x)$ caractérise complètement les perturbations du canal : elle donne la loi régissant la transformation des symboles de $\mathcal{X}$ en symboles de $\mathcal{Y}$
- On peut donc représenter le canal par un schéma fléché ou une matrice $\{P_{Y|X=x}(Y = y|X = x)\}_{x\in\mathcal{X}, y\in\mathcal{Y}}$ qu'on appelle **matrice de transfert**.

Canal discret sans mémoire

Illustration



- Matrice de transfert :

$Y\|X$	$Y = A$	$Y = B$	$Y = C$
$X = A$	$\frac{3}{4}$	0	$\frac{1}{4}$
$X = B$	$\frac{1}{3}$	$\frac{2}{3}$	0
$X = C$	0	0	1

- Plus les perturbations sont importantes, plus les symboles de la sortie ont tendance à être différents de ceux de l'entrée.

Canal discret sans mémoire

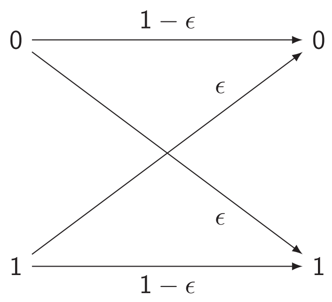
Canal binaire sans bruit

Supposons qu'on envoie un bit (0 ou 1) à travers un canal sans bruit et sans mémoire. On suppose que la source est également sans mémoire. Dans ce cas, on a :

- Pour modéliser le canal, il suffit de considérer l'émission, la transmission et la réception d'un seul bit.
- $X \in \{0, 1\}$ et $Y \in \{0, 1\}$
- Comme il n'y a aucune perturbation, on sait que si un 0 a été émis, c'est un 0 qui a été reçu (et idem pour le 1). La loi conditionnelle $p_{Y|X}(y|x)$ modélisant l'action du canal peut donc s'écrire :

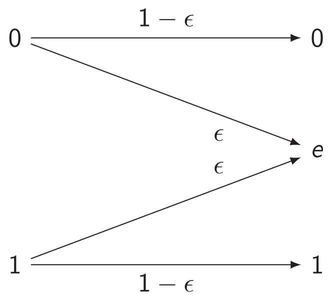
$Y \mid X$	$Y = 0$	$Y = 1$
$X = 0$	1	0
$X = 1$	0	1

Canal binaire symétrique



- Canal binaire : $\mathcal{X} = \{0, 1\}$ et $\mathcal{Y} = \{0, 1\}$
- Symétrique : même loi conditionnelle pour les deux symboles d'entrée
- $\epsilon \in [0, 1]$: probabilité d'erreur

Canal avec perte



- Entrée dans : $\mathcal{X} = \{0, 1\}$ mais trois symboles de sortie $\mathcal{Y} = \{0, e, 1\}$
- Soit le bit est bien transmis, soit il est perdu (il devient alors le symbole e)
- $\epsilon \in [0, 1]$: probabilité de perte de données.

Sommaire

Cadre du cours

Dans le cadre de ce cours, on considérera uniquement :

- Des sources discrètes sans mémoire où les symboles envoyés successivement sont indépendants et identiquement distribués
- Des canaux discrets sans mémoire
- Un système de communication sera donc entièrement déterminé par :
  - Les alphabets $\mathcal{X}$ et $\mathcal{Y}$ des symboles d'entrée et de sortie
  - La loi de probabilité $p_X(x)$ de la source
  - La loi de probabilité conditionnelle $p_{Y|X}(y|x)$: matrice de transfert du canal