Algorithmes de Transmission et de Recherche de l'Information dans les Réseaux de Communication

Philippe Robert
INRIA Paris-Rocquencourt

Le 2 juin 2010

Présentation

– **Directeur de recherche à l'INRIA**
Institut de recherche en Informatique
et en Mathématiques Appliquées.
Responsable de l'équipe de recherche
"Réseaux, Algorithmes et Probabilités"

Mathématicien
Spécialité : Probabilités.

– **Professeur Chargé de Cours**
à l'École Polytechnique.

Problématique générale

La raison d'être d'un réseau :

– **Diffuser**

– **Rechercher**

l'information.

Quelques problèmes classiques

Utilisation des ressources

– Contrôle d'accès

– Conflits d'accès

– Partage

– Recherche de l'information

Plan de la conférence

1. Une brève histoire.
2. Protocoles d'accès.

   Gérer les conflits d'accès à une ressource.
3. Transmission de données dans les réseaux.

   Circulation des données dans Internet.
4. Recherche d'information.

   Google
5. Conclusion.

1. Histoire

Un bref aperçu historique

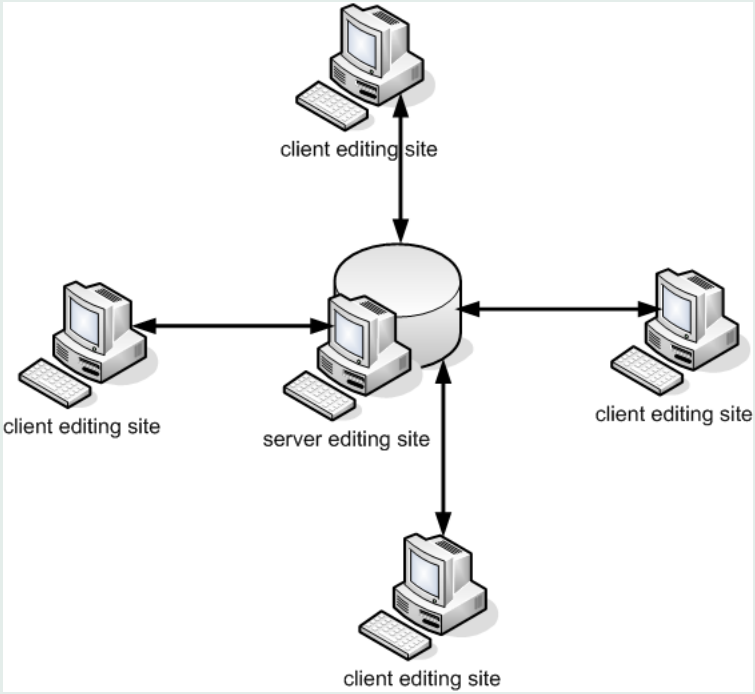
– $\simeq$ 1900 : **Réseaux téléphoniques.**

Un bref aperçu historique

- $\simeq$ 1900 : Réseaux téléphoniques.
- 1960 : Réseaux informatiques.
  - Serveur central.

IBM System/360

Le Modèle du Serveur Central



Un bref aperçu

- 1909 : **Réseaux téléphoniques.**
- 1960 : **Réseaux informatiques.**
  - **Serveur central.**

Un bref aperçu

- 1909 : Réseaux téléphoniques.
- 1960 : Réseaux informatiques.
  - Serveur central.
- 1980 : Systèmes distribués.
  - Réseaux Locaux.
  - Internet.
  - Réseaux Mobiles.

Système Distribué

– **Ensemble de Machines communicantes.**

– **Pas de Contrôle Central.**

– **Chacune agit de façon autonome.**

Système Distribué

– **Ensemble de Machines communicantes.**

– **Pas de Contrôle Central.**

– **Chacune agit de façon autonome.**

Exemples

– **Internet**
– **Réseaux Pair à Pair**
– **...**

Problématique

Fonctionnement d'un Système Distribué

Pb : Trouver une procédure, un Algorithme tel que

– Chaque machine utilise la même procédure.

– Globalement le réseau effectue la fonction demandée.

Problématique

Fonctionnement d'un Système Distribué

Pb : Trouver une procédure, un Algorithme tel que

– Chaque machine utilise la même procédure.

– Globalement le réseau effectue la fonction demandée.

Exemples

– Internet : Transmission des données.

– Réseaux Pair à Pair
Recherche et Stockage des données
Téléphone (Skype). . . .

Les Systèmes Distribués dans la Nature

– **Bancs de Poisson.**

– **Fourmilière, Essaims.**

Les Systèmes Distribués dans la Nature

- **Bancs de Poisson.**
- **Fourmilière, Essaims.**
- **Cerveau.**
- **...**

Innovations : Évolutions

Progrès technologiques :

– **Vitesse des processeurs, des composants.**

– **Nouveaux matériaux, . . .**

Innovations : Évolutions

Progrès technologiques :

– **Vitesse des processeurs, des composants.**

– **Nouveaux matériaux, . . .**

Important mais n'est plus la source dominante d'innovation.

Innovations : Évolutions

Progrès technologiques :

– **Vitesse des processeurs, des composants.**

– **Nouveaux matériaux, . . .**

Important mais n'est plus la source dominante d'innovation.

Conception d'Algorithmes

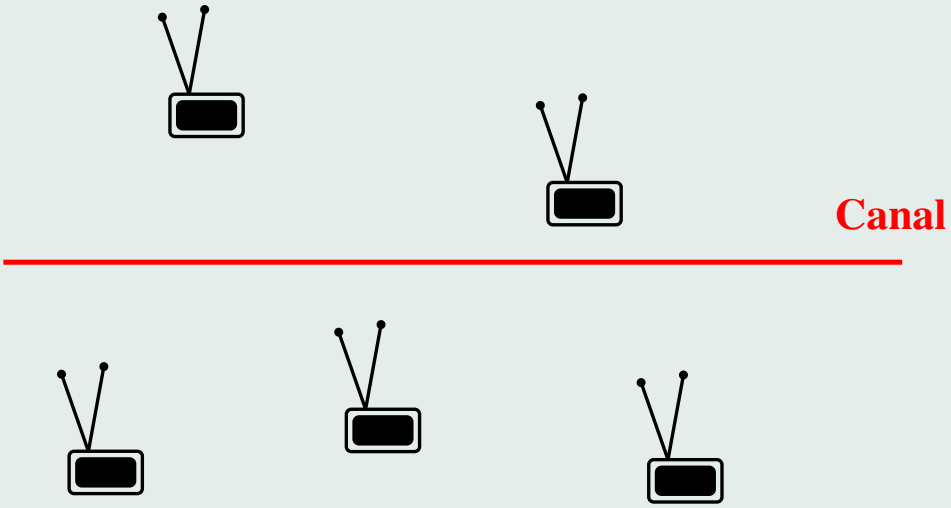
– **Langages/Programmes Informatiques.**

– **Modélisation mathématique.**

– **. . .**

Importance croissante.

2. Protocoles d'accès


Canal

Le cadre

- N émetteurs/stations dispersés dans la nature.
- Un seul canal de communication.
- Une station ayant un message doit le transmettre sur le canal.
- Deux émissions sur le canal en même temps $\Rightarrow$ échec.
- Transmission d'un message : une unité de temps.

Exemples

- **Réseaux locaux ;**
- **Réseau câblé ;**
- **Réseaux sans fil (Wifi, Bluetooth...) ;**

Problème

Contexte : Chaque unité de temps il arrive en moyenne λ nouveaux messages.

– Comment assurer la transmission des messages de chaque station ?

Algorithme de transmission

- Basé uniquement sur l'écoute du canal.
- Début de chaque unité de temps, chaque station avec un message : Tentative de transmission ou non.
- Toutes les stations utilisent la même politique.

Algorithme de transmission

Une station présente depuis n unités de temps :
Si $O_1, \ldots, O_n$, l'état du canal vu par celle-ci

$$1 \leq i \leq n, \qquad O_i \in \{0, 1, 2\}.$$

Algorithme de transmission

Une station présente depuis n unités de temps :
Si $O_1, \ldots, O_n$, l'état du canal vu par celle-ci

$$1 \leq i \leq n, \qquad O_i \in \{0, 1, 2\}.$$

Décision à $t = n + 1$: f_n tel que
$f_n(O_1, O_2, \ldots, O_n) \in \{0, 1\}$

Information d'une station

– Chaque station écoute le canal :
Information ternaire

0 — un blanc
pas d'essai de transmission sur le canal.

1 — un succès
un seul émetteur transmet sur le canal.

2 — une collision
au moins deux émetteurs essaient une transmission.

Problématique

Trouver un algorithme $\mathcal{P}$ tel que :

Existence de $\lambda_c(\mathcal{P}) > 0$ tel que si $\lambda < \lambda_c(\mathcal{P})$
λ taux d'arrivée des nouveaux messages

1. **Tous les messages sont transmis.**
2. **Stabilité :**
   si $L(t)$ nb de messages non transmis à t

$$\sup_{t \geq 0} \mathrm{E}(L(t)) < +\infty$$

Problématique

Trouver un algorithme $\mathcal{P}$ tel que :

Existence de $\lambda_c(\mathcal{P}) > 0$ tel que si $\lambda < \lambda_c(\mathcal{P})$
λ taux d'arrivée des nouveaux messages

1. **Tous les messages sont transmis.**
2. **Stabilité :**
   si $L(t)$ nb de messages non transmis à t

$$\sup_{t \geq 0} \mathrm{E}(L(t)) < +\infty$$

Nécessairement $\lambda_c(\mathcal{P}) \leq 1$.

Aloha

Aloha

Contexte historique (1967) :

– **Terminaux sur des îles reliés au serveur central de l'Université d'Hawaï.**

– **Connexion radio sur une seule fréquence.**

Pb. : Rapatriement des données sur le serveur central.

Aloha

Abramson (1967).

Au début de chaque unité de temps :

Chaque émetteur lance une pièce de monnaie de biais $p =$ proba Pile :

– Pile : tentative de transmission ;

– Face : pas de tentative de transmission.

Caractéristiques

Algorithme probabiliste :

L'aléatoire réduit les collisions répétées.

Caractéristiques

Algorithme probabiliste :

L'aléatoire réduit les collisions répétées.

N **messages à transmettre :**
probabilité d'un succès : $Np(1-p)^{N-1}$.

Caractéristiques

Algorithme probabiliste :

L'aléatoire réduit les collisions répétées.

N **messages à transmettre :**
probabilité d'un succès : $Np(1-p)^{N-1}$.

Algorithme instable : $\lambda_c(\mathcal{A}) = 0$.

Ethernet

L'algorithme

Metcalf (Harvard) 1973

Chaque émetteur a une variable "compteur" C.

– À l'arrivée sur le canal : $C = 0$.

– À chaque échec de transmission $C \to C + 1$.

L'algorithme

Metcalf (Harvard) 1973

Chaque émetteur a une variable "compteur" C.

– À l'arrivée sur le canal : $C = 0$.

– À chaque échec de transmission $C \to C + 1$.

Si compteur égal à k :
– Tentative de transmission avec proba $1/2^k$

– Pas de tentative de transmission sinon.

Caractéristiques

Un émetteur ayant subi $C = k$ échecs
essaie de transmettre avec proba $1/2^k$:

Caractéristiques

Un émetteur ayant subi $C = k$ échecs
essaie de transmettre avec proba $1/2^k$:

- **Prise en compte de l'histoire du canal (un peu) avec la variable C.**
- **Privilégie les nouveaux messages.**
- **Diminue la proba de collisions répétées.**

Instabilité d'Ethernet

Théorème (Aldous, 1987)

Si $\lambda > 0$, alors l'algorithme est instable

$$L(t) \to +\infty \text{ quand } t \to +\infty$$

Instabilité d'Ethernet

Théorème (Aldous, 1987)

Si $\lambda > 0$, alors l'algorithme est instable

$$L(t) \to +\infty \text{ quand } t \to +\infty$$

Instabilité "subtile".

Conclusions sur Ethernet

– **Protocole ouvert (Xerox PARC).**

Conclusions sur Ethernet

– **Protocole ouvert (Xerox PARC).**

– **Succès industriel. Norme IEEE 802.3.**

Conclusions sur Ethernet

– **Protocole ouvert (Xerox PARC).**

– **Succès industriel. Norme IEEE 802.3.**

– **Algorithme instable.**

Conclusions sur Ethernet

– **Protocole ouvert (Xerox PARC).**

– **Succès industriel. Norme IEEE 802.3.**

– **Algorithme instable.**

– **Instabilité théorique.**

Conclusions sur Ethernet

– **Protocole ouvert (Xerox PARC).**

– **Succès industriel. Norme IEEE 802.3.**

– **Algorithme instable.**

– **Instabilité théorique.**

– **Instabilité en pratique.**

L'algorithme en arbre

Capetanakis (MIT), 1979

Tsybakov et Mikhailov (Moscou), 1979

L'algorithme (I)

Chaque émetteur a une variable "compteur" C.

– Si $C = 0 \Rightarrow$ essai de transmission.

1. **Si succès, c'est terminé.**
2. **Si collision, tirage pile ou face :**
   si pile, $C = 0$, sinon, $C = 1$.

L'algorithme (II)

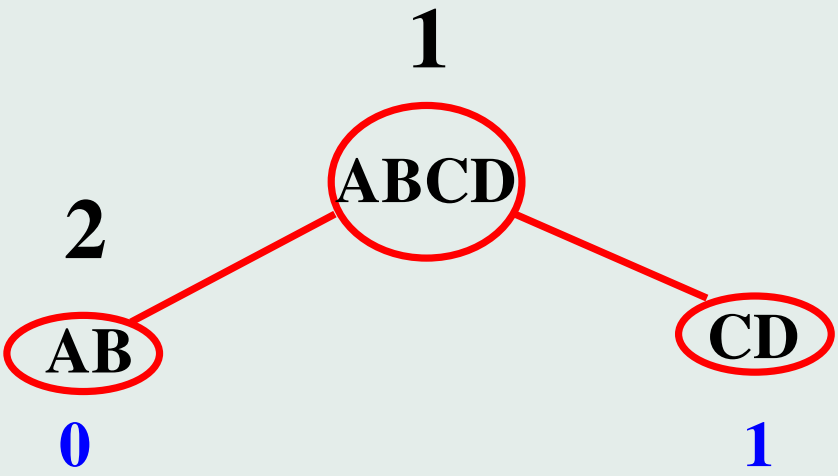
– Si $C > 0 \Rightarrow$ pas d'essai de transmission.

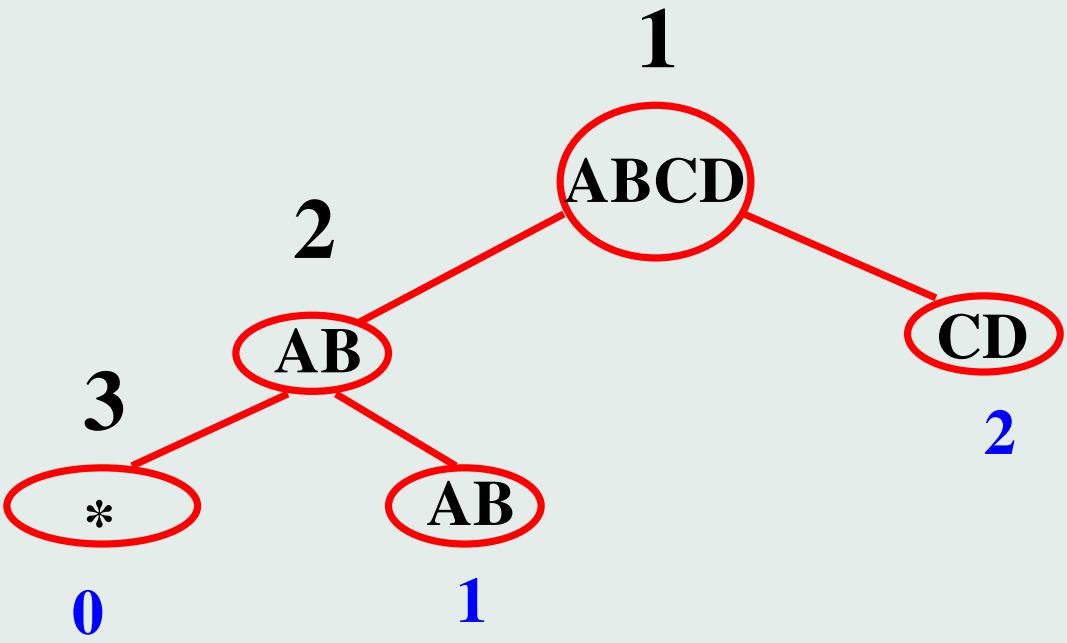
Écoute du canal :

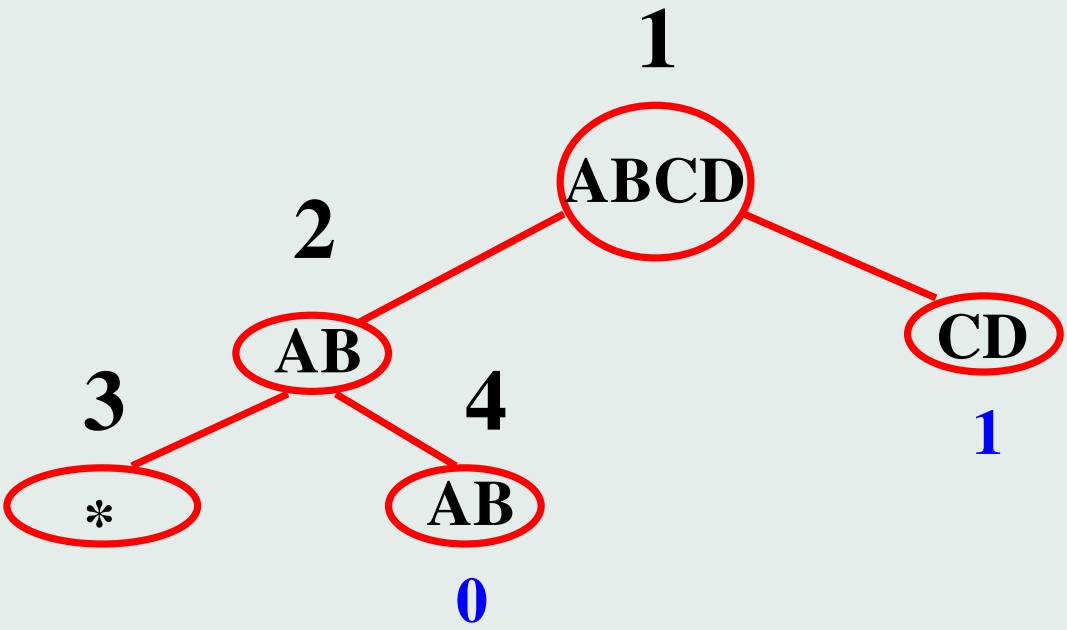
1. Si succès ou silence,
   $C \rightarrow C - 1$.
2. Si collision sur le canal,
   $C \rightarrow C + 1$.

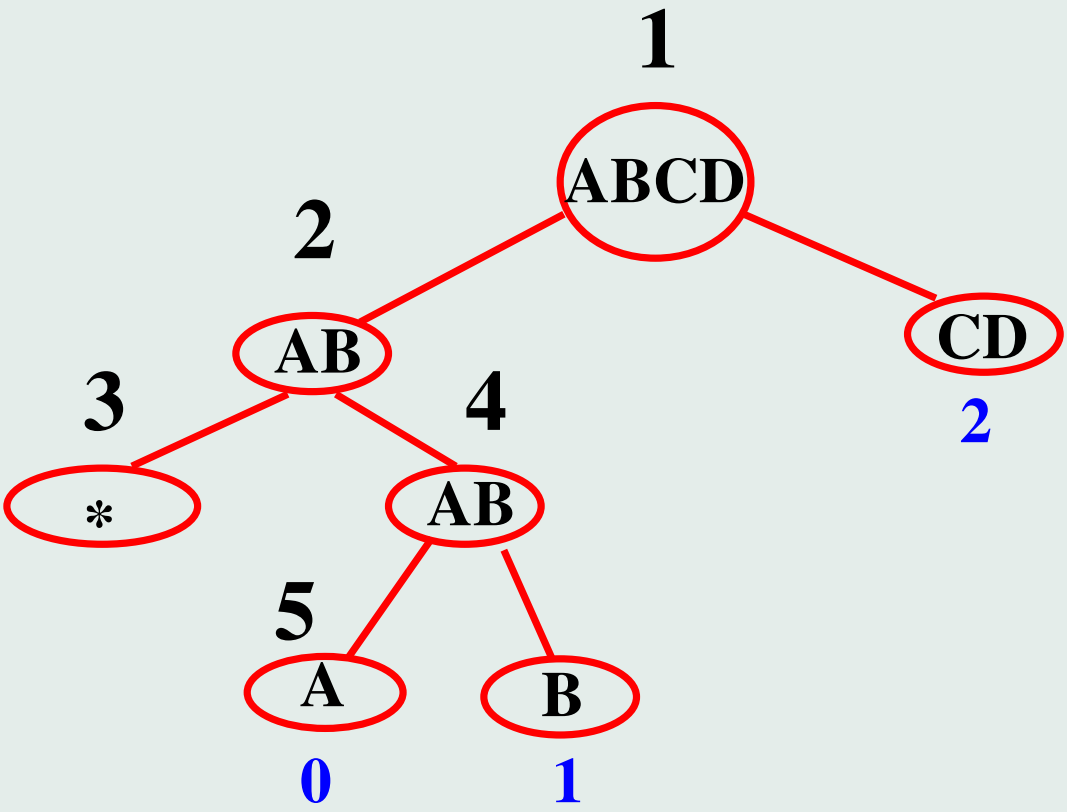
1

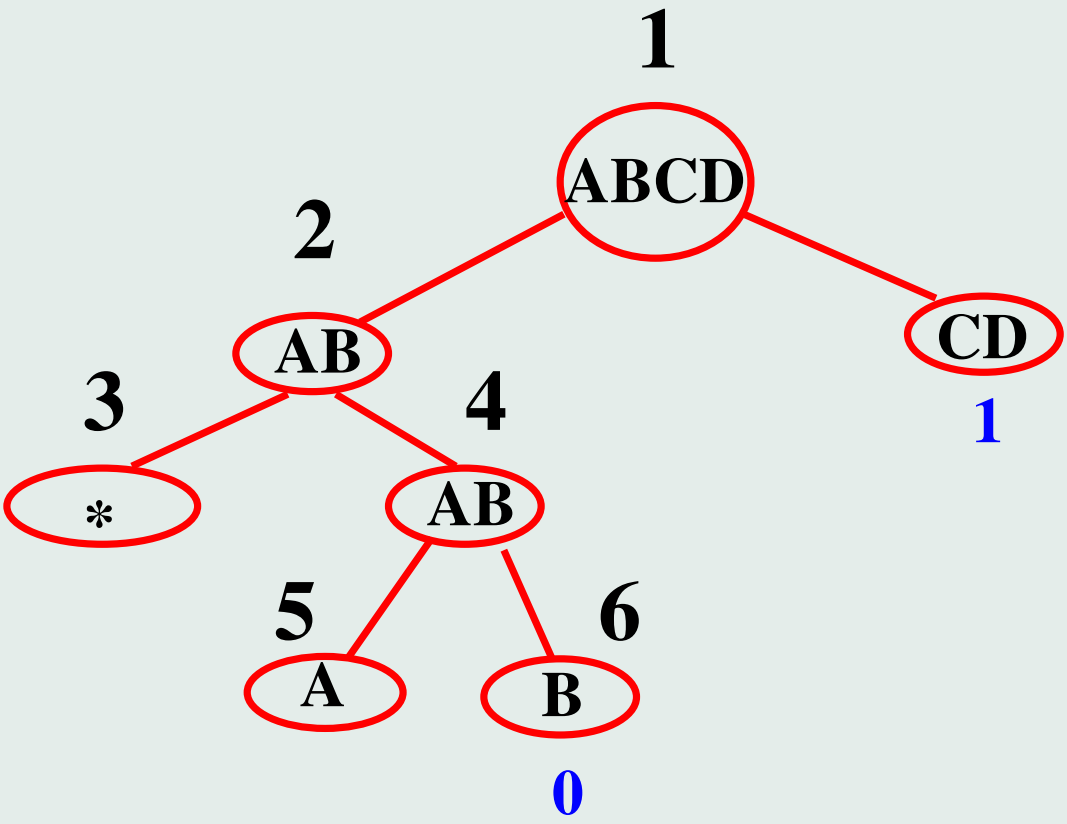
0

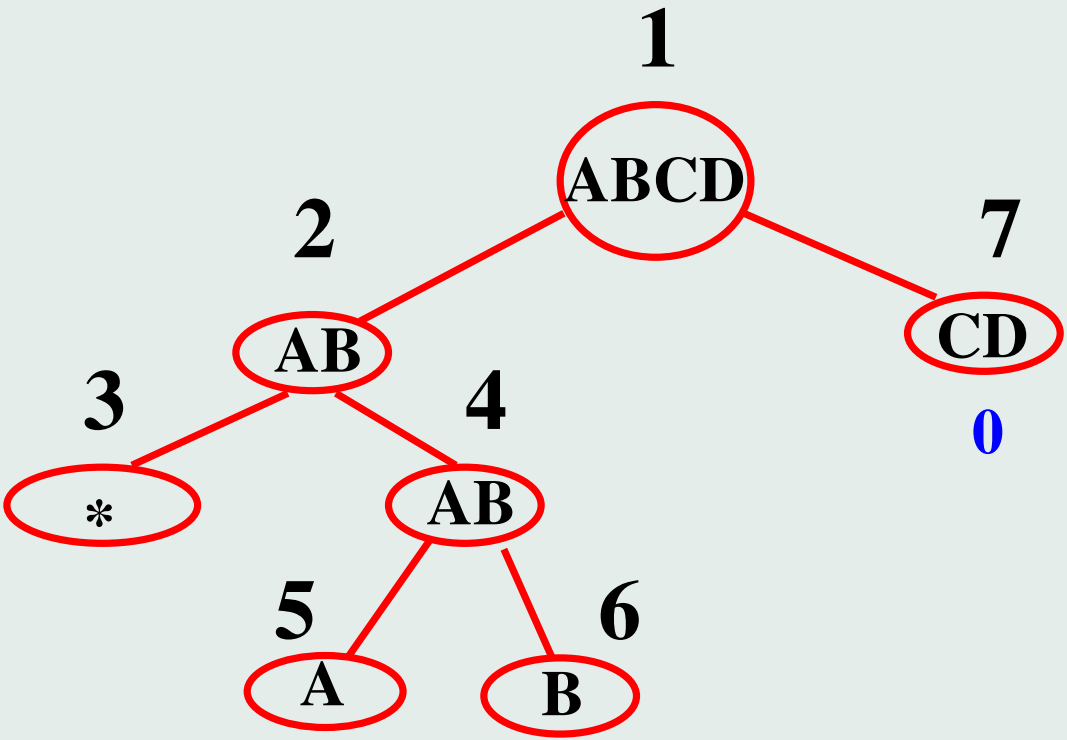

1
ABCD
2
AB
CD
0
1

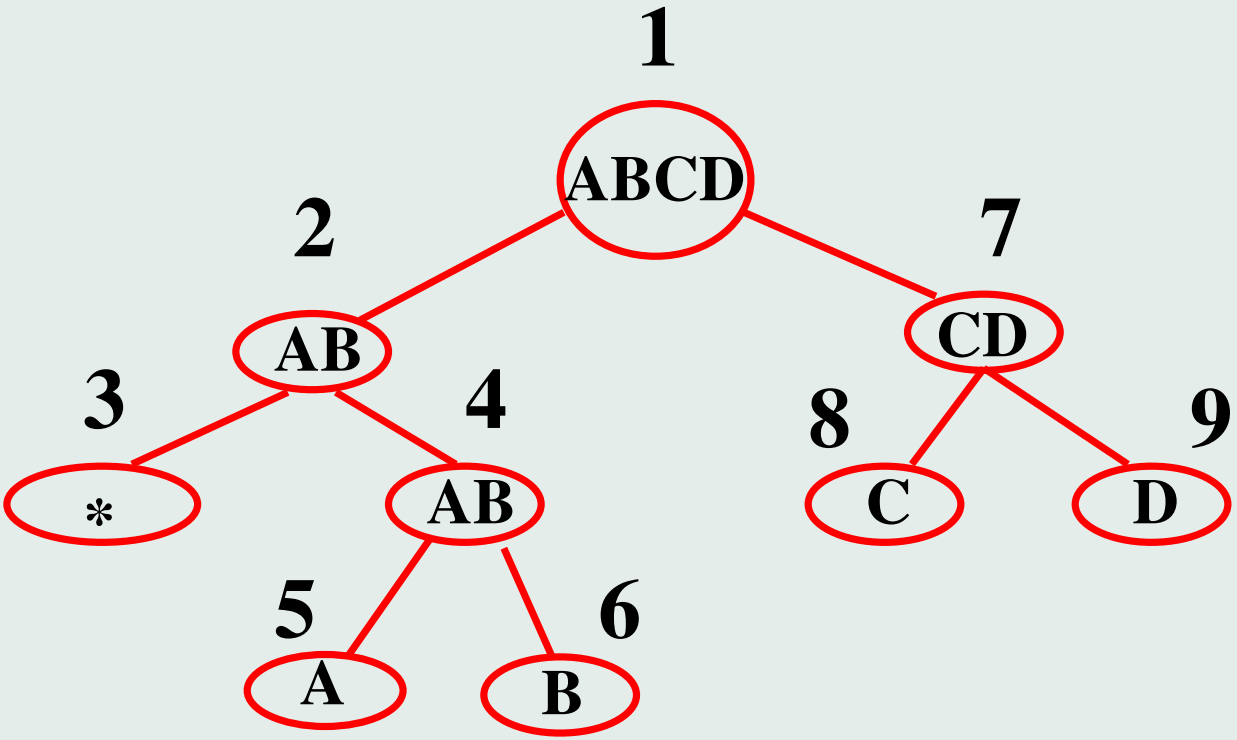

1
ABCD
2
AB
CD
2
3
*
AB
0
1


1
ABCD
2
AB
CD
1
3
4
*
AB
0


1
ABCD
2
AB
CD
2
3
4
*
AB
5
A
B
0
1


1
ABCD
2
AB
CD
1
3
4
*
AB
5
6
A
B
0


1
ABCD
2
AB
7
CD
0
3
*
4
AB
5
A
6
B


1
ABCD
2
AB
7
CD
3
*
4
AB
8
C
9
D
5
A
6
B

Caractéristiques

- **Écoute continue du canal.**

- **Prise en compte de l'information ternaire même en cas de non-transmission.**

Analyse mathématique du protocole en arbre

R_n : temps de transmission de n messages.

$\mathrm{E}(R_n)/n$
Temps moyen de transmission d'un message.

Analyse mathématique du protocole en arbre

R_n : temps de transmission de n messages.

$\mathrm{E}(R_n)/n$
Temps moyen de transmission d'un message.

Débit

$$\lambda_c = \lim_{n \to +\infty} \frac{n}{\mathrm{E}(R_n)}$$

Parenthèse Mathématique

En fait la limite

$$\lambda_c = \lim_{n \to +\infty} \frac{n}{\mathrm{E}(R_n)}$$

n'existe pas !

Parenthèse Mathématique

En fait la limite

$$\lambda_c = \lim_{n \to +\infty} \frac{n}{\mathrm{E}(R_n)}$$

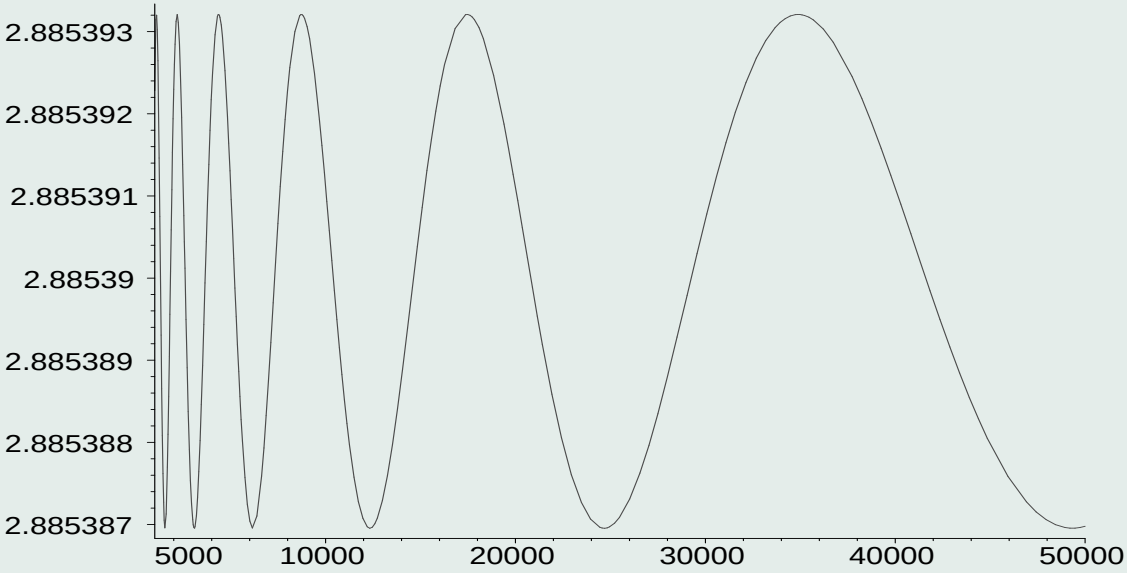
n'existe pas !

Mais

$$\liminf_{n \to +\infty} \frac{n}{\mathrm{E}(R_n)} \sim 0.34657 > 0.$$

Aloha et Ethernet : $\lambda_c = 0$.

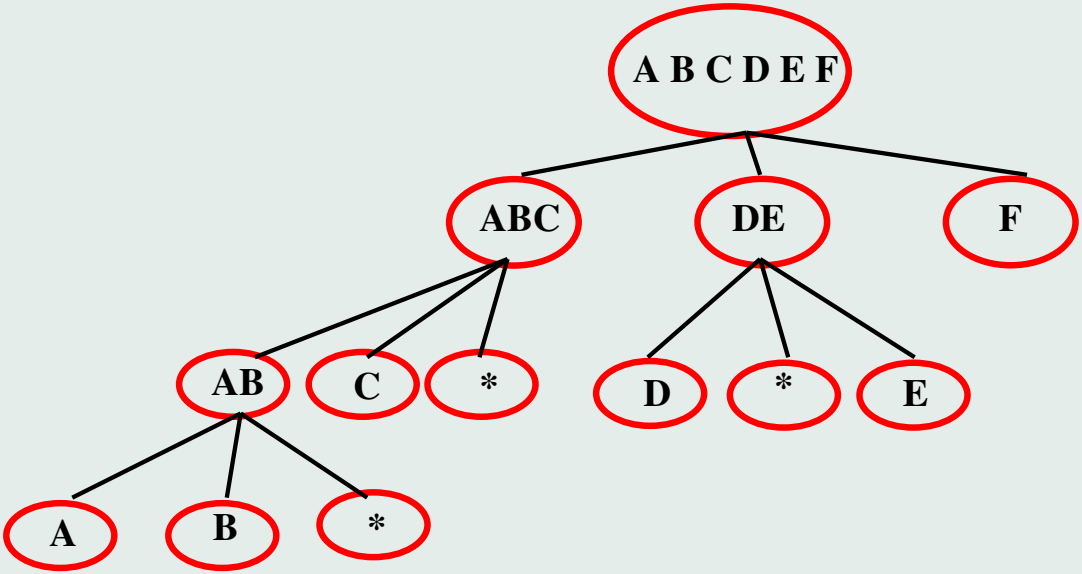
Évolution de $n \to \mathrm{E}(R_n)/n$



Stabilité du protocole en arbre

Théorème Si $\lambda < 0.36017$ alors le protocole en arbre est **stable**.

Améliorations : Protocole en arbre ternaire



Améliorations

Arbre d-aire

– Séparation en d groupes.

Améliorations

Arbre d-aire

– **Séparation en d groupes.**

– **d grand : Séparation plus rapide.**

Améliorations

Arbre d-aire

– **Séparation en d groupes.**

– **d grand : Séparation plus rapide.**

– **d grand : Beaucoup de silences.**

Améliorations

Arbre d-aire

– Séparation en d groupes.

– d grand : Séparation plus rapide.

– d grand : Beaucoup de silences.

Valeur optimale $d = 3$ Si $\lambda < 0.40159$ alors le protocole en arbre ternaire est stable.

Bornes théoriques

Protocole hybride basé sur l'arbre ternaire

Il existe un protocole stable dès que $\lambda < 0.487$.

Bornes théoriques

Protocole hybride basé sur l'arbre ternaire
Il existe un protocole stable dès que $\lambda < 0.487$.

Débit maximal : $\lambda_{max} \leq 0.56$.

Bornes théoriques

Protocole hybride basé sur l'arbre ternaire
Il existe un protocole stable dès que $\lambda < 0.487$.

Débit maximal : $\lambda_{max} \leq 0.56$.

Conjecture : $\lambda_{max} \leq 0.52$

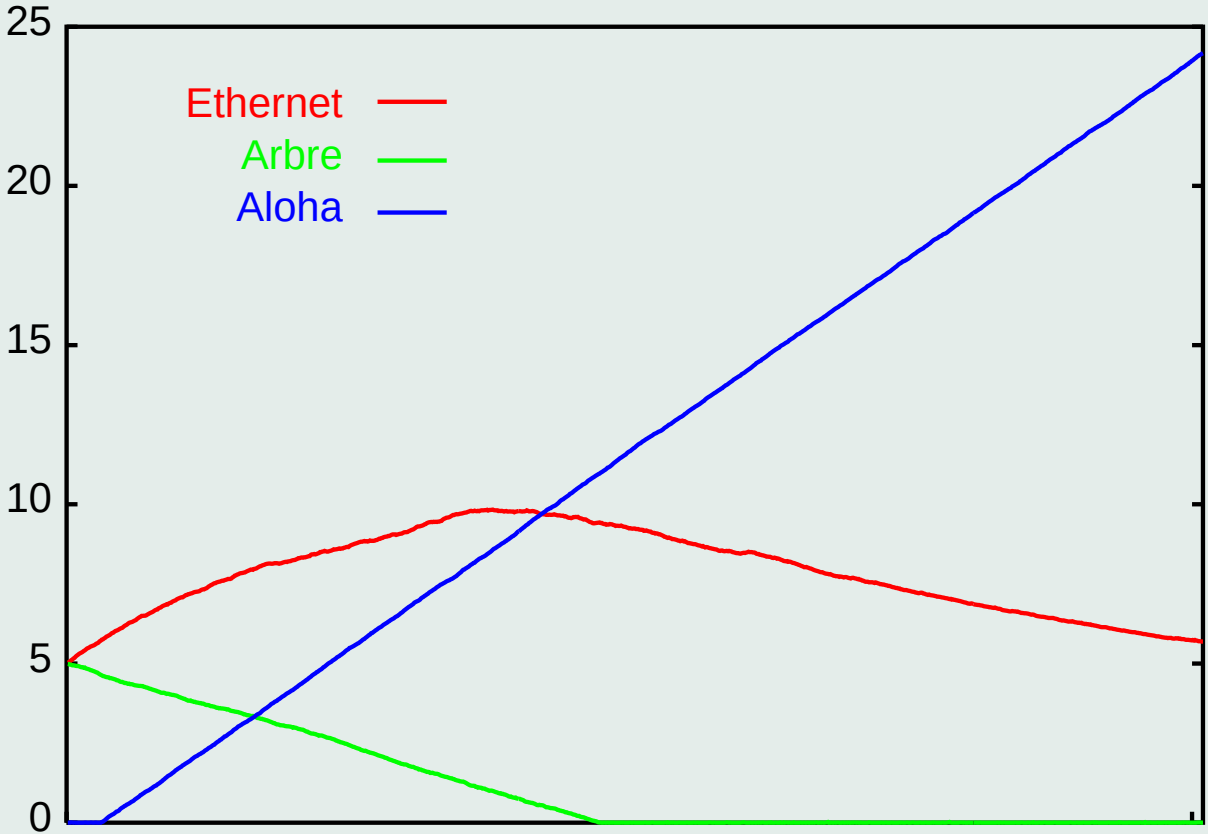
Bornes théoriques

Protocole hybride basé sur l'arbre ternaire
Il existe un protocole stable dès que $\lambda < 0.487$.

Débit maximal : $\lambda_{max} \leq 0.56$.

Conjecture : $\lambda_{max} \leq 0.52$

Protocole en arbre :
Utilisation quasi-optimale de l'écoute du canal.


25
20
15
10
5
0
Ethernet
Arbre
Aloha

Le protocole en arbre

Algorithme quasi-optimal.

Le protocole en arbre

Algorithme quasi-optimal.

Standard Ethernet établi :
$\Rightarrow$ Faible impact industriel.

Le protocole en arbre

Algorithme quasi-optimal.

Standard Ethernet établi :
⇒ Faible impact industriel.

Algorithmique générique.
Algorithmes “Diviser pour Régner”.

Conclusions

- **Algorithmes probabilistes**
  pour gérer les conflits d'accès aux ressources.

- **Poids des standards dans la conception**
  d'algorithmes.

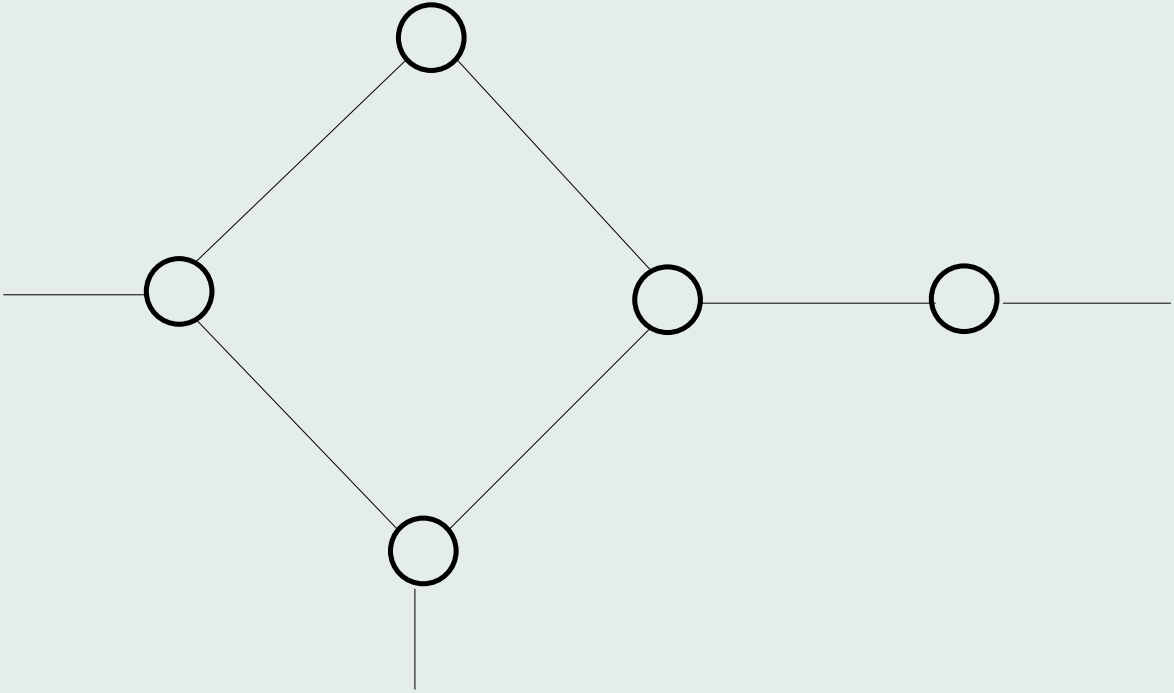
- **Algorithmes en arbre**
  Classe "universelle".

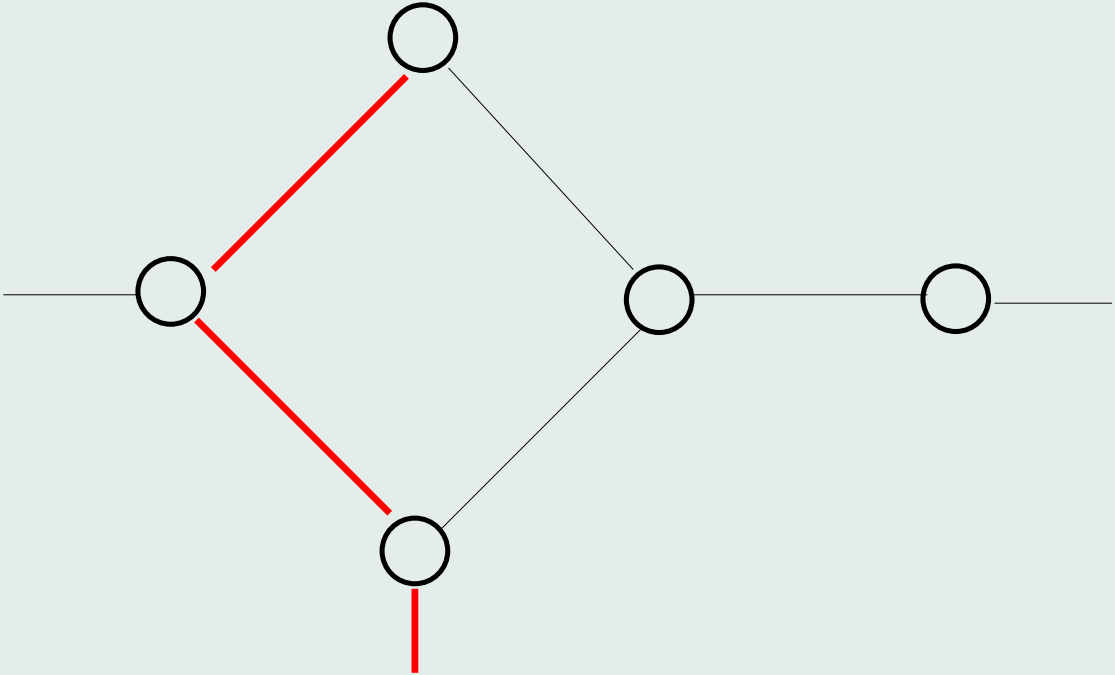
3. Transmission de données dans les réseaux

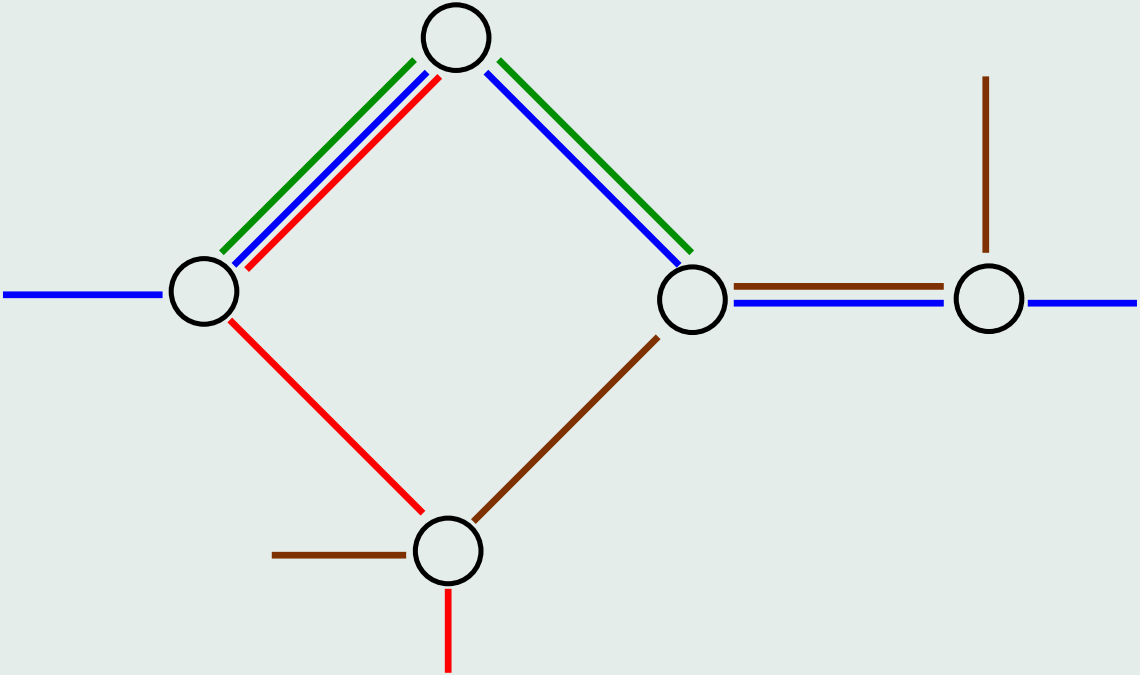
Deux Modèles de Réseaux

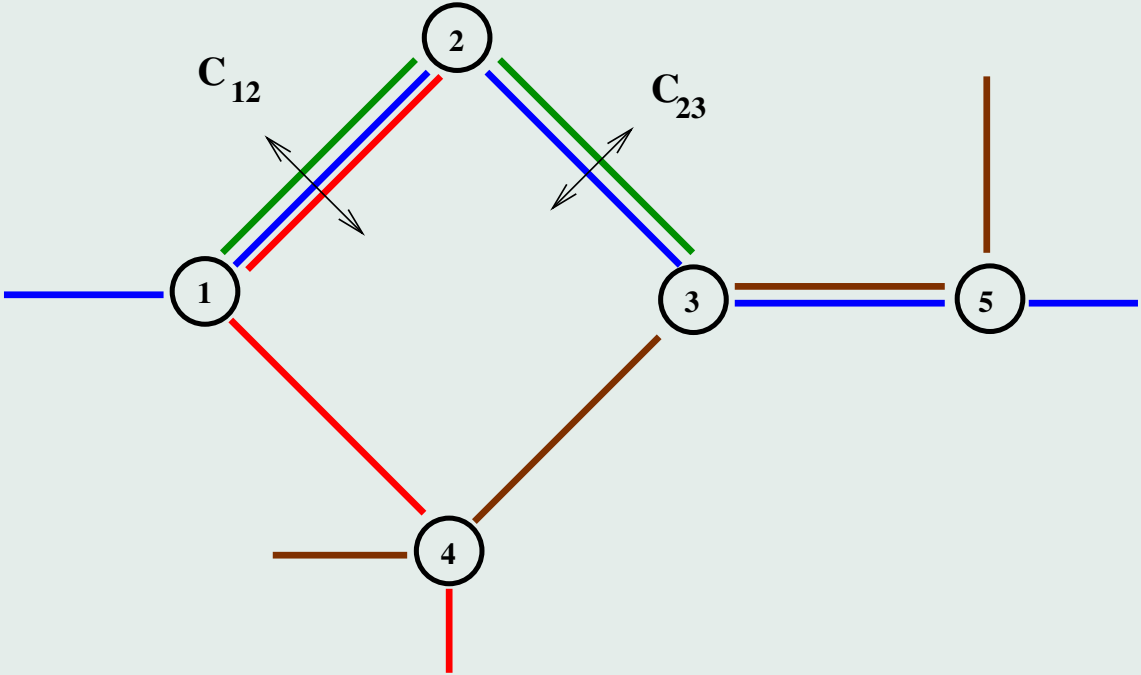
- Réseaux Téléphoniques
- Internet

Les Réseaux Téléphoniques








C 12
C 23
1
2
3
4
5

– **Le réseau téléphonique :**
Réseau à commutation de circuits.

– **Le réseau téléphonique :**
Réseau à commutation de circuits.

– **$\neq$ Réseaux à commutation de paquets.**
(Internet).

Le réseau téléphonique

Un réseau à commutation de circuits :

- **Avantages**
  - **Réservation de ressources.**
  - **Garantie de service.**

Le réseau téléphonique

Un réseau à commutation de circuits :

- **Avantages**
  - **Réservation de ressources.**
  - **Garantie de service.**

- **Problèmes**
  - **Système de type un peu centralisé.**
  - **Extensibilité difficile.**

Internet

L'unité d'information de l'Internet : LE PAQUET

Un paquet : $\simeq$ 15Ko
un entête + les données

- **L'entête : 64 octets**
  contient entre autres l'adresse de la machine qui doit recevoir le paquet.
- **Les données : une partie du contenu du fichier transféré.**

L'unité d'information de l'Internet : LE PAQUET

Un paquet : $\simeq$ 15Ko
un entête + les données

- **L'entête : 64 octets**
  contient entre autres l'adresse de la machine qui doit recevoir le paquet.
- **Les données : une partie du contenu du fichier transféré.**

Exemples
- **Un CD mp3 : 400 000 paquets.**
- **Un film : 4 000 000 paquets.**

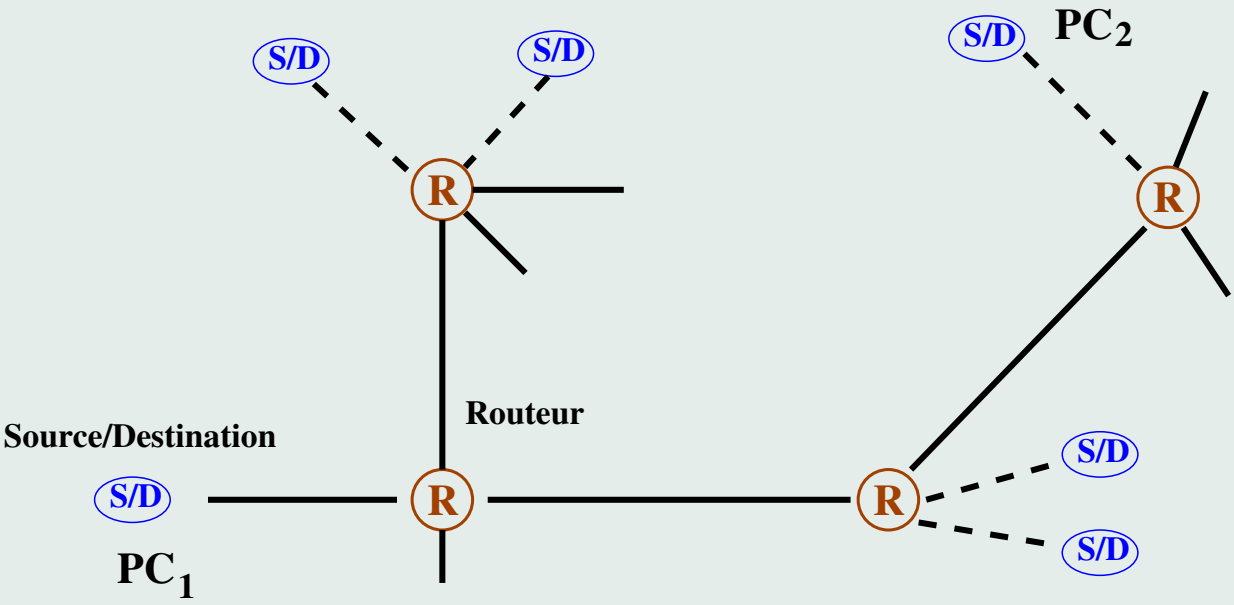
Internet

Un réseau à commutation de paquets

- Messages divisés en paquets.
- Paquets acheminés individuellement.

- Avantages
  - Système distribué.
  - Flexibilité : Évolution facile.

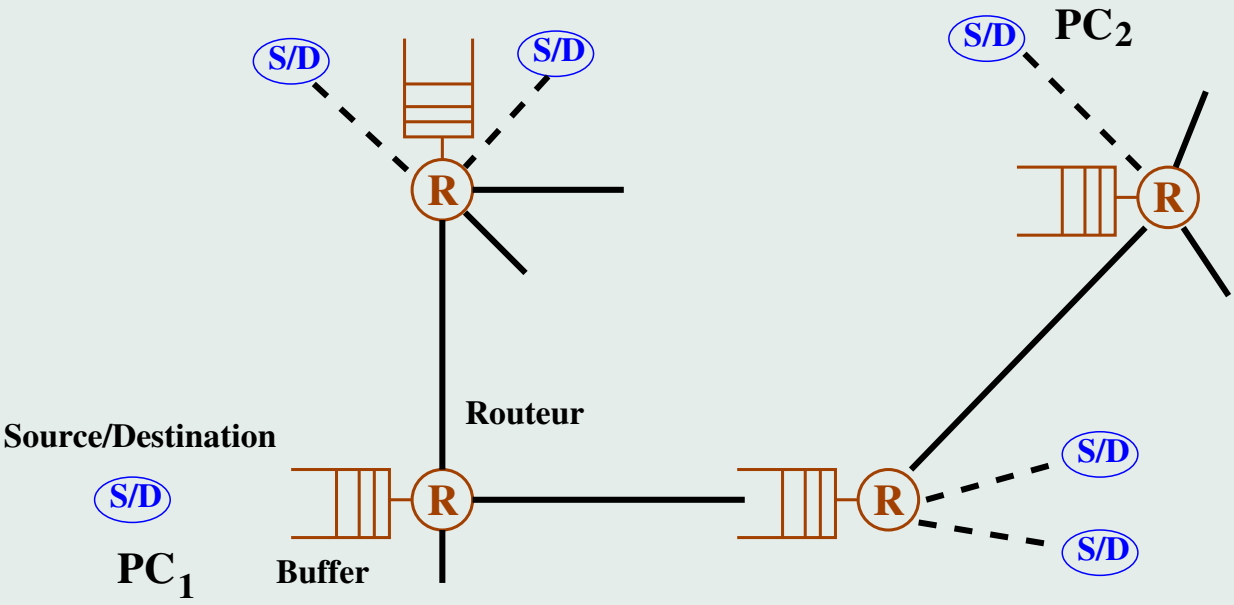
Internet : Une vue simplifiée



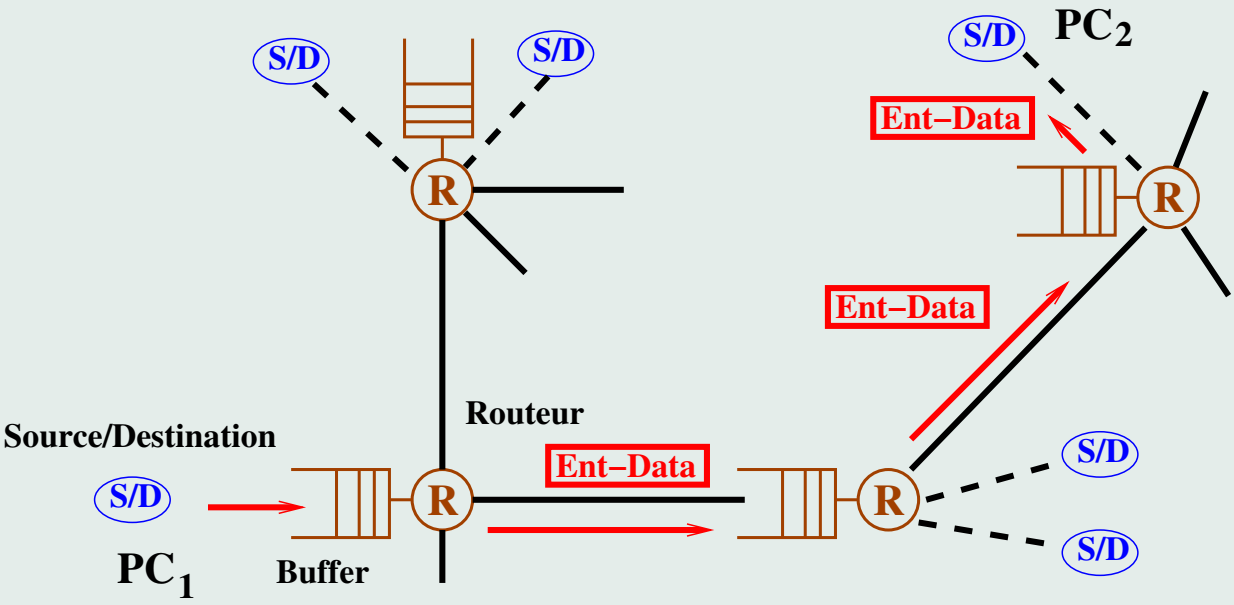
Voyage d'un paquet de Paris à Stanford (Californie)

1	@work	France
2	rocq-renater-gw.inria.fr	
3	gi2-0-inria.cssi.renater.fr	
4	gi8-10-paris1-rtr-021.cssi.renater.fr	
5	vl89-te0-0-0-3-paris1-rtr-001.cssi.renater.fr	
6	renater.rt1.par.fr.geant2.net	
7	so-3-0-0.rt1.lon.uk.geant2.net	UK
8	so-2-0-0.rt1.ams.nl.geant2.net	Hollande
9	198.32.11.50	États-Unis
10	so-0-0-0.0.rtr.wash.net.internet2.edu	Washington
11	so-0-0-0.0.rtr.atla.net.internet2.edu	Atlanta
12	so-3-2-0.0.rtr.hous.net.internet2.edu	Houston
13	so-3-0-0.0.rtr.losa.net.internet2.edu	Los Angeles
14	hpr-lax-hpr–i2-newnet.cenic.net	
15	svl-hpr–lax-hpr-10ge.cenic.net	
16	oak-hpr–svl-hpr-10ge.cenic.net	
17	hpr-stan-ge–oak-hpr.cenic.net	
18	bbrb-i2.Stanford.EDU	Stanford

Internet : Une vue simplifiée



Internet : Une vue simplifiée



Transfert du fichier F de la machine PC_1 vers la machine PC_2

– **Sur chaque machine : un programme contrôle l'échange.**

Transfert du fichier F de la machine PC_1 vers la machine PC_2

– **Sur chaque machine : un programme contrôle l'échange.**

– **PC_1 segmente en n paquets une copie de F et envoie le numéro 1, puis 2, . . .n à PC_2.**

Transfert du fichier F
de la machine PC_1 vers la machine PC_2

– **Sur chaque machine : un programme contrôle l'échange.**

– **PC_1 segmente en n paquets une copie de F et envoie le numéro 1, puis 2, . . .n à PC_2.**

Mémoire des routeurs finie : en cas de congestion
$\Rightarrow$ Le réseau perd des paquets.

Transfert du fichier F
de la machine PC_1 vers la machine PC_2

– **Sur chaque machine : un programme contrôle l'échange.**

– **PC_1 segmente en n paquets une copie de F et envoie le numéro 1, puis 2, . . .n à PC_2.**

Mémoire des routeurs finie : en cas de congestion
$\Rightarrow$ Le réseau perd des paquets.

Problème : Comment transmettre de façon fiable dans un réseau qui ne l'est pas ?

Transmission de Données sur Internet

TCP : Transmission Control Protocol.

– **Algorithme de transmission de données.**

– **$> 95\%$ du traffic Internet contrôlé par TCP.**

Les principes de base TCP

Cerf and Kahn (1973)

– **Accusé de réception des messages.**

– **Régulation des envois : à un instant une source a au plus W paquets en circulation dans le réseau.**

W : Taille de la fenêtre de congestion.

Les principes de base TCP (II)

Contrôle de la congestion Jacobson (1987)

– **Transmission de W paquets OK :**
$W \to W + 1$

– **Un paquet est perdu :**
$W \to W/2$.

Exemple simplifié

Transfert de F entre les machines PC_1 et PC_2 :

– **PC_2 : envoie un paquet pour demander F.**
– **PC_1 envoie les paquets n° 1, 2, ...W à PC_2**

Exemple simplifié

Transfert de F entre les machines PC_1 et PC_2 :

- **PC_2 : envoie un paquet pour demander F.**
- **PC_1 envoie les paquets n° 1, 2, ...W à PC_2**
- **PC_2 reçoit le $n^\circ 1 \Rightarrow$ envoie paquet OK–1 à PC_1 "j'ai bien reçu 1"...**

Exemple simplifié

Transfert de F entre les machines PC_1 et PC_2 :

– PC_2 : envoie un paquet pour demander F.
– PC_1 envoie les paquets n° 1, 2, ...W à PC_2
– PC_2 reçoit le $n^\circ 1 \Rightarrow$ envoie paquet OK–1 à PC_1
"j'ai bien reçu 1"...
– ...

– Si tout va bien :
PC_1 a reçu $OK-1$, ..., $OK-W$
$\Rightarrow$ PC_1 Envoie les $W+1$ paquets suivants
n° $W+1$, ...$W+W+1 = 2W+1$.

Exemple simplifié

– **Une perte :**
le paquet $n^\circ i$ a été perdu
$\Rightarrow$ PC_1 renvoie le paquet $n^\circ i$
PC_1 envoie les $\lfloor W/2 \rfloor$ paquets
n° $W+1, \ldots W+\lfloor W/2 \rfloor$.

– **etc. . .**

Conclusion sur TCP

+++ Adaptation aux conditions de trafic.

-- Pas de garantie de débit, d'accès, ...

Conclusion sur TCP

+++ Adaptation aux conditions de trafic.

-- Pas de garantie de débit, d'accès, ...

Remarquables propriétés d'auto-stabilisation

4. Google

Un peu d'histoire

- **1995 : Premiers moteurs de recherche**
  AltaVista

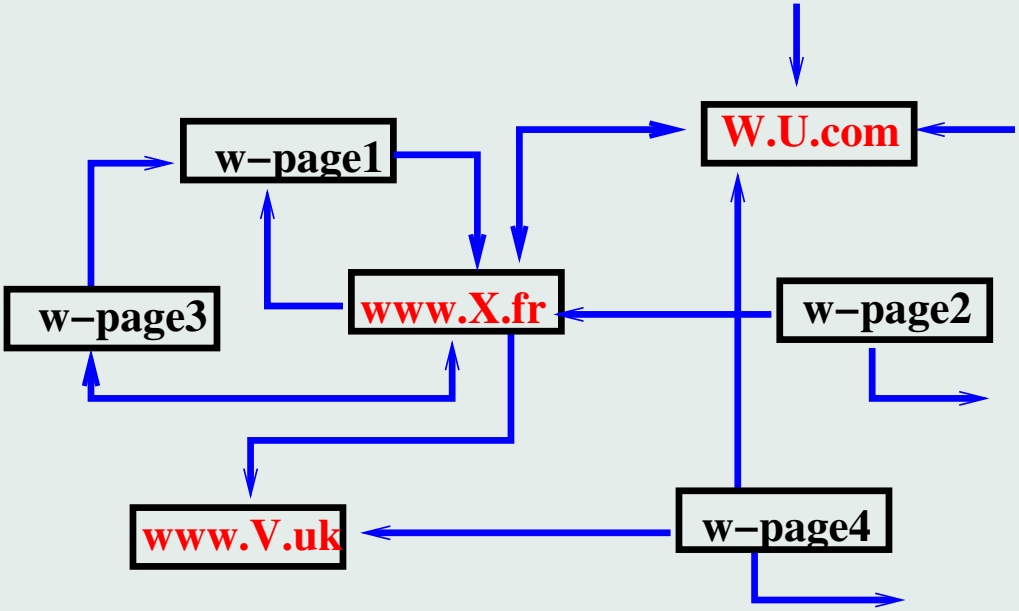
- **1998 : Article Brin et Page**
  "The anatomy of a largescale hypertextual web search engine"

Un peu d'histoire

- **1995 : Premiers moteurs de recherche**
  AltaVista

- **1998 : Article Brin et Page**
  "The anatomy of a largescale hypertextual web search engine"

  - **Introduction de la notion de "Page rank".**
  - **Algorithme pour estimer celui-ci.**



Le web : un graphe orienté

25 milliards de pages web indexées par Google

Comment Marche un Moteur de Recherche ?

Problème : Recherche d'un site web ayant une information sur le sujet "*XYZ*"

Comment Marche un Moteur de Recherche ?

Problème : Recherche d'un site web ayant une information sur le sujet "*XYZ*"

- **Première étape (facile)**
  Recherche de l'ensemble des sites web ayant ce mot "XYZ"

Comment Marche un Moteur de Recherche ?

Problème : Recherche d'un site web ayant une information sur le sujet "XYZ"

- Première étape (facile)
  Recherche de l'ensemble des sites web ayant ce mot "XYZ"

- Deuxième étape
  Quel est le site web le plus pertinent ?

Comment Marche un Moteur de Recherche ?

Principe : Il faut trouver une fonction π telle que :
À une page web p on associe $\pi(p) \in [0, 1]$.

p plus pertinent que q si $\pi(p) > \pi(q)$.

Comment Marche un Moteur de Recherche ?

Principe : Il faut trouver une fonction π telle que :
À une page web p on associe $\pi(p) \in [0, 1]$.

p plus pertinent que q si $\pi(p) > \pi(q)$.

$$\mathcal{E}_{XYZ} = \{p : p \text{ page web contenant “}XYZ''\}$$

Comment Marche un Moteur de Recherche ?

Principe : Il faut trouver une fonction π telle que : À une page web p on associe $\pi(p) \in [0, 1]$.

p plus pertinent que q si $\pi(p) > \pi(q)$.

$$\mathcal{E}_{XYZ} = \{p : p \text{ page web contenant “}XYZ''\}$$

Action : Afficher les adresses des pages $q_1, \ldots q_{10}$ ayant les 10 plus grandes valeurs pour π sur $\mathcal{E}_{XYZ}$.

gevrey chambertin

Rechercher

Recherche avancée
Préférences

Rechercher sur le Web Rechercher les pages en français

Web

Résultats **1** - **10** sur un total d'environ **659 000** pour **gevrey chambertin** (0,

site de l'office de tourisme de **Gevrey-Chambertin** en Bourgogne
site de l'office de tourisme de **gevrey-chambertin** en Bourgogne terre des grands vins et crus.
www.ot-**gevreychambertin**.fr/ - 2k - En cache - Pages similaires

Gevrey-Chambertin - Wikipédia
Gevrey-Chambertin est une commune française viticole, située à 15 km au sud de Dijon du département de la Côte-d'Or de la région Bourgogne. **...**
fr.wikipedia.org/wiki/**Gevrey-Chambertin** - 58k - En cache - Pages similaires

Communauté de Communes de **Gevrey** - **Chambertin**
Présentation des communes des Hautes-Côtes, tourisme, scolaire, développement économique, eau et assainissement, réserve naturelle de la combe Lavaux Jean **...**
www.cc**gevrey-chambertin**.com/ - 14k - En cache - Pages similaires

Vin de **Gevrey Chambertin** avec le Guide des vins de France en Bourgogne
Gevrey Chambertin est la plus grande appellation de Côte de Nuits en Bourgogne.
www.terroirs-france.com/region/bourgogne_**gevrey**.htm - 31k - En cache - Pages similaires

1855: **Gevrey-chambertin**, vin de Bourgogne
Gevrey-chambertin - Cette appellation produit l'un des vins rouges les plus réputés de la Côte de Nuits. Ils sont puissants, denses, intenses, bien plus que **...**
www.1855.com/bourgogne/app/69/fr/**Gevrey-Chambertin** - 121k -
En cache - Pages similaires

Gevrey-Chambertin : vins au catalogue
Gevrey-Chambertin : le sommelier de 75cl.com a sélectionné pour vous, une gamme de vins de l'appellation **Gevrey-Chambertin**.

La fonction π pour Google

Brin et Page (1997)

$\pi(q)$: Importance de la page q.

$$\mathcal{L}_q = \{p : q \text{ a un lien vers la page web } p\}$$

La fonction π pour Google

Brin et Page (1997)

$\pi(q)$: Importance de la page q.

$$\mathcal{L}_q = \{p : q \text{ a un lien vers la page web } p\}$$

Principe :
Importance de p "transmise/héritée" de q :

$$\pi(q) M(q,p)$$

avec

$$M(q,p) = \frac{1}{\text{Card}\mathcal{L}_q}$$

La fonction π pour Google

Importance de q transmise à p

$$\pi(q)M(q,p)$$

La fonction π pour Google

Équation linéaire pour π

$$\pi(p) = \sum_{q:p\in\mathcal{L}_q} \pi(q)M(q,p)$$

La fonction π pour Google

Équation linéaire pour π

$$\pi(p) = \sum_{q:p\in\mathcal{L}_q} \pi(q) M(q,p)$$

Le système est singulier de rang $\mathrm{Card}\mathcal{S} - 1$

$$\sum_{p:p\in\mathcal{L}_q} M(q,p) = \sum_{p:p\in\mathcal{L}_q} \frac{1}{\mathrm{Card}\mathcal{L}_q} = 1.$$

$\mathcal{S}$: ensemble de toutes les pages web.

La fonction π pour Google

Si $\mathcal{M} = (M(p,q), p, q \in \mathcal{S})$, il existe un unique vecteur $(\pi(p), p \in \mathcal{S})$ tel que

$$\pi = \pi\mathcal{M}$$

et

$$\sum_{p \in \mathcal{S}} \pi(p) = 1.$$

La fonction π pour Google

Si $\mathcal{M} = (M(p,q), p, q \in \mathcal{S})$, il existe un unique vecteur $(\pi(p), p \in \mathcal{S})$ tel que

$$\pi = \pi\mathcal{M}$$

et

$$\sum_{p \in \mathcal{S}} \pi(p) = 1.$$

Propriété : $\pi(p) \in (0,1), \forall p \in \mathcal{S}.$

La fonction π pour Google

Si $\mathcal{M} = (M(p,q), p, q \in \mathcal{S})$, il existe un unique vecteur $(\pi(p), p \in \mathcal{S})$ tel que

$$\pi = \pi\mathcal{M}$$

et

$$\sum_{p \in \mathcal{S}} \pi(p) = 1.$$

Propriété : $\pi(p) \in (0,1), \forall p \in \mathcal{S}.$

Un détail : $\text{Card}\mathcal{S} = 24$ milliards !

La fonction π pour Google en pratique

Analyse numérique :

– **Produits matrice/vecteurs.**

– **Techniques d'uniformisation.**

La fonction π pour Google

Une interprétation probabiliste

La fonction π pour Google

Une interprétation probabiliste

Modèle mathématique : surfeur aléatoire S

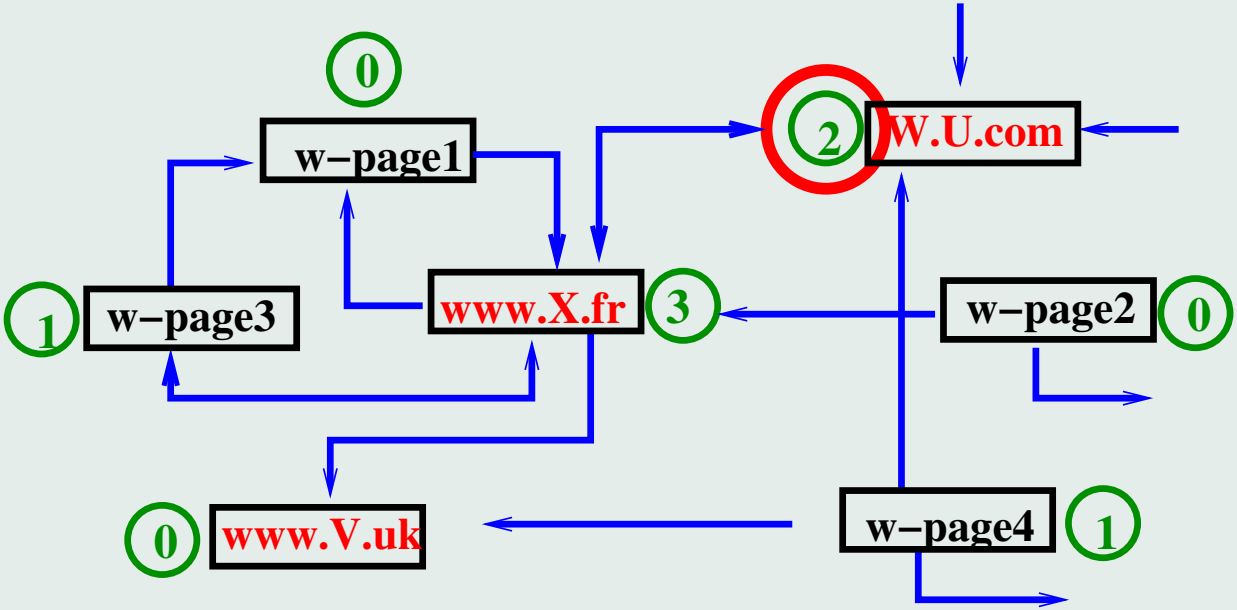
La fonction π pour Google

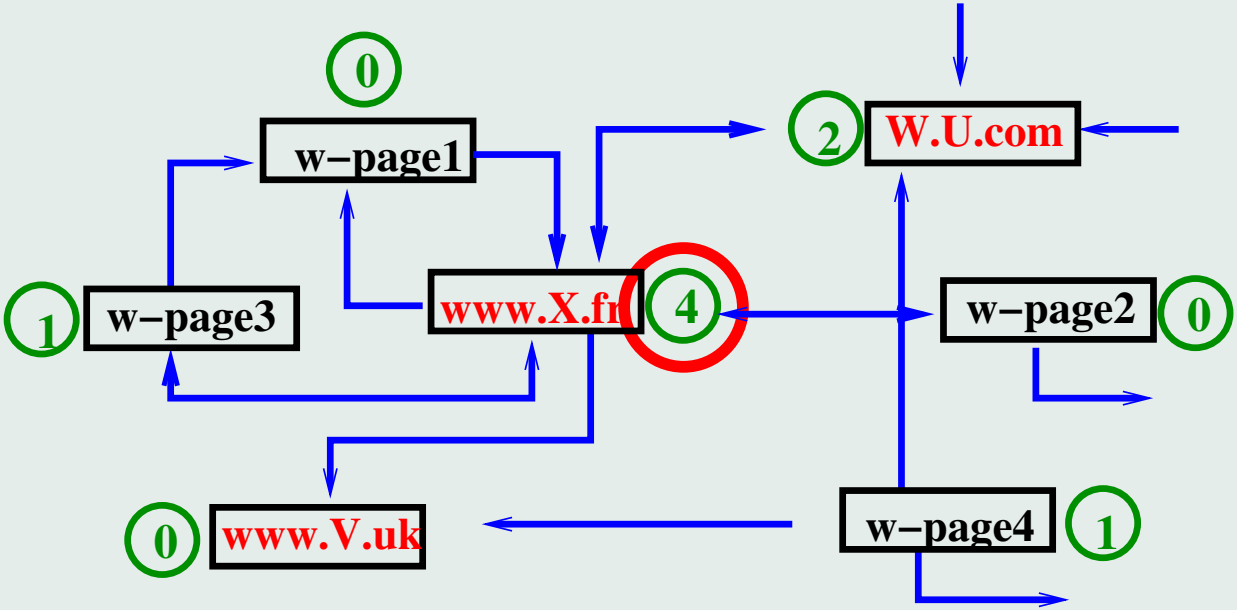
Une interprétation probabiliste

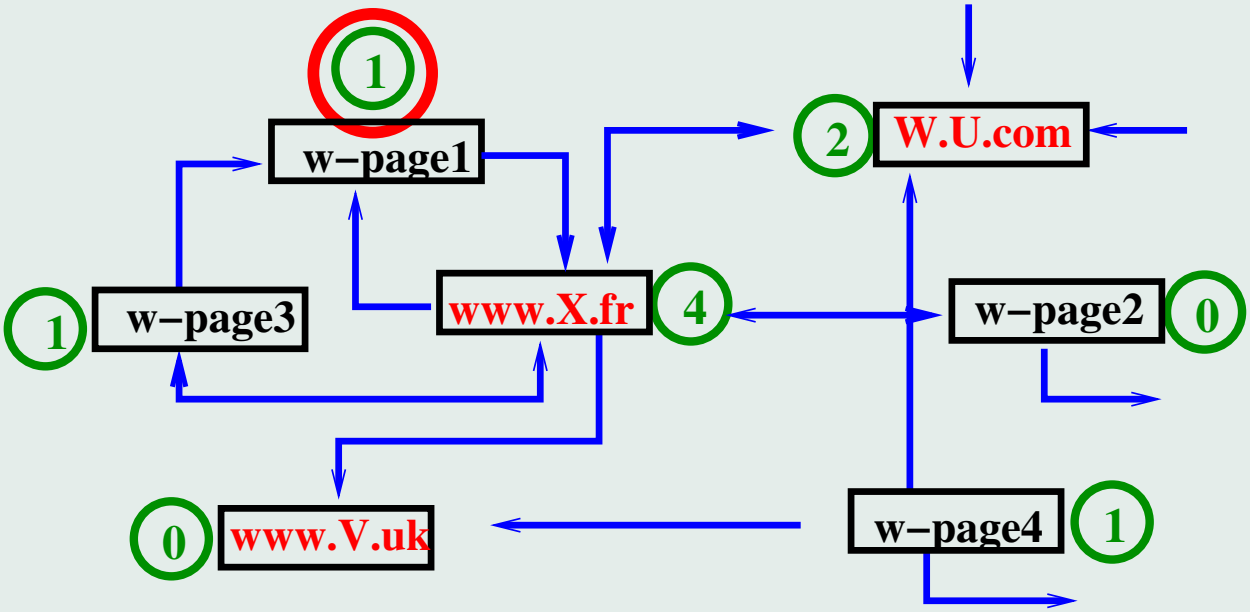
Modèle mathématique : surfeur aléatoire S

– S navigue au hasard sur le web :

Si S sur une page web à l'instant t
à $t+1$, S choisit au hasard un lien de cette page
et va sur la page web correspondante, etc...


0
w−page1
2
W.U.com
1
w−page3
www.X.fr
3
w−page2
0
0
www.V.uk
w−page4
1


0
w–page1
2
W.U.com
1
w–page3
www.X.fr
4
w–page2
0
0
www.V.uk
w–page4
1


1
w−page1
2
W.U.com
1
w−page3
www.X.fr
4
w−page2
0
0
www.V.uk
w−page4
1

La fonction π pour Google

Durée du surf $T = 10^{12}$ (par exemple).

Si p est une page web,

$$f_T(p) = \frac{1}{T} N_T(p),$$

où $N_T(p)$: nb de passages à p entre 0 et T.

La fonction π pour Google

Durée du surf $T = 10^{12}$ (par exemple).

Si p est une page web,

$$f_T(p) = \frac{1}{T} N_T(p),$$

où $N_T(p)$: nb de passages à p entre 0 et T.

p_1 plus pertinent que p_2 si $f_T(p_1) > f_T(p_2)$.

La fonction π pour Google

Durée du surf $T = 10^{12}$ (par exemple).

Si p est une page web,

$$f_T(p) = \frac{1}{T} N_T(p),$$

où $N_T(p)$: nb de passages à p entre 0 et T.

p_1 plus pertinent que p_2 si $f_T(p_1) > f_T(p_2)$.

Problèmes

– Dépend de T ?

La fonction π pour Google

Durée du surf $T = 10^{12}$ (par exemple).

Si p est une page web,

$$f_T(p) = \frac{1}{T} N_T(p),$$

où $N_T(p)$: nb de passages à p entre 0 et T.

p_1 plus pertinent que p_2 si $f_T(p_1) > f_T(p_2)$.

Problèmes

– Dépend de T ?

– Dépend du point de départ du surfeur ?

La fonction π pour Google

Durée du surf $T = 10^{12}$ (par exemple).

Si p est une page web,

$$f_T(p) = \frac{1}{T} N_T(p),$$

où $N_T(p)$: nb de passages à p entre 0 et T.

p_1 plus pertinent que p_2 si $f_T(p_1) > f_T(p_2)$.

Problèmes

– Dépend de T ?
– Dépend du point de départ du surfeur ?
– Dépend du choix du surfeur ?

Résultats de Maths : Théorème ergodique

La limite existe :

$$\pi(p) = \lim_{T \to +\infty} \frac{N_T(p)}{T}$$

Résultats de Maths : Théorème ergodique

La limite existe :

$$\pi(p) = \lim_{T \to +\infty} \frac{N_T(p)}{T}$$

– $\pi(p)$ **ne dépend donc pas de** T **(si** T **assez grand).**

Résultats de Maths : Théorème ergodique

La limite existe :

$$\pi(p) = \lim_{T \to +\infty} \frac{N_T(p)}{T}$$

– $\pi(p)$ **ne dépend donc pas de** T **(si** T **assez grand).**
– $\pi(p)$ **ne dépend pas du point de départ.**

Résultats de Maths : Théorème ergodique

La limite existe :

$$\pi(p) = \lim_{T \to +\infty} \frac{N_T(p)}{T}$$

– $\pi(p)$ **ne dépend donc pas de** T **(si** T **assez grand).**
– $\pi(p)$ **ne dépend pas du point de départ.**
– $\pi(p)$ **ne dépend pas du choix du surfeur.**

Résultats de Maths : Théorème ergodique

La limite existe :

$$\pi(p) = \lim_{T \to +\infty} \frac{N_T(p)}{T}$$

– $\pi(p)$ **ne dépend donc pas de** T **(si** T **assez grand).**
– $\pi(p)$ **ne dépend pas du point de départ.**
– $\pi(p)$ **ne dépend pas du choix du surfeur.**

π **est l'unique solution de**

$$\pi = \pi \mathcal{M} \text{ et } \sum_{x \in \mathcal{S}} \pi(x) = 1.$$

Conclusions

Idées brillantes de Brin et Page :

– **Modélisation :**
Représentation mathématique du “page rank”.

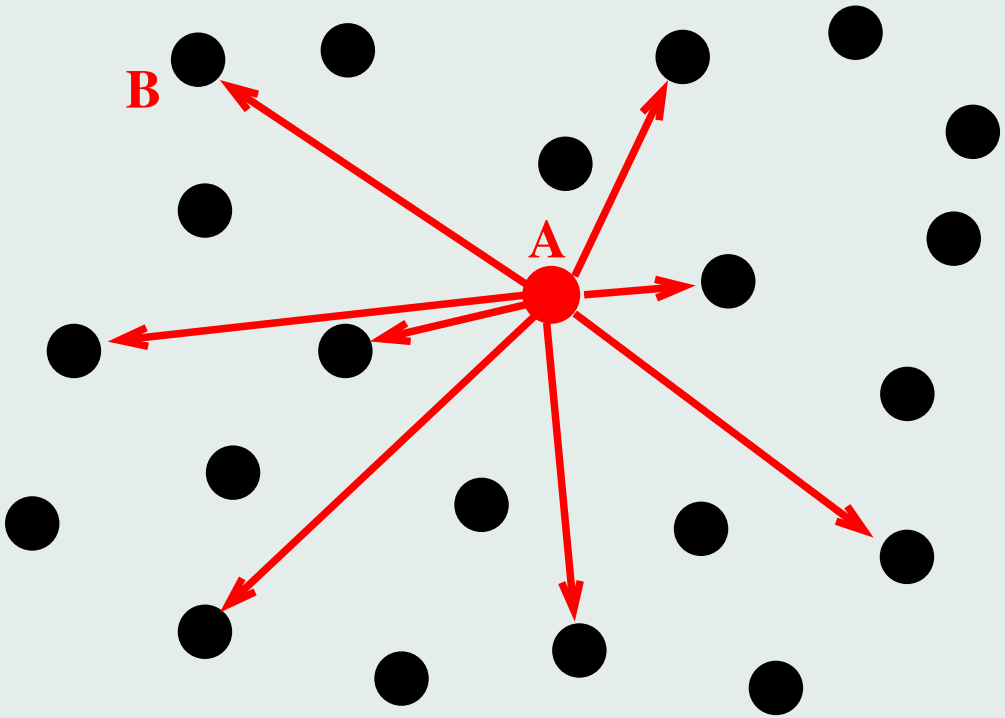
– **Algorithme de calcul de π.**

5. Les réseaux sociaux

Réseaux sociaux

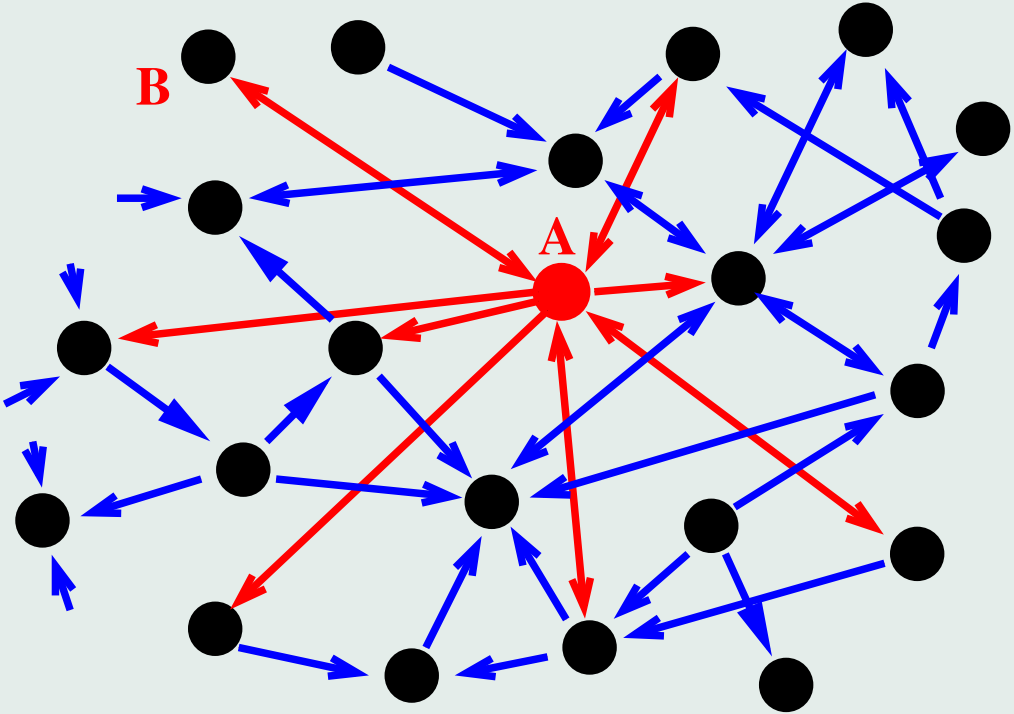
Réseaux sociaux

– **Réseaux par affinité : A ami de B :**
un lien de A vers B.


B
A

Réseaux sociaux

- **Réseaux par affinité : A ami de B : un lien de A vers B.**
- **Très grand nombre de nœuds.**
  Facebook : 500 millions


B
A

Réseaux sociaux

Problème :
Comment extraire de l'information de ces réseaux ?

Réseaux sociaux

Problème :
Comment extraire de l'information de ces réseaux ?

Un domaine actif de recherche

- **Data Mining (Fouille de données)**
  Algorithmes pour structurer les données.
- **Typologie des Réseaux Sociaux**
  Caractérisation/Estimation des graphes
- **Navigation**
  Algorithmes pour se déplacer.

Réseaux sociaux

Problème :
Comment extraire de l'information de ces réseaux ?

Un domaine actif de recherche
- **Data Mining (Fouille de données)**
  Algorithmes pour structurer les données.
- **Typologie des Réseaux Sociaux**
  Caractérisation/Estimation des graphes
- **Navigation**
  Algorithmes pour se déplacer.

Enjeux économiques

What's this?
TimesPeople Lets You Share and Discover the Best of NY...

Facebook's Users Ask Who Owns Information

By BRIAN STELTER
Published: February 16, 2009

Reacting to an online swell of suspicion about changes to Facebook's terms of service, the company's chief executive moved to reassure users on Monday that the users, not the Web site, "own and control their information."

COMMENTS (92)
E-MAIL
SEND TO PHONE
PRINT
REPRINTS
SHARE

ARTICLE TOOLS
SPONSORED BY

Related

The online exchanges reflected the uneasy and evolving balance

Les Modèles Économiques des Réseaux

1960 — 1985 IBM.
Machine+programme Informatique spécifique

Les Modèles Économiques des Réseaux

1960 — 1985 IBM.
Machine+programme Informatique spécifique

1985 — 201 ? Microsoft.
Programme Informatique généraliste

Les Modèles Économiques des Réseaux

1960 — 1985 IBM.
Machine+programme Informatique spécifique

1985 — 201 ? Microsoft.
Programme Informatique généraliste

2000 — ? Google.
Recherche sur Internet.

Les Modèles Économiques des Réseaux

1960 — 1985 IBM.
Machine+programme Informatique spécifique

1985 — 201 ? Microsoft.
Programme Informatique généraliste

2000 — ? Google.
Recherche sur Internet.

? — Facebook, Myspace, Twitter ?
? ? ?

6. Conclusion

Les algorithmes

Pas uniquement

– de la programmation

– des schémas numériques

Les algorithmes

Pas uniquement

– de la programmation

– des schémas numériques

Cadre générique :

– Résoudre un problème avec des règles du jeu.

Les algorithmes

Pas uniquement

– de la programmation

– des schémas numériques

Cadre générique :

– Résoudre un problème avec des règles du jeu.

– $\Rightarrow$ connaître les règles du jeu :
Définir le cadre algorithmique.
Formaliser/Modéliser.

La Fin