

**Systemes pair-à-pair
et
diffusion épidémique d'information**

Laurent Massoulié

**Thomson
Laboratoire de recherche parisien**

Plan

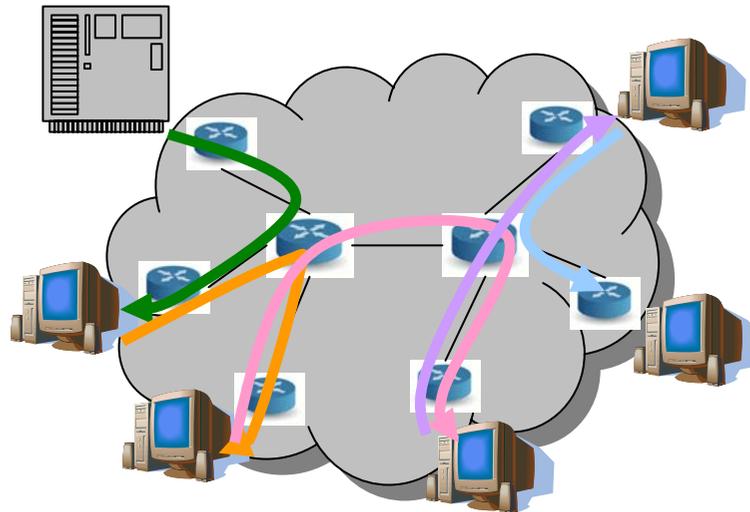
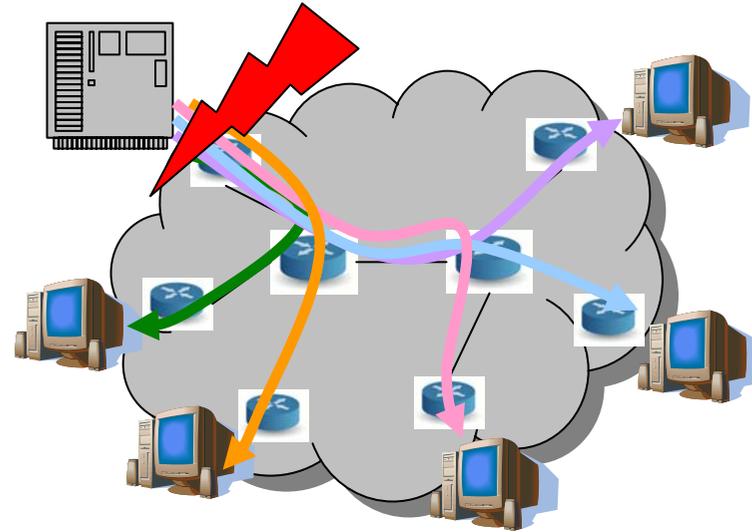
- **Le pair-à-pair aujourd'hui**
- **Exemple de diffusion épidémique: un virus informatique**
- **Epidémies et diffusion de flux d'information**
- **Conclusions et Perspectives**

Plan

- **Le pair-à-pair aujourd'hui**
- Exemple de diffusion épidémique: un virus informatique
- Epidémies et diffusion de flux d'information
- Conclusions et Perspectives

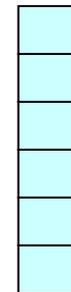
Principe: Mutualisation des ressources

Architecture client-serveur
(Telechargements; Podcast)



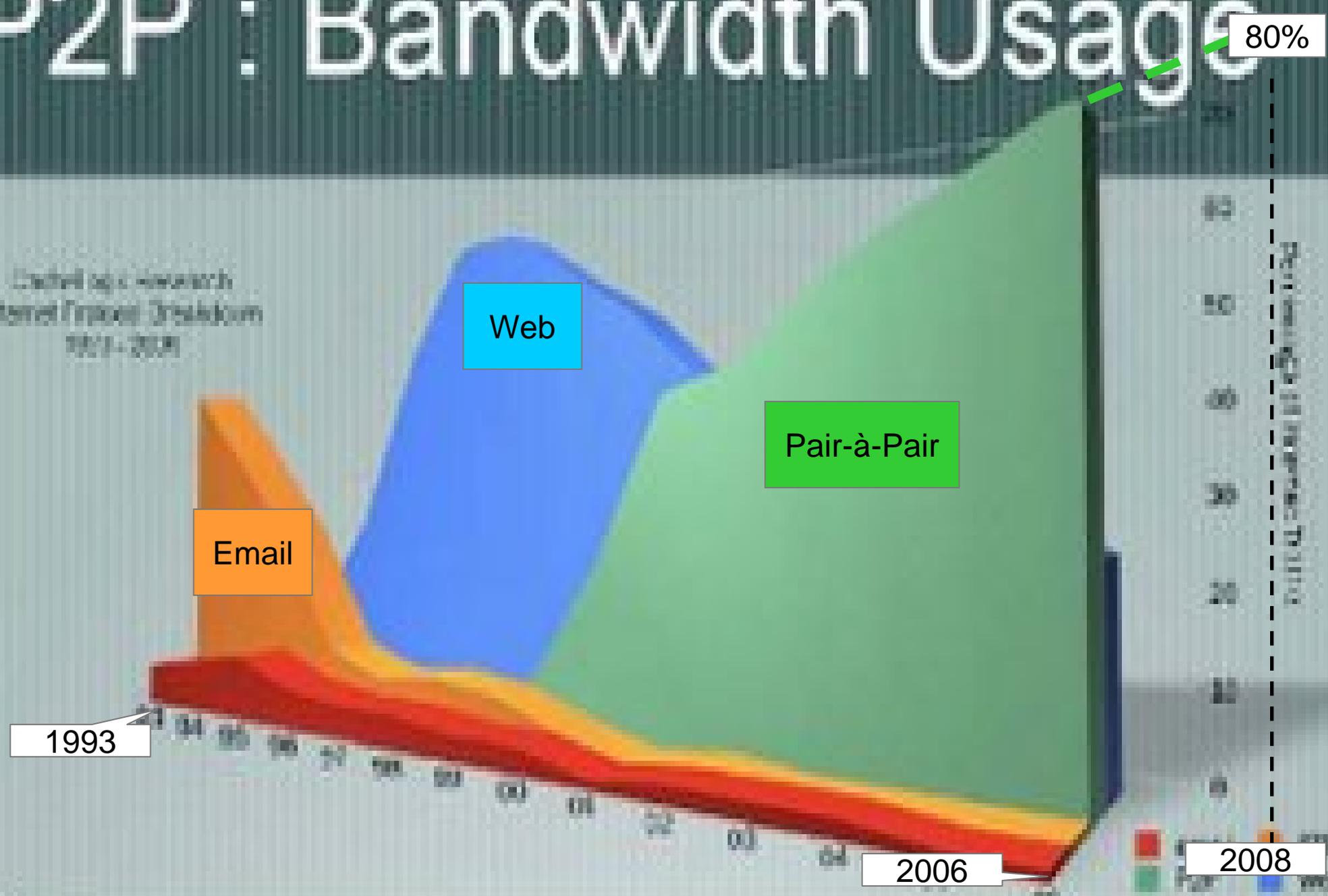
Architecture pair-à-pair

Capacité



P2P : Bandwidth Usage

Crutcher et al. Research
Internet Traffic Breakdown
1993-2008



Et pourtant...

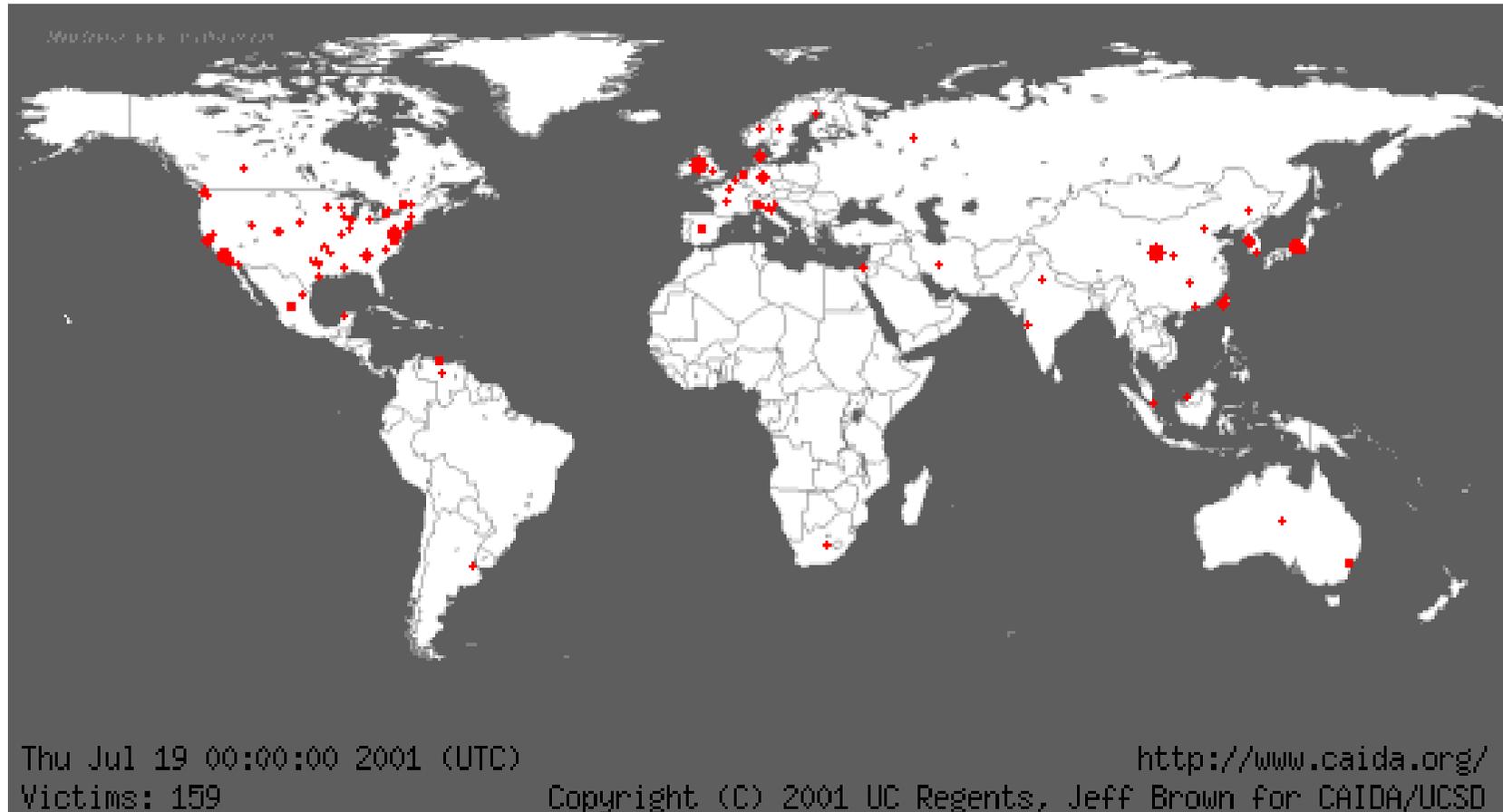
- Le pair-à-pair est là pour rester
- Incontournable pour ces applications :
 - Partage de fichiers
 - Vidéo à la demande
 - Diffusion de flux vidéo en temps réel (télévision sur Internet)



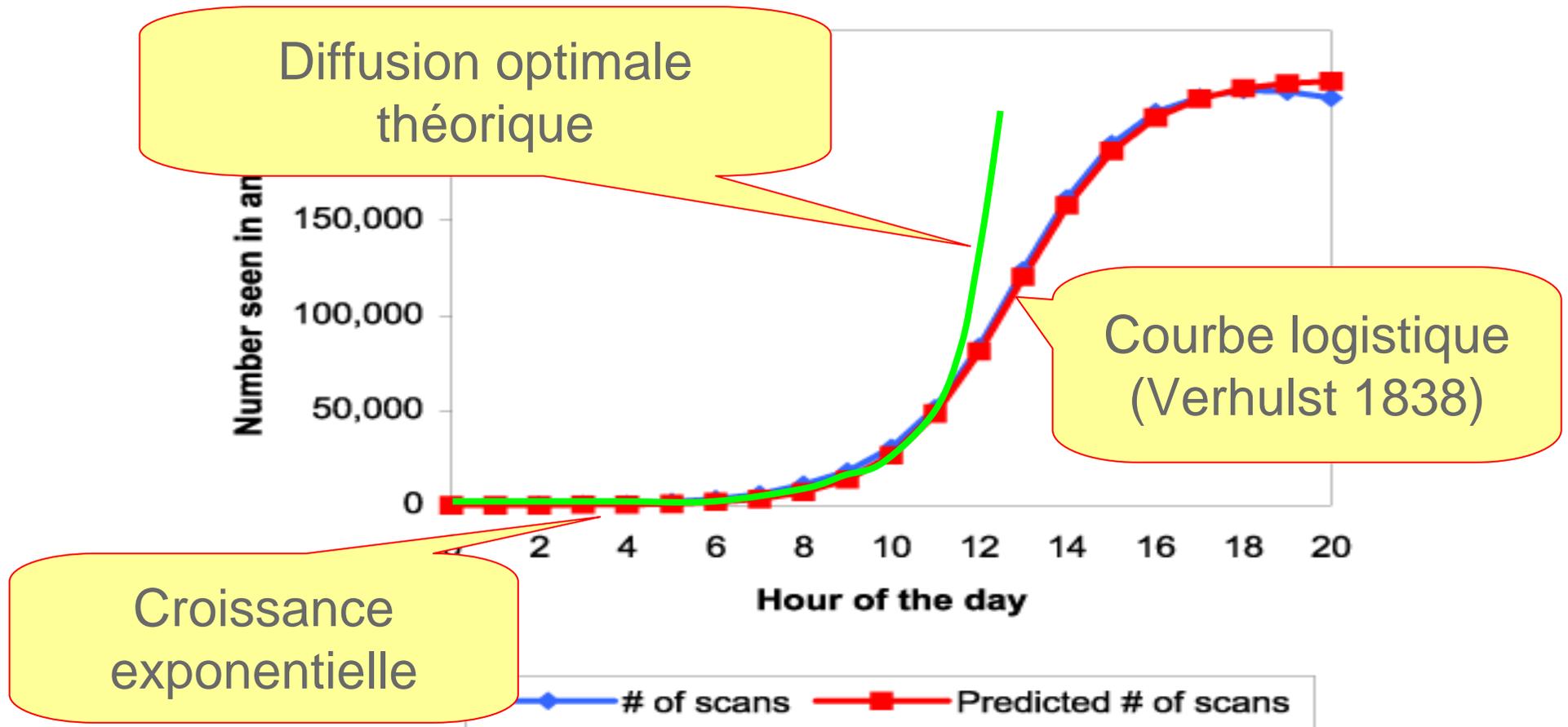
Plan

- Le pair-à-pair aujourd'hui
- **Exemple de diffusion épidémique: un virus informatique**
- Epidémies et diffusion de flux d'information
- Conclusions et Perspectives

Diffusion du virus « Code Red »



Diffusion du virus « Code Red »

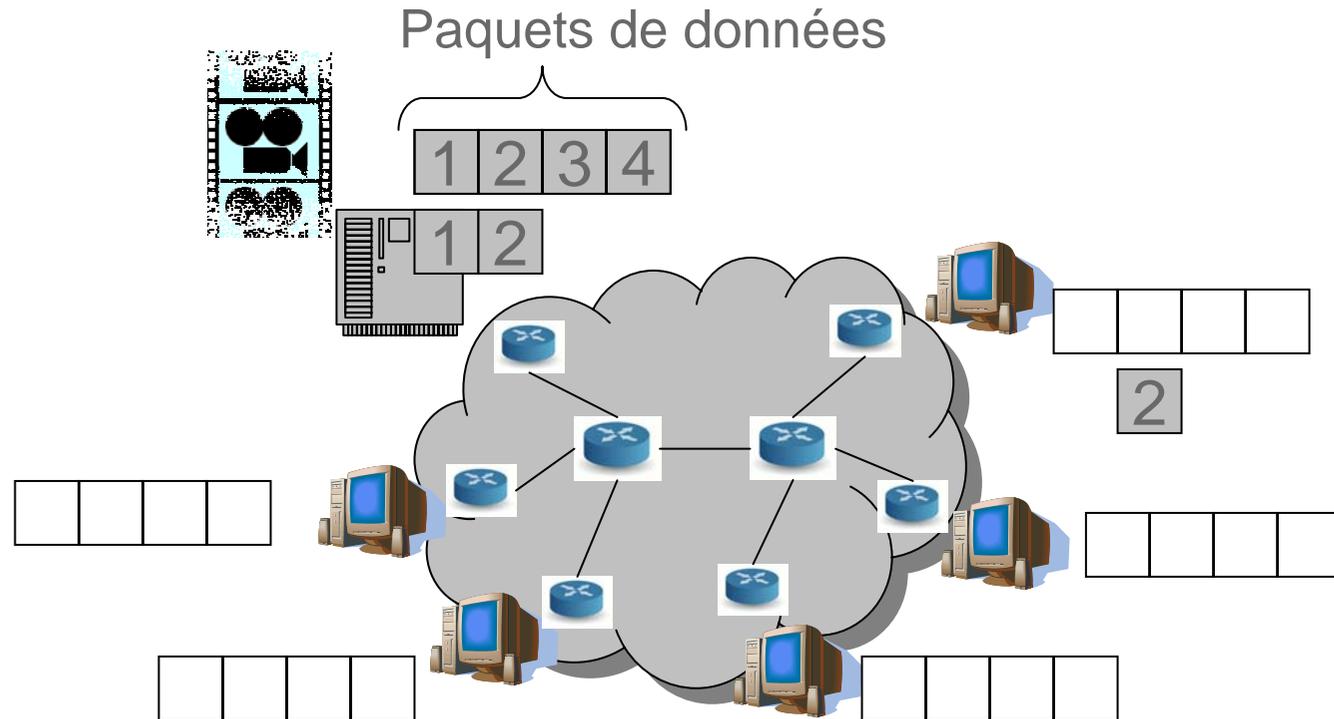


Temps d'infection complète « optimal »
(logarithmique en nombre de cibles)

Plan

- Le pair-à-pair aujourd'hui
- Exemple de diffusion épidémique: un virus informatique
- **Epidémies et diffusion de flux d'information**
- Conclusions et Perspectives

Epidémies pour la diffusion d'information



Spécification du mécanisme: règle de sélection de

- nœud cible
- paquet à transmettre

→ **Des épidémies (une par paquet) en compétition**

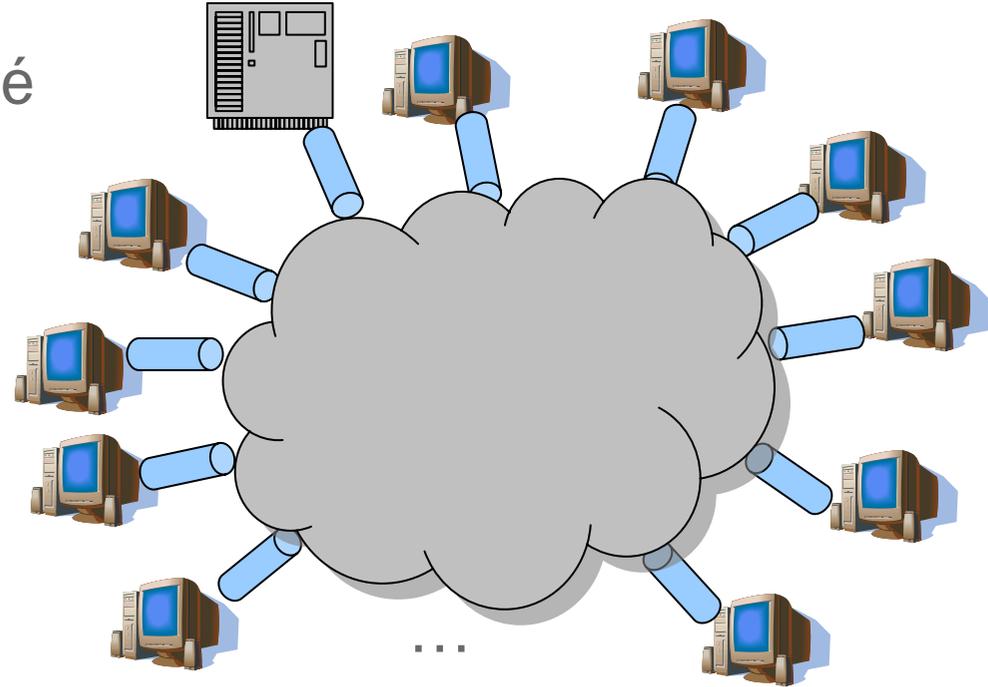
Pourquoi une approche épidémique

- **La centralisation est inopérante :**
 - Grandes populations d'utilisateurs ($10^3 - 10^6$)
 - Déconnexions et reconnexions fréquentes
 - **La décentralisation épidémique fonctionne :**
 - Protocoles très simples (à bien choisir)
 - Performances quasi-optimales en débit et délai
- **Systemes déjà déployés: tous de type épidémique**

Scénario 1: contraintes d'accès

Ressource limitée: Capacité d'accès

- Modélise contraintes de connexion Internet (ADSL, cable,...)
- Normalisée:
1 paquet / seconde

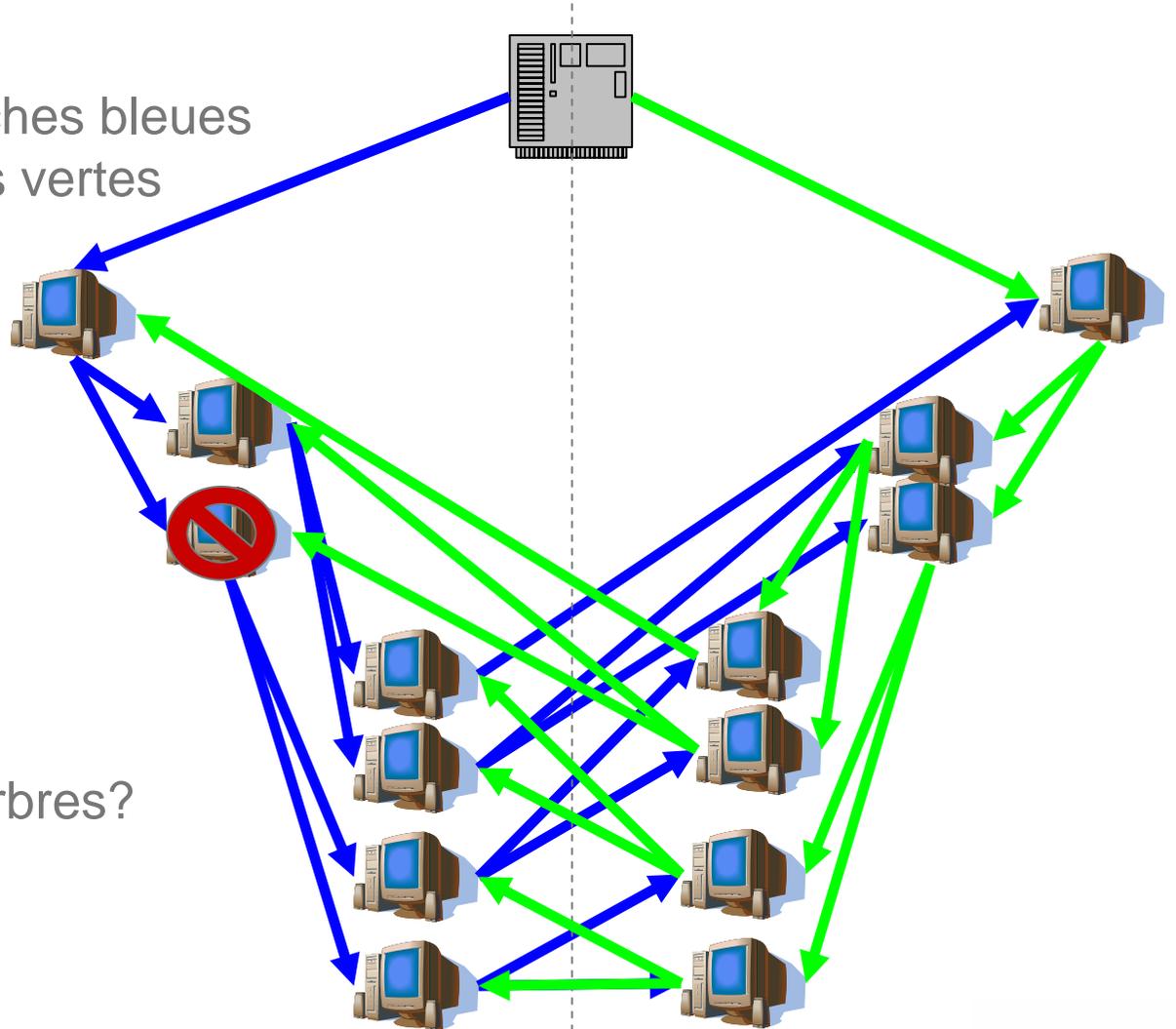


Performance de diffusion optimale:

- Débit = 1 (paquet / seconde)
- Délai = $\log_2(N)$ où N: nombre de nœuds

La version centralisée

Paquets pairs:
Envoyés le long des flèches bleues
Paquets impairs: flèches vertes



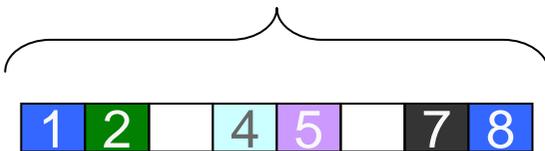
En cas de départ,
comment reconstruire les arbres?

Décentralisation – Approche naïve

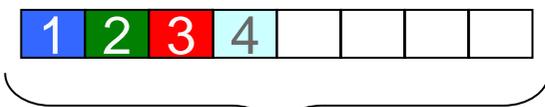
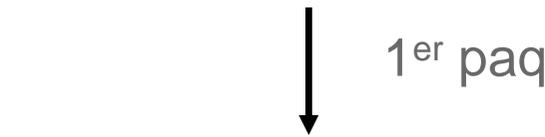
Fraction de nœuds atteints

- Choix de cible au hasard
- Choix du « premier paquet utile
- ➔ Privilégier les intérêts du receveur

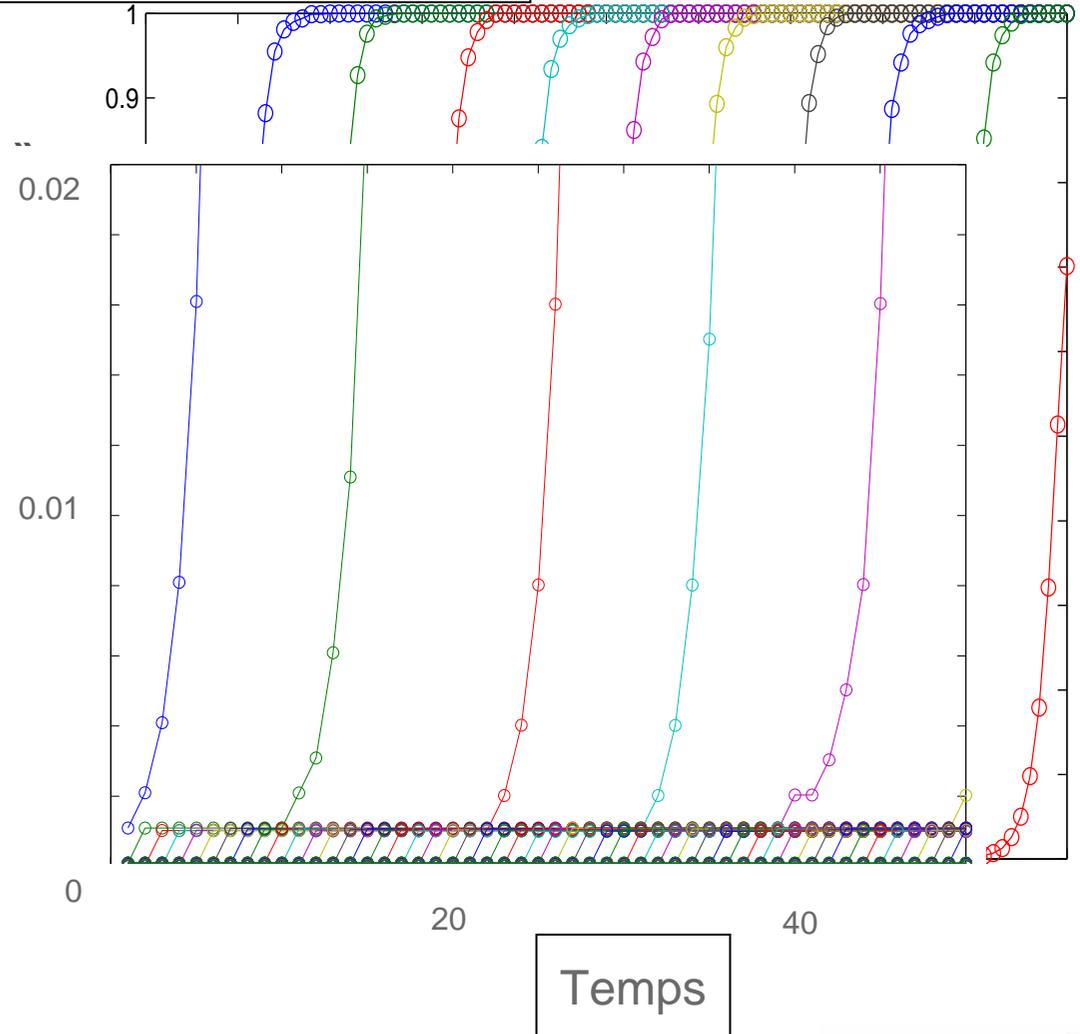
Paquets de l'envoyeur



1^{er} paquet utile



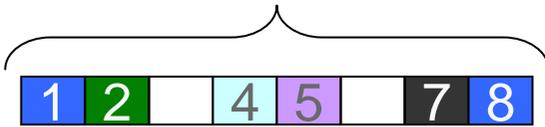
Paquets du receveur



Mieux: cible au hasard / dernier paquet

Fraction de nœuds atteints

Paquets de l'envoyeur

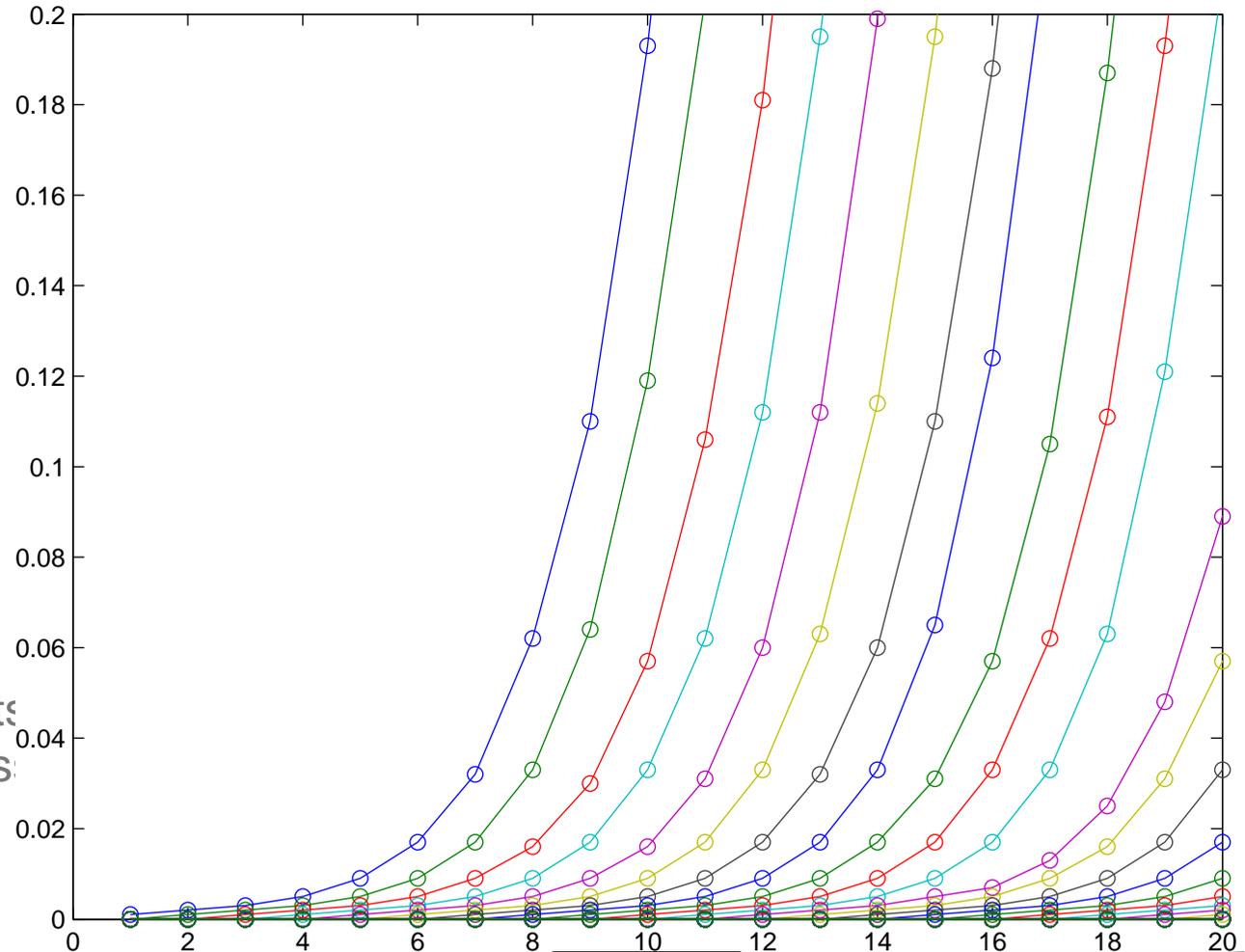


Dernier paquet ↓



Paquets du receveur

Privilégie les bénéfices indirects
Potentiel de nouvelles transmises



Temps

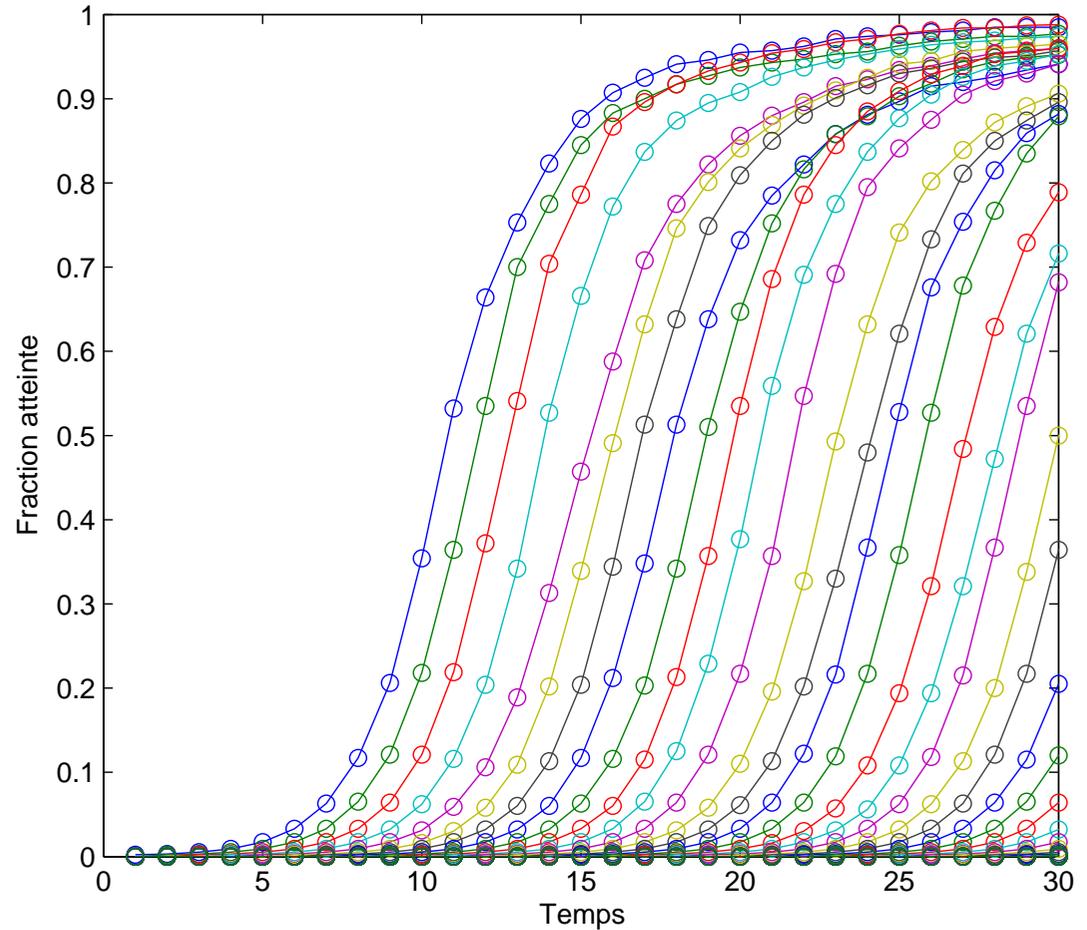
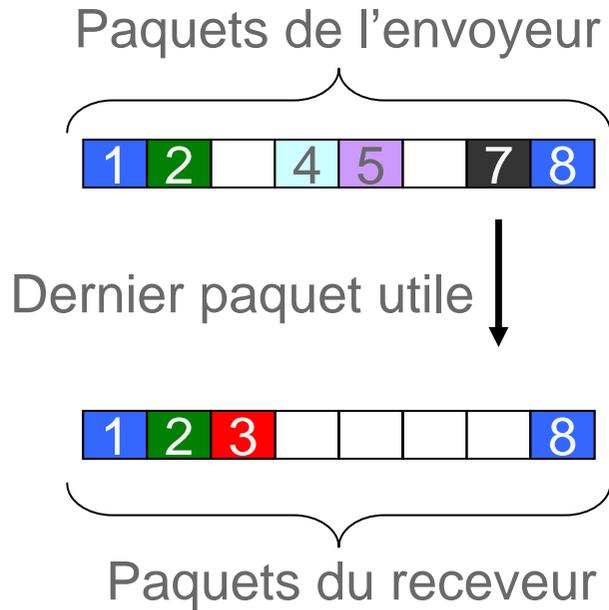
Mieux: cible au hasard / dernier paquet

chaque nœud reçoit $1-1/e \approx 63\%$ des paquets
en temps optimal (logarithmique)

→ Diffusion à débit de 63% de l'optimal, avec délai optimal,
est faisable

(en formant par codage à la source des paquets
supplémentaires, contenant de l'information redondante)

Encore mieux: cible au hasard / dernier paquet utile



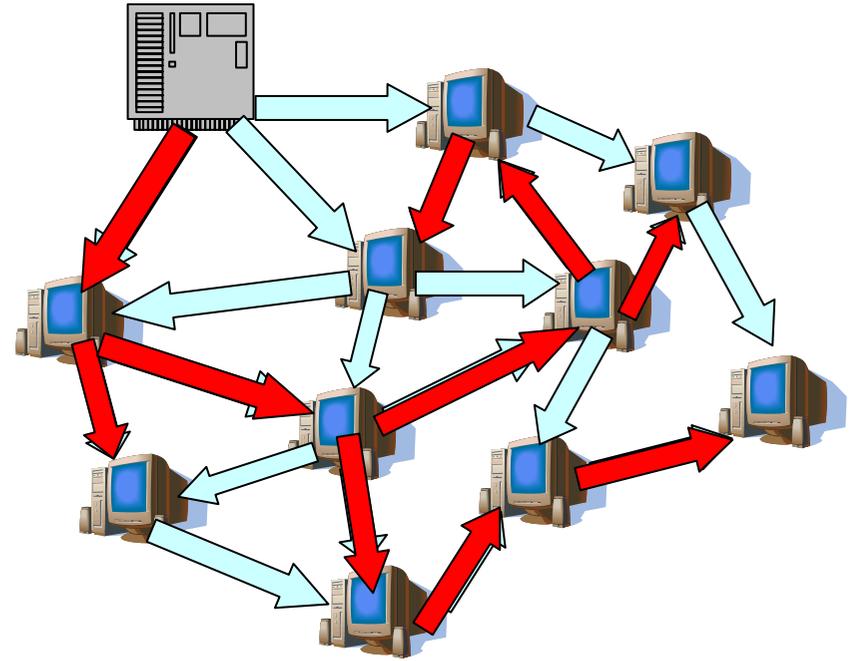
Encore mieux: cible au hasard / dernier paquet utile

Pour un taux d'injection $\lambda < 1$, pour tout $x > 0$, chaque nœud reçoit une fraction $1 - 1/x$ des paquets en temps $\log_2(N) + O(x)$.

En français: Diffusion à débit arbitrairement proche de l'optimal, avec délai optimal

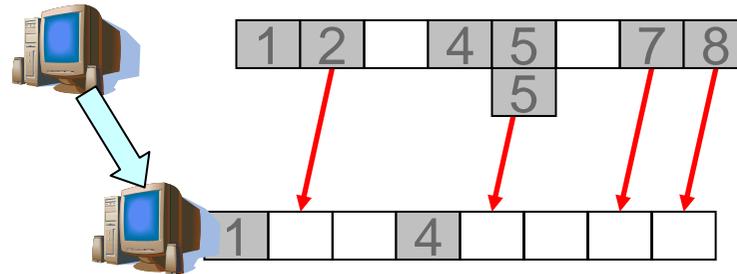
Scénario 2: contraintes de réseau

- Graphe entre les nœuds
- Capacités associées aux arcs



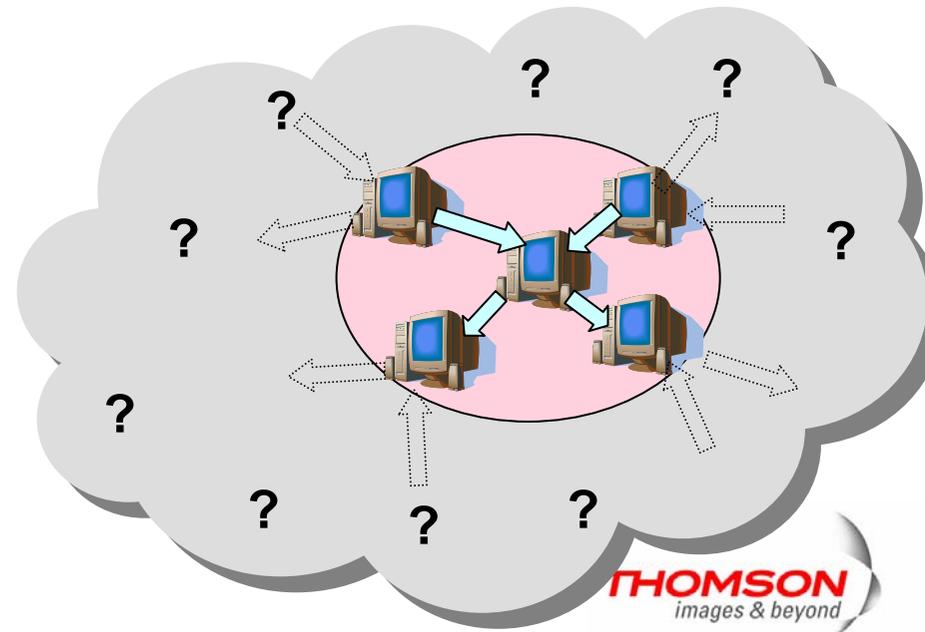
Débit de diffusion [Edmonds, 73]:
nombre maximal d'arbres couvrants arc-disjoints
qu'on peut inclure dans le graphe

Politique « paquet utile au hasard » et théorème d'Edmonds



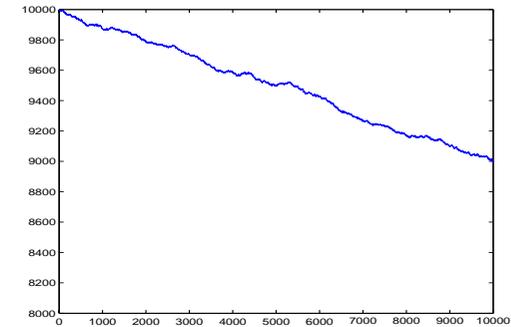
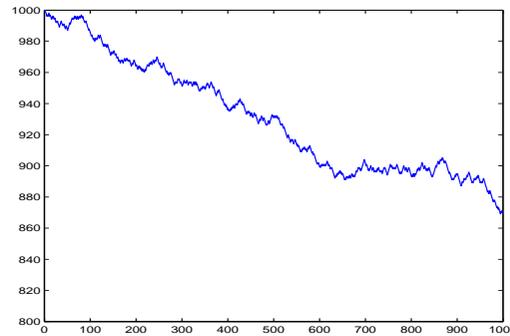
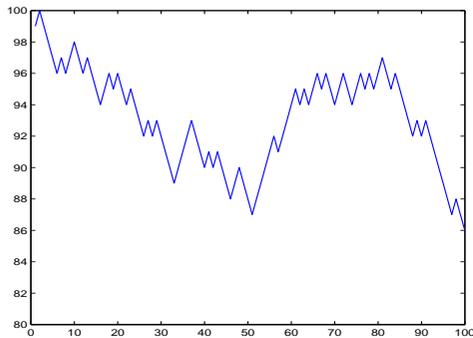
Cette politique permet un débit de diffusion optimal,
pour un graphe arbitraire

- Fondée sur des informations locales
- Ne requiert pas la construction d'arbres couvrants



Comment ça se prouve :

Comportement du système après renormalisation en temps et en espace :



→ Evolution régie par des équations différentielles (plus d'aléa)

On montre que les solutions convergent vers un point d'équilibre (avec une fonction de Lyapunov adaptée)

Corollaire: Une preuve *analytique* du résultat *combinatoire* d'Edmonds

Plan

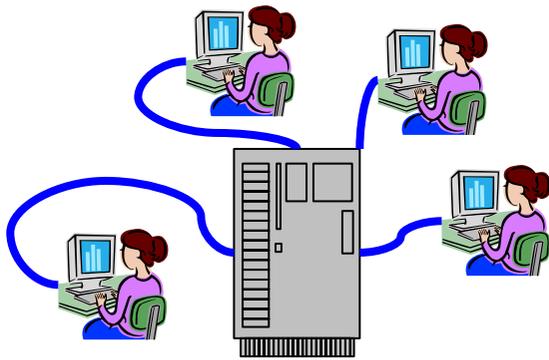
- Le pair-à-pair aujourd'hui
- Exemple de diffusion épidémique: un virus informatique
- Epidémies et diffusion de flux d'information
- **Conclusions et Perspectives**

Conclusions

- **Diffusion épidémique allie**
 - simplicité d'implémentation
 - Efficacité dans l'utilisation des ressources
- **Optimum global réalisable sur la base de décisions**
 - Aléatoires
 - Décentralisées
- ➔ **Comme en physique statistique, sauf qu'on peut choisir à volonté les règles d'interaction locale !**

Perspectives

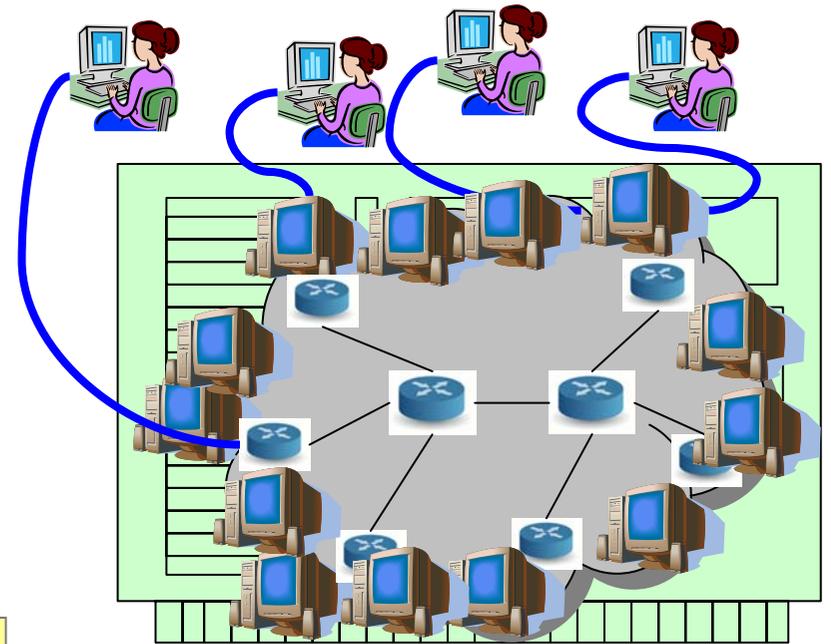
- Recherche de mécanismes efficaces pour :
 - La vidéo à la demande [ressources : **communication et mémoire**]
 - Les calculs distribués [ressource : **capacité des processeurs**]



L'ère du super-ordinateur



L'ère du « Personal Computer »



L'ère de l' « Internet Computer »