La Dissémination De Données

TIPE

Plan

- Présentation de la dissémination de données
- Résultats théoriques
- Présentation des algorithmes d'ordonnancement
- Résultats pratiques: un programme simulant un serveur Internet

Présentation de la dissémination de données

Intérêt de la dissémination?

- Environnements asymétriques
- Intérêt pour Internet:

 -Diminuer la bande passante

 -Alléger le serveur

 -SURTOUT: diminuer le temps d'attente des clients

Fonctionnement

- Approche habituelle:
  - -Traitement au cas par cas
  - -On sert le premier client de la liste d'attente
- Dissémination de données:
  - -Diffusion de l'information dans un canal
  - -On ne se préoccupe pas des clients actuellement connectés

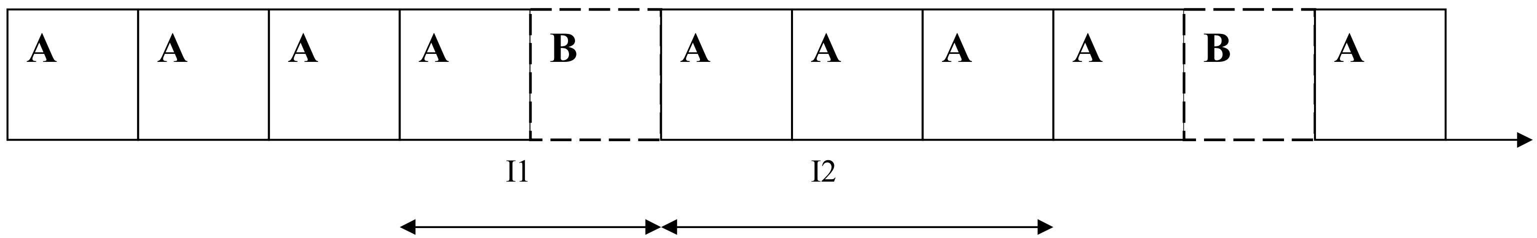
Un exemple pour comprendre

On a deux messages A et B à diffuser.

A est quatre fois plus populaire que B

Comment minimiser le temps d’attente des clients?

Diffuser A quatre fois plus que B

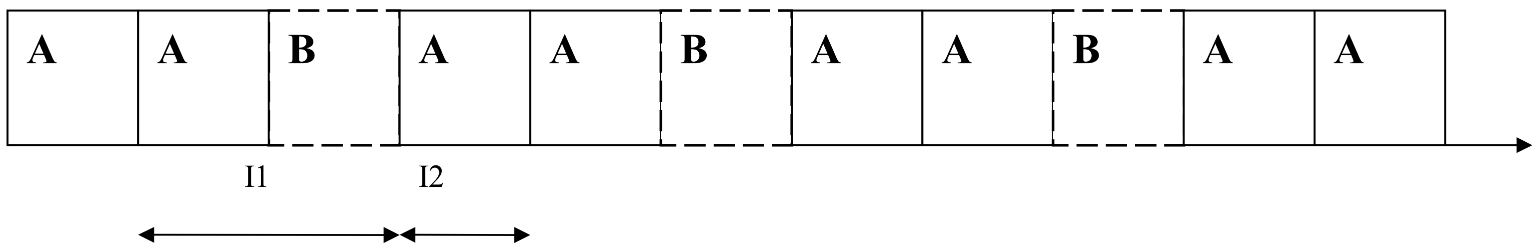


TS(B)=1+5/2=3.5

TS(A)=1+(3/5*0.5)+(2/5*1)=1.7

Temps d’attente moyen: 2.06 secondes

Diffuser A deux fois plus que B



TS(B)=1+3/2=2.5

TS(A)=1+(2/3*1)+(1/3*0.5)=1.83

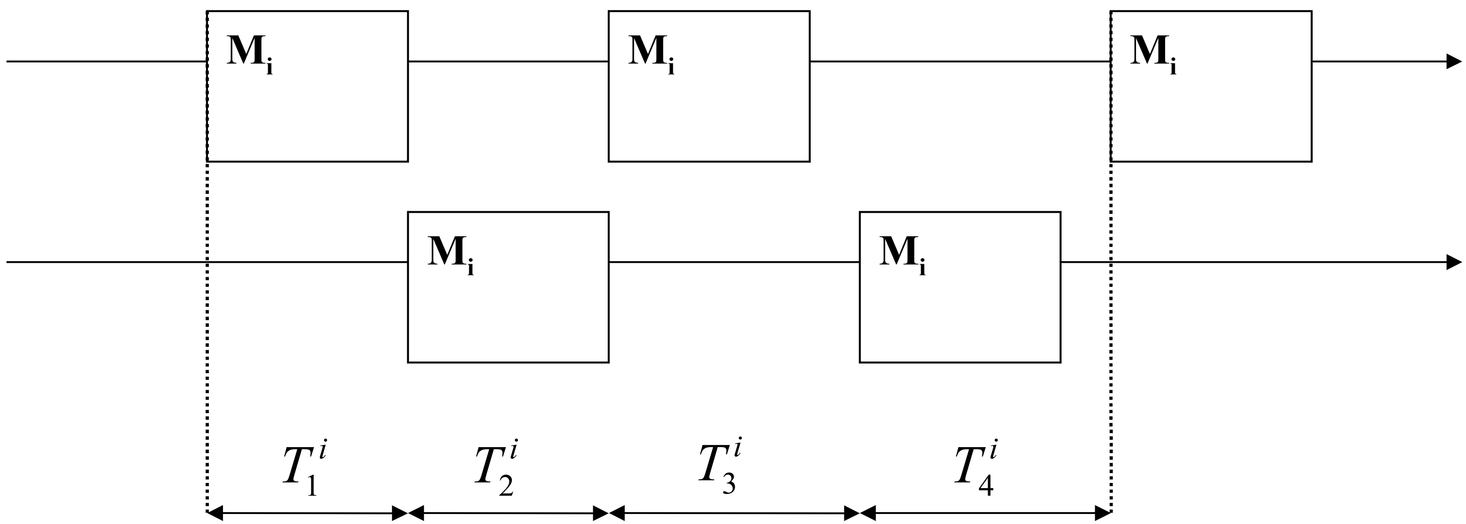
Temps d'attente moyen: 1.97 secondes

Résultats théoriques

Formalisation du problème

- m messages M_1, M_m, de mêmes longueurs à diffuser
- On dispose de W canaux de diffusion.
- Chaque message M_i a une popularité de p_i.

Définition d'un ordonnancement périodique



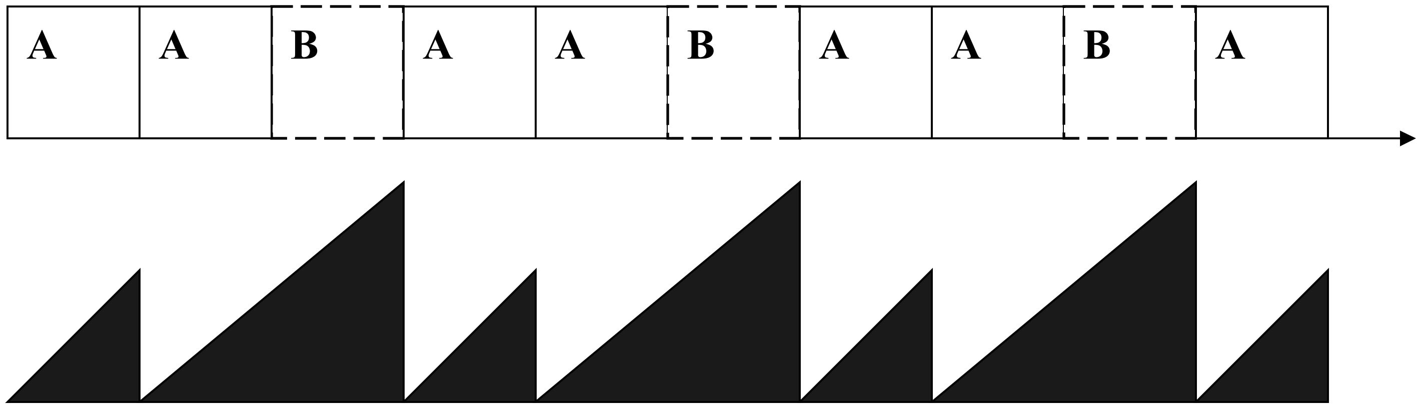
- n_i est le nombre de diffusions de M_i dans la période.
- T_j^i est l'intervalle de temps qui sépare les débuts de la $(j\text{-}1)^{\text{ième}}$ et la $j^{\text{iéme}}$ diffusion de M_i.

Calcul du temps d'attente moyen

Le temps d'attente moyen TS d'un ordonnancement de période T est

$$1 + \frac{1}{2T} \sum_{i=1}^{m} p_i \sum_{j=1}^{n_i} (T_j^i)^2$$

Représentation graphique du temps d’attente



Un minorant du temps de service

Soit

$$Min1 = \min(1 + \sum_{i=1}^{m} \left(\frac{p_i \tau_i}{2} \right)) \quad \textit{sur le domaine} \begin{cases} \forall i \ \tau_i \geq 1 \\ \sum_{i=1}^{m} \frac{1}{\tau_i} \leq W \end{cases}$$

- $Min1$ est un minorant de tout ordonnancement périodique
- Il est atteint pour un unique τ^*.
- Un ordonnancement périodique S a un coût égal à ce minorant ssi il diffuse M_i tous les τ_i^*.

Calcul du minorant

- Le minorant est atteint pour $\tau_i' = \dfrac{K}{\sqrt{p_i}}$

Où $K = \sum_{i=1}^{m} \dfrac{\sqrt{p_i}}{W}$

- Le minorant vaut donc $1 + \left(\sum_{i=1}^{m} \sqrt{p_i}\right)^2 / 2W$

Complexité

- Le problème de l'ordonnancement de m messages avec coûts de diffusion sur W canaux est NP-difficile (réduction du problème de coloriage des graphes).
- Le problème reste ouvert si on enlève les coûts de diffusion

Présentation des algorithmes

Algorithme aléatoire

- On diffuse le message M_i avec une probabilité proportionnelle à $1/\sqrt{p_i}$
- Le temps d'attente est de

$$1/2+\left(\sum_{i=1}^{m}\sqrt{p_i}\right)^2/W$$

- On obtient donc une 2-approximation

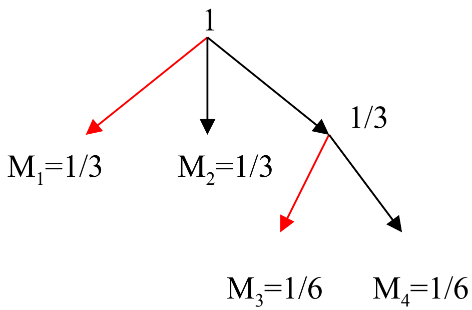
Algorithme glouton

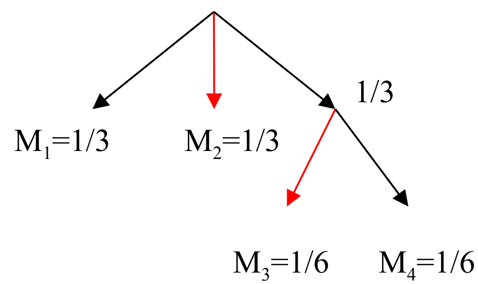
- Soit σ_i le temps qui s'est écoulé depuis le début de la dernière diffusion du message M_i
- Le glouton diffuse le message M_i tel que $p_i \sigma_i \tau_i^*$ est maximal
- Donne toujours de meilleurs résultats que l'algorithme aléatoire

Algorithme basé sur le nombre d’or

- $\varphi = \frac{1+\sqrt{5}}{2}$ est l’irrationnel dont les multiples modulo 1 découpent [0,1] le plus uniformément possible
- Ceci permet de construire un ordonnancement respectant les fréquences de diffusion en $1/W\tau_i^*$
- C’est une 9/8-approximation

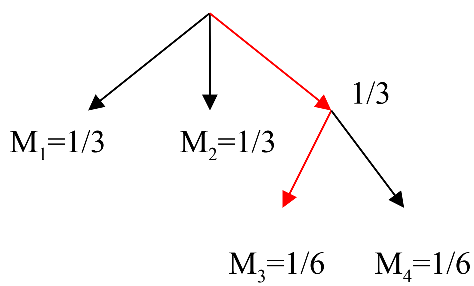
Algorithme basé sur un arbre

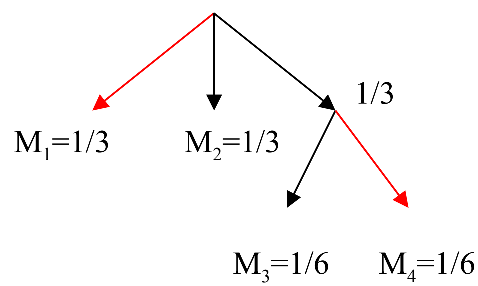




1) On diffuse M_1

2) On diffuse M_2





3) On diffuse M_3

4) On diffuse M_1…

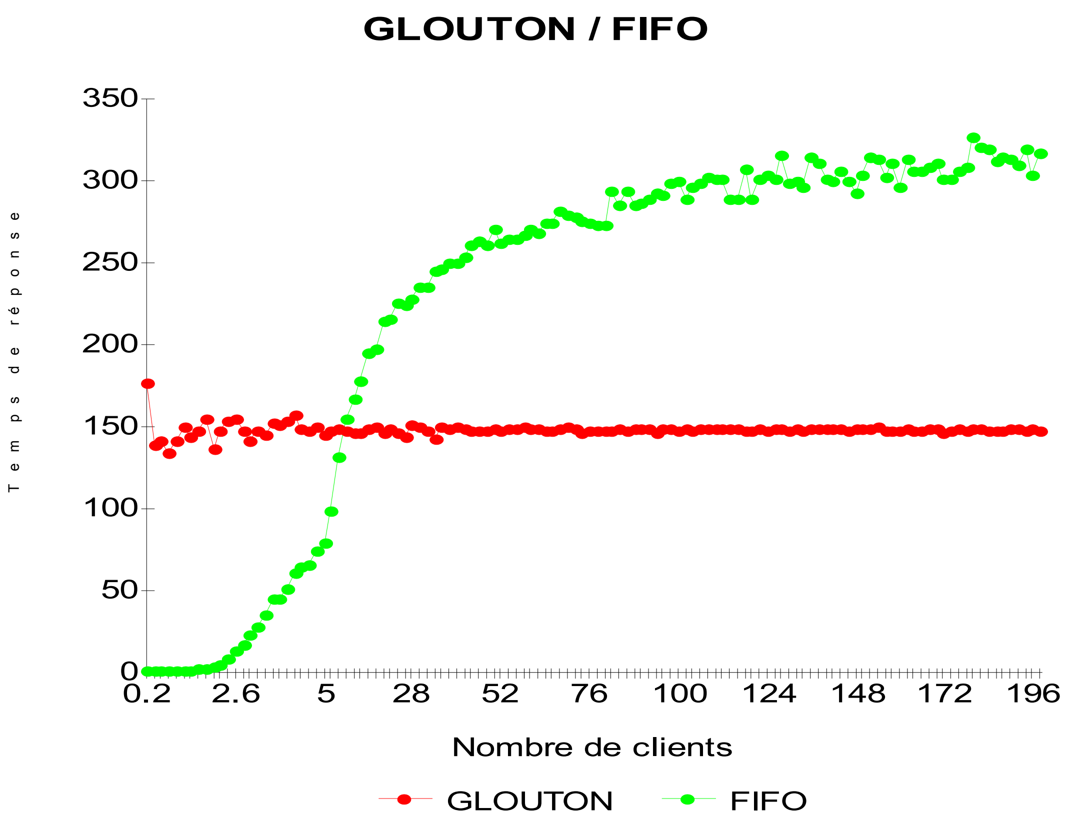
Résultats expérimentaux

Un programme simulant un serveur utilisant les algorithmes de dissémination

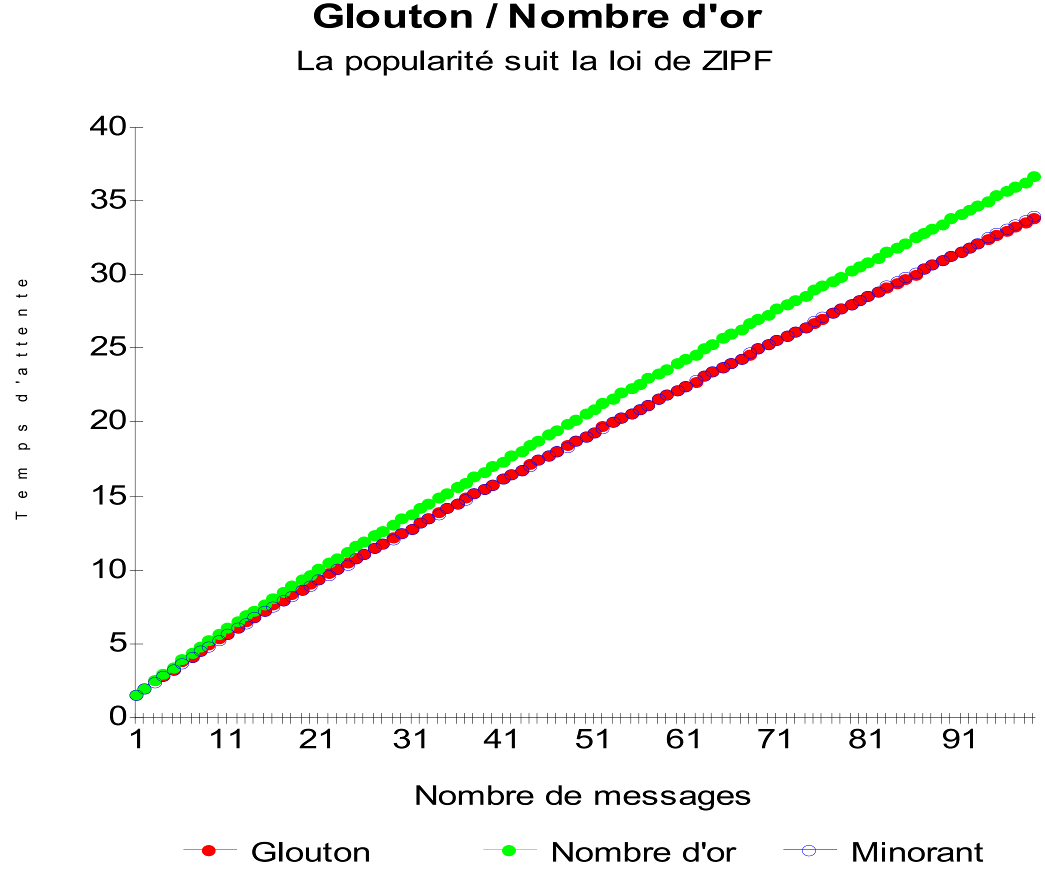
Description du programme

- Calcul du minorant et des fréquences idéales
- Implémentation des algorithmes précédents
- Simulation de clients se connectant à un serveur diffusant les messages suivant un des algorithmes précédents

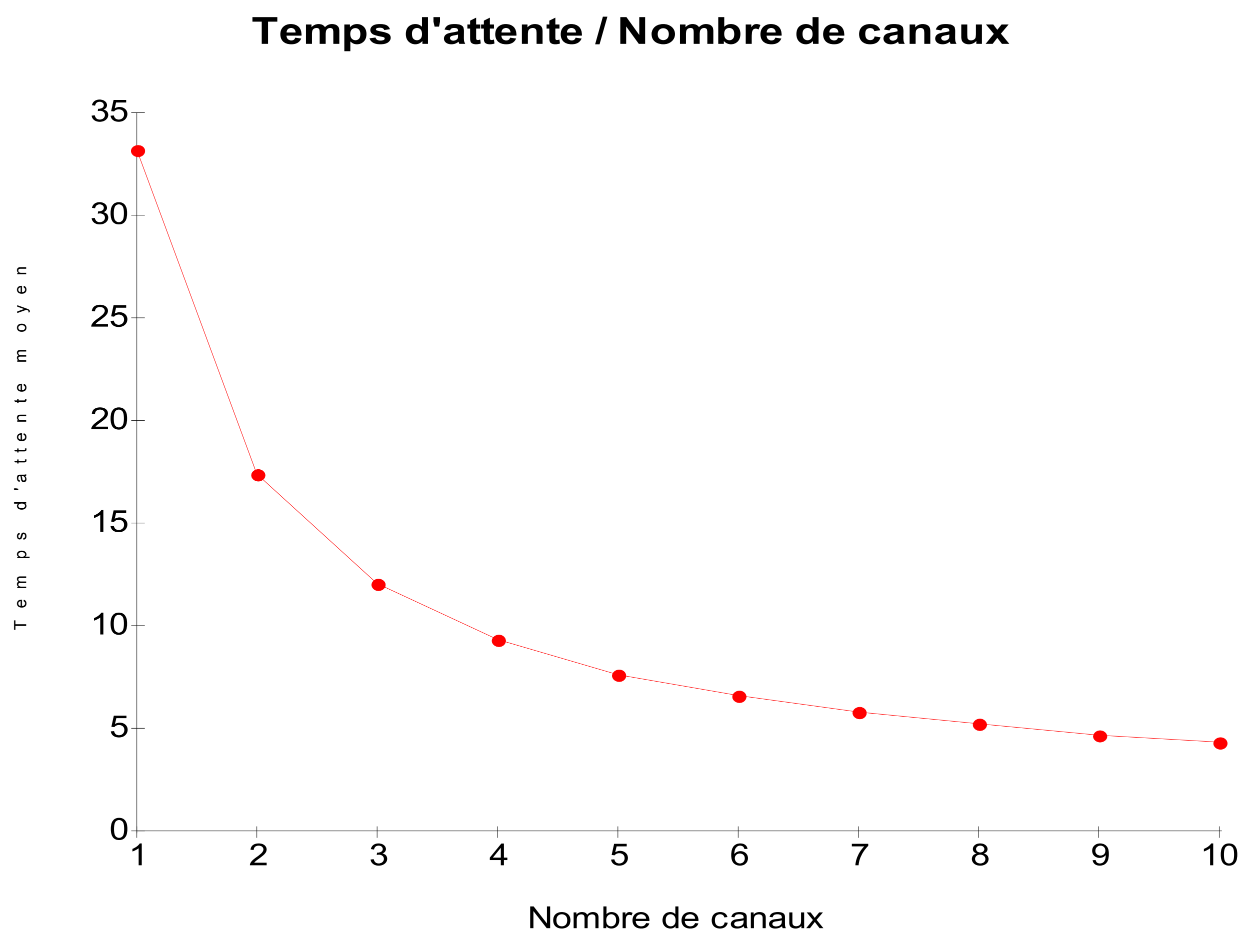
Comparaison: méthode habituelle contre dissémination



Comparaison des algorithmes de dissémination entre eux



Variation du temps d'attente en fonction du nombre de canaux



Conclusion

Résumé du travail

- Etude théorique partielle
- Validation expérimentale, avantages de la dissémination de données:
    - Diminuer la bande passante
    - Alléger le serveur
    - Diminuer le temps de service

Perspectives-Mise en place pratique

- Diffuser les messages les plus populaires par ce biais, les autres de la manière habituelle
- Problèmes de mise en place:

  -Comment créer un tel canal de diffusion?

  -Comment connaître la popularité des messages
- La solution: un programme dynamique