

Construction distribuée de groupes pour le calcul à grande échelle

O. Beaumont, N. Bonichon, P. Duchon, L. Eyraud-Dubois, H. Larchevêque

Plan

- 1 Les plates formes à grande échelle
- 2 Bin Covering avec diamètre contraint
 - Bin Covering
 - Introduction de la contrainte de distance
- 3 Premiers Résultats
- 4 Extension du facteur d'approximation aux dimensions supérieures
 - Facteur d'approximation avec augmentation de ressources
 - Algorithme glouton et preuve du facteur d'approximation
 - Version économe
 - Contre exemple
- 5 Perspectives

Les plates formes à grande échelle

- Exploitation de ressources disponibles chez les utilisateurs
- Ressources reliées entre elles via Internet
- Résolution de problèmes trivialement parallélisables (seti@home, genome@home, folding@home, ...)

Fonctionnement

- Chaque utilisateur est un client
- Un serveur central distribue de petites tâches indépendantes sur les ressources disponibles
- Récupère et traite les résultats

Nouvelles perspectives

- Décentralisé : suppression du serveur
- Tâches indépendantes plus grosses : intraitable par un hôte seul
- Regroupement de plusieurs hôtes pour travail en commun

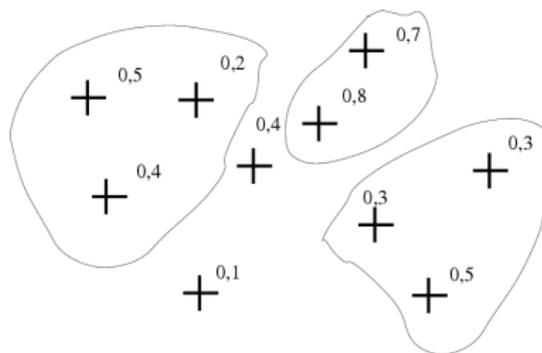
Regroupement des hôtes

- Basée sur une notion de proximité "géographique"
- Notion de distance basée sur les temps de communications

Problèmes :

- Espace à plusieurs dimensions ¹
- Distance entre deux points varie au cours du temps
- Nécessité d'une structure qui nous permette de connaître nos voisins

On se ramène à un problème de *bin covering*.



¹F. Dabek *et al.*, Vivaldi : A Decentralized Network Coordinate System, 2004

Plan

- 1 Les plates formes à grande échelle
- 2 Bin Covering avec diamètre contraint
 - Bin Covering
 - Introduction de la contrainte de distance
- 3 Premiers Résultats
- 4 Extension du facteur d'approximation aux dimensions supérieures
 - Facteur d'approximation avec augmentation de ressources
 - Algorithme glouton et preuve du facteur d'approximation
 - Version économe
 - Contre exemple
- 5 Perspectives

Plan

- 1 Les plates formes à grande échelle
- 2 Bin Covering avec diamètre contraint
 - Bin Covering
 - Introduction de la contrainte de distance
- 3 Premiers Résultats
- 4 Extension du facteur d'approximation aux dimensions supérieures
 - Facteur d'approximation avec augmentation de ressources
 - Algorithme glouton et preuve du facteur d'approximation
 - Version économe
 - Contre exemple
- 5 Perspectives

Qu'est ce que c'est ?

- Un ensemble d'éléments
- Chacun possède un poids

Le but :

- Les assembler en groupes de poids total dépassant une certaine valeur T
- Maximiser le nombre de groupes créés à partir d'une liste donnée

On normalise les poids par rapport à T , pour que la valeur à dépasser soit 1

NP-Complétude et approximations

- Problème NP-Complet
- Algorithmes d'approximation : obtenir un nombre de sacs le plus proche possible de l'optimal
- En centralisé : $2/3$, $3/4$ ²
- *PTAAS*, "Polynomial Time **Asymptotic** Approximation Scheme" ³

²S.F. Assmann *et al.*, On a dual version of the one-dimensional bin packing problem, 1984

³J. Csirik *et al.*, Better approximation algorithms for bin covering 

Plan

- 1 Les plates formes à grande échelle
- 2 Bin Covering avec diamètre contraint
 - Bin Covering
 - Introduction de la contrainte de distance
- 3 Premiers Résultats
- 4 Extension du facteur d'approximation aux dimensions supérieures
 - Facteur d'approximation avec augmentation de ressources
 - Algorithme glouton et preuve du facteur d'approximation
 - Version économe
 - Contre exemple
- 5 Perspectives

Qu'est ce que c'est ?

- Reprend le principe du *bin covering* : ensemble d'éléments pondérés à emballer
- Chaque hôte hébergeant possède des coordonnées
- Contraintes sur la formation des groupes :
 - Poids supérieur à 1
 - Diamètre inférieur à d_{max}

Définition (Diamètre d'un groupe)

Plus grande distance entre deux membres d'un même groupe

Cadre de notre étude

Définition (Convexité d'un groupe)

Groupe convexe \Leftrightarrow Aucun élément non membre dans l'enveloppe convexe de ce groupe

- Problème jamais étudié
- Premiers résultats en dimension 1, création de groupes convexes
- En dimension 1, groupe convexe \Leftrightarrow Intervalle

Plan

- 1 Les plates formes à grande échelle
- 2 Bin Covering avec diamètre contraint
 - Bin Covering
 - Introduction de la contrainte de distance
- 3 Premiers Résultats
- 4 Extension du facteur d'approximation aux dimensions supérieures
 - Facteur d'approximation avec augmentation de ressources
 - Algorithme glouton et preuve du facteur d'approximation
 - Version économe
 - Contre exemple
- 5 Perspectives

Théorème

Toute solution créant des groupes ayant les trois propriétés suivantes est une solution à un facteur de $1/3$ par rapport à l'optimal :

Propriété (P1)

Chaque groupe est convexe et de poids supérieur ou égal à 1

Propriété (P2)

Chaque groupe a un poids strictement inférieur à 2

Propriété (P3)

Il n'existe pas, entre une paire de groupes convexes, d'intervalle "constructibles", de taille inférieure à d_{max} , dans lequel la somme des poids est supérieure ou égale à 1

Résultat

Théorème

On a un algorithme avec facteur d'approximation $1/3$ qui permet de faire du bin covering avec diamètre contraint en dimension 1, de manière distribuée en $O(\log^2 n)$ étapes

Plan

- 1 Les plates formes à grande échelle
- 2 Bin Covering avec diamètre contraint
 - Bin Covering
 - Introduction de la contrainte de distance
- 3 Premiers Résultats
- 4 Extension du facteur d'approximation aux dimensions supérieures
 - Facteur d'approximation avec augmentation de ressources
 - Algorithme glouton et preuve du facteur d'approximation
 - Version économe
 - Contre exemple
- 5 Perspectives

Plan

- 1 Les plates formes à grande échelle
- 2 Bin Covering avec diamètre contraint
 - Bin Covering
 - Introduction de la contrainte de distance
- 3 Premiers Résultats
- 4 Extension du facteur d'approximation aux dimensions supérieures
 - Facteur d'approximation avec augmentation de ressources
 - Algorithme glouton et preuve du facteur d'approximation
 - Version économe
 - Contre exemple
- 5 Perspectives

Principe

- Trouver une solution qui s'autorise à former des groupes de diamètre plus grand que d_{max}
- Le comparer à l'optimal contraint à d_{max}

On choisit, en dimension d , d'autoriser notre groupe à avoir un diamètre de $3d_{max}$

- On utilise un algorithme glouton en deux phases
- Cet algorithme glouton construit des groupes de poids inférieurs à 2

Impact sur le facteur d'approximation

Théorème

Il est possible d'obtenir un facteur d'approximation avec diamètre contraint à $3d_{max}$ de $1/3$ par rapport à l'optimal contraint à d_{max} en utilisant un algorithme glouton à plusieurs phases vérifiant les propriétés suivantes

Propriété (P1)

Chaque groupe a un poids supérieur ou égal à 1

Propriété (P2)

Chaque groupe a un poids strictement inférieur à 2

Propriété (P3)

Il n'existe pas, dans l'espace privés des groupes créés, d'ensemble de points "constructibles", de diamètre inférieur à $3d_{max}$ et dans lequel la somme des poids est supérieure ou égale à 1

Plan

- 1 Les plates formes à grande échelle
- 2 Bin Covering avec diamètre contraint
 - Bin Covering
 - Introduction de la contrainte de distance
- 3 Premiers Résultats
- 4 Extension du facteur d'approximation aux dimensions supérieures
 - Facteur d'approximation avec augmentation de ressources
 - Algorithme glouton et preuve du facteur d'approximation
 - Version économe
 - Contre exemple
- 5 Perspectives

L'algorithme glouton

- Phase 1 :** Tant qu'il existe un groupe constructible, de diamètre $d \leq d_{max}$ et de poids P avec $1 \leq P < 2$, le construire. Oter ses éléments de l'espace et continuer. On crée ainsi n_1 groupes en phase $P1$
- Phase 2 :** Même chose avec les éléments restant en créant des groupes de diamètre $d \leq 3d_{max}$. n_2 groupes sont créés en phase $P2$

Quelques définitions

Rappels :

Définition (Diamètre d'un groupe)

Plus grande distance entre deux membres d'un même groupe

Définition (Element groupable)

Un élément est dit groupable si il appartient à un groupe constructible, de poids $P \geq 1$ et de diamètre $d \leq d_{max}$

Définition (Zone étendue)

La zone étendue d'un groupe G est l'ensemble des points y tels que $d(G; y) \leq d_{max}$ (comprenant donc les points de G)

Lemmes de base

Lemme (Diamètre d'une zone étendue)

Le diamètre de la zone étendue d'un groupe créé en phase P1 est inférieure ou égale à $3d_{max}$

Lemme (Lemme 2)

Tout groupe formé par une solution optimale contrainte à d_{max} est totalement inclus dans la zone étendue d'un groupe formé par l'algorithme glouton en phase P1

Idée de la preuve

On compte le nombre de groupes créés optimalement : les groupes à former étant de poids $P \geq 1$, cela revient à compter le poids total des éléments groupables :

$$\text{card}(\text{Opt}(d_{\max})) \leq \sum_{e \in \text{elem groupables}} w(e)$$

Idée de la preuve

On a donc :

$$\text{card}(\text{Opt}(d_{\max})) \leq \sum_{e \in \text{elem groupables}} w(e)$$

$$\text{card}(\text{Opt}(d_{\max})) \leq$$

$$\sum_{g1 \text{ construit en } P1} w(g1) + \sum_{g2 \text{ construit en } P2} w(g2) + w_{\text{perdu}}$$

$$\text{card}(\text{Opt}(d_{\max})) < 2n_1 + 2n_2 + n_1$$

$$\text{card}(\text{Opt}(d_{\max})) < 3(n_1 + n_2)$$

Donc :

$$n_1 + n_2 > 1/3 \text{card}(\text{Opt}(d_{\max}))$$

Plan

- 1 Les plates formes à grande échelle
- 2 Bin Covering avec diamètre contraint
 - Bin Covering
 - Introduction de la contrainte de distance
- 3 Premiers Résultats
- 4 Extension du facteur d'approximation aux dimensions supérieures
 - Facteur d'approximation avec augmentation de ressources
 - Algorithme glouton et preuve du facteur d'approximation
 - **Version économe**
 - Contre exemple
- 5 Perspectives

Utilisation d'un glouton économe

Le glouton économe fonctionne de la même façon que le glouton classique, en deux phases, mais la première phase est plus précise :

Phase 1 : Tant qu'il existe un groupe constructible, de diamètre $d \leq d_{max}$ et de poids P avec $P \geq 1$, le construire.

En ôter ensuite tous les éléments qui ne sont pas nécessaires à la validité du groupe.

Enfin, ôter les éléments de ce groupe de l'espace et continuer. On crée ainsi n_1 groupes en phase $P1$

Phase 2 : Même chose avec les éléments restant en créant des groupes de diamètre $d \leq 3d_{max}$. n_2 groupes sont créés en phase $P2$

Conséquences d'une telle économie

En agissant ainsi, on est sûr d'avoir le lemme principal suivant :

Lemme

Si un groupe de phase 1 a pour poids $1 + w$, $0 \leq w < 1$, alors tout élément de ce groupe a un poids strictement supérieur à w , et ce groupe a un nombre d'éléments $e < 1 + 1/w$.

Et d'avoir celui-ci :

Lemme

Si un groupe de phase 1 a N éléments alors $w(\text{groupe}) < 1 + 1/(N - 1)$.

Notamment si $N = 3$, $w(\text{groupe}) < 3/2$. Avec cette propriété, on prouve un facteur d'approximation de $2/5$.

Plan

- 1 Les plates formes à grande échelle
- 2 Bin Covering avec diamètre contraint
 - Bin Covering
 - Introduction de la contrainte de distance
- 3 Premiers Résultats
- 4 Extension du facteur d'approximation aux dimensions supérieures
 - Facteur d'approximation avec augmentation de ressources
 - Algorithme glouton et preuve du facteur d'approximation
 - Version économe
 - **Contre exemple**
- 5 Perspectives

Contre exemple du facteur $1/2$

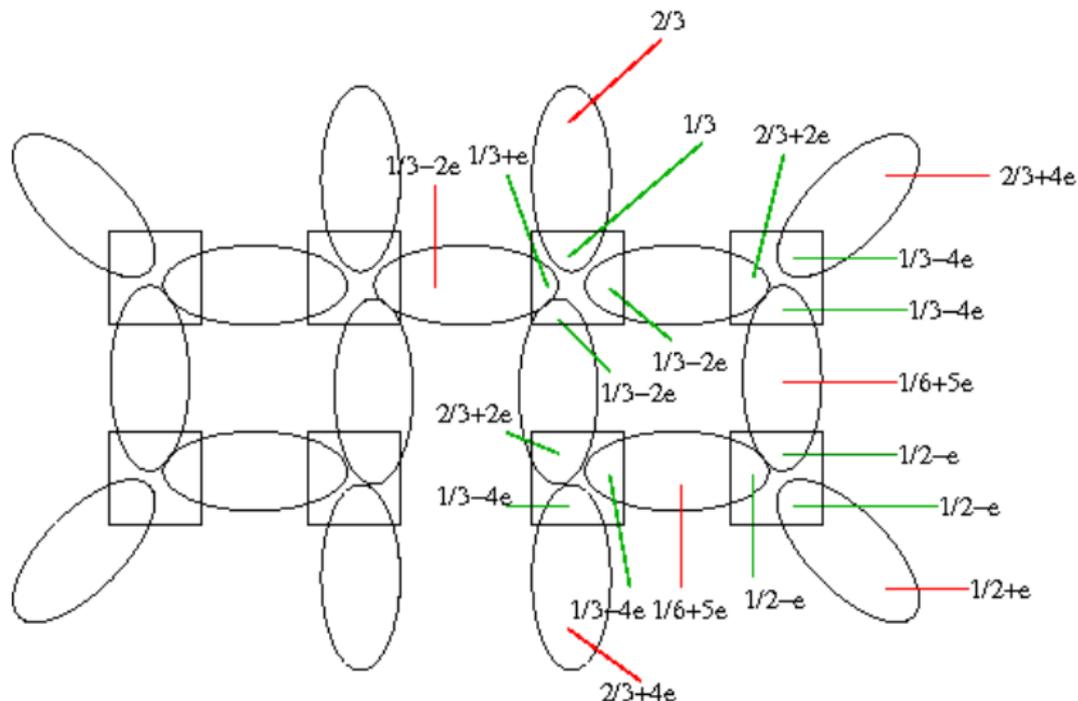


Fig.: $e > 0$. L'optimal fait 17 groupes (ovales), le glouton, 8 (carrés)

Plan

- 1 Les plates formes à grande échelle
- 2 Bin Covering avec diamètre contraint
 - Bin Covering
 - Introduction de la contrainte de distance
- 3 Premiers Résultats
- 4 Extension du facteur d'approximation aux dimensions supérieures
 - Facteur d'approximation avec augmentation de ressources
 - Algorithme glouton et preuve du facteur d'approximation
 - Version économe
 - Contre exemple
- 5 Perspectives

Perspectives

- On a un encadrement du facteur d'approximation F avec augmentation de ressources : $2/5 < F < 8/17$
- Extension de nos résultats au *bin packing* contraint
- Algorithmes distribués traduisant nos résultats
- Dynamicit  des h tes
- Dynamicit  des poids des h tes : probl me inh rent aux plates formes   grande  chelle. Complic    g rer m me en une dimension