RAIRO. RECHERCHE OPÉRATIONNELLE

A. BILLIONNET
S. ELLOUMI

Placement de tâches dans un système distribué et dualité lagrangienne

RAIRO. Recherche opérationnelle, tome 26, n° 1 (1992), p. 83-97
<http://www.numdam.org/item?id=RO_1992__26_1_83_0>

Recherche opérationnelle/Operations Research
(vol. 26, n° 1, 1992, p. 83 à 97)

PLACEMENT DE TÂCHES DANS UN SYSTÈME DISTRIBUÉ ET DUALITÉ LAGRANGIENNE (*)

par A. BILLIONNET ([1]) et S. ELLOUMI ([1])

Résumé. — *On considère un ensemble de tâches devant s'exécuter sur un réseau totalement maillé de processeurs hétérogènes disposant de ressources limitées. Un placement des tâches sur les processeurs doit tenir compte des contraintes de ressources et engendre un coût égal à la somme des coûts d'exécution des tâches sur les processeurs auxquels elles sont affectées et des coûts de communication entre les tâches. Nous formulons, dans cet article, le problème du placement optimal comme la minimisation d'une fonction pseudobooléenne quadratique sous des contraintes linéaires puis nous étudions en détail le dual lagrangien de ce dernier problème dans le but de fournir des indications sur la qualité de la borne inférieure du primal obtenue par cette technique.*

Mots clés : Systèmes distribués; placement de tâches; optimisation quadratique en 0-1; dualité lagrangienne.

Abstract. — *This paper addresses the problem of allocating the tasks of a distributed program to the heterogeneous processors of a distributed system. The processors are linked by a completely meshed network and their capacities are limited. The purpose is to make the best uses of resources in the considered distributed system,* i. e., *for a given program to minimize the sum of processing and communication costs. After formulating the problem as the minimization of a quadratic pseudoboolean function subject to linear constraints, we thoroughly study the langrangean dual of this mathematical programming problem in order to obtain relevant information about the primal lower bound that can be obtained in this way.*

Keywords : Distributed system; Task allocation; Quadratic 0-1 optimization; Lagrangean duality.

1. INTRODUCTION

Un problème important qui se pose dans les systèmes distribués est celui du placement optimal des tâches sur les différents processeurs. Ce problème a été abondamment étudié et il existe de très nombreux modèles (*cf.*, par exemple, [AND,88], [BEN,89], [BIL,89], [BOK,87], [FER,89], [GAB,84],

(*) Reçu en octobre 1990.
([1]) C.E.D.R.I.C., Institut d'Informatique d'Entreprise, 18, allée Jean-Rostand, 91002 Évry.

Recherche opérationnelle/Operations Research, 0399-0559/92/01 83 15/$ 3.50

[LO,84], [MA,82], [MAG,89], [PRI,84], [TOW,86]). Cependant, dans la plupart des références, seules des méthodes heuristiques, sans garanties de performances, sont proposées. En effet, pour la majorité des modèles, on ne dispose pas, à l'heure actuelle, d'algorithmes permettant de déterminer les placements optimaux. Pour une présentation détaillée de ces problèmes de placement, le lecteur pourra se reporter aux deux synthèses citées en référence [AND,88] et [BIL,89].

Nous nous intéressons dans cet article au modèle classique suivant : un ensemble de tâches doit s'exécuter sur un réseau totalement maillé de processeurs hétérogènes disposant de ressources limitées. Une affectation des tâches aux processeurs doit tenir compte des contraintes de ressources et engendre un coût égal à la somme des coûts d'exécution des tâches sur les processeurs auxquels elles sont affectées et des coûts de communication entre les tâches. On cherche à déterminer une affectation de coût minimal, c'est-à-dire un placement des tâches sur les processeurs qui utilise au mieux les ressources du système (processeurs et canaux de transmission). Ce problème est NP-difficile [RAO,79] et, en pratique, on ne sait pas le résoudre dès qu'il y a plus d'une dizaine de tâches à placer sur quelques processeurs.

D'une façon plus formelle, soit $\mathscr{P}=\{p_1, p_2, \ldots, p_P\}$ un ensemble de P processeurs et $\mathscr{T}=\{t_1, t_2, \ldots, t_T\}$ un ensemble de T tâches. Nous disposons des données suivantes :

— q_{tp} : le coût d'exécution de la tâche t quand elle est affectée au processeur p;

— $c_{tt'}$: le coût de communication entre les tâches t et t' quand elles ne sont pas affectées au même processeur. Notons que ces coûts sont souvent représentés par un graphe dit graphe de communication. Les sommets de ce graphe sont les tâches et une arête $[t, t']$ existe ssi $c_{tt'}>0$;

— s_t : la quantité de ressource requise par la tâche t. On pose $\mathscr{S}=\sum_{t=1}^{T} s_t$;

— n_p : la capacité du processeur p, c'est-à-dire la quantité totale de ressource dont il dispose. On supposera qu'il existe au moins un placement possible compte tenu des valeurs des s_t et des n_p.

$x=(x_{tp})_{(t=1,\ldots T;\, p=1,\ldots P)}$ est le vecteur des variables de décision bivalentes, valant 1 si et seulement si la tâche t est affectée au processeur p. On pose $\bar{x}_{tp}=1-x_{tp}$.

En posant

– $l_{tp}=q_{tp}+\sum_{t'=t+1}^{T} c_{tt'}$ pour $t=1, \ldots, T-1; p=1, \ldots, P$;

– $l_{Tp}=q_{Tp}$ pour $p=1, \ldots, P$,

le problème se formule comme le programme non linéaire en variables bivalentes suivant

$$
(\mathrm{P1})\quad \left\{
\begin{array}{ll}
\text{Min } F(x)=\sum_{t=1}^{T}\sum_{p=1}^{P} l_{tp}x_{tp}-\sum_{t=1}^{T-1}\sum_{t'=t+1}^{T}\sum_{p=1}^{P} c_{tt'}x_{tp}x_{t'p} & \\
\text{sous les contraintes :} & \\
\sum_{p=1}^{P} x_{tp}=1; \qquad t=1, \ldots, T & (1)\\
\sum_{t=1}^{T} s_t x_{tp}\leqq n_p; \qquad p=1, \ldots, P & (2)\\
x\in\{0,1\}^{T\times P} & (3)
\end{array}
\right.
$$

F est ici une fonction pseudobooléenne (c'est-à-dire de $\{0,1\}^{T\times P}$ vers $\mathbb{R}$) dont les coefficients des termes quadratiques sont tous négatifs. On sait que la minimisation d'une telle fonction, sans contraintes, revient à calculer un flot maximum dans un graphe bi-parti dit graphe de conflit (*cf.* [RHY,70] et, par exemple, [BIL,85]). Une approche par relaxation lagrangienne apparaît alors comme naturelle puisque les contraintes sont linéaires.

Dans la section 2, nous écrivons le problème dual lagrangien obtenu en relâchant les contraintes (1) et (2) et nous montrons qu'il peut se poser comme un programme linéaire continu qui est en fait la relaxation continue d'une linéarisation classique (en variables bivalentes) du problème primal. Une conséquence de cette propriété est que l'optimum du dual peut se calculer polynomialement.

Dans la section 3, nous montrons que la valeur à l'optimum du dual lagrangien ne peut pas dépasser une certaine combinaison des coûts d'exécution et nous en déduisons qu'il existe des problèmes de placement avec un saut de dualité aussi grand que désiré. Nous montrerons de plus qu'il est possible de construire une famille de problèmes très généraux ayant cette propriété et nous donnons une interprétation de ce résultat.

Dans la section 4, nous démontrons que la valeur optimale du dual est supérieure ou égale à la borne inférieure du primal qu'on obtiendrait en supposant que tous les coûts de communication sont nuls et en relâchant les contraintes d'intégrité sur les variables.

Enfin, dans la conclusion, nous proposons de considérer la détermination d'une solution duale comme une première étape dans le calcul d'une borne inférieure du primal, la deuxième étape consistant à améliorer cette borne inférieure par la minoration d'une posiforme quadratique.

2. LE DUAL LAGRANGIEN

Considérons le programme (P1), que nous appellerons le primal. Associons aux contraintes (1) le vecteur des multiplicateurs de Lagrange $\alpha=(\alpha_t\in\mathbb{R})_{(t=1,\ldots,T)}$ et aux contraintes (2) le vecteur des multiplicateurs de Lagrange $\beta=(\beta_p\geqq 0)_{(p=1,\ldots,P)}$. Nous allons nous intéresser au dual lagrangien de (P1).

La fonction de Lagrange s'écrit :

$$\mathcal{L}(x,\alpha,\beta)=-\sum_{t=1}^{T}\alpha_t-\sum_{p=1}^{P}n_p\beta_p$$
$$+\sum_{t=1}^{T}\sum_{p=1}^{P}(l_{tp}+\alpha_t+s_t\beta_p)x_{tp}-\sum_{t=1}^{T-1}\sum_{t'=t+1}^{T}\sum_{p=1}^{P}c_{tt'}x_{tp}x_{t'p}$$

La fonction duale est :

$$L(\alpha,\beta)=\min_{x\in\{0,1\}^{T\times P}}\mathcal{L}(x,\alpha,\beta)$$

et le problème dual est :

$$(\mathrm{D})\qquad\left\{\begin{array}{ll} & \mathrm{Max}\ L(\alpha,\beta)\\ \alpha_t\in\mathbb{R} & (t=1,\ldots,T)\\ \beta_p\geqq 0 & (p=1,\ldots,P)\end{array}\right.$$

Nous allons maintenant proposer une linéarisation du problème primal (P1) et montrer que la valeur optimale du dual lagrangien (D) est égale à la relaxation continue de cette linéarisation. Pour cela, remplaçons dans (P1) chaque produit $x_{tp}x_{t'p}$ par la variable bivalente $y_{tpt'}$. Il n'est pas difficile de vérifier que la valeur de la solution optimale de (P1) est égale à la valeur de la solution optimale du programme linéaire en variables bivalentes suivant :

$$
(\mathrm{LP1})\left\{
\begin{array}{ll}
\operatorname{Min} F(x)=\sum_{t=1}^{T}\sum_{p=1}^{P} l_{tp}\,x_{tp}-\sum_{t=1}^{T-1}\sum_{t'=t+1}^{T}\sum_{p=1}^{P} c_{tt'}\,y_{tpt'} & \\
\text{sous les contraintes :} & \\
\sum_{p=1}^{P} x_{tp}=1; \qquad t=1,\ \dots,\ T & (1)\\
\sum_{t=1}^{T} s_t\,x_{tp}\leqq n_p; \qquad p=1,\ \dots,\ P & (2)\\
\left.\begin{array}{c} y_{tpt'}\leqq x_{tp}\\ y_{tpt'}\leqq x_{t'p}\\ y_{tpt'}\in\{0,1\}\end{array}\right. & (4)\\
t=1,\ \dots,\ T-1; \qquad t'=t+1,\ \dots,\ T; \qquad p=1,\ \dots,\ P & \\
x_{tp}\in\{0,1\}; \qquad t=1,\ \dots,\ T-1; \qquad p=1,\ \dots,\ P & (3)
\end{array}\right.
$$

PROPRIÉTÉ 1 : La valeur optimale de (D), dual lagrangien de (P1), est égale à la valeur optimale de la relaxation continue de (LP1).

Preuve : Calculer la fonction duale revient à minimiser une fonction pseudo-booléenne quadratique sous-modulaire. Dans ce cas on sait que le problème peut se poser comme un programme linéaire continu (*cf.* [RHY,70] et, par exemple, [HAM,84]). Écrivons tout d'abord une linéarisation de la fonction duale :

$$
\left\{
\begin{array}{l}
L(\alpha,\beta)=\operatorname{Min}\left\{
\begin{array}{l}
-\sum_{t=1}^{T}\alpha_t-\sum_{p=1}^{P} n_p\beta_p+\sum_{t=1}^{T}\sum_{p=1}^{P}(l_{tp}+\alpha_t+s_t\beta_p)\,x_{tp}\\
\qquad -\sum_{t=1}^{T-1}\sum_{t'=t+1}^{T}\sum_{p=1}^{P} c_{tt'}\,y_{tpt'}
\end{array}\right\}\\
\text{sous les contraintes (3) et (4)}
\end{array}\right.
$$

La matrice des contraintes obtenue par cette linéarisation est totalement unimodulaire. On peut alors remplacer les variables bivalentes par des variables continues. La fonction duale $L(\alpha, \beta)$ est donc aussi égale à la valeur de la solution optimale du même problème de minimisation en variables continues, c'est-à-dire $y_{tpt'}\geqq 0$ et $0\leqq x_{tp}\leqq 1$ à la place de $y_{tpt'}\in\{0,1\}$ et $x_{tp}\in\{0,1\}$.

Le problème dual (D), $\{\operatorname{Max} L(\alpha, \beta) : \alpha \in \mathbb{R}^{T}, \beta \in \mathbb{R}_{+}^{P}\}$ peut donc être aussi considéré comme le dual lagrangien du *programme linéaire* suivant (par « dualisation » des contraintes (1) et (2)) :

$$
\text{(RLP1)}\left\{
\begin{array}{ll}
\displaystyle \operatorname{Min} \sum_{t=1}^{T} \sum_{p=1}^{P} l_{tp} x_{tp} - \sum_{t=1}^{T-1} \sum_{t'=t+1}^{T} \sum_{p=1}^{P} c_{tt'} y_{tpt'} & \\
\text{sous les contraintes :} & \\
\displaystyle \sum_{p=1}^{P} x_{tp} = 1; \qquad t=1, \ldots, T & (1) \\
\displaystyle \sum_{t=1}^{T} s_t x_{tp} \leqq n_p; \qquad p=1, \ldots, P & (2) \\
\left.\begin{array}{l} y_{tpt'} \leqq x_{tp} \\ y_{tpt'} \leqq x_{t'p} \\ y_{tpt'} \geqq 0 \end{array}\right. & (5) \\
t=1, \ldots, T-1; \qquad t'=t+1, \ldots, T; \qquad p=1, \ldots, P & \\
x_{tp} \geqq 0; \qquad t=1, \ldots, T-1; \qquad p=1, \ldots, P & (6)
\end{array}
\right.
$$

La valeur optimale du problème dual (D) est donc égale à la valeur de la solution optimale de ce dernier problème qui est la relaxation continue de (LP1) et que nous appelons (RLP1). ■

Remarquons d'après la propriété 1 que le saut de dualité qui nous intéresse, c'est-à-dire l'écart entre les valeurs des solutions optimales de (P1) et (D) est égal au saut d'intégrité relatif aux programmes linéaires (LP1) et (RLP1), c'est-à-dire à l'écart entre les valeurs des solutions optimales de ces deux programmes linéaires.

3. UNE BORNE SUPÉRIEURE DE L'OPTIMUM DU DUAL

PROPRIÉTÉ 2 : Soit

$$\mathbb{K} = \left\{ k \in \mathbb{N}^P / \sum_{p=1}^{P} k_p = \mathcal{S} \text{ et } \forall p \in \{1, \ldots, P\}, k_p \leqq n_p \right\}$$

et soit

$$Q_p = \sum_{t=1}^{T} q_{tp} \qquad \text{pour} \quad p=1, \ldots, P.$$

$B_{\sup}=(1/\mathscr{S})\ \underset{k\in\mathbb{K}}{\operatorname{Min}}\left(\sum_{p=1}^{P} k_p Q_p\right)$ est une borne supérieure de la valeur optimale du dual; cette borne peut être atteinte.

Preuve : Soit $k\in\mathbb{K}$. Montrons que $x_{tp}=k_p/\mathscr{S}\ (t=1, \ldots, T;\ p=1, \ldots, P)$ et $y_{tpt'}=k_p/\mathscr{S}\ (t=1, \ldots, T-1;\ t'=t+1, \ldots, T;\ p=1, \ldots, P)$ est une solution admissible de (RLP1).

$$\sum_{p=1}^{P} x_{tp}=\frac{1}{\mathscr{S}}\sum_{p=1}^{P} k_p=1 \qquad (t=1, \ldots, T)$$

$$\sum_{t=1}^{T} s_t x_{tp}=\frac{k_p}{\mathscr{S}}\left(\sum_{t=1}^{T} s_t\right)=k_p\leqq n_p \qquad (p=1, \ldots, P)$$

On vérifie immédiatement que les contraintes (5) et (6) sont aussi satisfaites. Calculons maintenant la valeur de cette solution admissible. Elle est égale à

$$\frac{1}{\mathscr{S}}\left(\sum_{t=1}^{T}\sum_{p=1}^{P} \mathrm{l}_{tp} k_p-\sum_{t=1}^{T-1}\sum_{t'=t+1}^{T}\sum_{p=1}^{P} c_{tt'} k_p\right)$$

$$=\frac{1}{\mathscr{S}}\left(\sum_{t=1}^{T}\sum_{p=1}^{P}\left[q_{tp}+\sum_{t'=t+1}^{T} c_{tt'}\right]k_p-\sum_{t=1}^{T-1}\sum_{t'=t+1}^{T}\sum_{p=1}^{P} c_{tt'} k_p\right)$$

$$=\frac{1}{\mathscr{S}}\left(\sum_{p=1}^{P} k_p Q_p\right)$$

Nous venons de montrer que, pour tout $k\in\mathbb{K}$, on peut construire une solution admissible de (RLP1) dont la valeur est $1/\mathscr{S}\left(\sum_{p=1}^{P} k_p Q_p\right)$. Il existe donc une solution admissible de (RLP1) dont la valeur est $1/\mathscr{S}\ \underset{k\in\mathbb{K}}{\operatorname{Min}}\left(\sum_{p=1}^{P} k_p Q_p\right)$ et cette valeur est donc supérieure ou égale à la valeur d'une solution optimale de (RLP1) et donc à la valeur optimale du dual (D).

Montrons maintenant que cette borne peut être atteinte. Considérons le problème de placement de 3 tâches sur 2 processeurs, défini par la figure ci-dessous.

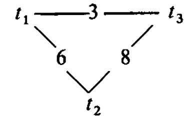
$\mathbb{K}=\{(1,2);\ (2,1)\}$, $B_{\sup}=1/3\ \underset{k\in\mathbb{K}}{\operatorname{Min}}(2k_1+14k_2)=6$ [en prenant $k=(2,1)$].

On vérifie aisément qu'en prenant $\alpha=(-6,\ -4,\ -4)$ et $\beta=(4,0)$, on obtient une solution du dual de valeur 6. Remarquons que dans ce cas il y a un saut

Les coûts d'exécution

q	p_1	p_2
t_1	2	0
t_2	0	8
t_3	0	6

Le graphe de communication
et les coûts correspondants



avec, $s_1=s_2=s_3=1$ et $n_1=n_2=2$

Un exemple de problème de placement de tâches.

de dualité puisque l'optimum du dual vaut 6 alors que l'optimum du primal vaut 9.

Notons que cette borne peut être atteinte pour des problèmes de placement comportant un nombre quelconque de tâches et de processeurs. C'est le cas si tous les processeurs sont identiques (*voir* section 4). ■

La propriété 2 fournit une borne intéressante de la valeur optimale du dual. En particulier cette borne peut être atteinte. Nous allons montrer ci-dessous pourquoi, dans ce type de problème, la fonction duale est bornée par une expression qui ne dépend que des coûts d'exécution.

Considérons le programme mathématique

$$(\mathrm{P}) \qquad \left\{ \begin{array}{l} \mathrm{Min}\, f_1(x)+f_2(x) \\ \text{sous les contraintes :} \\ (a)\ g_i(x)=0 \qquad i=1, \ldots, I \\ (b)\ h_j(x)\leqq 0 \qquad j=1, \ldots, J \\ x\in S \end{array} \right.$$

et soit $\alpha=(\alpha_i\in\mathbb{R})_{(i=1,\ldots,I)}$ [resp. $\beta=(\beta_j\geqq 0)_{(j=1,\ldots,J)}$] le vecteur des multiplicateurs de Lagrange associés aux contraintes (a) [resp. (b)].

Si, pour tout $k\in\{1, \ldots, J\}$, il existe $x^{(k)}\in S$ tel que

$$f_2(x^{(k)})=0$$
$$g_i(x^{(k)})=0 \qquad (i\in\{1, \ldots, I\})$$
$$h_j(x^{(k)})\leqq 0 \qquad (j\in\{1, \ldots, J\},\ i\neq k)$$
$$h_k(x^{(k)})>0$$
$$\sum_{j=1}^{J} h_j(x^{(k)})\leqq 0$$

alors pour tout (α, β), la valeur de la fonction duale $L(\alpha, \beta)$ associée au programme (P) est bornée par une expression qui ne dépend que de $f_1(x)$.

En effet,

$$L(\alpha, \beta)=\operatorname*{Min}_{x \in S}\left(f_1(x)+f_2(x)+\sum_{i=1}^{I} \alpha_i g_i(x)+\sum_{j=1}^{J} \beta_j h_j(x)\right)$$

et, en choisissant $x=x^{(k)}$ tel que $\beta_k=\operatorname*{Min}_{j \in \{1, \ldots, P\}} \beta_j$, on obtient

$$L(\alpha, \beta) \leqq f_1(x^{(k)})+\sum_{j=1}^{J} \beta_k h_j(x^{(k)}) \leqq f_1(x^{(k)})$$

Si par ailleurs toute solution admissible x de (P) vérifie $f_2(x)>0$, il est clair qu'en prenant une constante positive M suffisamment grande, le saut de dualité associé au problème $(\mathrm{P_M})$, obtenu à partir de (P) en remplaçant la fonction économique par $f_1(x)+M f_2(x)$, sera aussi grand que désiré.

Dans le cas du problème de placement de tâches considéré, $f_1(x)$, $f_2(x)$, $g_i(x)$ $(i=1, \ldots, I)$, $h_j(x)$ $(j=1, \ldots, J)$ et S correspondent respectivement aux coûts d'exécution, aux coûts de communication, aux contraintes de type affectation, aux contraintes de capacité et à l'ensemble $\{0, 1\}^{T \times P}$. La solution $x^{(k)}$ consiste à placer toutes les tâches sur le processeur k.

Construisons maintenant une famille de problèmes de placement qui réalise des sauts de dualité aussi grands que désirés. Supposons que le graphe de communication entre les tâches est connexe et qu'aucun des P processeurs ne peut exécuter l'ensemble des T tâches ($n_p<\mathcal{S}$ pour $p=1, \ldots, P$). Dans tout placement possible des tâches sur les processeurs, au moins une paire de tâches auront donc à communiquer. Multiplions les coûts de communication par un paramètre strictement positif λ. Le problème de placement se formule alors

$$\left\{\begin{array}{l}\operatorname{Min} F(x)=\sum_{t=1}^{T} \sum_{p=1}^{P} l_{tp} x_{tp} \\ \quad -\sum_{t=1}^{T-1} \sum_{t'=t+1}^{T} \sum_{p=1}^{P} \lambda c_{tt'} x_{tp} x_{t'p} \\ \text{sous les contraintes (1), (2) et (3)}\end{array}\right.$$

Il est clair que le coût de tout placement est supérieur ou égal à (par exemple) $\left[\sum_{t=1}^{T}(\operatorname{Min}_p q_{tp})+\lambda c_{t_0 t'_0}\right]$ où $c_{t_0 t'_0}$ est le coût minimal de communication entre deux tâches communicantes. D'après la propriété 2, pour tout $\Delta>0$, le saut de dualité est supérieur à Δ si λ est supérieur à la constante

$$\left[\Delta+\frac{1}{\mathscr{S}}\operatorname*{Min}_{k\in\mathbb{K}}\left(\sum_{p=1}^{P}k_p Q_p\right)-\sum_{t=1}^{T}(\operatorname*{Min}_{p} q_{tp})\right]\Big/ c_{t_0 t'_0}.$$

Remarque : Considérons le modèle cité par plusieurs auteurs (*cf.*, par exemple, [BIL,88], [LO,84], [LU,84], [PRI,84], [SIN,87], [SUT,89]) dans lequel on suppose que chacun des P processeurs peut exécuter l'ensemble des T tâches ($n_p=\mathscr{S}$ pour $p=1, \ldots, P$). $\mathrm{B}_{\mathrm{sup}}$ est alors égale à $\left(\operatorname*{Min}_{p}\left(\sum_{t=1}^{T} q_{tp}\right)\right)=\sum_{t=1}^{T} q_{tp_0}$. Dans ce cas, $\mathrm{B}_{\mathrm{sup}}$ est aussi le coût du placement de toutes les tâches sur le processeur p_0. Ainsi, $\mathrm{B}_{\mathrm{sup}}$ est une borne supérieure de la valeur optimale du primal et la propriété 2 n'apporte pas d'information supplémentaire quant à l'écart entre la valeur optimale du dual et la valeur optimale du primal. Notons cependant que l'étude expérimentale effectuée dans [SUT,89] montre que les sauts de dualité sont très rares. Cette étude expérimentale a porté sur un grand nombre de jeux d'essai dont la taille est, typiquement, de 20 processeurs, 50 tâches et 200 paires de tâches communicantes.

4. UNE BORNE INFÉRIEURE DE L'OPTIMUM DU DUAL

PROPRIÉTÉ 3 : Soit $\mathrm{B}_{\mathrm{inf}}$ la valeur de l'optimum du programme linéaire suivant :

$$(\mathrm{P})\qquad \begin{cases} \operatorname{Min} \sum_{t=1}^{T}\sum_{p=1}^{P} q_{tp}x_{tp} \\ \text{sous les contraintes (1), (2) et (6)} \end{cases}$$

$\mathrm{B}_{\mathrm{inf}}$ est une borne inférieure de la valeur optimale du problème dual (D) et cette borne peut être atteinte.

Preuve : La valeur optimale du problème dual est égale à la valeur d'une solution optimale de (RLP1) qui est elle-même supérieure ou égale à la valeur d'une solution optimale du problème suivant :

$$\left\{\begin{array}{l} \operatorname{Min} \sum_{t=1}^{T} \sum_{p=1}^{P} l_{tp} x_{tp} - \sum_{t=1}^{T-1} \sum_{t'=t+1}^{T} \sum_{p=1}^{P} c_{tt'} y_{tpt'} \\ \text{sous les contraintes (1), (2), (6) et} \\ y_{tpt'} \leqq x_{tp}; \quad t=1, \ldots, T-1; \quad t'=t+1, \ldots, T; \quad p=1, \ldots, P \end{array}\right.$$

La valeur d'une solution de ce dernier problème est égale à la valeur d'une solution optimale du problème suivant :

$$\left\{\begin{array}{l} \operatorname{Min} \sum_{t=1}^{T} \sum_{p=1}^{P} l_{tp} x_{tp} - \sum_{t=1}^{T-1} \sum_{t'=t+1}^{T} \sum_{p=1}^{P} c_{tt'} x_{tp} \\ \text{sous les contraintes (1), (2) et (6)} \end{array}\right.$$

Il suffit ensuite de remarquer que d'après la définition des l_{tp} :

$$\sum_{t=1}^{T} \sum_{p=1}^{P} l_{tp} x_{tp} - \sum_{t=1}^{T-1} \sum_{t'=t+1}^{T} \sum_{p=1}^{P} c_{tt'} x_{tp} = \sum_{t=1}^{T} \sum_{p=1}^{P} q_{tp} x_{tp}$$

Pour montrer que la borne B_{inf} peut être atteinte, nous allons considérer le cas où tous les processeurs sont identiques, c'est-à-dire que le coût d'exécution d'une tâche est le même quel que soit le processeur qui l'exécute :

$$q_{tp} = q_t \qquad (t=1, \ldots, T; p=1, \ldots, P).$$

Par conséquent, $Q_p = \sum_{t=1}^{T} q_{tp} = \sum_{t=1}^{T} q_t = Q \, (p=1, \ldots, P)$. Il est facile de voir que, dans ce cas, l'optimum du problème (PL) est égal à Q. Donc, Q est une borne inférieure de la valeur optimale du dual. Par ailleurs, une borne supérieure de la valeur optimale du dual est $B_{\text{sup}} = Q\,(1/\mathscr{S}) \operatorname{Min}_{k \in \mathbb{K}} \left(\sum_{p=1}^{P} k_p \right) = Q$.

Donc, la valeur optimale du dual est égale à Q.

5. CONCLUSION ET SUGGESTIONS DE RECHERCHES FUTURES

Nous nous sommes intéressés dans cet article à un problème d'optimisation combinatoire difficile : le placement de tâches dans un système distribué avec contraintes de ressources. Pour pouvoir résoudre ce problème par une procédure « branch and bound », il faut savoir calculer, en chaque nœud de l'arbre de recherche, une borne inférieure de bonne qualité. Une technique classique pour essayer d'obtenir une telle borne est celle de la relaxation lagrangienne. Les résultats obtenus dans ce travail fournissent des indications intéressantes concernant la qualité de la borne que l'on peut espérer obtenir par cette technique. Si les résultats de la section 3 sont assez négatifs, il faut cependant noter qu'ils n'interdisent pas l'obtention de très bonnes bornes par dualité lagrangienne pour certains problèmes de placement. C'est le cas lorsqu'on ne considère pas de contraintes de ressources sur les processeurs (*cf.* section 3). Notons également que les résultats de la section 4 fournissent une certaine garantie sur l'optimum du dual et que si le calcul de la solution duale est fait dans l'étape d'évaluation d'une procédure « branch and bound », on peut vérifier que la valeur du dual augmente avec les fixations de variables.

Par ailleurs, la détermination d'une solution duale peut être considérée comme une première étape dans le calcul d'une borne inférieure du primal. En effet, pour tout (α, β) solution admissible du problème dual et pour tout x solution admissible de (P1), on peut écrire $F(x) \geqq L(\alpha, \beta) + \mathscr{F}(x, \bar{x})$ où $\mathscr{F}(x, \bar{x})$ est une posiforme quadratique. Ce résultat se démontre aisément en remarquant que la fonction de Lagrange $\mathscr{L}(x, \alpha, \beta)$ est elle-même une fonction pseudobooléenne quadratique sous-modulaire. Elle peut donc s'écrire, d'après les travaux de Hammer, Hansen et Simeone sur le « roof dual » (*cf.* [HAM,84]), $\mathscr{L}(x, \alpha, \beta) = L(\alpha, \beta) + \mathscr{F}(x, \bar{x})$. Une voie intéressante pour améliorer la borne du primal fournie par le dual est de calculer une borne inférieure b de la fonction $\mathscr{F}(x, \bar{x})$ sous les contraintes de (P1) (par une autre méthode que la relaxation lagrangienne). On obtient alors une nouvelle borne inférieure pour le problème primal, égale à $L(\alpha, \beta) + b$. Notons que cette dernière propriété est vraie pour toute solution (α, β) du dual et en particulier pour la solution optimale (α^*, β^*).

Exemple : Reprenons l'exemple de la figure. Nous avons déjà montré que $\alpha = (-6, -4, -4)$ et $\beta = (4, 0)$ forment une solution optimale de valeur 6 pour le problème dual. Grâce à cette solution du dual, on peut écrire, pour tout x solution admissible du problème de placement, $F(x) \geqq 6 + g(x, \bar{x})$ où

$$g(x, \bar{x}) = 6x_{11}\bar{x}_{21} + 3x_{11}\bar{x}_{31} + 8x_{21}\bar{x}_{31} + 3x_{12}\bar{x}_{32} + 6x_{22}\bar{x}_{12} + 6x_{22}\bar{x}_{32} + 3x_{32}\bar{x}_{22}$$

On peut ensuite s'intéresser au problème

$$(\mathrm{R}) \quad \left\{ \begin{array}{c} \operatorname{Min} g(x, \bar{x}) \\ \text{sous les contraintes :} \\ x_{11}+x_{12}=1 \\ x_{21}+x_{22}=1 \\ x_{31}+x_{32}=1 \\ x_{11}+x_{21}+x_{31} \leqq 2 \\ x_{12}+x_{22}+x_{32} \leqq 2 \\ x \in \{0, 1\}^{3 \times 2} \end{array} \right.$$

pour lequel on peut aisément montrer (par exemple par substitution des variables) que la valeur optimale est au moins égale à 2. On améliore ainsi la borne inférieure qui passe de 6 à 8. ■

Dans le cas où chaque processeur du réseau dispose de suffisamment de ressources pour exécuter la totalité des tâches, on peut écrire $F(x)=L(\alpha, \beta)+\mathscr{F}(x, \bar{x})$. Cette propriété a été utilisée par Sutter ([SUT,89]) pour prouver l'existence – ou la non existence – d'un saut de dualité. Il propose en effet une méthode pour résoudre l'équation $\mathscr{F}(x, \bar{x})=0$, les variables x étant soumises à des contraintes de type « affectation ».

On peut également comparer la voie que nous proposons aux méthodes utilisées par Bourjolly ([BOU,86]) et Sutter ([SUT,89]) pour calculer une borne inférieure du minimum d'une fonction pseudobooléenne quadratique $f(x)$ sans contraintes. Ces deux auteurs écrivent la fonction considérée sous la forme $f(x)=c+\psi(x, \bar{x})$, ψ étant une posiforme quadratique et c la plus grande constante possible, puis ils calculent une borne inférieure de ψ. Les expériences de calcul effectuées par Bourjolly et Sutter prouvent l'efficacité d'une telle approche. Bien que ces méthodes concernent l'optimisation d'une fonction pseudobooléenne quadratique *sans contraintes*, on peut les rapprocher de celle que nous proposons pour le problème de placement de tâches puisque la constante c utilisée, qui est en fait la valeur optimale du « roof dual » ([HAM,84]), peut s'interpréter comme la valeur optimale d'un problème dual lagrangien associé à la minimisation de f ([ADA,90]).

REMERCIEMENTS

Nous remercions les lecteurs pour leurs commentaires et suggestions qui nous ont aidés à améliorer la première version de cet article.

BIBLIOGRAPHIE

[ADA,90] W. P. Adams, A. Billionnet et A. Sutter, Unconstrained 0-1 Optimization and Lagrangean Relaxation, *Discrete Appl. Math.*, 1990, *29*, n^{os} 2-3, p. 131-142.

[AND,88] F. André et J. L. Pazat, Le placement de tâches sur les architectures parallèles. *Tech. Sci. Inform.*, 1988, *7*, n° 4, p. 385-401.

[BEN,89] D. Benhamamouch, Résolution approchée d'un problème d'affectation de tâches dans les systèmes informatiques distribués, *Thèse de doctorat*, Université d'Oran Es Senia, 1989.

[BIL,85] A. Billionnet et M. Minoux, Maximizing a Supermodular Pseudoboolean Function: a Polynomial Algorithm for Supermodular Cubic Functions, *Discrete Appl. Math.*, 1985, *12*, p. 1-11.

[BIL,88] A. Billionnet, M. C. Costa et A. Sutter, An Efficient Algorithm for a Task Allocation Problem, *A.C.M.* (à paraître).

[BIL,89] A. Billionnet, M. C. Costa et A. Sutter, Les problèmes de placement dans les systèmes distribués, *Tech. Sci. Inform.*, 1989, *8*, n° 4, p. 307-337.

[BOK,81] S. Bokhari, A Shortest Tree Algorithm for Optimal Assignments across Space and Time in a Distributed Processor System, *I.E.E.E. Trans. Software Engrg.*, Nov. 1981, *SE-7*, n° 6, p. 583-589.

[BOK,87] S. Bokhari, Assignment Problems in Parallel and Distributed Computing, *Kluwer Academic Publishers*, Norwell, Massachusets, USA, 1987, 155 p.

[BOU,86] P. M. Bourjolly, Integral and Fractional Node-Pakings, and Pseudoboolean Programming, *Thesis* for the degree of Doctor of Philosophy in Combinatorics and Optimization, Université de Waterloo, Ontario, Canada, 1986.

[EFE,82] K. Efe, Heuristic Models for Task Assignment Scheduling in Distributed Systems, *Computer*, juin 1982, *15*, p. 50-56.

[FER,89] D. Fernandez-Baca, Allocating Modules to Processors in a Distributed System, *I.E.E.E. Trans. Software Engrg.*, 1989, *SE-15*, n° 11, p. 1427-1436.

[GAB,84] A. Gabrielian et D. B. Tyler, Optimal Object Allocation in Distributed Systems, *Proc. of the Int. Conf. on Distributed Computing Systems*, San Francisco, California, 14-18 May 1984, p. 88-95.

[GON,85] M. Gondran et M. Minoux, Graphes et algorithmes, *Eyrolles*, Paris, 1985, 524 p.

[HAM,84] P. L. Hammer, P. Hansen et B. Simeone, Roof Duality, Complementation and Persistency in Quadratic 0-1 Optimization, *Math. Programming*, 1984, *28*, p. 121-155.

[LO,84] V. M. Lo, Heuristic Algorithms for Task Assignment in Distributed Systems, *Proc. of the Int. Conf. on Distributed Computing Systems*, San Francisco, California, 14-18 May 1984, p. 30-39.

[LU,84] M. I. Lu et J. B. Sinclair, Module Assignment in Distributed Systems, *Proc.* 1984 *Computer Network Symposium*, Gathersburg, MD, p. 105-111.

[MA,82] P. R. Ma, E. Y. S. Lee et M. Tsuchiya, A Task Allocation Model for Distributed Computing Systems, *I.E.E.E. Trans. Comput.*, Jan. 1982, *C-31*, n° 1, p. 66-78.

[MAG,89] V. F. Magirou et J. Z. Milis, An algorithm for the multiprocessor assignment problem. *Oper. Res. Lett.*, 1989, *8*, p. 351-356.

[PRI,84] C. C. Price et S. Krishnaprasad, Software Allocation Models for Distributed Computing Systems, *Proc. of the Int. Conf. on Distributed Computing Systems*, San Francisco, California, 14-18 May 1984, p. 40-48.

[RAO,79] G. S. Rao, H. S. Stone et T. C. Hu, Assignment of Tasks in a Distributed Processor System with Limited Memory, *I.E.E.E. Trans. Comput.*, April 1979, *C-28*, n° 4, p. 291-299.

[RHY,70] J. Rhys, A Selection Problem of Shared Fixed Costs and Networks, *Management Sci.*, 1970, *17*, p. 200-207.

[SIN,87] J. B. Sinclair, Efficient Computation of Optimal Assignments for Distributed Tasks, *J. Parallel Dist. Comput.*, 1987, *4*, p. 342-362.

[SUT,89] A. Sutter, Programmation non linéaire en variables 0-1, Application à des problèmes de placement de tâches dans un système distribué, *Thèse de doctorat du C.N.A.M.*, juin 1989.

[TOW,86] D. Towsley, Allocating Programs Containing Branches and Loops within a Multiple Processor System, *I.E.E.E. Trans. Software Engrg.*, 1986, SE-12, n° 10, p. 1018-1024.