Evaluation et modélisation des communications concurrentes

Vienne Jérôme, Martinasso Maxime

Laboratoire d'Informatique de Grenoble
Projet Mescal
51, avenue Jean Kuntzman
38330 Montbonnot

1. Introduction

Les applications parallèles provenant du calcul scientifique s'efforcent d'exploiter au maximum les ressources d'une architecture parallèle. Pour ce faire, elles utilisent la totalité des unités de traitement disponibles afin de rendre leur exécution la plus rapide possible. Une exécution simultanée de plusieurs processus sur une machine génère des comportements particuliers causés par les accès concurrents aux ressources. Pour évaluer les performances d'une application, il devient donc nécessaire d'identifier et de comprendre les effets du partage de ressources. Avec l'augmentation de la puissance de calcul et du nombre d'unités de calcul par processeur, les performances des réseaux en terme de bande passante et latence continuent d'être un facteur important limitant l'efficacité des applications parallèles. En ajoutant des unités de calcul aux noeuds d'une grappe, les applications parallèles, lors des phases de communication, créent des accès concurrents aux ressources du noeuds et du réseau sous-jacent.
Ces nouveaux comportements de partage de ressources, ainsi produits, sont difficiles à interpréter et à prédire. Dans nos travaux [3, 4], nous étudions le problème du partage du réseau sur des grilles utilisant des réseaux dédiés à haute performance tels que Gigabit Ethernet, Myrinet, Quadrics ou Infiniband.
L'exécution simultanée des tâches d'une application entraîne des accès concurrents sur la ressource réseau. Leurs effets conduisent à une perte de performance qui découle du partage de la bande passante réseau entre communications. Suivant ce contexte, nous présentons, dans une première partie, comment nous avons analysé finement des comportements concurrents sur les architectures : Infiniband, Myrinet et Gigabit Ethernet. Cette analyse nous à conduit à la définition de modèles prédictifs basés sur la notion de partage de la bande passante comme nous le verrons dans la seconde partie. En outre, nous montrerons, dans une troisième partie, que l'intégration de ces modèles dans une simulation permet de prédire les impacts dûs à la concurrence entre communications MPI résultantes de l'éxécution d'applications scientifiques.

2. Analyse des comportements concurrents

Pour étudier ces phénomènes de contention, nous avons développé un logiciel permettant de connaître les pénalités obtenues en fonction de schémas de communication. La pénalité, appelée aussi coefficient de ralentissement, étant le rapport, pour une communication, entre le temps obtenu lors du conflit et un temps de référence obtenu lors d'une communication sans partage de la ressource réseau .
La figure 1 montre quelques exemples de comportement sur 3 architectures réseaux différentes. Le premier schéma représente une communication sans concurrence avec, donc, une pénalité de 1.

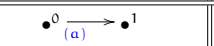
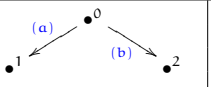
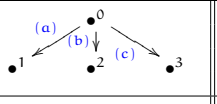
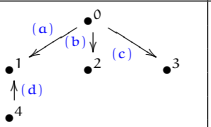
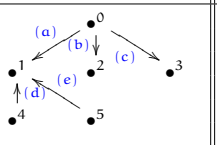
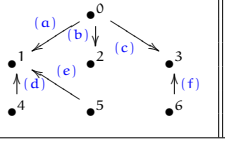
Schema de Communication	Architecture réseau		
	Gigabit Eth.	Myrinet	Infiniband
	a = 1	a = 1	a =1
	a = 1.5 b = 1.5	a =1.9 b = 1.9	a = 1.725 b = 1.725
	a = 2.25 b = 2.25 c = 2.25	a = 2.8 b = 2.8 c = 2.8	a = 2.61 b = 2.61 c = 2.61
	a = 2.15 b = 2.15 c = 2.15 d = 1.15	a = 2.8 b = 2.8 c = 2.8 d = 1.45	a = 2.61 b = 2.61 c = 2.61 d = 1.14
	a = 4.4 b = 2.6 c = 2.6 d = 2.6 e = 2.6	a = 4.4 b = 4.2 c = 4.2 d = 2.5 e= 2.5	a = 3.66 b = 3.66 c = 3.66 d = 2.035 e = 2.035
	a = 4.4 b = 2.0 c = 3.3 d = 2.6 e = 2.6 f = 1.4	a = 4.5 b = 4.5 c = 4.5 d = 2.5 e= 2.5 f = 1.3	a = 3.935 b = 3.935 c = 3.935 d = 1.995 e = 1.995 f = 1.01

FIG. 1 – Comparaison des pénalités sur différentes architectures pour un ensemble de schémas de communication

En augment le nombre de communication, on voit clairement que les pénalités sont différentes suivant le réseau et que les modèles connus de communication tel que LogP[2] ou LogGP[1] ne permettent pas de prédire ce genre de phénomène.

3. Modélisation des communications

En se basant sur l'étude précédente, plusieurs modèles de communication concurrente ont été définis. Ces modèles décrivent le partage de la bande passante réseau entre communications concurrentes et les pertes de performance qu'occasionnent ce processus de partage. Ils sont basés soit sur une compréhension des mécanismes de contrôle de flux, soit sur une description des effets du partage de la bande passante. Actuellement, deux architectures réseaux sont modélisées [4] : Myrinet (2000) et Gigabit Ethernet. Une modélisation de l'architecure réseau Infiniband (Infinihost III) est en cours de développement et devrait être disponible prochainement.

4. Simulation des modèles de communication

Afin d'employer ces modèles sur des applications scientifiques et de permettre ainsi une analyse fine de leurs performances réseaux, nous avons réalisé un simulateur prenant en entrée différents paramètres (trace de l'application, paramètres du noeuds, placement des tâches, modèle réseau). Pour valider ce simulateur, nous avons évaluer les modèles sur des graphes synthétiques puis l'application HPL qui est utilisée pour établir le classement du TOP500.

La figure 2 montre un exemple du résultat obtenu par le simulateur, pour 16 coeurs, en utilisant l'application HPL, avec trois politiques de placement de tâche différentes, à savoir Round Robin par Noeud (RRN), Round Robin par Processeur (RRP) et une politique aléatoire. Nous voyons d'un coté le temps de communication mesuré et celui prédit et, d'un autre, l'erreur absolue mesurée.

Les modèles proposés prédisent avec une erreur faible d'environ 10% les temps des communications concurrentes. Les résultats pour Gigabit Ethernet sont légèrement moins précis que pour Myrinet. Ces écarts peuvent s'expliquer par la plus grande variabilité des comportements sur le réseau Gigabit Ethernet.

5. Conclusion

Nous avons présenté ici la méthode utilisée pour l'évaluation et la modélisation des communications concurrentes à partir d'expérimentation. La prédiction des comportements concurrents donne lieu, au niveau scientifique et technique, à une meilleure connaissance des besoins des applications et, au niveau industriel, à la proposition de solutions de grappes adaptées aux besoins de leur clientèles.

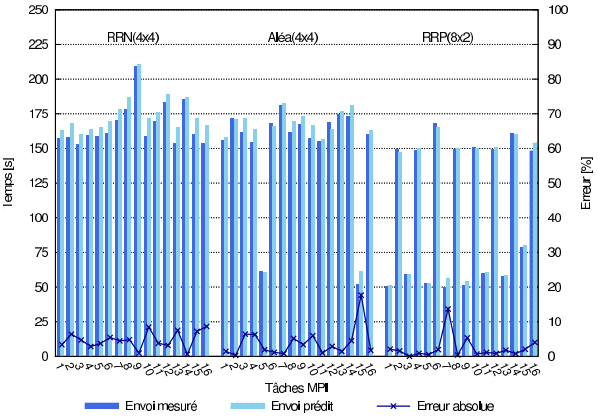


FIG. 2 – Résultat de la simulation d'un benchmark HPL de taille 20500 sur le réseau Myrinet

Bibliographie

1. Albert Alexandrov, Mihai F. Ionescu, Klaus E. Schauser, and Chris Scheiman. LogGP : Incorporating long messages into the LogP model for parallel computation. *Journal of Parallel and Distributed Computing*, 44(1) :71–79, 1997.
2. David E. Culler, Richard M. Karp, David Patterson, Abhijit Sahay, Eunice E. Santos, Klaus Erik Schauser, Ramesh Subramonian, and Thorsten von Eicken. Logp : a practical model of parallel computation. *Commun. ACM*, 39(11) :78–85, 1996.
3. M. Martinasso. Modèles de communications concurrentes sur des grappes smp. In *Perpi'2006, Actes de la conférences Renpar'17*, pages 132–139, Canet en Roussillon, October 2006.
4. Maxime Martinasso. *Analyse et modélisation des communications concurrentes dans les réseaux haute performance*. PhD thesis, Université Joseph Fourier, BP 53 - 38041 Grenoble Cedex 9, France, May 2007. 195 pages.