



HAL
open science

Anticipation des communications réseau grâce à la connaissance du futur dans le parallélisme à tâches

Tanguy Chatelain

► **To cite this version:**

Tanguy Chatelain. Anticipation des communications réseau grâce à la connaissance du futur dans le parallélisme à tâches. COMPAS 2025 - Conférence francophone d'informatique en Parallélisme, Architecture et Système, Jun 2025, Bordeaux, France. ⟨hal-05147860⟩

HAL Id: hal-05147860

<https://hal.science/hal-05147860v1>

Submitted on 7 Jul 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

Anticipation des communications réseau grâce à la connaissance du futur dans le parallélisme à tâches

Tanguy Chatelain *

Centre Inria de l'Université de Bordeaux,
200, avenue de la Vieille Tour
33405 Talence CEDEX - France
tanguy.chatelain@inria.fr

Résumé

Les architectures actuelles des calculateurs haute performance présentent des nœuds multi-cœurs ou GPU reliés par un réseau haute performance. Ce type d'architecture pousse à l'utilisation de modèles de programmation parallèle hybride multi-thread et distribuée. Par ailleurs, sur certains de ces réseaux haute-performance, une étape préalable à l'échange de données est l'enregistrement mémoire qui permet à la carte réseau d'accéder directement aux adresses mémoire enregistrées sans passer par l'OS. Or cette étape allonge la durée des communications car elle est réalisée juste avant l'envoi ou réception de données. Par conséquent, une pratique commune est de réaliser un cache d'enregistrement appelé *rcache*. Cependant, certaines applications utilisent des buffers de communications systématiquement différents et n'en bénéficient donc pas. Ainsi, notre objectif est de réaliser l'enregistrement mémoire plus tôt, ce qui est possible car cette étape n'utilise que les adresses mémoire et peut donc être réalisée avant que les contenus des buffers de réception ou d'envoi ne soient prêts.

Pour y parvenir, nous souhaitons mettre à profit l'utilisation des modèles de programmation parallèle hybride comme StarPU, qui est un système d'exécution s'appuyant sur un parallélisme à base de tâches. Dans sa version distribuée, StarPU s'appuie sur une bibliothèque MPI, ou bien sur NewMadeleine, et exécute les communications inter-nœuds de manière implicite en les inférant depuis le graphe de tâches. Ainsi, en utilisant la connaissance du futur fournie par le graphe de tâche maintenu par StarPU, il est possible d'être prévenu lorsque la dernière tâche de calcul utilisant un buffer d'envoi ou de réception commence. De cette manière, nous pouvons réaliser l'enregistrement mémoire pendant l'exécution de cette tâche, et éviter la latence induite par cette étape. En retirant la phase d'enregistrement mémoire du chemin critique de l'exécution des communications nous espérons réduire la durée des communications réseau, ce qui devraient conduire à de meilleures performances des applications distribuées utilisant StarPU.

Mots-clés : programmation parallèle hybride, réseau haute performance, programmation par tâches

*. Ces travaux sont encadrés par Alexandre Denis, Samuel Thibault et Philippe Swartvagher