2019-01437 - PhD Position F/M Static Allocation Algorithms for Scheduling High-Performance Applications

Niveau de diplôme exigé : Bac + 5 ou équivalent
Autre diplôme apprécié : master degree in Computer Science (or maths)
Fonction : Doctorant
Niveau d'expérience souhaité : Jeune diplômé

A propos du centre ou de la direction fonctionnelle

An important force which has continued to drive HPC has been to focus on frontier milestones which consist in technical goals that symbolize the next stage of progress in the field. After the step of the teraflop machine in the 1990s, the HPC community envision the use of generalist petaflop supercomputers and soon coming exaflop machines in the 2020s. For application codes to sustain petaflop and more using a few millions of cores or more will be needed, regardless of processor technology. Currently, a few HPC simulation codes easily scale to this regime and major code development efforts are critical to achieve the potential of these new systems. Scaling to at this rate will involve improving physical models, mathematical modelling, super scalable algorithms that will require paying particular attention to acquisition, management and vizualization of huge amounts of scientific data.

In this context, the purpose of the HiePACS project is to perform efficiently frontier simulations arising from challenging research and industrial multiscale applications. The solution of these challenging problems require a multidisciplinary approach involving applied mathematics, computational and computer sciences. In applied mathematics, it essentially involves advanced numerical schemes. In computational science, it involves massively parallel computing and the design of highly scalable algorithms and codes to be executed on future petaflop (and beyond) platforms. Through this approach, HiePACS intends to contribute to all steps that go from the design of new high-performance more scalable, robust and more accurate numerical schemes to the optimized implementations of the associated algorithms and codes on very high performance supercomputers.

Contexte et atouts du poste

Within the framework of a partnership (you can choose between)

- Projet Région Nouvelle-Aquitaine 2018-1R50119 “HPC scalable ecosystem”.
- SysNum cluster https://sysnum.labex.u-bordeaux.fr/en/
- ANR DASH https://project.inria.fr/dash/
- The thesis will take place in Hiepacs Inria Team Project http://hiepacs.bordeaux.inria.fr/ in close collaboration with Tadaam Team Project https://team.inria.fr/tadaam/

Mission confiée

In recent years, task based runtime schedulers (such as StarPU [1] or PARSEC) have shown their efficiency for scheduling complex HPC applications on complex platforms, where each node consists of multicore CPUs and accelerators like GPUs or FPGAs. These runtime schedulers take allocation decisions at runtime, based on the actual state of the platform and on the set of ready tasks. The flexibility of such runtimes comes from the explicit expression of complex scientific calculations as graphs of dependent tasks, and this allows to obtain better and more portable performance than monolithic applications based directly on OpenMP and MPI. Indeed, it is now possible both to implement sophisticated allocation strategies that would be too complex for monolithic implementations, and to specify dynamic strategies which can react to unexpected runtime events or incorrect processing and communication time predictions. In the context of a single node, it has also been proved that it is even possible to achieve better performance by using intermediate mixed strategies [2,3,4,5], where some static knowledge about the computational graph is injected into the runtime scheduler.

The goal of this thesis is to extend these mixed strategies to the context of large scale platforms with distributed memory. In this context, we can assume that the graph of the application is known in advance and the initial allocation of data is clearly crucial in order to avoid slow inter node communications. This initial allocation, computed before the execution, should be able

(i) to minimize the transfer or generation cost to obtain the initial data distribution

(ii) to balance the load between computing resources while minimizing the communication overhead and

(iii) to be robust to small changes in the performance of the ressources.

(iv) to be able to make use of distributed storage devices such as Burst Buffers

In all cases, the possibility to replicate data and to replicate computations will be explored.

Then, at runtime, the system must be able to decide (i) when to move a given task from one resource to another, or to regenerate this data onto another resource and (ii) when to perform a more complex expensive data redistribution.

These two problems will be formulated as combinatorial optimization problems, and we will design algorithms to solve them with provable performance guarantees, possibly in restricted settings (for specific task graphs that may come for linear algebra, tensor computations or deep learning or specific computing platforms)

Principales activités

Main activities :

- Algorithm design
- Optimality proofs
- Modelling of applications and computing platforms

Compétences

Technical skills and level required : good level in algorithmic design and modelling

Languages : Python, C, C++

Avantages

- Subsidized meals
- Partial reimbursement of public transport costs
- Possibility of teleworking (after 6 months of employment) and flexible organization of working hours
- Professional equipment available (videoconferencing, loan of computer equipment, etc.)
- Social, cultural and sports events and activities
- Access to vocational training
- Social security coverage

Rémunération

- 1982€ / month (before taxs) during the first 2 years
- 2085€ / month (before taxs) during the third year

Informations générales

- **Thème/Domaine :** Calcul distribué et à haute performance
  Calcul Scientifique (BAP E)
- **Ville :** Talence
- **Centre Inria :** CRI Bordeaux - Sud-Ouest
- **Date de prise de fonction souhaitée :** 2019-10-01
- **Durée de contrat :** 3 ans
- **Date limite pour postuler :** 2019-05-31

Contacts

- **Equipe Inria :** HIEPACS
- **Directeur de thèse :**
  Beaumont Olivier / olivier.beaumont@inria.fr

A propos d'Inria

Inria, l'institut national de recherche dédié aux sciences du numérique, promeut l'excellence scientifique et le transfert pour avoir le plus grand impact. Il emploie 2400 personnes. Ses 200 équipes-projets agiles, en général communes avec des partenaires académiques, impliquent plus de 3000 scientifiques pour relever les défis des sciences informatiques et mathématiques, souvent à l'interface d'autres disciplines. Inria travaille avec de nombreuses entreprises et a accompagné la création de plus de 160 start-up. L'institut s'efforce ainsi de répondre aux enjeux de la transformation numérique de la science, de la société et de l'économie.

L'essentiel pour réussir

This thesis requires some taste for modelling, algorithmic design and combinatorial problems. The goal will be to work first on toy task graphs and platform models to understand the problems, their complexity and possible approaches. Then, we will move to actual applications (coming from Linear Algebra, Tensor Computations, Learning) and implement our solutions on top of runtime systems such as StarPU.

Consignes pour postuler

Thank you to send CV + cover letter + support letter

Sécurité défense :
Ce poste est susceptible d'être affecté dans une zone à régime restrictif (ZRR), telle que définie dans le décret n°2011-1425 relatif à la protection du potentiel scientifique et technique de la nation (PPST). L'autorisation d'accès à une zone est délivrée par le chef d'établissement, après avis ministériel favorable, tel que défini dans l'arrêté du 03 juillet 2012, relatif à la PPST. Un avis ministériel défavorable pour un poste affecté dans une ZRR aurait pour conséquence l'annulation du recrutement.

Politique de recrutement :
Dans le cadre de sa politique diversité, tous les postes Inria sont accessibles aux personnes en situation de handicap.

Attention: Les candidatures doivent être déposées en ligne sur le site Inria. Le traitement des candidatures adressées par d'autres canaux n'est pas garanti.