2018-00722 - Memory-aware scheduling for distributed large-scale platforms [Phd campaign]

Niveau de diplôme exigé : Bac + 5 ou équivalent
Fonction : Docteurant

A propos du centre ou de la direction fonctionnelle

Inria the French national institute for research in computer science and control, is dedicated to fundamental and applied research in information and communication science and technology (ICST). Inria has a workforce of 3,800 people working throughout its eight research centers established in seven regions of France.

Grenoble is the capital city of the French Alpes. Combining the urban life-style of southern France with a unique mountain setting, it is ideally situated for outdoor activities. The Grenoble area is today an important centre of industry and science (second largest in France). Dedicated to an ambitious policy in the arts, the city is host to numerous cultural institutions. With 60,000 students (including 6,000 foreign students), Grenoble is the third largest student area in France.

Contexte et atouts du poste

The PhD will take place in the ROMA Inria project-team which aims at designing models, algorithms, and scheduling strategies to optimize the execution of scientific applications. It is lead by Frédéric Vivien and is organized around six themes: static algorithms for dynamic environments, direct solvers for sparse linear systems, memory-aware algorithms, linear algebra on post-petascale multicore platforms, multi-criteria optimization and combinatorial scientific computing.

Key words : Scheduling ; High-Performance ; Computing Memory-aware

Mission confiée

This PhD will study the problem of scheduling scientific applications under memory constraints. The goal is to extend the results of the team in this area, which are for now limited to shared-memory platforms. The objective is to derive algorithmic strategies on a larger scale, and especially for distributed memory platforms.

Principales activités

Scientific applications are usually modeled as Directed Acyclic task Graphs (DAGs of tasks) and they often involve data of large sizes [1]. However, the computing platforms on which these applications are to be processed have limited memory capabilities, which are distributed among the available computing nodes. Moving data among these nodes is costly, and reading/writing data from/to disks (usually of much less limited size) is even more time consuming.

There have been a number of studies focusing on data locality to optimize the mapping and scheduling of such applications. In particular, in the context of the past ANR project SOLHAR, the ROMA team has studied the problem of scheduling such task graphs with limited memory, and has proposed algorithms to bound the memory requirement of an execution [2,3,4]. However, most of these studies are limited to the case of a single memory, shared among the computing cores. The objective of this thesis is to extend these results to larger-scale platforms, composed of several nodes equipped with their own memory. There are several challenges that have to be taken into account when designing scheduling strategies for such platforms:

* One could think of the different distributed memories as chunks of a large, shared memory, as one node may access the memory of another node. However, this type of access is much slower than accessing a local memory bank, hence transfer times have to be taken into account to avoid moving time back and forth.

* When going large-scale, we cannot rely anymore on a centralised scheduler: most scheduling decisions have to be made locally and to avoid the frequent synchronisation of all nodes.

* Task graphs to be processed on such large-scale platforms are likely to include a large number of tasks. However, scheduling decisions have to be made in a very short time. Thus, scheduling...
algorithms cannot rely on complex graph algorithms, as proposed in current solutions. Nevertheless, part of the scheduling process can be performed beforehand, by pre-computing important information on the graph, or as a background task, while other cores are busy computing tasks.

A promising concept to tackle all three challenges is the use of hierarchical scheduling: tasks of the graph may be clustered into coarser-grain groups of tasks. Such coarse-grain groups may be scheduled onto the distributed nodes, and the schedule may be locally refined when opening the coarse-grain groups to reveal the original tasks. The thesis will consist in exploring such a hierarchical approach as well as other possible approaches, such as adapting classical work-stealing scheduling algorithms for memory-bounded scenarios.


Compétences
The candidate will be required to have a solid background in algorithms, be familiar with scheduling problems and a good level in programming with usual languages (C, C++, python).

Avantages sociaux
- Subsidised catering service
- Partially-reimbursed public transport
- Social security
- Paid leave
- Flexible working hours
- Sports facilities

Rémunération

Monthly salary after taxes : around 1596,05€ for 1st and 2nd year. 1678,99€ for 3rd year. (medical insurance included).