Contribution ID: **81**   Type: **Oral**

Distributed evolutionary optimization framework on the Grid and HEP use case

Wednesday, 14 April 2010 14:20 (20 minutes)

This paper describes an optimization framework which utilizes a distributed evolutionary algorithm with the ability of adaptation to the complex physical structure of the grid fabric. The algorithm is divided into several partitions (demes) which are located on physical clusters. This architecture allows simultaneous use of all the available resources, regardless of their geographical location, within one application. The paper also presents an example use case of this technique, which is track reconstruction optimization for COMPASS experiment.

Detailed analysis

Numerical optimization by means of EA (Evolutionary Algorithms) became popular in many disciplines of technology and science because it is more resistant to local optima than classical, gradient-based methods. On the other hand, EA require much more computational power, and this limits their usage. The grid environment, with its huge number of CPUs,seems to be a perfect environment for this kind of applications. The main difficulty here is how to use Grid resources, coming from many clusters, in a coordinated way within one program, without huge performance losses on communication. In this application we adopted the multiple-deme architecture of EA in which every cluster is occupied by separate master-slave populations. These populations (called demes) occasionally exchange their best solutions, but most of the communication takes place within the cluster. For internal communication each population use MPI messages, and for external (between demes), files on the Storage Elements accessed via GFAL. This way, there is no need for non-standard extensions to grid middleware, and the performance penalty caused by slow access to files is marginal (due to the low intensity of this kind of communication)

Conclusions and Future Work

The tests performed on the example application show that even big differences between processing speeds of the clusters hosting individual demes have little influence on the final solution quality, making the method very useful. However, slow communication between demes may be the limiting factor when using the program to solve problems with less computationally expensive objective functions (and thus faster swapping of generations). To extend the functionality we plan to introduce faster communication methods to the migration operator, like PACX-MPI, or DIANE messages.

Impact

In the scientific world there is a strong need for general-purpose optimization frameworks. This is confirmed by a vast number of tools of this kind available in the Internet, offering broad variety of optimization algorithms, including EA. These tools usually work in abstraction from the hardware layer, making it possible to run on many platforms. This comes at a price. It is virtually impossible to run such a general-purpose tool in a highly heterogeneous grid environment and utilize all the available resources in a coordinated way. Our program is aware of the structure of the underlying fabric and is specifically designed to minimize the

losses caused by communication bottlenecks. This allows to use EA-based optimization for the problems too big to fit on one supercomputer. Development of our program was initially driven by one application, which is track reconstruction optimization in COMPASS experiment. Still, a considerable effort has been put into maintaining the modular structure of the program. It is thus possible to easily adapt it to other problems, changing only a couple of options in the configuration file. For this reason the framework can be useful for other users.

Keywords

Evolutionary Algorithms, Computational Grids

URL for further information

http://wiki.polgrid.pl/index.php/DEAG

Primary author: Mr PADEE, Adam (Warsaw University of Technology)

Co-author: Prof. ZAREMBA, Krzysztof (Warsaw University of Technology)

Presenter: Mr PADEE, Adam (Warsaw University of Technology)

Session Classification: Computer Science

Track Classification: Scientific results obtained using distributed computing technologies