Laboratoire d'Informatique de Grenoble

UMR 5217 - CNRS, INPG, INRIA, UJF, UPMF

Adresse : Inria Grenoble Rhône-Alpes
Inovallée
655 avenue de l'Europe
Montbonnot Saint Martin
38334 St Ismier Cedex
France

Téléphone : (+33) 4 76 61 20 64
Télécopie : (+33) 4 76 61 20 99
Adresse électronique : mailto:arnaud.legrand@imag.fr

Arnaud Legrand, CNRS researcher
INRIA **MESCAL** project, LIG Laboratory

Proposal for a M2R internship
Fair Response Time Optimization of Multi-User BOINC Projects

Advisors: Arnaud Legrand
Required Skills:

- C programming, UNIX, shell, ssh, git
- Basics of experiment analysis/workload characterization with R and fundamentals of scheduling are a plus

1 Context

Recent evolutions in inter-networking technology and the decreasing cost-performance ratio of computing components have enabled Volunteer Computing (VC) systems. Among these platforms, BOINC (Berkley Open Infrastructure for Network Computing) [And04] is the most representative, with over 580,000 hosts that deliver over 2,300 TeraFLOP per day and many projects deployed. In such platforms, volunteers usually donate their machines' CPU idle time to scientific projects and indicate some preferences on how to share their processing power among them. BOINC projects usually have hundreds of thousands of independent CPU-bound tasks and are interested in overall throughput, i.e., maximize the total number of tasks completed per day. Each project has its own server which is responsible for distributing work units to clients, recovering results and validating them.

To handle this complexity, BOINC implements a simple and robust distributed protocol which ensures that local shares and volunteer priorities are respected. This protocol has been used for many years now and its efficiency and fairness have been assessed in the context of throughput oriented projects [AMV07, KAM07, AR09].

Lately, the BOINC workload has evolved in the following two noticeable ways:

- First, to answer the burden of deploying and maintaining BOINC servers, multi-user projects (also called *umbrella* projects) have appeared. Such BOINC project host several scientific projects (in the later, we will talk about users rather than scientific projects to avoid confusion with BOINC projects), each with their own application, which provides a much larger audience to every user than if they had deployed their own server. For example, the World Community Grid, funded by IBM, has been one of the first such BOINC projects, helping research on drug search for Leishmaniasis, clean water localization, childhood cancer, ... More recently, CAS@home a Chinese project has been set up to help both protein structure, nanotechnology, cancer genomics and high energy physics research. Likewise Ibercivis a Spanish-Portuguese initiative has been set up to provide a huge BOINC umbrella Project for scientists in those countries and Yoyo@home helps research on harmonious trees, elliptic curve factorization, and DNA evolution among others. In such BOINC projects, demand for computing power generally exceeds supply, hence a need for a mechanism to divide the supply among users.
- Second, although in the past most users had an unlimited supply of jobs to process (throughput-oriented users), the popularization of BOINC has attracted other type of users. In CAS@home, Many users have campaigns (batches) made of fewer tasks and are interested in response time of campaigns. For such users, throughput is not meaningful and response time should be minimized (response-time oriented users).

In such multi-user projects, each such user will have a numeric quota representing the fraction of the computing resource they should get. What are the precise semantics of quotas? Ideally the mechanism should accommodate the previous two classes of users. The GridBot project [SSGS09, Sil11, BYSS+12] recently made use of hybrid computing platform composed of grid, clusters and VC to execute workload resulting from mix of throughput and response time oriented applications. Such initiatives are very encouraging but somehow *ad hoc* and does not consider the problem of streams of batches nor the resource sharing issue. We think that regular BOINC servers can be improved to handle seamlessly such workloads.

2 Environment

The members of the MESCAL team focus their research on large scale systems and parallel applications. They have a strong expertise regarding parallel algorithms, environment for parallel programming, performance evaluation of large scale distributed systems, middleware for clusters and grids and scheduling. The MESCAL team is associated with the **Berkeley** laboratory in which the BOINC middleware is developed. We have thus very strong relations with the main BOINC developers and with the project leader, David Anderson.

The MESCAL team also has a strong expertise in simulation [DCLV10], scheduling [KCC07, DLG11] and workload characterization [KTB+04, KJIE10, JKVA11] of such volunteer computing systems. Hence, the MESCAL team is involved in the recent ANR French National project SONGS (Simulation Of Next Generation Systems) whose goal is to design a unified and open simulation framework for performance evaluation of next generation systems. The SONGS projects is a follow-up of the ANR USS-SimGrid project, which is based on SimGrid, an open-source toolkit that provides core functionalities for the simulation of distributed applications in heterogeneous distributed environments.

3 Goal

In the last BOINC workshops (2011, 2012, 2013) members from the Mescal team, the BOINC team, the World Community Grid project and the OurGrid project have brainstormed to design a new server architecture that would enable to fairly and efficiently handle complex workloads such as the ones we previously described. Yet, many parts of this architecture (batch prioritizing, batch completion estimation, client reliability estimation, ...) remain to be specified.

The goal of this internship is thus to first implement a simulation of such prototype and to evaluate its performance under some realistic workload from the World Community Grid or from CAS@home. Such evaluation will then guide the design of different scheduling and estimation mechanisms. The project could involve visits to Berkeley, California to present results, which will later be implemented in the BOINC server and deployed on the World Community Grid and in CAS@home.

References

[AMV07] David P. Anderson and John McLeod VII. Local Scheduling for Volunteer Computing. In *Proc. of the Intl. Parallel and Distributed Processing Symp. (IPDPS 2007)*. IEEE Intl., 2007.

[And04] D. P. Anderson. BOINC: a system for public-resource computing and storage. In *Proc. of the 5th IEEE/ACM Intl. Workshop on Grid Computing*, 2004.

[AR09] David P. Anderson and Kevin Reed. Celebrating Diversity in Volunteer Computing. In *System Sciences, 2009. HICSS '09. 42nd Hawaii*, January 2009.

[BYSS+12] Orna Agmon Ben-Yehuda, Assaf Schuster, Artyom Sharov, Mark Silberstein, and Alexandru Iosup. Expert: Pareto-efficient task replication on grids and a cloud. In *IPDPS*, pages 167–178, 2012.

[DCLV10] Bruno Donassolo, Henri Casanova, Arnaud Legrand, and Pedro Velho. Fast and Scalable Simulation of Volunteer Computing Systems Using SimGrid. In *Workshop on Large-Scale System and Application Performance (LSAP)*, 2010.

[DLG11] Bruno Donassolo, Arnaud Legrand, and Claudio Geyer. Non-Cooperative Scheduling Considered Harmful in Collaborative Volunteer Computing Environments. In *Proceedings of the 11th IEEE International Symposium on Cluster Computing and the Grid (CCGrid'11)*. IEEE Computer Society Press, may 2011.

[JKVA11] Bahman Javadi, Derrick Kondo, Jean-Marc Vincent, and D.P. Anderson. Discovering Statistical Models of Availability in Large Distributed Systems: An Empirical Study of SETI@home. *IEEE Transactions on Parallel and Distributed Systems*, 99(preprint), 2011.

[KAM07] Derrick Kondo, David P. Anderson, and John VII McLeod. Performance evaluation of scheduling policies for volunteer computing. In *Proceedings of the 3rd IEEE International Conference on e-Science and Grid Computing e-Science'07*, Bangalore, India, dec 2007.

[KCC07] Derrrick Kondo, Andrew A. Chien, and Henri Casanova. Scheduling task parallel applications for rapid application turnaround on enterprise desktop grids. *Journal of Grid Computing*, 5(4), 2007.

[KJIE10] D. Kondo, B. Javadi, A. Iosup, and D. Epema. The Failure Trace Archive: Enabling Comparative Analysis of Failures in Diverse Distributed Systems. In *Proc. of the IEEE Int. Symp. on Cluster Computing and the Grid (CCGrid)*, 2010.

[KTB+04] Derrrick Kondo, M. Taufer, C. Brooks, Henri Casanova, and Andrew A. Chien. Characterizing and Evaluating Desktop Grids: An Empirical Study. In *Proc. of the Intl. Parallel and Distributed Processing Symp. (IPDPS)*, 2004.

[Sil11] Mark Silberstein. Building an online domain-specific computing service over non-dedicated grid and cloud resources: The superlink-online experience. In *11th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing, CCGrid 2011, Newport Beach, CA, USA, May 23-26, 2011*, pages 174–183. IEEE, 2011.

[SSGS09] Mark Silberstein, Artyom Sharov, Dan Geiger, and Assaf Schuster. Gridbot: execution of bags of tasks in multiple grids. In *Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis*, SC '09, New York, NY, USA, 2009. ACM.