Parallel and Distributed Simulation of Large-Scale Distributed Applications

Executive summary: The context of this project is to allow the efficient parallel and distributed simulation of large systems within the SimGrid framework. The proposed work will improve the existing parallel simulation mode, and propose a novel distributed simulation mode. We target a simulation comprising millions of heavy computational nodes on a much smaller cluster.

Key skill required: System programming and networking programming in C on Linux.

Research Team:	AlGorille (http://www.loria.fr/equipes/algorille/)
Research Unit:	Nancy – Grand Est
Intern Tutor:	Martin Quinson (http://www.loria.fr/~quinson/)
Intern level:	Master student (or PhD student)
Internship duration:	4 to 6 months
Followed by a PhD:	possible (but not mandatory)

Context

Recent and foreseen technical evolution allow to build information systems of unprecedented dimensions. The potential power of the resulting distributed systems offers new possibilities in terms of applications, be them scientific such as multi-physic simulations in High Performance Computing (HPC), commercial in the Cloud with the data centers underlying the Internet, or public in very large peer-to-peer systems. For example, ExaScale systems in the HPC area are expected to aggregate millions of high end compute nodes by the end of this decade for unpreceded scientific computations.

Evaluating computer systems of this extreme scale raises severe methodological challenges. Simply executing them is not always possible as it requires to build the complete system beforehand (what is not possible for ExaScale systems for example), and it may not even be enough when uncontrolled external load prevents reproducibility. Simulation is an appealing alternative to study such systems. It may not capture the whole complexity of every phenomena, but allows to easily capture some important trends, while ensuring the controllability and reproducibility of experiments.

SimGrid[1] (developed by the AlGorille team in collaboration) is a toolkit providing core functionalities for the simulation of distributed applications in heterogeneous distributed environments. The specific goal of the project is to facilitate research in the area of distributed and parallel application scheduling on distributed computing platforms ranging from simple network of workstations to Computational Grids.

This framework was shown orders of magnitude faster than concurrent simulators such as GridSim or PeerSim, and can simulate up to a few million lightweighted P2P processes on a single node [LAM+12]. This falls however short to simulate ExaScale systems, as these systems are expected to count dozen of millions of heavy processes. Both CPU and memory

[1] The SimGrid Project: http://simgrid.gforge.inria.fr/

limitations must be overtaken to scale the simulation further. In a previous work, we shown that parallel simulation can improve the computational performance in some cases [QRT12], but the memory limitation claim for the distribution of the simulation to leverage the memory of several computational nodes.

[LAM+12] Laurent Bobelin, Arnaud Legrand, David Marquez, Pierre Navarro, Martin Quinson, Frédéric Suter, Christophe Thiéry. *Scalable Multi-Purpose Network Representation for Large Scale Distributed System Simulation.* 12th Intl Symposium on Cluster Computing and the Grid (CCGrid'12), 2012. http://hal.inria.fr/hal-00650233

[QRT12] Martin Quinson, Cristian Rosa, Christophe Thiéry. *Parallel Simulation of Peer-to-Peer Systems.* 12th ACM/IEEE Intl Symposium on Cluster Computing and the Grid (CCGrid'12), Canada, May 2012. http://hal.inria.fr/inria-00602216

Precise Work Description

Even if the existing *parallel* execution mode described in [QRT12] is several orders of magnitude faster than the concurrent simulators, it is still improvable in several ways. More specifically, the goals of this internship are:

- The current parallel mode is implemented using futex primitives for thread synchronization. As these construct only exist on Linux, it is necessary to provide a backup implementation using portable constructs (such as POSIX primitives). The performance loss induced by this portability should then be properly evaluated.
- Our novel approach to parallel discrete-event simulation was only compared to high-level applicative simulators; Packet-level network simulators are believed to exhibit lower performance, but this should be properly evaluated anyway.
- Also, these contribution were only evaluated using the Chord P2P protocol as a workload. It is believed that such a fine grain workload constitutes the worst case for our contribution, but this should be evaluated using other workloads too, possibly from differing research domains such as HPC and Grids.

A new *distributed* execution mode will be designed during this internship to overcome memory limitations in very large scenarios.

- Several designs are possible to that extend. The intern is expected to develop several proof of concepts to understand their relative advantages. S/he will then select the best design through a careful evaluation.
- A complete framework toward distributed simulation, based on these proof of concepts, is then to be proposed by the intern.
- This work will be evaluated experimentally and compared to state of the art solutions.
- The ultimate goal is to run a typical HPC application (such as linpack, used for the Top 500 ranking[2]) using a sizable portion of the Grid'5000 experimental facility[3].
- Finally, the intern is also expected to participate to a publication presenting the findings of this research project.

Skills required

In addition to the skills that can reasonably be expected from Master-level students, the applicant should have a strong knowledge of system programming in C, and of modern Unix-based Operating Systems such as Linux.

[2] Top 500 SuperComputing ranking: http://www.top500.org/
[3] The Grid'5000 Scientific Instrument: https://www.grid5000.fr/