Simulation for Large-Scale Distributed Computing Research

Abstract

This tutorial will provide attendees with clear perspectives on the challenges for experimental research in the area of parallel and large-scale distributed computing, and on current technology for conducting experiments with real-world testbeds, emulated testbeds, or simulated testbeds. The first part of the tutorial will present and contrast current experimental methodologies, giving attendees in-depth understanding of the scientific and technological issues at hand. The second part of the tutorial will focus on simulation, giving a state of the art of current simulation technology and discussing challenges for the development of sound simulation models. The tutorial will use the SimGrid simulation framework as an exemplar since it implements sophisticated and validated simulation models. The last part of the tutorial will focus on an in-depth presentation of the different simulation approaches enabled by SimGrid, each with its specific range of applications and goals. SimGrid has been used to obtain results published in over 50 research articles and has thus emerged as one of the key tools for simulation in the area of parallel and large-scale distributed computing. Tutorial attendees will have the opportunity to gain some hands-on experience with SimGrid, by witnessing step-by-step development of small simulation projects. By the end of this tutorial attendees should have a clear understanding of current technology and best practice for experimental parallel large-scale distributed computing research, and in particular on the use of simulation.

Intended audience

This tutorial is intended for attendees who are computer scientists with knowledge and interest in parallel and large-scale distributed computing. This tutorial should be extremely relevant for researchers interested in parallel and distributed algorithms (e.g., for application scheduling, resource management, data management), which corresponds to a large subset of the researchers attending PDCAT each year. Reasonable programming skills are expected, either in C or Java, so that attendees can follow example source codes.

Outline and schedule

- **Introduction** (30 minutes)
  - Needs and challenges for experimental large-sale distributed computing research
  - Comparison of current methodological approaches
    * real-world execution
    * emulation
    * simulation
    * hybrid approaches

- **Simulating parallel and large-scale distribution platforms and applications** (50 minutes)
  - State of the art of relevant simulation tools
  - Challenges for developing sound simulation models
  - Case in point: simulation models in the SimGrid framework and their validation
  - Instantiation of simulated platforms

- **Using SimGrid** (5 + 3 * 25 minutes)
  - Overview of SimGrid components and APIs
  - The SimDAG interface: simulating the execution of DAG-structured applications
    * Goals of the interface
    * An example simulation
  - The MSG interface: simulating asynchronous communicating processes
    * Goals of the interface
    * Hands-on experience: step-by-step development of a small simulation project
The GRAS interface: development of distributed applications for simultaneous simulated and real-world execution

- Goals of the interface
- Hands-on experience: step-by-step development of a small distributed application

**Conclusion (20 minutes)**
- Summary of the tutorial
- SimGrid success stories
- Perspectives for the next 5 years

**Possible tutorial speakers**

If the tutorial gets accepted by the organizer, the presence of Martin Quinson as tutorial speaker is confirmed, but the presence of the two other members of the team is still pending confirmation.

**Arnaud Legrand** is a junior CNRS researcher in the computer science laboratory of Grenoble (LIG). He works in the INRIA MESCAL project and his research focus on scheduling and performance evaluation of large-scale distributed systems. He has published over 50 research articles in peer-reviewed journals and conferences. He obtained his M.S from the Ecole Nationale Superieure of Lyon, France in 2000, and his Ph.D. from the Ecole Nationale Superieure of Lyon, France in 2003. Arnaud Legrand is one of the main developers of the SimGrid project since 1999.

**Martin Quinson** is an Associate Professor in the School of Computer Science and Applications of Lorraine at University of Nancy. His research interests are distributed, grid and internet computing. In particular his research emphasizes the development of distributed services over large-scale distributed platforms, assessing the quality of distributed applications and the experimental evaluation of distributed algorithms. He has published over 15 research articles in peer-reviewed journals and conferences. He obtained his B.S from Universite Jean Monet of Saint Etienne, France in 1999, his M.S from the Ecole Nationale Superieure of Lyon, France in 2000, and his Ph.D. from the Ecole Nationale Superieure of Lyon, France in 2003. He is a program committee member of the SIMUTools conference since 2008 and of the CCGrid 2009 conference. Martin Quinson is one of the main developers of the SimGrid project since 2002.

**Henri Casanova** is an Associate Professor in the Information and Computer Sciences Department at the University of Hawai'i at Manoa. His research interests are in the area of parallel, distributed, grid, and internet computing. In particular, his research emphasizes the modeling and the simulation of distributed platforms and applications, as well as both the theoretical and practical aspects of scheduling problems. He has published over 70 research articles in peer-reviewed journals and conferences. He obtained his B.S. from the Ecole Nationale Superieure d'Electronique, d'Electrotechnique, d'Informatique et d'Hydraulique de Toulouse, France in 1993, his M.S. from the Universite Paul Sabatier, Toulouse, France in 1994, and his Ph.D. from the University of Tennessee, Knoxville in 1998. Prior to joining the University of Hawaii, he was a Research Scientist at the San Diego Supercomputer Center and an Adjunct Professor in the Dept. of Computer Science and Engineering at the University of California, San Diego. Henri Casanova is the original designer and developer of the SimGrid project.

**Contact person**

Martin Quinson ([martin.quinson@loria.fr](mailto:martin.quinson@loria.fr) – (+33/0) 3 83 59 20 98)
LORIA / Campus Scientifique
BP 239 - 54506 Vandoeuvre-lès-Nancy Cedex France
http://www.loria.fr/~quinson

**Tutorial history**

Previous versions of this tutorial were presented at the PDCS’2007 conference by Martin Quinson (30 participants) and at the CCGrid’2008 conference by Martin Quinson and Arnaud Legrand (40 participants).

A draft of the tutorial is available at: [http://www.loria.fr/~quinson/articles/simgrid-tutorial.pdf](http://www.loria.fr/~quinson/articles/simgrid-tutorial.pdf)