Chapter 28
A General Framework for the Modeling and Simulation of Grid and P2P Systems

Ciprian Dobre
University "Politehnica" of Bucharest, Romania

ABSTRACT

The field of modeling and simulation was long seen as a viable alternative to develop new algorithms and technologies and to enable the development of large-scale distributed systems, where analytical validations are prohibited by the nature of the encountered problems. The use of discrete-event simulators in the design and development of large-scale distributed systems is appealing due to their efficiency and scalability. In this chapter we focus on the challenge to enable scalable, high-level, online simulation of applications, middleware, resources and networks to support scientific and systematic study of Grid and P2P applications and environments. We describe alternatives to designing and implementing simulators to be used in the validation of distributed systems, particularly Grid and P2Ps.

DOI: 10.4018/978-1-61520-686-5.ch028

INTRODUCTION

In the broad area of distributed systems researchers often ask questions such as which scheduling algorithm is best suitable for deploying an application on a given Grid or which caching strategy serves better a community of users that are working in distributed data analysis. The answers to these questions can be obtained in several ways. A solution is to develop purely analytical or mathematical models, but this often leads to NP-complete problems, such as routing, partitioning or scheduling, for which no analytical solution can be found. Another solution consists in conducting live experiments. Unfortunately there are no standard approaches to conduct live experiments on large-scale distributed systems. Real-world experiments can be time-intensive, since the execution of applications could last for hours, days, month or even more. In the same time real-world experiments can be labor-intensive, since the entire application needs to be built and functional. The choosing of the right experimental testbed is yet another problem. Real platforms can experience failures that may disrupt the experiments and the
platform configurations may change drastically while experiments are being conducted. Real-life experiments in most cases are uncontrolled and unrepeatable. And to make things worse, live experiments are limited to testbeds (the particular capabilities of the testbed can affect the outcome) and the obtained results can not be reproduced by others (which is in fact the basis for scientific advances).

Because of such problems simulation can prove to be a more elegant solution to conduct experiments. This chapter focuses on the challenge to enable scalable, high-level, online simulation of applications, middleware, resources and networks to support scientific and systematic study of Grid and P2P applications and environments. The field of modeling and simulation was long-time seen as a viable alternative to develop new algorithms and technologies and to enable the development of large-scale distributed systems, where analytical validations are prohibited by the nature of the encountered problems. The use of discrete-event simulators in the design and development of large scale distributed systems is appealing due to their efficiency and scalability. Their core abstractions of process and event map neatly to the components and interactions of modern-day distributed systems and allow the design of realistic scenarios. Compared with the alternative of implementing a new technology directly in real-world to demonstrate its viability, the simulation of distributed systems is a far better alternative because it achieves faster validation results, minimizing the costs involved by the deployment process. Several key advantages make it attractive when compared with conducting live experiments: simulation assumes that there is no need to build a real system, the simulated experiments are conducted in a controlled and repeatable fashion, the simulated scenarios have no limitations, and anyone can reproduce the obtained results.

Throughout this we describe new alternatives to designing and implementing simulators to be used in the validation of distributed system technologies, particularly Grid and P2P related. The chapter is organized as follows. We first present the characteristics of large scale distributed systems such as Grids and P2Ps. We present their specific requirements in terms of modeling and simulation. We then present a comparison study of the most important simulation projects involved in the modeling of distributed systems. The critical analysis is based on the categories proposed for an original taxonomy for comparing simulation instruments for Grid and P2P systems. The proposed taxonomy is particularly focused on the simulation of such systems, hence it introduces categories such as motivation and specific components that are particularly designed to properly categorize the special family of Grid and/or P2P simulation instruments.

BACKGROUND

Distributed systems have become very useful especially for complex scientific applications, involving the processing of very large data volumes, in a very short amount of time, as well as the storage of these data. Taking into account the tremendous popularity of complex distributed systems, favored by the rapid development of computing systems, high speed networks, and the Internet, it is clear that it is imperative, in order to achieve adequate performances in the utilization of these systems, to pick optimal architectures and solutions for designing and deploying distributed systems.

Currently there are many forms of distributed system architectures. Computing in the late 1990s has reached the state of Web-based distributed computing. A basis of this form of computing is distributed computing which is carried out on distributed computing systems. Distributed network systems are now used everywhere. As an example we can recall the well-known network file systems based on the client-server model, the Sun Micro-soft’s Network File System, known as NFS, that