Grid and Cyber-infrastructure Activities at UPRM

Jaime Seguel
Electrical and Computer Engineering Department
University of Puerto Rico at Mayagüez

Abstract. The College of Engineering of the University of Puerto Rico at Mayagüez (UPRM) is pioneering efforts for harnessing distributed and collaborative computing environments, and using them for solving problems of national relevance. Most remarkable among these efforts is the Latin American Grid (LA Grid), an IBM sponsored consortium of Hispanic universities and supercomputing centers aimed at advancing grid computing and its applications. This short paper outlines LA Grid aims and structure, and some of research projects that are expected to be performed on it both, at the computing systems and at the applications layer.

Cyber-infrastructure

The term *Cyber-infrastructure* was coined by NSF to describe emerging research environments in which computing capabilities are made available to researchers in an interoperable computer network. Cyber-infrastructure environments are built on the basis of high quality, service-oriented distributed computer architectures. Advanced information technologies for supporting “intelligent” searching, semantic integration, and visualization of information as well as scientific databases are essential to configure a cyber-infrastructure. Access to high performance computing platforms is a crucial ingredient, as well. The expected result of a cyber-infrastructure deployment is a mature and robust integrative environment fostering and enabling transformative advances in research and education.

Science and education are still pretty much isolated activities, with little disciplinary and even less cross-disciplinary interaction. In fact, it is quite fair to visualize scientists and educators as a distributed system, distributed both, geographically and intellectually. And although the benefits of integration and cross-disciplinary collaboration are well-understood, the actual creation and sustenance of collaborative environments is often difficult to achieve in practice. Overcoming a number of sometimes subtle institutional, disciplinary, or cultural barriers require a long time effort and a significant investment in time and resources which often renders low cost-effectiveness. Information technology promises to change this by offering educators and researchers a potential to exploit an uncharted territories and trying new approaches to education and research. At the foundations of these potentially new approaches are the old libraries, archives, and museums that have preserved information per centuries; the bibliographies, finding aids and citation systems that have been making information retrievable; the journals that distribute the information; and the

editors, librarians, archivists, and curators that link the operation of this structure to the scholars and educators who use it. Cyber-infrastructure not only recasts these structures built over centuries in a modern digital realm but offers an unprecedented opportunity to automate the editors, librarians, archivist, and curators work. This is certainly not an easy task. Enabling virtual laboratories, distributed research organizations, technology-enabled education environments, requires still significant advances in networking systems and distributed computing. This challenge and its opportunities are outlined in the Blue Ribbon Committee on Cyberinfrastructre Report (report to NSF, 2003): *"We are at a threshold where a collaboratory or grid community can become "the place" where a research community interacts with colleagues, data, literature and observational systems together with very powerful computational models and services. Although many technical, social, and economical challenges remain, the potential exists for facilitating both, deeper and broader scientific and engineering research and education."*

The Latin American Grid

Aware of cyber-infrastructure challenges and opportunities, and quite conscious of their institutional responsibilities with the Hispanic population, faculty from the School of Computer Science of the Florida International University, and from the doctoral program in Computer Sciences and Engineering of the University of Puerto Rico at Mayagüez join efforts with IBM to deploy the Latin American Grid (LA Grid).

The Hispanic minority group is the largest and fastest growing ethnic segment in the United States. Currently, Hispanics constitute about 14% of the US population, a percentage that is expected to reach about 25% by the year 2050. However, the Hispanic participation in Computer Science and Engineering (CSE) is disproportionately low. Only 3.9% of bachelors conferred in 2003, no more 1.3% of Masters and less than 1% of PhDs turn out to be Hispanics. The Latin American Grid (LA Grid) aims at closing this gap by providing a well-focused approach towards Hispanic CSE research and professional development. LA Grid is developing a multi-university Computer Grid for education, research, collaboration and promotion of Hispanic Technical Professionals, Researchers and Educators in the emerging and challenging computer science field of grid computation and in its applications to science, business, and engineering.

The Latin American Grid has grown rapidly. The consortium is currently formed by Florida International University, University of Miami, Monterrey Technology Institute (Mexico), and University of Puerto Rico at Mayagüez, together with the Barcelona Supercomputing Center and the IBM Research Division. Additional Latin American institutions, especially Brazilian, Venezuelan, and Chilean universities, are expected to join LA Grid in the medium term. The consortium is governed by an Advisory Board which provides business, technology, and diversity direction, a Governance Board, which provides operational guidance,

and a Technical/Executive Board in charge of developing and maintaining the operation.

The Puerto Rican component

Under the provisions of its 1952 constitution, Puerto Rico is a commonwealth freely associated with the United States. The Puerto Rican government maintains control over local issues, but the island is required to comply with most federal legislation. Puerto Ricans are citizens of the United States.

Before WW II Puerto Rican economy was based on agriculture. With the help of the U.S. federal government, in 1947 the Puerto Rican government established Operation Bootstrap, a program of governmental support for industry through tax breaks. The government hoped to attract industries that would import goods to the island to be finished for export. The Puerto Rican government also worked to develop the tourist industry.

Puerto Rico greatly improved its educational institutions throughout the 20th century. By 2003, 94.3 percent of the adult population was literate, compared with some 67 percent in 1940. In 1998-1999 the commonwealth had 51 private and 14 public institutions of higher education with a combined enrollment of 164,000 students.

The University of Puerto Rico (UPR) was founded in 1903. It is the oldest institution of higher education in Puerto Rico. Currently the UPR has an enrollment of about 74,000 students distributed among its 11 campuses, including the three graduate degree granting campuses: Rio Piedras, Medical Sciences, and Mayagüez (UPRM). The UPRM also homes most of the graduate programs in Information Technology in the UPR system: Masters in Computer Engineering (ECE Department), a Masters in Scientific Computing (Math Department) and the Doctoral Program in Computing and Information Sciences and Engineering (jointly offered by ECE and Math Departments). These programs concentrate most of the research activity in Information Technology in the Island, as well.

Since 1983, as manufacturing declines, Puerto Rican governments have launched initiatives for fostering a transformation of the Island's economic base to what is often termed knowledge based economy. Singapore, Ireland are mentioned as models for this change. Collaboration between academia, government and industry is regarded as the means to produce this transformation. Efforts concentrate in creating a human and intellectual capital; this is well-trained innovators in science and engineering together with a well-rounded portfolio of inventions; and effective technology transfer mechanisms. LA Grid is destined to play a mayor role as enabler in all these efforts.

Scheduling: a sample research problem

Harnessing the power of grid computing strongly depends on the quality of the method employed to map tasks into the computer nodes, and the time at which the tasks are scheduled for execution. In most practical single-node situations, the optimization of a scheduling mapping is intractable, *i.e.* NP-Hard. Large, wide-area distributed, heterogeneous systems can only worsen the situation. Orchestrating different hardware, operating systems, scheduling and administration policies; and anticipating unexpected node or communication failures, or communication congestions is beyond the capability of traditional theoretical analysis. Furthermore, large, computer grids are likely to receive jobs on-line, from different sources, bringing about a significant degree of uncertainty regarding the size of the job, resources available and other information that is necessary to optimize the classical scheduling problem. As a consequence, traditional scheduling optimization techniques, which assume a perfect knowledge of the scheduling problem information, are not longer applicable.

We propose to explore a variant of what is known as steady-state scheduling. This approach produces a periodic schedule that optimizes the operation of the system in steady-state. Although is still a research subject, strong theoretical results indicate that optimal scheduler are obtainable in polynomial time for important problems such as master-worker tasking. The polynomial time attribute comes from the fact that the scheduler can be characterized as an optimization problem in rational variables, which in turn, is solvable by linear programming. Unfortunately, extensions of the steady-state optimization methodology to other scheduling problems do not preserve this formulation. In some cases efficient heuristics have been found but there are still others cases in which steady-state scheduling remains NP-hard. Our research explores a steady-state scheduling variant that uses game theory and competitive analysis for optimizing the system's utilization. The objective two-folded: on one hand we hope to provide a general framework for a large class of scheduling problems. On the other hand, we expect to make the scheduler adaptive to unexpected network congestions and inaccuracies or variance in the estimation of the speed of processing of a computer node. This will be done by measuring some system's parameters within a period, computing the adjustment, and using the results to schedule the next period. Currently, our efforts concentrate in modeling this variant, which we have called adaptive steady-state scheduling.