When the Grid becomes Pervasive: A Vision on Pervasive Grids

Manish Parashar
Rutgers University, USA

Jean-Marc Pierson
IRIT, Université Paul Sabatier, France

Grid computing has emerged as the dominant paradigm for wide-area distributed computing [3]. The goal of the original Grid concept is to combine resources spanning many organizations into virtual organizations that can more effectively solve important scientific, engineering, business and government problems. Over the last decade, significant resources and research efforts have been devoted towards making this vision a reality and have lead to the development and deployment of a number of Grid infrastructures targeting a variety of applications.

However, recent technical advances in computing and communication technologies and associated cost dynamics are rapidly enabling a ubiquitous and pervasive world - one in which the everyday objects that surround us have embedded computing and communication capabilities and form a seamless Grid of information and interactions. As these technologies weave themselves into the fabrics of everyday life [6], they have the potential of fundamentally redefining the nature of applications and how they interact with and use information. This is rapidly leading to a new revolution in the original Grid concept and leading the way to a ***Pervasive Grid*** vision. The Pervasive Grid vision is motivated by the advances in Grid technologies and the proliferation of pervasive systems, leading to a seamless integration of sensors, light devices, together with classical high performance systems in a common framework to offer the best immersion of users in the global environment. This is, in turn, leading to the emergence of a new generation of applications that use pervasive and ambient information as an integral part to manage, control, adapt and optimize [7]. These include a range of application areas including crisis management, homeland security, personal healthcare, predicting and managing natural phenomenon, monitoring and managing engineering systems, optimizing business processes, etc.

Note that it is reasonable to argue that in concept, the vision of *Pervasive Grids* was inherent in the visions of "computing as a utility" originally by Corbató et al [1] and later by Foster et al [2]. In this sense, *Pervasive Grids* are the next significant step towards realizing the metaphor of the power grid. However, while *Pervasive Grids* present new opportunities, they also present many challenges that must be addressed by Grid technologies, which are highlighted next.

The *Pervasive Grid* environment is inherently large, heterogeneous and dynamic, globally aggregating large numbers of independent computing and communication resources, data stores, instruments and sensing/actuating devices. The result is an unprecedented level of uncertainty that is manifested in all aspects of the Pervasive Grid: System, Information and Application [4, 5]. **System uncertainty** reflects in its structure (from flat to hierarchical to P2P ...), in the dynamism of its components (entities may enter, move or leave independently and frequently), in the heterogeneity of its components (connectivity, reliability, capabilities, …), in the lack of guarantees, and more importantly, in the lack of common knowledge of numbers, locations, capacities, availabilities and protocols used by its constituents. **Information uncertainty** is manifested in its quality, availability, compliance with common understanding and semantics as well as trust in its source.

Finally, **application uncertainty** is due to the scale of the applications, the dynamism in application behaviors, and the dynamism in its compositions, couplings and interactions (services may connect to others on a dynamic and opportunistic way).

The scale, complexity, heterogeneity, and dynamism of *Pervasive Grid* environments and the resulting uncertainty requires that the underlying technologies, infrastructures and *Pervasive Grid* applications must be able to detect and dynamically respond during execution to changes in the state of the execution environment, the state and requirements of the application and the overall context of the application. This requirement suggests that [4]: (1) Applications should be composed from discrete, self-managing components which incorporate separate specifications for all of functional, non-functional and interaction-coordination behaviors. (2) The specifications of computational (functional) behaviors, interaction and coordination behaviors and non-functional behaviors (e.g. performance, fault detection and recovery, etc.) should be separated so that their combinations are composable. (3) The interface definitions of these components should be separated from their implementations to enable heterogeneous components to interact and to enable dynamic selection of components. Given these features, a *Pervasive Grid* application requiring a given set of computational behaviors may be integrated with different interaction and coordination models or languages (and vice versa) and different specifications for non-functional behaviors such as fault recovery and QoS to address the dynamism and heterogeneity of the application and the underlying environments.

We believe that addressing these challenges requires new paradigm for realizing the *Pervasive Grid Infrastructure* and its technologies that is founded on **semantic knowledge** and **autonomic mechanisms** [4, 5]. Specifically, (1) static (defined at the time of instantiation) application requirements, system and application behaviors to be relaxed, (2) the behaviors of elements and applications to be sensitive to the dynamic state of the system and the changing requirements of the application and to be able to adapt to these changes at runtime, (3) common knowledge to be expressed semantically (ontology and taxonomy) rather than in terms of names, addresses and identifiers, and (4) the core enabling middleware services (e.g., discovery, coordination, messaging, security) to be driven by such a semantic knowledge. Further the implementations of these services must be resilient and must scalably support asynchronous and decoupled behaviors.

References:

1. F. J. Corbató and V. A. Vyssotsky, Introduction and overview of the Multics system, Proc. AFIPS 1965 FJCC, 27(1), 1965, 185-196.
2. I. Foster, C. Kesselman and S. Tuecke, The Anatomy of the Grid: Enabling Scalable Virtual Organizations, The International Journal of High Performance Computing Applications, 15(3), 2001, 200-222.
3. M. Parashar, and C. A. Lee, Scanning the Issue: Special Issue on Grid Computing, Proceedings of the IEEE 93(3), 2005, 479-484.
4. M. Parashar and J.C. Browne, Conceptual and Implementation Models for the Grid, Proceedings of the IEEE, 93(3), 2005, 653-668.
5. M. Parashar, Autonomic Grid Computing – Concepts, Requirements, Infrastructures, Autonomic Computing: Concepts, Infrastructure and Applications, Editors: M. Parashar and S. Hariri, CRC Press, ISBN 0-8493-9367-1, 49-70, 2006.
6. M. Weiser, The computer for the 21st century, Scientific American, vol 265 (3), pp. 66-75, 1991.
7. J-M. Pierson, A Pervasive Grid, from the data side. Research Report, LIRIS Laboratory RR-08-2006:http://liris.cnrs.fr/publis/?id=2436