Mobile Grid Computing: Changes and Challenges of Resource Management in a Mobile Grid Environment

Antonios Litke[1], Dimitrios Skoutas[1], Theodora Varvarigou[1]

[1] Department of Electrical and Computer Engineering, National Technical University of
Athens, 9 Heroon Polytechniou Str.,
15773 Athens, Greece
{ali, dskoutas, dora}@telecom.ntua.gr
http://www.ntua.gr

Abstract. Grid computing has recently migrated from traditional high perform-
ance and distributed computing to pervasive and utility computing based on the
advanced capabilities of the wireless networks and the lightweight, thin de-
vices. This has as result the emergence of a new computing paradigm which is
the Mobile Grid. This paper provides a discussion on Mobile Grid definition,
the basic motivation and benefits that can stem from this new technology, espe-
cially as far as it concerns collaborative engineering, industry and the everyday
life of the "mobile" citizen. In the sequel it presents a set of major technical
topics that influence the traditionally basic Grid Infrastructure services under
the new Grid environment involving Mobile Grids.

1 Introduction

One of the most critical things for understanding and realizing *Mobile Grid comput-
ing* is to have a consistent and accurate definition, or at least determination of what a
Mobile Grid is. There are many attempts for the accurate definition of the Grid. How-
ever the various approaches that have been made address in a high degree of accuracy
the term Grid. The Grid can be viewed as a distributed, high performance computing
and data handling infrastructure, that incorporates geographically and organization-
ally dispersed, heterogeneous resources (computing systems, storage systems, instru-
ments and other real-time data sources, human collaborators, communication sys-
tems) and provides common interfaces for all these resources, using standard, open,
general-purpose protocols and interfaces [1]. However, it is also the basis and the
enabling technology for pervasive and utility computing due to the ability of being
open, highly heterogeneous and scalable.

Mobile Computing is a generic term describing the application of small, portable,
and wireless computing and communication devices. This includes devices like Lap-
tops with Wireless Local Area Networks technology, mobile phones, and Personal
Digital Assistants (PDAs) with Bluetooth or Infrared Data Association (IrDA) inter-
faces. The Mobile Computing focuses on the requirement of providing access to in-
formation, communications and services everywhere, anytime and by any available
means. The technical solutions for achieving this are not always easy to implement.

In fact, mobile computing requires the creation of communication infrastructures and the modification of computer networks, operating systems, and application programs [2]. The mobility issue implies some constraints that must be addressed since they limit the capability of a moving resource in contrast to a fixed one.

Mobile Grid, in relevance to both Grid and Mobile Computing, is a full *inheritor* of Grid with the additional feature of supporting *mobile users* and *resources* in a *seamless*, *transparent*, *secure* and *efficient* way. It has the ability to deploy underlying ad-hoc networks and provide a self-configuring Grid system of mobile resources (*hosts and users*) connected by wireless links and forming arbitrary and unpredictable topologies.

2 Why Mobile Grid?

Grid is already being successfully used in many scientific applications where huge amounts of data have to be processed and/or stored. Such demanding applications have created, justified and diffused the concept of Grid among the scientific community. As the amount of potential Grid users is really enormous, the accumulated data processing and storage requirements are at least comparable. Wireless devices (laptops and PDAs), with currently limited resources (low processing power, finite battery life and constrained storage space), would benefit from the opportunity of using a considerable amount of resources made available by all the other devices connected to the network [3]. In particular, mobile users might be the future users of this new technology. Moreover, we have nomadic users who travel and work only seldom at their offices [2].

Mobile Grid enables both the mobility of the users requesting access to a fixed Grid and the resources that are themselves part of the Grid. Both cases have their own limitations and constraints that should be handled [5]. In the first case the devices of the mobile users act as interfaces to the Grid enabling job submission, monitoring and management of the activities in an 'anytime, anywhere' mode, while the Grid provides them with a high reliability, performance and cost-efficiency. Physical limitations of the mobile devices make necessary the adaptation of the services that Grid can provide to the users' mobile devices. In those cases mobile Grid has the meaning of '*gridifying*' the mobile resources. In the second case of having mobile Grid resources, we should underline that the performances of current mobile devices are significantly increased. Laptops and PDAs can provide aggregated computational capability when gathered in hotspots, forming a Grid on site. This capability can advantage the usage of Grid applications even in places where this would be imaginary.

However, there will always be the question on why a Grid solution should be adopted in comparison to any other non-Grid Information Technology (IT) solution. Grid is not intended to be the 'panacea' to the problems related to IT domain. It is a promising emerging technology that has the ambition to provide more efficient and beneficial solution than its 'competitors' by enabling the simple 'connect and share' approach in the same manner as the current Internet search engines apply the 'connect and acquire information' concept. By this, Grids and mobile Grids can be the ideal solution for many large scale applications that are of dynamic nature and require transparency for users. Grid will increase the job throughput and performance of the

involved applications and will increase utilization rate of resources by applying efficient mechanisms for resource management in the vast amount of its resources. It will enable advanced forms of cooperative work by allowing the seamless integration of resources, data, services and ontologies.

3 What is 'mobile' in a mobile Grid?

Mobile Grid is a platform that should address mobility issues by means of enabling both fixed and mobile users to have access to both fixed and mobile Grid resources utilising transparently and efficiently the underlying technologies. Mobility involves a set of issues for both users and communication/collaboration mode that will be presented shortly in the sequel.

Nomadic users move around between different terminals and ask session movement (*user mobility*). For example, a user can open a session on a workstation at his office and then ask its movement to a personal computer at home. Roaming users, instead, moving around with their own wireless devices, want to continue their sessions when they cross different localities, as it happens today in cell phones telecommunications (*terminal mobility*). In both cases the middleware should hide the details of the dynamic utilization of service components, when the session movement is required, besides it should manage all the session hand-off transparently to the user and possibly to the application components. Another kind of mobility is the *service mobility* which is meant as the ability of the infrastructure to maintain the services for a user while he is mobile. These services should be independent from the devices and from the location of the user with respect to the network he is accessing. [9]

Another aspect of mobility that should be considered is the distinction between wireless networking and mobile user interfaces. In wireless networking the problem relies in the fact that the connection is highly variable both in quality and in price, and moreover that disconnections happen unexpectedly, thus constituting synchronous interaction uncertain. In case of mobile user interfaces the challenges come from the restrictions imposed by the small device size. Small displays can only present limited amount of information, while absence of keyboard makes data entry hard [5].

It should be noticed that mobility stresses the fact of the synchronous mode of communication. In a collaboration environment the most challenging aspect is to enable a synchronous mode of co-operation between the peer entities. While asynchronous mode is easier to be implemented in unreliable and complex environments, the synchronous mode introduces a set of constraints that are by nature difficult to handle. The shared objects are no more workspaces and databases where many users act on, but moreover mechanisms have to be applied for enabling shared applications and concurrent schemes for the session handling.

4 What will change?

The mobile Grid will introduce changes to the general Grid concept. New functionalities of the Grid will be needed since the old ones will not make use of all the capabilities that will be available. These functionalities will involve end-to-end solutions

with emphasis on Quality of Service (QoS) and security, as well as interoperability issues between the diverse technologies involved. Enhanced security policies and approaches to address large scale and heterogeneous environments will be needed. Additionally, the volatile, mobile and poor networked environments have to be addressed with adaptable QoS aspects which have to be contextualized with respect to users and their profiles.

Mobile Grids will make use of the value that many mobile users perceive due to the advanced capabilities of their mobile devices. These advanced capabilities refer to the comparison of today's mobile devices with the ones that existed in the past. Although mobile devices are subject to physical constraints due to their nature they have the ability of having computational and storage capabilities similar to PCs, high quality displays, multiple interfaces (for instance Bluetooth, Ethernet adapters, USB, Infrared) etc. This value can be converted into revenue for service providers. This implies changes in various business models and policy issues. Complex workflows for businesses will be needed and Virtual Organizations (VO) will be enriched with the opportunity for automatic federations and resource sharing schemes. Valuations of the diverse services and especially with respect to the Service/Sharing level agreement have to be adapted in relation to QoS aspects. Enterprises have to issue policies that will handle the conflict between public rights to their users and viable models for operation. Moreover the fair use of Grid must be determined by reconciling rights of public access to resources and private ownership of infrastructure and resources.

In the sequel we present a short discussion on some of the most important challenges of Mobile Grids with respect to the resource management topic. Of course many things will change and raise their own challenges to be addressed in this new context. Our study will focus in resource management topic which is a very critical subject for the efficient utilization of the Grid infrastructure functional entities.

5 Challenges of Resource Management in Mobile Grids

5.1 Resource Discovery and Selection

The volatile and dynamic environment of the mobile Grid makes necessary the use of sophisticated mechanisms for resource discovery and selection. The list of the authorized machines and resources that are available in the Grid infrastructure is continuously updated. The selection of those resources that are expected to meet the time or cost constraints imposed by the user has to be based not only on deterministic criteria but also on stochastic parameters, thus providing complex probabilistic models for the topology of the Grids and their capabilities. Some of the required parameters are resource accessibility, system workload, network performance, etc. Not to forget that the target audience of the mobile Grids is the broad spectrum of the citizens who would like to use the Grid for solving problems of their everyday life. Thus, a financial criterion of the resources used should be under consideration for the proper resource selection.

5.2 Job Scheduling

The Job Scheduling problem is classified as an NP-complete problem [6] [7] due to the fact that an algorithm has to be applied that would allocate jobs to resources in an efficient and cost effective manner, so as to minimize the resource utilization gaining the maximum profit and satisfying at the same time the user constraints (Security, Quality of Service, Fault tolerance etc). In a mobile Grid environment there are many more constraints that would make the job scheduling problem more complicated. The cost of resources is in that case a metric that is subject to context parameters. Imagine what can happen in a big conference room where the aggregated performance due to the many notebooks linked in wireless hotspots can result in a very cost effective solution for "instant" high performance computing. So the optimization criteria for a job scheduling mechanism should take under consideration not only the cost and the performance of each resource, but also the current availability of resources in the context and their reliability in providing the specific resources for the execution of the job under the given QoS constraints.

5.3 Job Replication, Migration and Monitoring

In accordance to the previous Job Scheduling issue, we claim that the mobile grid is a dynamic system where environmental conditions are subject to unpredictable changes: system or network failures, system performance degradation, addition of new machines, variations in the cost of resources, etc. In such context job migration and re-scheduling, as well as job replication and co-scheduling are both (from different perspectives) efficient ways to guarantee the completion of the jobs according to the restrictions set by the users. Job monitoring and checkpointing (especially in the case of long running tasks) is difficult in dynamic environments. Especially job monitoring is responsible for detecting alert situations that could initiate a migration and alternatively identify if a job has been completed so as to suspend/stop/cancel the execution of the other replicas.

5.4 Replica Management for large Data Sets

Replica management is an important issue for a number of applications that utilize large data sets (like for instance experimental results for scientific applications, for simulations and prediction models) [8]. The complete data set of a VO may exist in one or possibly several physical locations but it is possible that no single organization has sufficient storage to hold a complete copy. Instead, the common practice is to have copies of the most relevant portions of the data set stored on local storage for faster access. Replica Management keeps track of where portions of the data set can be found by providing mappings between logical names for files and one or more copies of the files on physical storage systems. This mapping is of particular interest in mobile environments, since access to specific physical locations might be impossible. Storage devices are constantly changing their connection status and sometimes

their physical location, thus imposing different schemes for accessing their data, which includes context aware information. Moreover, mobile Grids have to address the issue of not only locating a Data Base containing the relative data, but also if this location is the most suitable and reliable (with respect to avoiding a failure) to be used.

6 Existing work

There are some efforts focusing on designing and developing the Mobile Grid. However they are in a preliminary form and are expected to provide the first results by end of 2005.

The K*Grid project [10] is an initiative in Grid research supported by Korean Ministry of Information and Communication. The main goal of the K*Grid project is to provide a powerful research environment to both industries and academia. In this project, a study on mobile grid technologies is elaborated that uses idle resources of a great number of mobile devices and the development of a mobile grid platform. The scope of the effort includes the analysis of wireless mobile networks, devices and technologies and the requirements for mobile grid, design and implementation of mobile grid platform which is based on PDA and wireless LAN technology.

The AKOGRIMO project [11] is a European funded project aiming to architect and prototype a blueprint of a Next Generation Grid based on Open Grid Services Architecture (OGSA) which exploits and closely co-operates with evolving Mobile Internet infrastructures based on IPv6. The concept of the project is to evaluate the derived Mobile Grid through testbeds that are chosen based on existing evolving applications from the domain of e-Health, e-Learning and Crisis management.

7 Conclusions

In this paper we presented a discussion on the Mobile Grid concept, the changes that this implies in the Grid technology and the challenges that have to be met in order to address the new computing paradigm, mainly focused on the aspect of resource management, which is the heart of every Grid functionality. This topic indicates the necessity for research in these domains so as to enable the implementation of the new Mobile Grid environment. Moreover, the paper presented existing work and projects on worldwide level which are promising for the elaboration of thorough study in the domain. As the Grid computing seems to gain more space in the technology areas, it will be the basis for utility and pervasive computing in the forthcoming years by leveraging the mobility advantage that it will be enhanced with. The future work in this area will investigate not only issues concerning the interoperability but mainly with respect to the efficiency of mobile Grid solutions so as to enable the deployment in commercial and everyday life applications.

References

1. I. Foster, "What is the Grid? A Three Point Checklist", GRID Today, July 20, 2002
2. S. Gai, "Internetworking IPv6 with Cisco Routers", avail. at http://www.ip6.com
3. D. Bruneo, M. Scarpa, A. Zaia, A. Puliafito, "Communication Paradigms for Mobile Grid Users", 3rd International Symposium on Cluster Computing and the Grid May 2003, Tokyo, Japan
4. M. Graf, "Grid collaboration in the mobile environment." IBM. Speech at the Collaboration@work Workshop on June 25th 2003.
5. S.-M. Park, Y.-B. Ko, and J.-H. Kim, "Disconnected Operation Service in Mobile Grid Computing," International Conference on Service Oriented Computing(ICSOC'2003), Trento, Italy, Dec 2003
6. R. Moreno, "Job scheduling and Resource Management Techniques in Dynamic Grid Environments" 1st European Across Grids Conference, February 2003, Spain
7. A. Abraham, R. Buyya, B. Nath, "Nature's Heuristics for Scheduling Jobs on Computational Grids", International Conference on Advanced Computing and Communications, December 2000, India
8. B. Allcock, J. Bester, J. Bresnahan, A. L. Chervenak, I. Foster, C. Kesselman, S. Meder, V. Nefedova, D. Quesnal, S. Tuecke "Data Management and Transfer in High Performance Computational Grid Environments". Parallel Computing Journal, Vol. 28 (5), May 2002, pp. 749-771.
9. H. Schulzrinne, and E. Wedlund "Application Layer Mobility using SIP", ACM Mobile Computing and Communications Review, vol. 4, no.3, July 2000, pp. 47-57.
10. The K*Grid project http://gridcenter.or.kr/MobileGrid/index.php
11. AKOGRIMO Integrated Project http://www.akogrimo.org
12. Open Grid Services Architecture http://www.globus.org/ogsa/