A Case for the World Wide Grid: Interlinking Islands of Grids to Create an Evolvable Global Cyberinfrastructure

Marcos Dias de Assunção1 and Rajkumar Buyya

Grid Computing and Distributed Systems (GRIDS) Laboratory
Department of Computer Science and Software Engineering
The University of Melbourne, Australia
http://www.gridbus.org

ABSTRACT

This paper presents a case for what we call the World Wide Grid (WWG) and identifies some research areas that deserve special attention to enable such a scenario. Over the last few years, several organizations have set up their own Grids to share resources such as computers, data, and instruments. Countries have created e-Science projects to enable resource sharing and collaboration among scientists from different parts of the globe. This approach has resulted in islands of isolated production Grids in different parts of the world with no true resource sharing and exchange among them. Nowadays, there is a profound interest on (a) federating Grid islands allowing users from different Grids to share resources; (b) the establishment of Virtual Organizations comprising individuals and resources from different Grids; (c) making the Grid more pervasive; and (d) applying Grid technologies to the commercial world as a means to manage IT infrastructure. Therefore, an infrastructure is necessary to enable the bartering of resources among such islands of Grids, to reallocate computing capacity when necessary, to promote the collaboration among members from such islands of Grids, and to allow the easy joining of new members to these infrastructures. We believe that these needs will take us to the next step in the evolution of the Grid, hereafter called World Wide Grid (WWG), which consists of mechanisms for linking these islands. However, in order to make it reality and link such islands of Grids, studies have to be carried out in some key areas. Just like it happened to well-known technologies such as the Internet, several technological, economical, social, and cultural aspects can speed up or slow down the development of such infrastructure. In this work we examine current global infrastructures, identify key problems to address and present a research agenda for it.

Keywords: Grid computing, World Wide Grid.

1. INTRODUCTION

Maturity of Internet-based communication, computing, storage and software technologies in recent past has resulted in emergence of the Grid computing paradigm that allows the secure and coordinated sharing of globally distributed computing and storage resources. Thus Grid computing being the enabler for Virtual Organizations (VO) [1] supports a range of e-Science and e-Business applications [2][3][4]. Over the last few years, several organizations have set up their own Grids to share resources such as computers, data, and instruments. Countries have created e-Science projects to enable resource sharing and collaboration among scientists from different parts of the globe. This approach has resulted in islands of isolated production Grids in different parts of the world with no true resource sharing and exchange among them. In addition, as the Grid is becoming pervasive, sensors, personal digital assistants, cell phones and other devices are being integrated into it.

Although the increasing interest about Grid is evident and many islands of production Grids exist, currently there are some additional concerns. There is no ecosystem that enables Grids to evolve in a similar manner to the Internet and the Web. There are needs to support interoperability and global collaboration among scientists and organizations from different islands of Grids and to provide the bartering of resources among islands of Grids. It allows participants to tap into resources from different Grid islands across administrative boundaries in a seamless manner and without replicating federation efforts. In addition, it is important to provide an ecosystem that permits Grids to evolve from local isolated islands to a global infrastructure and allows the peering between Grids under different

1 Corresponding author, marcosd@csse.unimelb.edu.au
administrative policies and political boundaries. This infrastructure has to evolve similarly to the World Wide Web and the Internet. Such infrastructure that links all the islands of Grids is here called World Wide Grid (WWG). For this, it is important to identify the key elements that are missing to support and enable a global and omnipresent Grid. Hence, some of the questions that need to be answered are:

- What are the architectural issues that prevent current Grid architectures to scale to a WWG?
- How do we put the islands of Grid together while ensuring interoperability and resource sharing among its members?
- What kind of structure should the WWG have to promote formation of Grids of Grids?
- What are the coordination mechanisms that we need to put in place to enable the WWG?
- What are the incentives for end users, laboratories, organizations, service providers and Grids in general to engage in such a global Grid?

To enable this vision, it is important to draw some lessons from existing infrastructures such as the Internet and the World Wide Web and social and biological systems. The Internet has rapidly grown and globally interconnected several scales of networks of different kinds of network technologies. It allows the cooperation of service providers who peer and exchange traffic based in several factors in order to provide Internet services to their clients. The World Wide Web has enabled the exchange of information on a globally basis and allows access to any Web page stored on any Web server in any part of the globe. Similarly, there are many examples of infrastructures in our society that present a sustainable growth. For example, it may also be important to learn from social and biological networks, by analyzing their patterns and characteristics, find out what can be applied to Grid computing and more importantly, analyze how they have evolved and grown without centralized control or interference and how they self-organize [5]. Such ecosystem is lacking in Grid. The structure of the Grid ecosystem is still not clear and we need to draw inspiration and lessons from the Internet, the Web, social and biological systems to define how a global Grid infrastructure should look like and how it should be organized.

Therefore, in this work, we present a case for the WWG. The WWG may have users with different Quality of Service (QoS) requirements yet interconnected by a set of common protocols in a similar fashion to the Internet [6]. Similar to the Web and the Internet, the WWG allows researchers or organizations to maximize the use of their resources, look for collaborators, and utilize other’s infrastructure to leverage their research or businesses. It may also be important to identify the business models that underlie current global infrastructures and analyze how they can be applied into the WWG [7][8]. However, in this paper we concentrate in drawing lessons from key global infrastructures, defining key areas to be researched to enable the World Wide Grid and in building a research agenda for the topic.

The rest of this paper is organized as follows. First, we present the motivation for the WWG as well as background ideas in Section 2. Section 3 contains a description and analysis of business models that form the basis of the Internet and the WWW. After that, we discuss about the structure of the WWG in Section 4. Section 5 presents a gap analysis of existing Grid technologies and further, in Section 6, we aim to draft a research agenda on the topic. Finally, Section 7 concludes the paper and presents our final consideration on the subject.

2. MOTIVATION AND BACKGROUND

In the origins of the World Wide Web (WWW), it was difficult to support that a global infrastructure would make information globally available through a Web browser to any person in any corner of the world where access to the Internet is available [9]. More than research, technological progress is the result of paradigms that are modified, habits that change and the use of technology by people in their day-by-day activities. Many things in our society present characteristic steps of evolution in which diverging solutions, individuals and ad hoc initiatives come to a stage in which bigger infrastructures or groups are built and the synergy of such isolated initiatives are explored.

In the Grid area, over the last years some national and global Grid infrastructures have also been set up as a way to enable eScience and to promote collaboration among organizations or individuals geographically dispersed [10][11][12][13][14]. Some of these Grids use middleware based on Globus [15] while others use their own toolkits for the establishment of Grids. Although there has been much effort on Grid, a unique Grid based global infrastructure linking these islands of Grids is still not
available. There is yet a lack of sharing and exchange among Grids due to some factors such as technological limitations, social and political barriers, lack of proven business models and incentives for joining such infrastructure.

In the scope of this work, we are more concerned with what we call social organizational Grids, such as regional Grids, state Grids, national Grids or Grids set up by different projects. In setting the grounds for this case, we divided the rest of our motivation into two parts. Firstly, a description of the WWG and its characteristics is presented. Secondly, to demonstrate the interest of the community in such global Grid infrastructure, we present some main research projects that aim at federating Grids or building global Grid infrastructures.

2.1 The World Wide Grid

The WWG must be an evolvable system that can expand from organizational Grids to a truly global Grid without major problems or scalability limitations. In an abstract view of the WWG presented in Figure 1, we can see that the Grid comprises of several islands of Grids and has a structure that grows from local isolated Grids to a unique global infrastructure. In the figure, several islands of Grids are linked. Such islands may use different Grid technologies, may adopt different distributed computing models, such as P2P or master-slave based, and can serve for different purposes. However, they are linked to a unique global infrastructure instead of replicating federation efforts. It should not be difficult to an existing Grid, an organization, or a new service provider to join the WWG, that is, they should join without taking part of hardly complex and multilateral negotiations involving many organizations.

Some desirable characteristics of the WWG are presented here, while some of the major challenges to achieve such goals are discussed in Section 5.

Standards based: The islands of Grid are put together by using an agreed upon architecture and a set of common interfaces such as OGSA and Web Services [16].

Varied connectivity support: Grids allow the sharing of resources in the application level. Therefore, complex topologies may be formed, which will not follow only a single approach. For example, linking of islands of Grids will not always happen either using strictly a peer-to-peer approach or a hierarchical one; it may be a mixture of both kinds. Coordination mechanisms that deal with different connectivity supports are put in place in the WWG.

Different or diverging policies: Grid islands may have their own policies for access control, usage of resources and accounting. Organizations can also implement different security mechanisms.
Incentives for joining and collaborating: Although the WWG is comprised of islands of Grids with competing organizations, countries with strict protectionism policies, and strong barriers regarding collaboration and business with other countries, Grid users have incentives to participate of the WWG.

New business models: On top of the infrastructure provided by the WWG, new opportunities for collaboration and new business models will emerge.

Pervasive: Currently, various wireless sensor networks, cell phones, personal digital assistances and other mobile devices are being integrated to Grid and will generate high amount of data [17]. The WWG may extend the power of such devices by providing means to process this data as well as to use these devices as actuators or interfaces.

Decentralized resource allocation: Decentralized approaches for resource allocation such as self-organizing economic models [18][19] are applied in the WWG.

Easy to use: Although the WWG presents an intricate topology, users with different requirements of Quality of Service [20] will be able to develop applications and use it irrespective the complexity that may exist.

2.2 Existing Projects

Recently, some initiatives to federate Grids and/or form virtual organizations comprising of several sites have been proposed. We mention a few here, with the details of some discussed in Section 5:

Enabling Grids for E-scienceE (EGEE): This project, funded by the European Union, provides an infrastructure to support academic users. It integrates national and regional Grid efforts across the Europe [21].

Open Science Grid: Open Science Grid Consortium aims at creating a marketplace for resource providers and consumers to barter for Grid resources. By following the rules of the charter, participants have services and interfaces to build new partnerships and to create new virtual organizations for different purposes. Currently, OSG comprises of several sites located in US, Korea and South America [22].

OurGrid: This open and free to join Brazilian Grid is based on the OurGrid middleware and aims at linking scientific laboratories through a peer-to-peer Grid [13]. Users of the sites participating in such a Grid donate and make use of resources through a mechanism called "network of favours", which avoids free riders.

TeraGrid: It is a cyber-infrastructure composed of several resource providers across the US. The Grid Infrastructure Group (GIG) is responsible for integrating such resources and for providing central services.

APACGrid: The Australian Partnership for Advanced Computing Grid Program targets at creating a national Grid infrastructure by integrating APAC National facility and partner facilities allowing Australian researchers to access computational and data resources. The integration with international initiatives is also a future goal [10].

Other National Grid Initiatives: There are also other national Grid initiatives such as the National e-Science Centre in UK [23] and K*Grid [24].

Although efforts in linking existing Grids have been done, a WWG as described earlier is not possible yet. Some reasons are that the aforementioned projects present different linking approaches, different software for enabling VOs, heterogeneous organizational models and adopt different policies for the approval and formation of VOs. In addition, some projects for example do not even follow the VO model. It is important to identify the key issues of current Grid technologies that do not allow them to evolve to such a WWG level. Therefore, questions that may arise and that need to be answered are:
• What are the key elements to provide truly global collaboration and integration of diverse Grid islands?
• Since such global integration is not a reality yet, what are the obstacles for it?
• Is the WWG problem just a matter of lack of technology or are there socio-economic factors that inhibit such global integration?
• Should the WWG be a self-organized, self-healing ecosystem?
• Will current developments allow the WWG scenario? If not, what are the shortcomings of current technologies?
• If the integration of such islands requires the establishment of agreements in an off-line fashion, what would be the policies for enabling such agreements?

We envision that the large-scale adoption of Grid is not just a technological shift, but also social, economic, and cultural one. It is worthwhile to look back at how key technologies such as the Internet and World Wide Web have become widely visible and ubiquitous, and try to apply some key ideas from these infrastructures to the Grid, learning from their experiences. Therefore, in the next section, we look at these technologies, their key elements and we try to identify what will be the elements in the WWG.

3. GLOBAL INFRASTRUCTURES AND THEIR BUSINESS MODELS

As pointed out in [25], many infrastructures for well-known services nowadays, evolved from isolated initiatives that were connected and put together. There intents have had a profound impact in current society. In a similar way to the beginnings of the Internet and data networks, there is a great interest on standardizing the diverse Grid toolkits to promote the integration and best practices. In this section, we examine existing infrastructures and draw some lessons from them to Grid computing.

3.1 The Internet

The Internet was initially a project from DARPA initiated in 1969, linking a few sites in US. Currently, millions of hosts compose the borders of the topology of Internet, which are connected to local Internet Service Providers (ISPs) through an access network (see Figure 2). In dial-up or broadband services, the local PSTN (Public Switched Telephone Network) loop is commonly used to provide users with access to the Internet. These local ISPs connect to regional ISPs, which in turn, connect to national and international ISPs, also known as holding backbones. Such national and international ISPs, also called National Service Providers (NSPs), represent the highest level of the Internet hierarchy and are connected to each other through Network Access Points (NAPs). Thus, the ISPs can provide services like access, backbone, content, application, and hosting. As we can see, the topology of Internet can grow quickly and without the endorsement of a central authority [26].

Nowadays, the Internet presents an intricate structure comprised of a vast number of physical connections established by commercial contracts such as peering agreements. Such agreements are legal contracts, which specify the details of how ISPs exchange traffic. Norton [27] highlights the difference between peering and transit. Peering is the relationship whereby ISPs provide connectivity to each other's transit customers. Transit on the other hand, is the relationship through which one ISP provides access to all destinations in its routing table (see Figure 3). The reasons for peering involve social, economical, and technological factors. ISPs can consider their policies, economical advantages and conflicts before establishing agreements. However, there is no common routine for choosing with whom to peer.

Such agreements can be of various types, such as private, via exchange points or in a relationship between customer and provider. They can contain policies regarding the amount of traffic and proportion because the traffic between peering ISPs can be asymmetric. Tier-1 providers, also known has having access to the global Internet, generally establish contracts not charging other Tier-1 providers, whereas charge for peering with smaller ISPs. Policies are used to divert traffic or to avoid some peers on the Internet, which are enforced by using protocols such as BGP (Border Gateway Protocol).
By analyzing the structure of the Internet, and how it has grown, some lessons can be learnt, such as:

- The use of a common set of simple protocols;
- Even though its topology is complex, it has shown that it can grow quickly because there is no need of agreements and negotiations involving multiple organizations;
- A self-healing structure, in which the failure of part of the network does not compromise the whole Internet;
- Albeit competing, ISPs have clear benefits for peering because both can reduce the amount of traffic across an expensive boundary and improve the efficiency for their users [27].
addition, its business model benefits end-users and compensates service providers;

- Routing protocols that allow traffic to be diverted when it is not allowed or it is not viable to cross a specific network.

### 3.2 The World Wide Web

Two applications have contributed to the rapid growth of the Internet: the electronic mail and the World Wide Web (WWW). Tim Berners-Lee originally conceived a hypertext-based system at CERN to meet the demand for automatic information sharing among scientists, which consequently became the WWW [9]. Currently, the Web is a merge of network, protocols, and hypertext, which has led to the development of a plethora of scientific and commercial applications.

An important contribution of the WWW is that of the URL (Uniform Resource Locator). A URL enables anything "on the Web" to be uniquely identifiable by single string of characters. Hypertext has provided a platform independent standard that allows the linking of web pages and an easy way to access the information they contain. It has enabled a range of applications and content on the web, which made it very attractive. The egalitarian approach of the WWW and the non-existence of central control are key features, since they allow new organizations to engage in such infrastructure without the need of any kind of approval. This simple architecture allowed the fast growing and evolution of the Web.

In addition, by allowing organizations and users to do business online, it enabled several business models and the migration of existing ones to the virtual world. Internet and the Web have benefited from each other. The Internet has played an important role in the expansion of the Web by providing the network infrastructure for it. The Web in turn, has allowed the expansion of the Internet because it has shown a key application that supports the use of the latter. In addition, organizations and users have various incentives for using the Web, such as accessing information, broadening the audience for their business, participate in communities, exchange experiences, take classes, work online, to cite just a few.

Some lessons to be drawn from the Web are as follows:

- Presence of the network effect and a cycle that allows the growth of the Web and that of the network infrastructure. Since there was information available on the web, applications to use it and the number of users and servers increased, it became easier to grow faster.
- The protocols that underlie it allow the interaction among entities, by using an existent network infrastructure, in a free to use fashion to access information or carry out business online.
- No central authority governs it.
- It is simple to use.
- Both providers and clients, whether they are organizations or end users, have incentives to participate in such a system. Commercial organizations for example are attracted to the WWW as a way to improve their business by selling goods online, or establishing their presence in the virtual world and thereby broadening the client-base and commercial coverage.
- It has improved people's quality of life by enabling a range of services that optimize people's time and made information much easier to access.

### 3.3 Peer-to-Peer Networks

Peer-to-peer networks allow the sharing of compute resources such as content, storage and CPU cycles without the need of a centralized server or authority [28]. Such kinds of networks generally maintain their own network mechanisms such as addressing, connection and routing, which overlay the Internet. The segment in which these technologies have had the greatest impact is the one of content distribution. The music distribution market has been suffered constant changes induced by such networks. The peer-to-peer approaches have been compared to Grid [29] and its outcomes have been used in some Grid computing technologies [30].

Peer-to-peer technologies have given good insights in technical terms to Grid computing, such as the existence of partially centralized control or the absence of centralized control in some cases, the self-organizing structure of some networks, the overlay mechanisms for routing, fault tolerance, and others. However, they lack on presenting a viable business model and we believe that they are suitable for
specific niches. For example, even though the peer-to-peer networks have revolutionized the content distribution sector, companies have faced problems to adjust to this new scenario, to build business models and to explore business opportunities that are suitable to them [31].

However, there are some aspects of peer-to-peer networks that worth to be mentioned due to their relevance to Grid computing:

- The decentralized architecture;
- Overlaying addressing and protocols for data replication and resource discovery;
- The self-organizing characteristics of some peer-to-peer networks;
- Overlay network-oriented approaches that build redundant and fault-tolerant infrastructures on top of the Internet;
- They are in general easy to join. No complex negotiations are required to a new member to join such networks.

4. GRID PROJECTS AND PROPOSED ARCHITECTURE

In this section, we analyze some existent approaches that aim to federate Grids and create global Grid infrastructures. By reviewing such approaches, we target at presenting our contribution and ideas in creating a truly global Grid. Therefore, in the second part of this section, we present our proposal for the WWG.

4.1 Current National and Global Grid Infrastructures

As highlighted in Section 2, currently there are many national Grid initiatives. In addition, there are some efforts in federating such national Grids in continental and global infrastructures. In this subsection, we briefly review two of these efforts and highlight the main aspects that underlie such efforts.

The Enabling Grids for E-sciencE (EGEE) [21] aims at federating resource centers in Europe for enabling a continental architecture for researchers. Its organization consists of an Operations Manager Centre (OMC) located at CERN, Regional Operations Centers (ROC) located in different countries, Core Infrastructures Centers (CIC) and Resource Centers (RC), which are responsible for providing resources to the Grid. A ROC is responsible for activities as providing support deployment and operations; negotiating SLAs within the region and organizing certification authorities. CICs are in charge of providing VO-services, such as maintaining VO-Servers and registration; VO-specific services such as databases, resource brokers and user interfaces; and other activities such as accounting and resource usage. The OMC aims at interfacing with international Grid efforts. It is also responsible for activities such as approving connection with new RCs, promote cross-trust among CAs, and enable cooperation and agreements with user communities, VOs and existing national and regional infrastructures.

EGEE uses the VO model. The process of getting involved for using such infrastructure for a new application or providing resources to it needs a formal request and assessment from special committees. Once the application has been considered suitable to EGEE, a VO can be formed. Accounting is to be carried out based on the use of resources by members of the VO. At present, EGEE relies on LCG-2/gLite Grid tools [32].

The Open Science Grid Consortium [22] also requires that users belong to a VO in order to make use of the resources. A VO includes at least one organization that is member of the consortium. The main aim is leveraging existent Grid islands under existing Grid middleware by organizing them in virtual organizations and providing services that allow the integration of such virtual organizations. These VOs are recursive. A resource can belong to different VOs and VOs can be composed of sub-VOs. In this case, VOs get together for a common purpose, which can consist in collaborating in a research project, for example.

For forming new VOS, some requirements must be satisfied, such as to install a VO Membership Service (VOMS) and to describe the intention of the VO. At the present, the formation and finalization of short-term VOs are not possible in the OSG.
4.2 Proposal for the World Wide Grid

We expect that the WWG will be pervasive as sensors networks of various types, wireless devices, large-scale processors, data storage devices, and scientific instruments are integrated into it. It can give rise to a high number of interaction patterns and imposes challenges in current mechanisms. In addition, the VO model may not be the only solution and the only model for Grid computing.

By analyzing previous infrastructures, we saw that the concept of organization presented by the Internet is missing in Grid computing (The Internet is a network of networks). In addition, the Internet aims at providing a set of simple protocols and simplicity; Grid is becoming a very complex architecture. Self-healing, peering and benefits from peering, such as reducing traffic, increasing revenues or using services, are reasons adopted by ISPs to peer with each other. From the Web, we can see that the lack of centralized control allowed its fast growth. Organizations have reasons for using it and in addition a range of business models are Web-enabled. Self-organization, self-healing, decentralization are characteristics of some P2P networks that should be implemented into Grid.

Based on communities and groups in our society and how they have formed, we see that such structures evolve from locally organized structures to ones that are more complex. For example, a group of individuals has a common interest on a given activity. Leaders of this group may look for another similar groups and may found interesting to interact with another groups. After the agreement to cooperate has been settled, interactions may take place, new links can be made and existing ones may be broken. Search engines, such as Google, have helped people to find their collaborators.

From the structures analyzed, we can see the following characteristics and needs:

- Small structures are linked to more complex ones through some access point. In the Internet, routers link networks, in groups, leaders start agreements or collaboration with other groups;
- In joining and forming communities, there are places where people publish not just their capabilities, but also their interests and needs;
- Search mechanisms that allow people to locate people or organizations that can fill their needs.

Based on existing works mentioned previously and the works in [33][34][13] we have elaborated a sketch of an architecture for the WWG (Figure 4). The Figure shows an evolvable and scalable architecture for linking islands of Grids. Like in the Internet, each island of Grid will have a peering arrangement with other Grids. This peering arrangement will be managed by a component called InterGrid Resolvers (IGRs). Some of the elements of such architecture are described as follows:

![Figure 4. Abstract view of the structure of the World Wide Grid.](image-url)
**InterGrid Resolvers (IGRs):** These entities are responsible for negotiating agreements with other IGRs, and for acting as site selectors. They may be in charge of establishing peering agreements when the establishment of some Virtual Organization is necessary or when there is a need for users from a Grid to tap into resources from another Grid; the peering agreements could be pre-defined in an off-line manner. Resolvers are also responsible for trading for resources in open markets. They publish the Grids’ requirements or resources offered in WWG directories or markets and also look for collaborators when there is a need.

**Grid Resolvers (GRs):** They are responsible for similar functions as IGR, but within a single Grid. Brokers or resource managers interact with resolvers when they need capabilities beyond what their internal Resource Providers (RP) are able to offer.

**IntraGrid Resource Managers (IRMs):** These are components responsible for the management of resources in an IntraGrid and can use local protocols in order to communicate with resources within an IntraGrid.

**Resource Providers (RP):** They are responsible for providing resources in the IntraGrid and to users from other Grids.

**Grid Resource Brokers (GRBs):** Users wanting to make use of Grid resources utilize GRB to do so. The broker uses resources from the Grid and WWG when the need surpasses the resources its Grid can offer. The GRBs should interact with IGR in order to obtain access to other resources from other Grid islands.

Although not described in the architecture, it is important to mention the need for the following components:

**Common Interfaces (IF):** The WWG uses common interfaces for accessing resources. Such common interfaces have to follow standards such as the ones proposed by GGF [6].

**WWG Directories (WDs):** What we mentioned here as a WWG Directory, is a database with information regarding Grids, Grid projects, their goals and capabilities, proposals for collaboration and requirements by Grid projects. The current facilitators for virtual organizations such as OSG and EGEE could maintain WWG Directories with information that could be shared such as the existing VOs and Grid projects involved into them.

## 5. ISSUES IN CURRENT GRID TECHNOLOGIES

When considering a large-scale system such as the WWG, problems arise, such as resource management among different Grid islands, varying usage and connectivity patterns, different security policies, resource reservation, QoS and service level agreements, formation and management of virtual organizations. Besides, users and providers need incentives to participate to such a Grid. This section presents a list of some of the problems that need to be addressed in order to have a World Wide Grid.

**Protectionism and collusion:** In an open global Grid, local protectionism problems may appear. For example, scientists in the **Country A** may be interested in developing a new technology or running a given set of applications. For doing so, they will look for Grid partners and use resources from the WWG. However, **Country B** has interests in slowing down the development of such technology by **Country A**. The reasons for that may be for example: (a) because it is also developing similar technology; (b) it does not consider right that **Country A** develops it and considers that it represents a threat to others; (c) any other political reason. Therefore, **Country B** imposes local barriers for using its resources and tries to persuade others to do the same. Such political, financial, and cultural issues are not solved in a global Grid scenario. Issues such as collusion and formation of groups to reduce competition should be investigated as well as the development of mechanisms to address such issues. As argued in [39], the potential non-cooperativeness should be modeled and studied at various levels in Grid computing.

**Incentives for collaboration and compensation of service providers:** In the Internet and World Wide Web, users have some incentives to participate and there are somehow projects in several
countries to take such technologies to the most of their populations. When considering a global Grid, except for the "coolness" of the technology, there are not, at present, convincing incentives for end users and organizations to participate and run critical applications on it. In addition, the WWG needs to provide ways to compensate resource suppliers. Some approaches to address the resource usage and incentives aim at bringing from economics the answer for these problems. However, in the WWG, this approach requires globally accessible services such as a Grid Bank, a common currency or token exchange mechanism, and the Grids involved need to trust these entities. Some institution may need to fund and maintain such globally available entities.

Lack of applications for ordinary users: In contrast to P2P networks and the WWW, the lack of appealing applications for ordinary users is a problem in Grid computing and certainly hinders its pervasiveness. There are not enough ordinary users of Grid computing because there are not applications for them. As pointed out in [35], the problem is cyclical, where there are not many applications because Grid technologies are not mature enough. In addition, Grid technologies are not mature enough because the lack of applications that require them to be migrated from the academic world to ordinary users in their day-by-day activities. In order to be used by normal users, Grid should present applications that, as example of the WWW, can improve people's quality of life.

Pricing of resources: Economic models are important to Grid because: (i) Grid resources are not shared for free and charging for their usage can provide incentives for resource providers to offer and share their resources in the Grid; (ii) the participants of the Grid can be divided into resource consumers and providers and the allocation is achieved through the economic behavior of these actors; (iii) markets can offer a decentralized approach for scheduling in which each participant acts in order to maximize its own utility (iv) market-based resource allocation provides incentives for users to truthfully reveal how much they value resources [48]. However, another important concern is how resources should be priced and how usage is measured. What would be the basic units of usage of a compute or a storage resource? How do resource providers adjust the price of their resources in a competitive Grid? What are the different price mechanisms in the Grid market, when considering the local pricing, a competitive market and collaboration among Grids? How do the price mechanisms affect the system?

Connectivity and interaction patterns: The integration of Grids can enable a high number of interaction patterns, which would be difficult to design in terms of middleware, scheduling, and resource allocation. It is advocated that overlay networks will be important in a large-scale Grid to tackle this heterogeneity and guarantee several interaction patterns [36]. Overlay networks are virtual networks that overlay the physical infrastructures such as the Internet and add value with some features and semantics. They can provide the infrastructure to enable various interaction models providing them in terms of APIs and abstracting the middleware from details of the underlying network.

Coordination mechanisms: As mentioned in [37], current approaches to resource allocation are non-coordinated. Such approaches can lead to inefficient schedules and worsen resource utilization. Coordination mechanisms that allow brokers and resource management systems to exchange information may need to be put in place. However, the main challenge is that the WWG may have Grids with different connectivity patterns. One question is what metaphors should such mechanism follow and how current mechanisms can be improved to suit to the WWG?

Policies for joining the WWG: Currently, organizations define agreements in an off-line basis. In the academic world, such contracts define the terms for the collaboration and the role of each organization in providing resources for such an endeavor. In addition, organizations may set SLAs with service providers by explicitly describing their objectives and the requirements in terms of Quality of Service and penalties that providers or consumers agree to pay in case of being disloyal with the contract. At present, the integration of campus Grids is made through multilateral agreements among the involved organizations, whereas joining a national Grid normally requires the approval by special committees. Joining the WWG should be simpler, such as joining the Internet or the WWW.

Standards: As presented in [38] there are two ways to adopt standards such as OGSA for Grids. The first way is to make every single Grid service OGSA compliant. The second way is to have all
service providers providing a standards compliant interface externally, while using their own protocols and interfaces internally. Regardless the approach used, the Grid community has to come up with a simple set of interfaces widely acceptable and easy to implement.

5.1 Issues in Virtual Organizations

The concept of Grid has evolved along the last years and at the present the main problem that underlies the Grid is the coordinate resource sharing and problem solving in dynamic and multi-institutional Virtual Organizations [1], hereafter called VO. A VO can be composed of a group of individuals and/or institutions that come together to share resources with a common purpose. The Grid is supposed to provide such an infrastructure and is considered the enabler for VOs.

According to [33], the life cycle of VOs can be divided into (a) the identification of business opportunities that require VOs to be formed; (b) their formation (c) their operation and management; and (d) termination of such VOs. However, some problems arise when considering these steps. Some of these problems are listed here:

**Formation of VOs:** Currently, organizations define the terms for formation of VOs through multilateral contracts and agreements. Such processes are done in an off-line basis. It is not possible to create VOs in an on-demand and dynamic way for security and policy related issues. Also, mechanisms for the negotiation and establishment of agreements for the dynamic formation of virtual organizations are an issue. Moreover, a framework for how the off-line and out-of-the band agreements are defined to compose the source network or physical infrastructure is another key aspect. In addition, some legal barriers for the formation of VO exist and this may require the change of laws and legal processes for the establishment of such VOs.

**Merging VOs:** In [33], it is pointed out the need of a broker who is responsible for examining the market and identifying market opportunities that require the formation of a VO. When the opportunity is identified, it will look for partners who may be interested in joining the VO. We believe that the formation of the VO is not the end of the story. Once a VO has been formed, it may be of common interest of **VO A**, for example, to merge or collaborate with **VO B**. Brokers can also be in charge of analyzing whether collaboration of such VOs can provide the resources and outcomes both VOs need. Similarly, **VO A**’s broker may decide that it is no longer viable to collaborate with **VO B**, even though the finalization of **VO A** will not take place at this moment. Automated decisions are desirable in a scenario that can evolve to several VOs and approaches used in the formation of communities and coalition formations can be applied in such a scenario [49].

**Resource allocation in VOs:** Providing a fair resource allocation in VOs is troublesome since resources can be part of multiple VOs and resource providers can be providing different amount of resources to different virtual organizations. Works in meta schedulers [39] and in adjusting meta schedulers to consider VOs have been made. Dumitrescu et al [40] for example, highlight that challenging usage policies can arise in VOs that comprise participants and resources from different physical organizations. Participants may want to delegate access to their resources to a VO while maintaining such resources under the control of local usage policies. In this context, they seek to address questions such as: how usage policies are enforced at the resource and VO levels; what are the mechanisms for a VO to ensure policy enforcement; how is the distribution of policies to the enforcement points carried out; and how are they made available to VO job and data planners. In [40] they have proposed a policy management model in which participants can specify the maximum percentage of resources delegated to a VO, a VO in turn can specify the maximum percentage of resource usage it wishes to delegate to a given VO’s group. However, such policies are defined in an off-line basis and may be quite complex to be reconciliated. We believe that resource allocation in static and dynamic VOs should use the metaphor of an organization that goes public, for example. Shareholders that hold the most of the shares have the right to take decisions regarding how resources are allocated in the VO. The decision takers are to be chosen in the formation of the VO or as it evolves. However, it is important to have in the VO somebody who plays the role of accounting and ethic committee to avoid abuse in the VO.
Security in VOs: Grid Security Infrastructure (GSI) provides the basis for security in the Grid at present. In a VO level, VO Membership Services (VOMS) offer support to manage users, groups, roles and capabilities in VOs. They allow a centralized control of VOs and extend Grid security concepts to a VO level by proving additional services, such as: (a) VOMS server that maintains information about users, the groups they belong to, roles and permissions; (b) a client that allows the user to create a VOMS proxy certificate; and (c) a VOMS admin service that allows the manager of the VO to setup roles and capabilities. There has been few works on automated generation and negotiation of access control policies in VOs [41]. However, such security models have to deal with ad hoc Grids and short-lived VOs. Issues regarding the negotiation of access control policies and the mapping of existing privileges from a source domain to a target domain have been investigated [47]. However, efforts are still necessary in this area in order to make Grid a robust infrastructure for commercial applications. On the other hand, public Grids have to get used to the fact that one provider, consumer or organization may not be reliable irrespective of the security mechanisms and therefore, it has to consider trust and reputation based mechanisms. Trust and reputation in peer-to-peer networks is an interesting research field and some key ideas should be borrowed from such systems.

Language to define actor's roles: For the process of dynamic formation of VOs, there is a need for a language or semantics for specifying the actor’s roles and responsibilities. In addition, since a VO has been formed, identifying the roles to be played by each actor is a problem. Understanding the behavior of such actors and how they interact with each other seems to be an interesting area.

Short-term VOs: Nowadays, cell phones and personal digital assistants are equipped with many facilities such as digital camera, audio recorder, mp3 player, video recorder and GPS. In addition, the connection of such devices has improved in recent past and now it is possible to use messaging services such as EMS and MMS, and access the WWW. We can imagine examples of VOs such as for Terrorist Detection. In this scenario, a citizen who has seen some suspicious person or fact in an area of public access, such as a tram, train station or a shopping mall, could initiate a short-term VO for analyzing and checking the potential attack. The formation process of the VO begins when the user contacts a counter-terrorism service. This service will be responsible for taking over the evolution of the VO and leading the necessary participants. The necessary people should be included by the counter-terrorism service for evaluating and checking the facts. Necessary police officers would be contacted to inspect or isolate the area. Evidences and identification of involved people are to be checked, which can require the access to other services. This scenario is still not possible due to rigid security mechanisms, lack of middleware and coordination mechanisms for integrating such devices into the Grid.

6. RESEARCH AGENDA FOR THE WWG

The following research agenda was built from our analysis of existing global infrastructures, their evolution and the lessons we can draw from them, and the issues identified. This agenda aims at grouping and defining some of the research topics that, in our opinion, deserve special attention in order to enable the vision of a World Wide Grid.

6.1 Pricing Resources in the Grid Economy

In [42], a discussion on Grid Economy is presented. The use of economic principles in Grid comes from the success of economic institutions in the real world as a sustainable model for regulating resources, goods and services. However, the adoption of such economic approaches requires the study of, for instance, pricing of Grid resources and/or agreement on pricing mechanisms. Therefore, if an economic approach is used by the WWG, detailed studies have to be done in areas such as resource pricing, modeling consumer's utility, resource provider's marginal cost, and benefit in providing resources.

In this field, some of the questions that need to be answered are the following:

- What resources should be free of charge and what resources should be priced in the market in the Grid? What are the policies that define in which circumstances resources should or should not be shared in the Grid? In this first stage, a Grid environment, such as the World Wide
Grid, should be modeled.

- How to price the resources in the Grid?
- What kinds of issues related to the price setting for the resources arise in the Grid?
- What is the relationship among the issues related to the price setting for the resources in the Grid?
- How do resource providers adjust the price of their resources in the Grid in order to achieve the price that best maximizes its profits in a competitive market, yet maintaining the equilibrium of supply and demand?
- What are the different price mechanisms in the Grid market, when considering the local pricing, a competitive market and collaboration among Grids? How do the price mechanisms impact the system?
- How to model the resource price variation process to predict the price of resource in the Grid?

When considering a World Wide Grid, the Grid economy can become quite complex. This way, the study of pricing of resources and its effect in the Grid economy should be worked out in different steps. Therefore, we divide the pricing and its study in different levels, as presented in Figure 5. As presented in this figure, we are not interested at this moment, in studying the impact of resources that are not priced, such as files in some P2P networks. Firstly, we manage to study the pricing of resources in a local level. In this level, resource providers calculate the price of resources by taking into account the cost of producing them and the marginal benefit of providing them to the Grid. The challenge at this level is how to define the cost of resources and how to calculate the benefit of providing them to the Grid. In this first scenario, a competitive market is not taken into account.

At the second level, a competitive market is taken into account. In this scenario, the resource providers must consider forms of adjusting their prices while maximizing their profits. In this level, the target price in the market should be the one that is the equilibrium of supply and demand and allows a fair allocation. In this regard, equilibrium theories are considered. The model developed in this stage should include resource providers of several types (e.g. providers with lower price and better resources; lower prices and same resources; lower prices and worse resources and vice-versa). Achieving price for resources that leverage a fair allocation is the goal of the model at this stage.

In the third level, it is studied the effect of resource providers collaborating in reducing or increasing their prices to have enough power to beat competitors. In this regard, mechanisms to avoid unexpected behavior in the economy and the emergence of monopoly or oligopoly would be studied. In the next stage, the fourth level, the impact of local protectionism or governmental entities that aim to control prices or impose barriers for using resources from other Grids would be considered. The macroeconomic aspects in price and exchange are taken into account. These characteristics can take us to a scenario in which the Grid is a complex and unstable system because of the competition and the chaos generated by such rules and protectionism. Achieving the equilibrium, the balance of such World Wide Grid in such a way that it becomes stable for some time, is the main goal. We aim at taking inspiration from macroeconomics in this case.

### Figure 5. Pricing of resources steps/issues in the Grid.

At the second level, a competitive market is taken into account. In this scenario, the resource providers must consider forms of adjusting their prices while maximizing their profits. In this level, the target price in the market should be the one that is the equilibrium of supply and demand and allows a fair allocation. In this regard, equilibrium theories are considered. The model developed in this stage should include resource providers of several types (e.g. providers with lower price and better resources; lower prices and same resources; lower prices and worse resources and vice-versa). Achieving price for resources that leverage a fair allocation is the goal of the model at this stage.

In the third level, it is studied the effect of resource providers collaborating in reducing or increasing their prices to have enough power to beat competitors. In this regard, mechanisms to avoid unexpected behavior in the economy and the emergence of monopoly or oligopoly would be studied. In the next stage, the fourth level, the impact of local protectionism or governmental entities that aim to control prices or impose barriers for using resources from other Grids would be considered. The macroeconomic aspects in price and exchange are taken into account. These characteristics can take us to a scenario in which the Grid is a complex and unstable system because of the competition and the chaos generated by such rules and protectionism. Achieving the equilibrium, the balance of such World Wide Grid in such a way that it becomes stable for some time, is the main goal. We aim at taking inspiration from macroeconomics in this case.

#### 6.2 Coordination Mechanisms and Self-Organization

Resource allocation approaches in Grid are currently non-coordinated and different domains have their own resource brokers, objectives, QoS requirements and so on. Such divergent approaches can take us
to a scenario with bad schedules and inefficient resource allocation. The problem to be addressed here is how to coordinate and organize such disparate resource brokers and schedulers. Scalable coordination mechanisms for the WWG may allow the fast growing of such ecosystem. On the other hand, such environment can become quite complex that can exhibit requirements for self-organization [19][43]. In this case, the fair global behavior emerges from local actors designed to interact locally without the sense of global control or a centralized system [44]. Therefore, there may be the need for engineering and developing economic self-organizing brokers and schedulers.

In this scenario, some questions that need to be answered are:

- What kind of coordination mechanisms need to be put in place to allow the growing of the Grid ecosystem?
- What are the requirements and issues of such coordination mechanisms? If the coordination of resource and service brokers spanning different administrative domains is necessary, are there protocols that can be used for exchanging information among such brokers? What are the drawbacks that avoid the scalability?
- Is it possible to achieve a fair global behavior through engineered entities that do not take into consideration the global system, yet present local self-organizing behavior?
- What are the metaphors and models that can be used for self-organization in Grid? How they can be applied in terms of development?

### 6.3 Infrastructure for Grid Economics

Economic approaches are useful for coping with problems as providing Grid resources to different users with diverging QoS requirements and compensate resource suppliers. However, it is not defined whether the Grid economy should use real money or virtual currency [45]. Economic models might also require globally trusted entities for several activities such as accounting, usage quota enforcement and charging. Although these requirements imposed serious constraints on the scalability of Grid technologies, slowly it may become feasible, mainly when trust federations, for example [46], are solving the problem of making such entities trustful. Trying to fill the gap of global trust, the International Trust Federation aims at promote harmonization and synchronization of regional Policy Management Authorities (PMA) policies.

However, the design of economic institutions for accounting, Grid banking and for charging for resource usage should be carried out. Whether each Grid island adopts its own virtual currency for resources, the study of a money exchange system and its impact could be an interesting line of research. In addition, electronic payments infrastructures for the WWG are also difficult, since countries may have different policies regarding the flow of money.

Resource exchange among virtual organizations and Grids is also promising. However, it is difficult to say, for example, how much storage is equivalent to 30 CPU hours. If that were a solved issue, enabling markets for resource exchange would be the next step. Institutions largely trusted should maintain and set rules for such markets.

### 6.4 Agents and Search Engines for Grids and VOs

In this work, we advocate the need of Grid Resolvers as entities aware of agreements among Grid islands and VOs, and for looking for partners when there is a need for forming a VO. Quality of service, service level agreements and guarantees are important among service providers and consumers. However, automatic ways of negotiating such agreements and establishing such links are necessary. Thus, decisions can be automated and therefore studies in agent-based techniques must be done. In addition, agent oriented software engineering should be applied in modeling these VOs.

Another aspect is the lack of search engines for the Grid. Although there exist directory services for the Grid, markets where Grids and VOs can publish their capabilities, interests and requirements are necessary. In addition, these markets can run auctions, in which bidders can bid to become member of a virtual organization, merge virtual organizations or to contract services to start a new virtual organization. Search mechanisms are important to look for auctions and to choose which auction the agents should bid for.
7. CONCLUSIONS

In this work, we presented a case for the World Wide Grid as an evolvable and sustainable Grid infrastructure. Such case started with the analysis of current global infrastructures, how they have evolved and what lessons can be drawn from them to Grid computing in order to enable the vision of the WWG.

Existing projects aiming at creating national and continental Grid infrastructures were discussed. An architecture for the World Wide Grid, based on existing works and aiming at minimizing the issues discussed in this work, was presented. A gap analysis of current technologies was carried out.

Current technologies do not allow the WWG vision due to conceptual and technological drawbacks such as the lack of coordination mechanisms. As argued in this work, the need of an architecture is necessary to allow Grids' structure to evolve from local to the World Wide Grid and to enable the easy development of Grid applications for e-Science, e-Business and e-A-lot-more-things. In addition, as shown in this work, in order to realize the vision of a WWG, many issues related to cultural, social, and political divergences have to be solved. Our contribution in this work was of identifying key problems to realize a truly WWG and in delineating a research agenda on the topic.

ACKNOWLEDGEMENTS

The authors are thankful to Krishna Nadiminti from the University of Melbourne for helping to improve the language and expression of this paper. We also thank Jianhong Xia from RMIT University, Fernando Koch from Utretch University and Vanessa Teague from the University of Melbourne for sharing their thoughts on the topic.

REFERENCES


Systems, Knowledge Management, Organizational Intelligence and Learning, and Complexity, 2003.


