Book Section:

Pham, T.V., Lau, L.M.S. and Dew, P.M. (2004) The integration of grid and peer-to-peer to support scientific collaboration. In: Michaelides, D. and Moreau, L., (eds.) Proceedings of GGF11 Semantic Grid Applications Workshop. Semantic Grid Community , Southampton , pp. 71-77.

The Integration of Grid and Peer-to-peer to Support Scientific Collaboration

Tran Vu Pham, Lydia M S Lau and Peter M Dew
{tranp, llau, dew}@comp.leeds.ac.uk.
School of Computing, University of Leeds, Leeds, UK.

Abstract. There have been a number of e-Science projects which address the issues of collaboration within and between scientific communities. Most effort to date focussed on the building of the Grid infrastructure to enable the sharing of huge volume of computational and data resources. The 'portal' approach has been used by some to bring the power of grid computing to the desk top of individual researchers. However, collaborative activities within a scientific community are not only confined to the sharing of data or computational intensive resources. There are other forms of sharing which can be better supported by other forms of architecture. In order to provide a more holistic support to a scientific community, this paper proposes a hybrid architecture, which integrates Grid and peer-to-peer technologies using Service Oriented Architecture. This platform will then be used for a semantic architecture which captures characteristics of the data, functional and process requirements for a range of collaborative activities. A combustion chemistry research community is being used as a case study.

1. Introduction

Research on infrastructure to support scientific collaboration has attracted the attention of many institutions and organisations worldwide. The UK e-Science programme and its research centres are amongst the most active participants [1]. E-Science projects, such as myGrid [2] or DAME [3], have mainly focused on developing middleware to support scientific collaboration commonly known as the Grid. The main purpose of the Grid is for "*coordinated resource sharing and problem solving in dynamic, multi-institutional virtual organisations*" [4]. Grid developer community recently has adopted a Service Oriented Architecture, OGSA [5], to enhance the interoperability of services on the Grid and the flexibility when using the services. However, the challenge is how to bring the benefits of the Grid to the wider scientific communities (i.e. those which may need the power of the Grid only occasionally). The current portal approach provides a gateway to the Grid power from a desktop machine, e.g. in [3], but its support for interactive collaboration amongst end users is rather limited.

Peer-to-peer, on the other hand, is a computing model, which has a more lightweight approach to the sharing of computing resources. This model has been proved to be successful in many commercial desktop file-sharing applications such as Napster[1] and currently Kazza[2]. The advantage of peer-to-peer is that its application is closer to end users, and when using the system, they have the sense of ownership over their shared resources. In addition, a peer-to-peer application often provides online means of communication to support collaborative work, therefore, not only computing resources but also scientific knowledge could be shared. It is anticipated that the world of scientific computing will be more and more decentralised into peer-to-peer model, where scientists' desktops will be edges of the network [6].

[1] Napster has been put out of service
[2] http://www.kazza.com

In the effort to improve the support for scientific collaboration, an integrated architecture combining the Grid and peer-to-peer concepts is proposed. The goal of this integration is to bring resources on the Grid more widely available to the whole scientific community. In order to increase automation and quality of collaboration within the new architecture, the use of Semantic Web technology is planned.

The domain of experimentation for this study comes from the Combustion Chemistry Research Community. Requirements capture has been an on-going activity. The next section of this paper will provide the background of collaborative requirements within the Combustion Chemistry Research Community. The proposed hybrid architecture will be presented in the third section. The potential offered by the Semantic Web technology and its challenges will be assessed.

2. Context: the Combustion Chemistry Research Community[3]

The centre of Combustion Chemistry research is building models of chemical reactions. This activity is time-consuming and requires expert knowledge. However, because the input data necessary for the building process, experimental data and reaction rate coefficients, is scattered around in the community and improperly evaluated, the model builders do not have access to all of these data, and hence, the accuracy of resulted models are limited to only particular conditions. Consequently, subsequent combustion processes that use these models are also limited to certain level of accuracy. In order to overcome these shortcomings and to get model users involved in building process, the combustion chemistry community is looking for a computing infrastructure, which has:

- A storage for storing knowledge relevant to compiling reaction models, such as experiment data, as well as reaction models.
- A collection of tools for enabling scientific collaboration amongst all distributed participants and tools for enabling processing, analysis and validation of data as well as assembly the data into models.

The ultimate goal of the Combustion Chemistry Research Community is a paradigm that enables building reaction models in a consistent and systematic way by incorporating all available data and expertise of all members of the community.

(The problems & requirements above identified by chemists in [7], [8] & [9])

3. The integration of Grid and Peer-to-peer

Integration between Grid and peer-to-peer has been considered in Peer-to-peer Grids [10], which mixes Grid and peer-to-peer concepts into a democratic architecture mediated by Web Services. This paper proposes a different way of integrating the Grid and peer-to-peer, which attaches peer-to-peer and the Grid together (figure 1) so that they can mutually support each other. This method of integration reduces the cost of implementation and management while still maintaining flexibility as characteristics of the Grid and of peer-to-peer will remain the same.

As shown in figure 1, the integrated architecture separates heavy computation and storage into the Grid side, computation layer, and lightweight collaboration into peer-to-peer side, collaboration layer. For example, in the context of Combustion Chemistry research, heavy computation can be simulations of reaction mechanism; sharing of experimental data and mechanisms amongst scientists are lightweight

[3] Information about combustion chemistry research is collected from discussions with chemists from Department of Chemistry at the University of Leeds and other work related to combustion chemistry

collaboration. The two layers can be connected by either Web or Grid Services that depends on the choice of implementation.

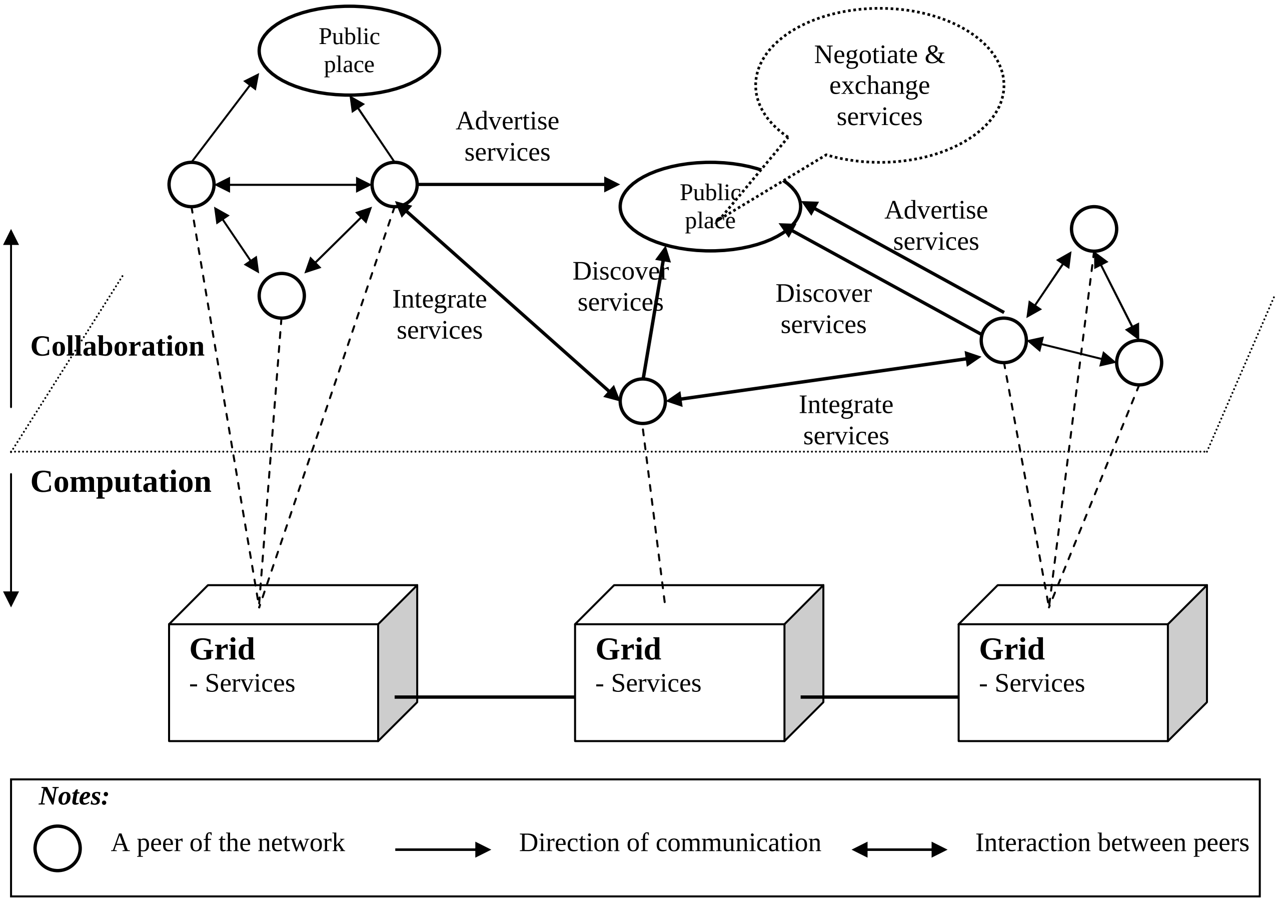

Figure 1: **The integration of Grid and Peer-to-peer**

Information about resource on the traditional Grid such as experimental data, simulation programs, is only known and accessible to authorised users. However, research communities are widely distributed, and not all members are granted access to Grid. Therefore, there might be a situation in which some scientists are desperate for resource, whereas the resource is readily available on the Grid with minimal use. The proposed integration is addressing this problem in the following way: (i) resources on the Grid are transformed into Web (Grid) services; (ii) the owners of these services publish service advertisements into their peer-to-peer communities; (iii) on receiving information about necessary services, scientists, who do not have access to Grid, can request service owners to execute the services on their behalf. In this model, the role of resource owners is to bridge the need of peer-to-peer communities and resources on the Grid. This role will be automated and can be extended to authorised Grid users, who are also members of peer-to-peer communities.

In addition to the above advantage, the separation of computation to the Grid and lightweight collaboration into peer-to-peer also reduces the management complexity of trivial collaboration on the Grid. Other features of peer-to-peer computing such as instant messaging, file sharing also add great values to the integrated architecture in supporting scientific collaboration.

4. Potential application of the Semantic Web

The above section has described in principle the interaction of Grid and peer-to-peer in the collaborative architecture. A semantic architecture would also be necessary to bring the architecture to its full potential as well as to satisfy strict requirements of modern scientific communities. For instance, Combustion Chemistry Research Community requires of higher level of automation in building chemical mechanism and locating validated source of scientific data. In order to achieve this, characteristics of scientific collaborative processes and data need to be captured and categorised in a proper format that is understandable to computer programs. This requires the use of ontology from the Semantic Web technology. One example from the architecture, the discovery of available services on the Grid in peer-to-peer environment will clearly benefit from the use of ontology.

The use of ontology in a decentralised peer-to-peer environment leads to a challenging problem. As ontology is a means of capturing scientific knowledge and knowledge in a scientific community evolves overtime, it will not be feasible to have only a common static ontology for a community. Ontology building has to be a continuous process. It will also not appropriate for some people to develop the ontology, and the other people using it. The one who uses the ontology is the one who has most knowledge of it, then that one should contribute to development of the ontology. In addition, with one piece of ontology, communities from different backgrounds may understand differently. Therefore, ontologies should be local to communities that it is built to support.

For all the above reasons, this paper is proposing an approach to be used within the integrated architecture, in which ontology within a community will be used and contributed by its members. In order to achieve this, the peer-to-peer environment will be organised as communities and sub-communities. Each community or sub-community will promote a leader or a chairperson, who leads the community, and agree on a common ontology to be used. A member of a community can make change to the ontology of the community by advertising definition of new terms and concepts to all other members of that community. Other members will give feedback if appropriate. When all the agreements are made on the new ontology, the chairperson will make it official to the community and notify other members about the decision.

5. Conclusion and work to be done

This paper has introduced an integrated architecture between Grid and peer-to-peer to support scientific collaboration, focusing more on the collaboration aspect with the aim to bring large-scale resources a step closer to end scientist users. The use and the management of ontologies have also been considered to exploit full potential of the collaboration within and amongst communities in peer-to-peer environment.

A prototype implementation of the integration in the context of Combustion Chemistry Research Community without ontology has been developed and proved to be successful. This implementation is using Grid Services from Globus Toolkit 3 [11] and JXTA [12] on Grid and peer-to-peer side respectively. The concept Peer Group and messaging mechanism of JXTA seems to match well with the integrated architecture. The next implementation will be incorporating ontology into collaboration layer to support resource discovery, focusing particularly on discovering services, and managing ontologies within communities. This implementation promises to bring a novel architecture into reality.

Reference:

1. National e-Science Centre, http://www.nesc.ac.uk/.
2. myGrid - directly supporting the e-Scientists, http://www.mygrid.org.uk.
3. DAME: Distributed Aircraft Maintenance Environment, http://www.informatics.leeds.ac.uk/pages/03_research/rp_dame.htm.
4. I. Foster, C. Kesselman and S. Tuecke. *The Anatomy of the Grid: Enabling Scalable Virtual Organisations.* **International J. Supercomputer Applications**, 15(3), 2001.
5. I. Foster, C. Kesselman, J. Nick, S. Tuecke. *The Physiology of the Grid: An Open Grid Services Architecture for Distributed Systems Integration.* Open Grid Service Infrastructure WG, Global Grid Forum, June 22, 2002.
6. I. Foster and A. Iamnitchi. *On Death, Taxes, and the Convergence of Peer-to-Peer and Grid Computing.* **2nd International Workshop on Peer-to-Peer Systems (IPTPS'03),** Berkeley, CA, Feb. 2003.
7. D. A. Dixon and L. A. Rahn. *Advancing Chemical Science: Future Networking Requirements*. **High Performance Network Planning Workshop**, Office of Science, US Department of Energy, August 2002. http://doecollaboratory.pnl.gov/meetings/hpnpw/finalreport/high-impact_science.pdf#page=41
8. R. Bair and et al. *Report of the High-Performance Network Planning Workshop*, **High Performance Network Planning Workshop**, Office of Science, US Department of Energy, August 2002. http://doecollaboratory.pnl.gov/meetings/hpnpw/finalreport/high-performance_networks.pdf.
9. Process Informatics Model, http://www.me.berkeley.edu/prime/.
10. G. Fox, D. Gannon, SH. Ko, S. Lee, S. Pallickara, M. Pierce, X. Qiu, X. Rao, A. Uyar, M. Wang and W. Wu. *Peer-to-peer Grids.* **Grid Computing - Making the Global Infrastructure a Reality**, John Wiley & Sons Ltd, 2003, pages: 471-490.
11. The Globus Toolkit, http://www-unix.globus.org/toolkit/.
12. Project JXTA: http://www.jxta.org.