Collaborative e-Science Architecture for Reaction Kinetics Research Community

Tran Vu Pham
School of Computing
University of Leeds
Leeds, LS2 9JT
UK
tranp@comp.leeds.ac.uk

Lydia MS Lau
School of Computing
University of Leeds
Leeds, LS2 9JT
UK
llau@comp.leeds.ac.uk

Peter M Dew
Informatics Institute
School of Computing
University of Leeds
Leeds, LS2 9JT
UK
dew@comp.leeds.ac.uk

Michael J Pilling
School of Chemistry
University of Leeds
Leeds, LS2 9JT
UK
m.j.pilling@leeds.ac.uk

Abstract

This paper presents a novel Collaborative e-Science Architecture (CeSA) to address two challenging issues in e-Science that arise from the management of heterogeneous distributed environments: (i) how to provide individual scientists an integrated environment to collaborate with each other in distributed, loosely coupled research communities where each member might be using a disparate range of tools; and (ii) how to provide easy access to a range of computationally intensive resources from a desktop. The Reaction Kinetics research community was used to capture the requirements and in the evaluation of the proposed architecture. The result demonstrated the feasibility of the approach and the potential benefits of the CeSA.

1. Introduction

Distributed global online collaborations play an increasingly important role in supporting international collaborations amongst scientific communities. This has given rise to a new subject called e-Science [21]. The kind of collaborations includes the exchange of scientific results amongst scientists from various disciplines and the sharing of powerful computer resources and potentially huge volume of data. Computer scientists have been actively engaged in the provision of suitable information technology infrastructures to the support for new online collaborative communities.

The collaborations required by a scientific community can be classified into two types: direct (explicit) and indirect (implicit) collaboration amongst the scientists (Fig. 1). Indirect collaboration happens when a scientist runs a simulation program or stores some data on computational resources provided by other scientists. The direct collaboration is the kind of collaboration where scientists communicate directly to each other. Such kind of collaboration is required when there is a need for instant and direct exchanging of day-to-day working data from one scientist to another. This kind of collaboration is seen as increasingly important to the support of any e-Science research community.

The traditional web-based model which is based on the centralised client/server approach has been widely used to develop collaborative working environments to support collaborations amongst the scientists, such as the VKP [30] and the BADC [2]. However, this approach has been seen as inflexible for distributed and loosely coupled communities as it does not well support spontaneous collaborations [28]. The centralised web-based model is especially not suitable for indirect collaborations on the sharing of large-scale resources, such as computational intensive scientific simulations. Further, it does not scale well as the number of users increases.

The Grid [7][8][9] is currently the most promising infrastructure for collaborations within distributed scientific communities. The indirect kind of collaborations being addressed by the Grid is the coordination of large-scale re-

0-7803-9043-1/05/$20.00 ©2005 IEEE

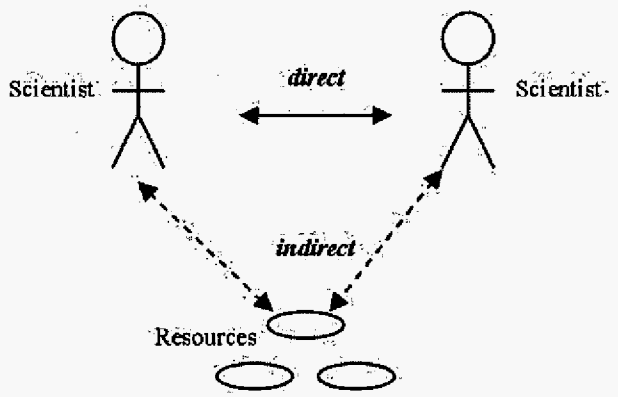

Figure 1. Types of collaborations between scientists

source sharing amongst distributed virtual organisations [8]. An end user's access to grid resources can be via web-based portals (e.g. [5]). Although the Grid provides the backend support for sharing large-scale resources, due to the use of web-based portals, it inherits the limitations of the traditional web-based model in supporting user collaborations within distributed and loosely coupled communities as discussed above.

Peer-to-peer is another computing model that also supports the sharing of computing resources, but on a smaller scale. This model has been proved to be successful in a number of commercial desktop file-sharing applications such as Napster [20] and currently Kazaa [15] or eMule [6]. This success has motivated the application of peer-to-peer computing model in some scientific projects such as Triana [27][25] or Chinook [12]. Although the term *peer-to-peer computing* is very popular at present, there is still no clear agreement on its definition. In this paper, peer-to-peer computing can be understood as *a network-based computing model for applications where computers share resources via direct exchanges between the participating computers* [3]. This definition stresses two fundamental properties of peer-to-peer computing: the direct communication and the sharing resources between peer users.

These two fundamental properties allow users in peer-to-peer environment to communicate directly with each other to dynamically and autonomously establish their own communities without being regulated by any third party administration. The ability to provide direct communication also allows the users to share resources in a timely manner, especially with the current advance of network bandwidth and personal computer processing power. As resources are shared directly from their computers, users still maintain the sense of ownership on the shared properties and have the right to revoke any resource from sharing anytime. On the technical aspect, peer-to-peer is a decentralised network-computing model, where computation is taken place at the edges. Hence, it is more scalable when the number of users increased. The bottleneck problem can also be avoided. Furthermore, peer-to-peer applications often provide means for real-time communications, which are highly suitable for direct collaborations amongst scientists. Therefore, not only computing resources but also scientific knowledge could be exchanged more spontaneously.

In this paper, a novel Collaborative e-Science Architecture (CeSA) is proposed. The aim is to bring together services from computational grids into a peer-to-peer environment to leverage advantages of both Grid and peer-to-peer technologies. The architecture focuses on the support for general collaborative activities for scientists as well as for the sharing of computational capability (e.g. for simulations and analyses) required by the community. This paper reports on the requirements and architecture for the Reaction Kinetics research community. A prototype has been developed to evaluate the architecture in a realistic environment. This has shown the feasibility of this approach and the resulting benefits to the users.

2. Problem Domain: Reaction Kinetics Research Community

A typical e-Science community can be demonstrated by using a Reaction Kinetics research community. This has been chosen because the authors have a close working relation with this community. Reaction Kinetics researchers study the elementary and complex chemical reactions and their applications. This is a multidisciplinary research subject, spanning across Combustion, Atmospheric and Environmental Studies. This is shown in Fig. 2.

A central object in an applied reaction kinetics problem is the chemical reaction mechanism. Different research activities are taking place on different aspects of reaction mechanisms. A research group in the community may undertake one or more of these activities. The allocation of activities in the groups often depends on the expertise and the resources at the institutions where the groups are based.

There is a wide range of applications of chemical reactions mechanisms in Combustion, Atmospheric and Environmental Studies. For example, in Combustion, knowledge about reaction mechanisms, i.e. process of burning fossil fuel, is necessary for designing combustion engines. Feedback on chemical reaction models from the application developers (or engineers) is also necessary for the chemists to refine the reaction mechanism and to concentrate resources on sensitive elements.

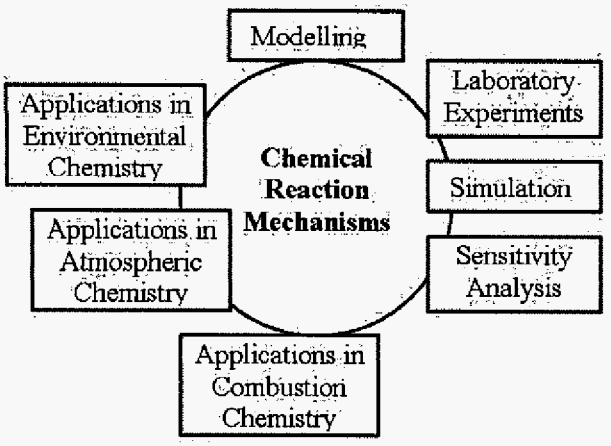

Figure 2. An example of research activities in the reaction kinetics research community

2.1. A Typical Collaborative Scenario

As research activities in Reaction Kinetics are interrelated and tied together by the common interests in chemical reaction mechanisms, as depicted in Fig. 2, research groups that undertake these activities also need to collaborate with each other. The following scenario shows in detail the dependencies that may exist amongst these activities.

A chemical reaction mechanism consists of a series of steps called *elementary reactions* in which chemical species are inter-converted. Essential steps in the construction of the mechanism are the identification of the species in the elementary reactions and measurement of parameters - *rate coefficients* - that determine the rate at which these reactions occur. These elementary reactions and their associated rate coefficients are investigated in the laboratory. It is also feasible to calculate some rate coefficients using quantum theory. The computing resource needed for this approach is substantial.

The elementary reactions are then collected together into a chemical reaction mechanism that describes the overall system. For example, it might describe the steps involved when a fossil fuel is converted into the products carbon dioxide and water, with the release of heat, via a large number of intermediates. This mechanism can then be used to construct a model that consists of a set of ordinary differential equations that represent the rates at which the concentration of each individual species in the mechanism changes with time. There may be several thousand such species.

The model can be tested in a variety of ways for example through experiments on a flame, in which the concentrations of some of the species are directly measured and checked against those simulated using the mechanism. A sensitivity analysis can also be conducted in which the sensitivity of an important observable to the mechanism components, e.g. the rate coefficients, is determined, thus allowing the experiments to be targeted at key features of the mechanism. This feedback is an essential element in the overall model development.

Some of the applications, in combustion for example, involve the interaction between chemistry and fluid dynamics. This adds a further stage of complexity and requires an additional set of scientists with specific expertise. This stage is essential in applications to real systems, such as the design of engines. It is often divorced from the chemical developments, for practical reasons, but it should ideally be incorporated in the overall set of interactions, so that feedback is feasible between all elements.

2.2. Characteristics of Reaction Kinetics Research Community

In general, there is a very close working relationship amongst the research groups in Reaction Kinetics research. Research in one area of Reaction Kinetics depends on data produced by several related areas. Data exchange within the Reaction Kinetics community is an iterative process. A delay caused by one individual or one research group might affect several other people and groups and eventually the whole community. Collaboration amongst individual scientists as well as research groups in Reaction Kinetics from many locations worldwide is very important. A range of computational intensive resources is also required for running simulations and analysis and for storage of huge volume of experimental data.

To support scientific user collaborations, the Atmospheric Chemistry research community, which is linked to the Reaction Kinetics research community, has developed a collaboratory environment for its own [2] based on the centralised model of the web. This collaborative environment addresses the need for sharing working data and collaborative activities within working groups. The centre of this environment is the atmospheric database. However, it has not made any provision for supporting scientists with heavy computation requirements. Furthermore, the use of a web-based interface with a centralised server for user collaborative activities is inflexible, and likely to suffer from the well-known bottleneck problem.

Through the collaboration with chemists working on Reaction Kinetics research group at the School of Chemistry, the University of Leeds, a new collaborative architecture for supporting the community needs to:

- Allow scientists who are working on the same or similar research activities to dynamically form working groups.

- Provide efficient support for timely collaborations within and across working groups in the community for sharing expert knowledge, day-to-day working data, such as experimental data, chemical reaction mechanisms and related input data for reaction modelling.

- Provide easy access to computational intensive resources for time and resource consuming simulations and analyses and for storage of large amount of experimental data.

3. Related Work

A number of projects are currently in progress to address the need for collaborations within scientific communities. These projects are using various computing architectures, such as Web, Grid, Services Oriented or a combination of them, for different types and scales of collaborations.

Dealing with the need for sharing computationally intensive resources, most research effort on the Grid to date has been spent on metadata and middleware infrastructure within the Grid. The NERC DataGrid project [22][31][17] has been working on metadata infrastructure to facilitate the discovery and delivery of data in environmental science, which are held in a loosely coupled federation of distributed locations. On the architectural issue, ICENI is a middleware infrastructure that federates the collaborations amongst underlying resources on computational grids [13][10]. Users can construct and define their applications by composing distributed components and services. By combining Services Oriented Architecture to Grid, myGrid project focuses on the middleware components that enable collaborations on Web Services and composition of these services into workflows [19][11]. Towards supporting collaborations amongst end users, collaborative visualisation projects such as [4][16] or Access Grid [1], have extended the Grid for user collaborations. However, the types of collaborations addressed by these projects were for specialised purposes. For example, the Access Grid was focussed on the provision for distributed videoconferencing. It did not cater for scientists to share or collaborate on shared processes such as a simulation.

Collaboratory for Multi-Scale Chemical Science (CMCS) project [5][18] is using a combination of Grid infrastructure, for management of large datasets, and web-based portal, for user's collaborations. CMCS uses PrIMe (Process Informatics Model) [23] as its pilot. This is aimed at the development of predictive reaction models for combustion a related research domain to Reaction Kinetics. Since a web-based model is adopted, CMCS is having the similar disadvantages regarding its lack of support for running remote simulations or analyses required by the community.

These projects can help to deal with problem from Reaction Kinetics research community to some extent. Using Grid-based and services-based technologies can assist in the sharing of computationally intensive resources and data. Web-based portals can provide end users access to these resources. However, there are limitations of using just these technologies as discussed above.

4. The Collaborative e-Science Architecture

As discussed previously, many research projects have attempted to address the need for collaborations within scientific research communities. However, none of the current work has efficiently supported collaborations at the user end. This new CeSA is aimed to provide better support for the scientific communities by extending Grid technologies, and Service Oriented Architecture, and by combining them with the advanced features of peer-to-peer computing.

4.1. Architecture Overview

As shown in Fig. 3, the CeSA consists of two layers: a computation layer and a collaboration layer.

Resources provided by the grid are often computationally intensive, whereas the collaboration in a peer-to-peer environment is more lightweight (e.g. exchanging of information about resources on the grid or small work-in-progress files). In the context of Reaction Kinetics research, intensive computation can be simulations of reaction mechanisms. The sharing of experimental data and mechanisms amongst scientists can be classified as lightweight collaboration.

The two layers are loosely integrated by services.

4.2. Computational Grids

The computational layer can have one or many computational grids acting as resource providers for the user community in a peer-to-peer environment.

Each of these computational grids has a service container called Grid Service Container (Fig. 4), which provides high-level services to the user community. These high-level services can be built by wrapping resources on the grid or they can be composite services which are composed of other services, including low-level services, available on the computational grids. The difference between these high-level services and the basic low-level Grid services, such as job submission or database services, is that the former aim to deliver output to the users, e.g. results from a simulation on chemical data, whereas the latter usually deliver output to another computer application or service.

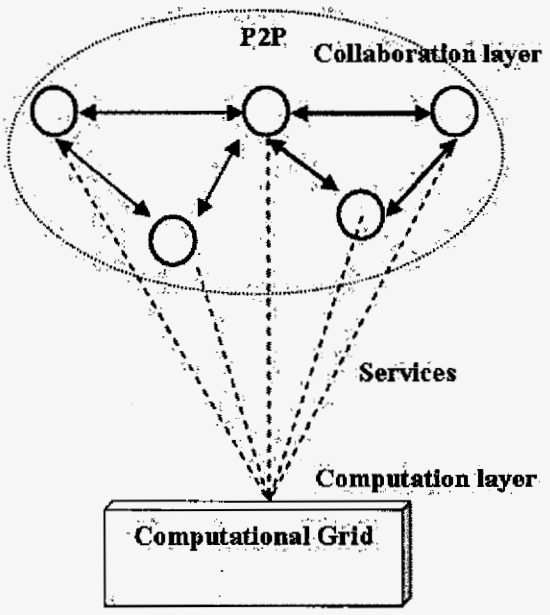

Figure 3. The Collaborative e-Science Architecture

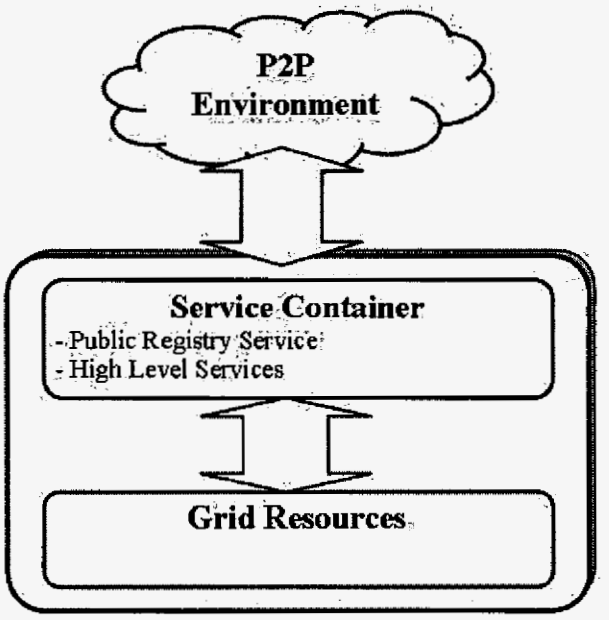

Figure 4. Grid Service Container

The Grid Service Container also contains a Registry Service, which allows the peer users (i.e. scientists) to make queries on services available on the grid. The information about the services can then be published within the peer-to-peer environment so that other peer users can locate and use these services on the grid.

4.3. Peer-to-Peer Environment

The collaboration layer contains peer-to-peer applications, which sit on the users' personal workstations. Each peer-to-peer application (Fig. 5) consists of a Grid Service Client, a Service Publication and Discovery Agent, a set of Community Services and a set of Peer Core Components.

The Grid Service Client is an interface component of a peer-to-peer application, which allows the peer to browse selected Grid Registry Services and to run services provided from grids. The Service Publication and Discovery Agent plays a very important role in the collaborative architecture. It provides two basic functions to the application: publishing information about services available on grids to the peer-to-peer environment and discovering information about services previously published by other peers. Information about a service published in peer-to-peer environment consists of the information such as service name, service URI (Universal Resource Indicator, the identifier that identifies a particular service from a service container), its required input data and its output data. The published information also includes information about the grid that provides the service. The structure of a discovery query is similar to the structure of published information about services. A matching algorithm between the published information and the discovery query will be used during a service discovery process. Because of its importance, the Service Publication and Discovery mechanism needs to be as effective as possible.

Community Services consists of service components for the day-to-day collaboration within a community. Examples include components for file sharing, community/group formation and instant messaging. Through these components, a user can set up a working group or community. Then, the user can establish sharing relationships with other users in his working group or community.

Peer Core Components consists of components that make it a peer in a peer-to-peer network. These include components for communication with other peers, peer identification and peer resources discovery. Service Publication and Discovery and Community Services rely on these core components to communicate with other peers in a peer-to-peer environment.

4.4. Services for Peer-to-Peer Applications

As discussed earlier in Section 4.2 high-level services provided by the grids enable peer users to perform tasks such as simulation or analysis. However, to use such ser-

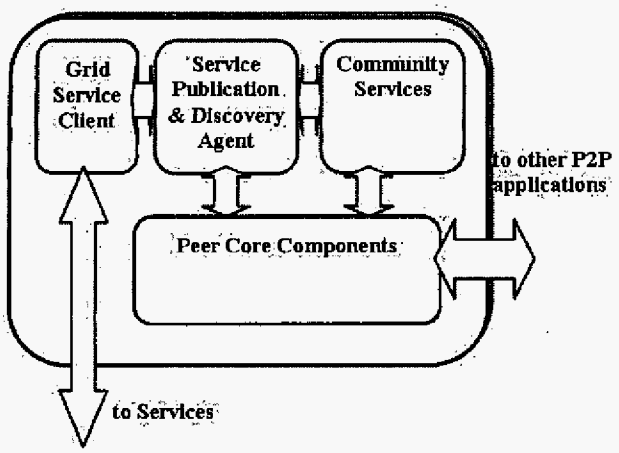

Figure 5. Components of Peer-to-peer Application

vices is often a rather challenging for the users. It is necessary to have a simple user interface for users to easily execute services from grids on their desktops. For example, it is not feasible to give them a service handler (an identification of a Grid service) and ask them to build a service client application to run it. This work is still difficult even for programmers who have no knowledge about Web Services or Grid Services.

Fortunately, simulations and analyses used in scientific communities share common characteristics. They often used files as input and output. Based on these commonalities, a simple unified service interface has been developed. More specifically, the unified service interface consists of the following operations:

- *List input required:* allows service client to query in advanced input files required.
- *Load input:* to upload input files required to server prior to execution.
- *Execute:* to run the services after all required input uploaded.
- *List output:* to query number of outputs produced by the service.
- *Transfer output:* to send back outputs to users.

Service clients can then interact with any service that conforms to this unified interface.

5. Prototype

The purpose of prototyping was to provide an insight into the technical challenges as well as testing the applicability of the CeSA for the user community. Hence, in order to proceed with the experiment, a prototype was developed for the Reaction Kinetics research community.

The prototype consisted of generic tools for collaborations, such as file sharing, chatting, group formation, publishing and discovering of information about Grid services. It also provided a Grid Service Client interface for chemists to execute simulations from a computational grid. With these generic tools and a common unified Grid service interface, the prototype can be applied to other scientific communities with minimal adaptation.

The peer-to-peer application was developed from JXTA peer-to-peer technology [14]. The Grid services and Service Container were based on Open Grid Service Architecture [9] version 3.0.2 [26].

5.1. Conversion of Existing Applications into Grid Services

A few applications used in the community for simulations and analyses of chemical reaction mechanisms were wrapped into Grid services. These applications use files as input and output, and could also produce console output. When wrapping these programs into Grid services using Java, input and output (including console output) were mapped to the input and output parameters of Grid services. These Grid services conformed to the unified service interface specified previously. After being wrapped into Grid services, these new services were deployed into a Grid Service Container provided in Globus Toolkit version 3.0.2. This Grid Service Container ran on one machine, played the role of a computational grid providing simulation and analysis services for the peer-to-peer community.

5.2. CeSA Prototype System

CeSA prototype system, which is a peer-to-peer application for the user community, was developed entirely in Java and was based on JXTA peer-to-peer core technology. It had the following components:

Grid Service Client: The Grid Service Client consisted of two parts: a service browser and a service executor.

The service browser had an interface that enabled the users to browse services available from a particular Grid Service Container. The users would need to provide the URI of the container's Registry Service. Through interaction with the Registry Service, the browser would display a list of service handlers of the services provided by the container.

The service executor could invoke a service from a service handler. The version of Globus Toolkit used for prototyping had two types of service handler for a Grid service. One type of handler was for the Factory Service and the other type was for the Grid service instance. The service executor can be used to generate new instances of a service from a factory service handler or to execute the service using the handler of a service instance. The service executor could interact with any the services developed previously for chemical reaction simulations as they all conformed to the unified service interface.

Service Publication and Discovery Agent: The method used for publication and discovery of service information was based on JXTA protocols [24].

All information about a service, such as service name, service provider, input, output, and so on, was enclosed in a JXTA advertisement. The advertisement about the service was then published in JXTA peer-to-peer network using JXTA discovery protocol.

The discovery of information about services, however, was not based on JXTA discovery, but using JXTA resolver protocol, because the default discovery mechanism provided with JXTA discovery protocol [29] was not flexible enough to deal with complex query requirements. In the prototype, service discovery was required to allow users to make query using any information about services or a combination of them, whereas, with the default mechanism provided by JXTA discovery protocol, only a few indexed attributes could be searched. With JXTA resolver protocol, a query could be distributed to the other peers in the environment. On receiving the query, a peer would flexibly search through its cache for service advertisements that matched the criteria specified in the query, such as service name, service provider, etc. The results would be sent back directly to the query issuer.

Community Services: Community Services consisted of various service components for collaboration within peer-to-peer environment.

The function for the formation of working groups or communities was developed upon JXTA Peer and Peer Group concepts. An individual peer user was mapped to a peer in the JXTA peer-to-peer network. Similarly, a working group or community is corresponding to a JXTA Peer Group. Tools for file sharing and instant messaging were built using JXTA pipe. To send a message or a file from one peer to another peer, a pipe between the two peers would be set up first. Then, the message or file would be sent over the pipe. Each peer had a unique pipe advertisement, which was the necessary information for setting up a pipe connection. This information of a peer was published to peer-to-peer environment when the peer enters the network.

There were also additional components for managing share relationships amongst peer users and working groups or communities and for searching for shared resources. Shared resources available on a peer, such as a working data file, were more dynamically managed, not as static as a shared service. A file could be set to share to a group at a particular time, but not at the other time. Therefore, the approach to resource discovery, more specifically file search, was different from the method used for publication and discovery information about services. There was no publication. As only the owner of the resources could say whether it had resources being shared for the query issuer, the query message had to be distributed to every potential resource owner. The current version of the prototype was using broadcasting method to distribute query messages. Scope of queries could be limited within particular working groups.

6. Experiment and Evaluation

6.1. Objectives

The main aim of this experiment was to collect feedback from a sample of the Reaction Kinetics research community on the functionalities of the prototype based on the CeSA. User comments were invited to compare this new way of working with their current practice. More specifically, the objectives of this experiment were:

- To evaluate the effect of using peer-to-peer collaborative functionalities provided by CeSA prototype system for collaborations within the user community. The kind of collaborations to be experimented includes activities for setting up working groups, sharing and exchanging of day-to-day working data.
- To assess how users can benefit from the access to remote simulations and analyses in Reaction Kinetics using Grid services. This includes the method for publication and discovery of information about services in peer-to-peer environment.
- To capture users general attitude to the new collaborative infrastructure.

6.2. The Experiment Process

The experiment was conducted at a Reaction Kinetics research laboratory at The University of Leeds. The researchers in this laboratory were potential users of the new system. There were three chemists from Reaction Kinetics research community participating in the initial experiment A few copies of CeSA prototype system were installed in the laboratory used for the experiment. The Grid Service Container was running on another machine elsewhere. Each participant was provided with a documentation that guided

him/her through the functionalities of the CeSA prototype system. The guide also outlined the following scenarios for the participants to work through:

- Share working files with other participants,
- Use remote services,
- Publish information about remote services to their working group and
- Discover information about remote services published by others.

All the participants were using the system at the same time so that they would have a better chance to collaborate with each other. As the software was installed on each users machine, the users had the opportunity to share their day-to-day working data.

A questionnaire was issued to each participant to record his/her feedback on the CeSA prototype system. The questionnaire consisted of a mix of open and closed questions and divided into sections corresponding to the experiment objectives. To assess the effect of using the prototype, there were questions about the comparison between users' current way of working and the method provided by functionalities of the prototype. Any potential benefit of using the prototype was also recorded. General feedback and impression on CeSA prototype system was collected at the end of the questionnaire. There were discussions during the experiment to clarify the questions in the questionnaire.

6.3. Results and Analysis

The following is a summary of the findings in relation to the objectives of this evaluation.

Firstly, on the use of peer-to-peer collaborative functionalities, all three participants had a need for sharing working data with their colleagues. All participants recognised the benefits of using the CeSA prototype system for sharing working data because of the following reasons:

- A file can be shared directly from the user machine. Therefore, there is no need to move a file around for sharing.
- Spontaneous sharing of file-in-progress is possible. This allows other users to be able to copy the latest version of the file.
- Users have control over shared data. They can choose to share with a group or a specific person. They can also easily revoke a file from sharing.

The most popular methods currently were using email, via a shared area on the laboratory's computer network, and uploading to or downloading from central/external websites. When sharing files using a central website, a file being shared might not be the most up-to-date as the provider might be deterred from uploading new versions frequently if the demand is unclear.

Concern about security of peer-to-peer environment was expressed by one participant. When the access controls are decentralised to individual machines, it will be important to provide a security mechanism to protect user's own computing resources and personal data.

There was also a worry about the stability of the peer-to-peer environment. As JXTA is a decentralised peer-to-peer environment, it takes some time for the community on JXTA network to reach a stable condition. This is actually a trade-off for the dynamics and flexibility of peer-to-peer computing.

A suggestion on documenting and tracking changes was made as there are times when different users might be working collaboratively on the same set of data.

Secondly, on the use of Grid Services for accessing remote simulations and analyses, the participants recognised the benefit of freeing their desktops' resources for other work, especially for jobs that required long execution time.

All participants had a need to run some kind of simulations or analyses for their research. It often took minutes to hours to complete a simulation. They ran their simulations on their local desktop computers and/or on remote computers via telnet. As the type of simulations they were using was suitable to run as Grid services, they all agreed that it would be very useful for them to run their simulations in this way.

When asked about the comparison between the use of the prototype to discover/use remote services and their current way of working, one participant explained that currently they would need to search for the tools, learn how to install, configure and use them. In this prototype, they could search and use the service straight away.

One participant also pointed out an advantage of allowing him to publish information about available Grid services directly to specific working groups in the peer-to-peer environment.

Finally, on the general feedback, the functionalities of the CeSA prototype system had made a good impression on all participants. The participating chemists believed that their research community would benefit from a system like this. Although there were concerns about security issues and the stability of peer-to-peer network, they would be happy to use it if the system was fully implemented and widely accepted by their community. This *acceptance by the community* is a vital condition that any peer-to-peer system has to satisfy. A user-friendly graphical interface was

also identified as an important characteristic to encourage the users to use the system.

6.4. Discussion

The result of this initial experiment has provided some positive feedback to the CeSA, especially on the use of peer-to-peer environment for collaborations within the user community and the access to remote simulations and analyses using Grid services. As a result of the evaluation, a number of new issues emerged: security in the peer-to-peer environment, user interface and tools for documenting and tracking changes were raised. These issues will be considered in the design of the next version of the architecture.

The Reaction Kinetics research community is typical of e-Science communities. The architecture therefore has much wider applicability. Broadly speaking, e-Science communities are potential application domain of this collaborative architecture.

In addition, today's research requires collaborations not only within a specific research community but also across the boundary to other related research communities. Interdisciplinary research subjects, such as Biochemistry, Bio Informatics, and so on, are becoming popular. In this respect, the CeSA can be exploited to its full potential. The power of peer-to-peer computing is its ability to provide seamless collaborations across organisational boundaries. Within a single research community, resources are limited to the domains that the community is working on. When the boundaries are open up, scientists will have access to huge amount of resources and knowledge contributed by scientists from their related research areas. Consequently, more scientists will then be attracted to join in the community.

7. Conclusion and Future Work

The paper has discussed the requirements for a collaborative architecture through the case of Reaction Kinetics research community. These requirements included the need for collaborations at the user end as well as the need for access to large-scale computing resources. A collaborative e-Science architecture has then been proposed. The new architecture leverages the advantages provided by Grid and peer-to-peer computing and directly addresses both of the requirements of Reaction Kinetics research community.

The result of the experiment on CeSA prototype system shows that peer-to-peer environment is more preferable for collaborations within user community than the centralised web-based approach. The reason is that the users have more control over shared resources. The users are also more willing to share as the resources can be shared directly from their computers without moving around to any third party servers. From the view of resource consumers, they can instantly get access to most updated resources shared by their partners. The ability to provide direct instant communication between users in a community is also an advantage of a peer-to-peer environment.

The result also shows that running scientific simulations on powerful remote computers via Grid services will release computing resources on small user desktop computers for other day-to-day work. Easy access to Grid services via a unified service interface is also an advantage of the architecture.

Despite of these advantages, issues about security, change documentation and, especially, resource discovery need to be carefully considered in the next version of the CeSA. As this prototype system was prioritised on functional aspects of the CeSA, less attention was really paid on the scale issues. In addtion, for the simplification of prototyping, the unified service interface for peer-to-peer applications is currently too simplistic. It needs to be revised for a fully working system.

When the number of users reaches a certain scale, the architecture will need a more efficient resource publication and discovery mechanism. The use of ontology will then be necessary. Ontology will serve two purposes. Firstly, it is for the annotation of shared resources. Secondly, the ontology will help to develop an efficient and scalable discovery mechanism in a peer-to-peer environment.

Acknowledgement

Many thanks to Dr. Andrew Rickard, Dr. Lisa Whalley, Jenny Stanton and other researchers working in Reaction Kinetics group at School of Chemistry, University of Leeds, who have been offering us great help on requirement collection, experiments and evaluation of this work.

References

[1] Access Grid. http://www.accessgrid.org/, 2005.

[2] BADC. http://badc.nerc.ac.uk/, 2005.

[3] D. Barkai. *Peer-to-peer computing: technologies for sharing and collaborating on the net.* Intel Press, 2001.

[4] K. W. Brodlie, D. A. Duce, J. R. Gallop, J. P. R. B., Walton, and J. D. Wood. Distributed and collaborative visualization. *Computer Graphics Forum*, 23:223–251, 2004.

[5] Collaboratory for Multi-Scale Chemical Science. http://cmcs.org/home.php, 2005.

[6] eMule. http://www.emule-project.org/, 2005.

[7] I. Foster and C. Kesselman. Computational grids. In I. Foster and C. Kesselman, editors, *The Grid: Blueprint for a New Computing Infrastructure*, chapter 2, pages 15–51. Morgan-Kaufman, 1999.

[8] I. Foster, C. Kesselman, and S. Tuecke. The anatomy of the grid: Enabling scalable virtual organisations. *International J. Supercomputer Applications*, 15(3), 2001.

[9] I. Foster, C. Kesselmana, J. Nick, and 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 2002.

[10] N. Furmento, W. Lee, A. Mayer, S. Newhouse, and J. Darlington. ICENI: An open grid service architecture implemented with jini. In *SuperComputing 2002*, Baltimore, USA, Nov. 2002.

[11] C. A. Goble, S. Pettifer, R. Stevens, and C. Greenhalgh. Knowledge integration: In silico experiments in bioinformatics. In I. Foster and C. Kesselman, editors, *The Grid: Blueprint for a New Computing Infrastructure.* Morgan-Kaufman, 2 edition, November 2003.

[12] GSC-Chinook. http://www.bcgsc.bc.ca/chinook/, 2005.

[13] ICENI - Imperial College e-Science Networked Infrastructure. http://www.lesc.ic.ac.uk/iceni/index.html, 2005.

[14] JXTA. http://www.jxta.org, 2003. *(Project JXTA website).*

[15] Kazaa. http://www.kazaa.com/, 2005.

[16] G. Kong, J. Stanton, S. Newhouse, and J. Darlington. Collaborative visualisation over the access grid using the ICENI grid middleware. In *UK e-Science All Hands Meeting*, pages 393–396, Nottingham, UK, Sep. 2003.

[17] A. Lawrence and et al. Googling secure data. In S. J. Cox, editor, *Proceedings of the U.K. e-science All Hands Meeting*, 2004.

[18] J. D. Myers and et al. A collaborative informatics infrastructure for multi-scale science. In *Proceedings of the Challenges of Large Applications in Distributed Environments (CLADE) Workshop*, Honolulu, USA, June 2004.

[19] myGrid. http://www.mygrid.org.uk, 2005. *(myGrid project website).*

[20] Napster. http://www.napster.com/, 2005.

[21] National e-Science Centre. http://www.nesc.ac.uk/, 2005. *(NeSC website).*

[22] NERC DataGrid. http://ndg.badc.rl.ac.uk/, 2005.

[23] PrIMe. http://www.me.berkeley.edu/prime/, 2005.

[24] Project JXTA. *JXTA v2.3.x: Java Programmer's guide, http://www.jxta.org/docs/JxtaProgGuide_v2.3.pdf.* Sun Microsystems, Inc., Apr. 2005.

[25] I. J. Taylor, M. S. Shields, I. Wang, and R. Philp. Distributed P2P Computing within Triana: A Galaxy Visualization Test Case. In *17th International Parallel and Distributed Processing Symposium (IPDPS 2003)*, pages 16–27. IEEE Computer Society, 2003.

[26] The Globus Alliance. The Globus Toolkit, http://www-unix.globus.org/toolkit/, 2005.

[27] The Triana Project. http://www.trianacode.org/index.html, 2005.

[28] Y. Tian, L. M. S. Lau, and P. M. Dew. A peer-to-peer knowledge sharing approach for a networked research community. In *Proceedings of the Fifth International Conference on Enterprise Information Systems*, pages 642–645, Angers, France, Apr. 2003.

[29] B. Traversat, M. Abdelaziz, and E. Pouyoul. *Project JXTA: A Loosely-Consistent DHT Rendezvous Walker.* Sun Microsystems, Inc., Mar. 2003.

[30] Virtual Knowledge Park. http://www.leeds.ac.uk/vkp/intro.html, 2005.

[31] A. Woolf and et al. Enterprise specification of the nerc datagrid. In S. J. Cox, editor, *Proceedings of the U.K. e-science All Hands Meeting*, 2004.