A review of medical grids and their direction - A Swiss/Japanese perspective

Nakai, Toshiharu
Neuroimaging and Informatics, National Center for Geriatrics & Gerontology, Ohbu, Japan (toshi@ncgg.go.jp)

Müller, Henning
Business Information Systems, University of Applied Sciences Western Switzerland, Sierre, Switzerland (henning.mueller@hevs.ch)

Bagarinao, Epifanio
Brain and Mind Research Center, Nagoya University, Nagoya, Japan (ebagarinao@met.nagoya-u.ac.jp)

Tomida, Koji
Nagoya Institute of Technology, Graduate School of Engineering, Nagoya, Japan (tomida.koji@nitech.ac.jp)

Shiraishi, Yoshiaki
Kobe University, Graduate School of Engineering, Kobe, Japan (zenmei@port.kobe-u.ac.jp)

Niinimaki, Marko
CERN, Helsinki Institute of Physics, Geneva, Switzerland (marko@cern.ch)

Received: 15 February 2015 Revised: 7 March 2015 Accepted: 4 May 2015
Available Online: 15 June 2015 DOI: 10.5861/ijrsc.2015.1109

Abstract

This paper presents a general and comparative review of the advances of grid technologies in medical sciences since 2000, with a special emphasis on Europe and Japan. The EU has over the years funded several big projects in the Grid and now the cloud area, covering besides the high energy physics domain also bioinformatics and medical applications as in image analysis. Bioinformatics and pharmaceutical design have been the major targets of grid applications in Japan. Grids and clouds share many of the same goals and techniques, although clouds are more commercially oriented and privately provided services and grids are rather publicly initiated resource sharing platforms between institutions without strong economic objectives. In the future, merging these two may potentially solve inter-operation problems of grids, which have been limiting the propagation of biomedical grids.

Keywords: medical grid; biomedical science; cloud computing; Switzerland; Japan; e-science
A review of medical grids and their direction - A Swiss/Japanese perspective

1. Introduction

The phenomenon described by (McFedries, 2011) as “data deluge” has affected many scientific fields. Almost since its beginnings, grid computing - sharing of resources of potentially heterogeneous clusters of standard computers across organizations - has sought to address the challenges intrinsic to large data sets as (Chervenak, Foster, Kesselman, Salisbury, & Tuecke, 2000). In this paper, we discuss the applications of grid technologies in the medical field and also highlight links and differences with cloud computing. As an example, we consider in particular the use of grid technologies in medical imaging. Images in medical diagnosis and treatment planning are produced in ever-increasing quantities and also in increasing modalities. For example, the radiology department in the Geneva University Hospitals (HUG) currently produces on average over 200,000 images per day in 2012, up from 70,000 in 2008. Many of them are large series of tomographic images (Niinimaki, Zhou, Depeursinge, Geissbuhler, & Müller, 2008), making it challenging for any clinician to fully understand the available information in all possible exams. Medical images are estimated to occupy 30% of world storage in 2010 (The High Level Expert Group on Scientific Data, 2010), another reason to highlight this area in this article.

The availability of medical images in digital form, often via the electronic patient record, has also transformed clinicians’ attitude towards the use of these data. As images are now readily accessible not only to experts (i.e. radiologists) but also to all clinicians with access to the electronic record, new tools are needed to handle the data and aid clinicians, who are not experts in the imaging modalities, in properly interpreting and optimizing the use of the available information. Content-based medical image analysis and retrieval (CBMIA/CBMIR) can be a technology to aid this step (Montagnat et al., 2008). With CBMIA/CBMIR, past cases similar to a current one, including the cases’ outcomes, can be retrieved by taking into account not only the textual metadata associated with the image but also the image content itself. In the context of case-based reasoning and evidence-based medicine, access to the scientific literature treating similar problems and using similar images can also be provided.

The effectiveness of the CBMIA/CBMIR technology, however, strongly depends on computing and networking infrastructures as well as on the available information resources. For instance to improve the quality of information retrieval, several types of visual features need to be extracted from medical images. However, extracting these features and testing them on standard benchmark databases to optimize the quality of the results is a computationally demanding task. This is where grid computing or other distributed computing could play a significant role. A computational grid is generally understood as a “system that coordinates resources that are not subject to centralized control, using standard, open, general-purpose protocols and interfaces to deliver nontrivial qualities of service” (Foster, 2002). Grid computing could therefore allow access to larger amounts of information as well as the needed computational resources to enable more sophisticated visual analysis.

Here, we encounter a problem specific to medical images. Medical images for clinical purposes are under a very strict access control to protect patients’ personal information, i.e. privacy of the patients. These images are therefore only accessible within the organization and are usually protected by a firewall. This requirement is different from the grid applications for e-science based on the original concept of grid computing to achieve seamless computing over the network. Thus, grid systems for clinical services are often organized as a “community GRID”; where grid technologies harness the resources within only one organization or administrative domain. This concept of community grids and the demands for CBMIA/CBMIR supported by distributed and globally accessible medical image resources are incompatible.

To investigate the concept of “bridging grids” in order to enable the linking of medical image resources
A review of medical grids and their direction - A Swiss/Japanese perspective

Across different organizations or administrative domains, a collaboration research between Switzerland and Japan to review the trends of biomedical grids both in the EU and Japan was formed. The discussion was initiated in 2009 during the SNSF/JSPS International Scientific Seminar held in Grimentz, Switzerland. This paper serves as a general review of the advances of grid technologies in medical sciences since 2000 and up until now, with a special emphasis on Europe/Switzerland and Japan.

In this work, we will review which grid technologies have been employed so far and what are the principal areas of their application (medical image analysis, data sharing, service integration, and genome research, among others). Many of the applications have been originally centered on a particular grid system or project. Large European projects like Enabling Grids for E-science (EGEE) (Jones, 2005) and the European Grid Initiative (EGI) (Laure et al., 2006) have established a continent-wide software and hardware infrastructure for grid applications. Several grid middlewares have been developed in Europe; one of the tasks of the EGI project is to build a unified middleware distribution (EGI_DS, 2008), adopting components from Advanced Resource Connector (ARC) (Ellert et al., 2007), gLite (Laure et al., 2006; Laure & Jones, 2008) and UNICORE (Erwin, 2002). In Switzerland the Swiss Multi Science Computing Grid (SMSCG) project is a connection of many existing resources in a number of organizations (Stockinger et al., 2009).

In Japan, the NAREGI (National Research Grid Initiative) project was started for development of multipurpose GRID infrastructure (Miura, 2006), which was dedicated to integrate super-computing resources to establish the environment for academic researches. The NAREGI GRID middleware collaborating with Globus Toolkit was developed to enable tera-floating-point instructions per second (TFLOPS) computation on SuperSINET network to promote collaboration among leading academic research institutes (Miura, 2006). Despite the fact that the term grids is less commonly used, the described applications and ideas have much in common with cloud computing or other distributed computing schedulers, also for example Hadoop/MapReduce (Markonis, Schaer, Eggel, Müller, & Depeursinge, 2012).

2. Developments of Medical Grids

2.1 GRID development and biomedical applications in Europe

Though medical data and computing collaborations have been organised earlier, to our knowledge the first medical application environment that used a grid middleware was introduced by von Laszewski (Foster, von Laszewski, Thiruvathukal, & Toonen, 1998). This "telemedical environment", however, focused on tools and infrastructure rather than actual applications or collaborations. Somewhat later, demonstrations of biomedical applications were set up using grid tools (Breton, Medina, & Montagnat, 2003) also in connection with the first HealthGrid conference in Europe, and large collaborations like caBIG (Saltz et al., 2006) in the U.S. started taking shape.

Not surprisingly, the most often cited aims of medical grids are data sharing within a particular collaboration, and distributed computing of medical data. As an example of data sharing, TRENCADIS (Espert, Garcia, Anastasio, & Quilis, 2009) is a software created for sharing DICOM images using grid technologies. A system to share and analyse 3D images generated from CT and MRI scans is discussed by Marovic and Jovanovic (2006). With a focus in distributed computing, a grid-enabled analysis of mammograms for breast cancer detection, CT-scans for lung disease detection and PET scans for Alzheimer disease detection is described by Bellotti et al. (2007).

MediGRID (Krefting et al., 2009) is a suite of applications and methods based on the resources of D-Grid. The emphasis is on the institution perspective of grids in medicine: most researchers work in university hospitals where the network is highly regulated by the hospital's policies and the users of the grid applications are not computer scientists or programmers. Using grids in hospital environments has also been discussed by Niinimaki et al. (2008). A summary of medical grid applications in the EU is shown in Table 1.
Table 1

<table>
<thead>
<tr>
<th>System</th>
<th>Middleware / Components</th>
<th>Novelty</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRENCADIS (Espert et al., 2009)</td>
<td>SRB, gridftp, VOMS</td>
<td>DICOM - ontology mappings</td>
</tr>
<tr>
<td>VIVE (Maronic et al., 2006)</td>
<td>EGEE-LCG-2</td>
<td>Web-based analysis of 3D images</td>
</tr>
<tr>
<td>MAGIC-5 (De Mitri &amp; Collaboration, 2005)</td>
<td>AliEn</td>
<td>distributing large-scale data processing</td>
</tr>
<tr>
<td>MediGrid (Krefting et al., 2009)</td>
<td>Globus4, SRB, VOMS</td>
<td>service integration, ease of use</td>
</tr>
<tr>
<td>arcGIFT (Niinimaki et al., 2008)</td>
<td>ARC, virtual machines on windows desktops</td>
<td>Grid network inside a hospital, desktop grid, for medical image analysis</td>
</tr>
</tbody>
</table>

2.2 GRID development and biomedical applications in Japan

In 2003, the NAREGI program (Miura, 2006) was started to integrate super-computing resources over the country with the goal of establishing a multipurpose environment for basic research. The NAREGI middleware employed in Super SINET is designed for co-allocation of resources from multiple sites and execution of complex workflows of jobs. The representative modules, based on the Open Grid Forum (OGF) standard, include Super Scheduler to handle the workflow over virtual organizations (VO), GridVM for computation resource management and the NAREGI information service (IS). SINET, operated by the National Institute of Informatics (NII), supplies the high-speed Internet backbone consisting of 45 edge nodes and 8 core nodes in SINET4 by using NAREGI middleware. This network is mostly dedicated to five research fields including high energy and nuclear fusion, space and astronomical science, supercomputer-interlocking distributed computing, nanotechnology and bio-informatics (genome information).

As applications in bio-science research utilizing these GRID infrastructures, two representative projects have been conducted in Japan. One of these is the Open Bioinformatics Grid (OBI Grid) (Konagaya, Konishi, Hatakeyama, & Satou, 2004) started in 2002. Its goal is to provide application tools such as distributed bioinformatics environments, scalable genome databases, genome annotation systems, and biochemical simulators, among others, to the social structure of virtual organization of the participating researchers and engineers. In this project, a Grid-oriented Genetic Algorithm (GOGA) Framework was proposed as a search engine to efficiently find the mutual interactions among genes from gene-expression time-course data by using master-worker model (Imade, Mizuguchi, Ono, Ono, & Okamoto, 2005).

Another project, Biogrid, initiated in 2002 as a testbed for in silico pharmaceutical research and development (Fukunishi, 2009) and for the establishment of a super computer network core technology to be used in other research fields with similar needs. This project attempted to expand the application of grid technologies to e-science topics such as remote use of the ultra-high voltage electron microscopy, remote X-ray observation system for accelerator (Spring 8) and brain function analysis (Mizuno-Matsumoto, Date, Kaishima, Kadobayashi, & Shimojo, 2002). Overall, bioinformatics and pharmaceutical design have been the major targets of grid applications as exemplified in the HPCI consortium dedicated to the development of analysis software and database system of bio-molecules to support industries. On the other hand, medical images have been mostly confined within the individual facility’s commercially available systems (e.g., picture archiving and communication system or PACS), and there have been no systematic attempts to share medical images across facilities or countries to augment their utility except in the medGRID project.

2.3 The medGRID project

The medGRID project was initiated in 2004 with the goal of developing a real-time processing system for neuroimaging datasets between Japan and the Philippines (Bagarinao, Tanaka, & Nakai, 2007). MedGrid in 2009
A review of medical grids and their direction - A Swiss/Japanese perspective

included four countries and aimed to achieve the following goals by using high performance GRID middleware (Bagarinao et al., 2007; Nakai et al., 2009): (1) to provide a neuroimaging data-sharing system among the participating organizations; (2) to supply high-performance computational services for data analysis; (3) to organize image analysis and visualization tools; (4) to provide a data fusion platform among different neuroimaging modalities; (5) to allow neuroimaging data retrieval based on semantic indexing; (6) to provide standardized imaging protocols and a standardized framework for data archiving and; (7) to assist in the communication among the participant sites. In order to support the management and processing of functional magnetic resonance imaging (fMRI) data sets within the medGRID testbed, a grid-enabled analysis software package called BAXGrid and an integrated neuroimaging database system called BAXSQL (Bagarinao, Matsuo, Nakai, & Tanaka, 2008) were developed. BAXSQL provided an infrastructure to retrieve fMRI data sets from data servers distributed across medGRID participating sites and applied temporal analysis of brain activation via BAXGrid.

BAXSQL was built on top of the NinfG, a reference implementation of the GridRPC API (Tanaka, Nakada, Sekiguchi, Suzumura, & Matsuoka, 2003). In the most recent version (version 5) of NinfG, remote procedure calls are made via various protocols and middleware such as SSH and the Globus Toolkit (Foster, 2005). Moreover, it no longer assumes a specific grid middleware as a prerequisite, unlike in previous versions. NinfG version 5 also works with non-Globus Toolkit environments. One of the features introduced is the implementation of the invoke server module. With this module, NinfG-based applications can use approaches, such as Globus Toolkit’s Web Services, Condor, or SSH, to invoke remote processes depending on the remote grid environment. This significantly simplifies the inter-operation of applications between different grid environments. Though NinfG is a part of the NAREGI middleware, BAXSQL uses only NinfG.

BAXSQL was implemented using a server-client architecture. The server component of BAXSQL runs on remote data servers and manages the back-end database system used for storing the metadata associated with the available neuroimaging data sets. The client component of BAXSQL can be used to upload data sets, to access both local and remote data servers to browse available data sets, edit metadata information, or download data sets. The client component can be also used to request from the remote server BAXSQL’s built-in support for common image pre-processing and basic statistics for fMRI.

The medGRID project was also part of an EU-Asia collaboration called ONCO-Media project organized as a consortium for developing a grid-distributed, contextual and semantic based, intelligent information access framework for medical images and associated medical reports (Brezillon & Racoceanu, 2007). BAXSQL has been applied to collaborative research across countries (Chen, Tseng, Nakai, Bagarinao, & Matsuo, 2008; Nakai et al., 2009).

3. Communication among GRIDs

As discussed in the previous section, early applications of grid technologies were often technology trials or concentrated on the quantitative side of processing – i.e., grid technologies enable analysis of bigger data sets in shorter time. However, instead of just using a grid middleware to run medical jobs, other projects have introduced workflows (Cao, Fingberg, Berti, & Schmidt, 2004) and ontologies (Jin, Sun, Zheng, He, & Zhang, 2009) in connection with grids. Certain applications using the technology such as in parametric studies (Soleman, Glatard, Veltman, Nederveen, & Olabarriaga, 2008) were also reported.

Given the confidentiality of medical data, authentication and authorisation in accessing data are compulsory. Grid applications need to be able to address this (Erberich, Bhandekar, Chervenak, Kesselman, & Nelson, 2007). Grid technologies commonly use X.509 certificates and a virtual organization-based authorisation. Kommeri et al. (2011) discuss a safe medical data storage system for medical images, along the guidelines of the MDM (Medical Data Manager) system (Montagnat et al., 2008). In this study, a safe metadata storage system was integrated with search functionality. After authentication and authorisation, the user is able to query the system.
3.1 Inter-operation between GRID middlewares

There have been several attempts to establish a connection between grid middlewares. Many approaches are based on job submission based on the OGF Job Submission Description Language (JSDL) standard (Anjomshoaa et al., 2005). The inter-operability between the NAREGI middleware and gLite was reported by Nakada et al. (2007). Modules developed for information exchange and job submission enabled inter-operation with some latency. No problems occurred for security layers such as certificates and virtual organization (VO) management. Inter-operability between gLite and the ARC (Advanced Resource Connector) middleware has been reported by Gronager et al. (2008).

The idea of “bridging GRIDs” is different from inter-operation between grid middlewares. We encounter two challenges, (1) the cost to newly develop modules for inter-operation and (2) the security assurance to handle medical data, if we attempt to connect community GRIDs by using inter-operation tools. Overcoming technical complexity is another question in bridging GRIDs. Olabarriaga et al. (2009) encountered a large scope of inter-operability questions when bridging the Dutch VLeMed and the German MediGRID system, though their architectures are quite similar.

3.2 Grids and Clouds

In contrast to grids mostly characterized by their middlewares, such as the Globus toolkit, gLite and ARC, a standard definition of Clouds has not been firmly adopted. Several characteristics to explain their functionality have been mentioned such as service delivery over the Internet, scalable and virtualized resources, on-demand network access to a shared pool of configurable computing resources (Mell & Grance, 2010). Although the concepts of both cloud and grid are similar, their usage is different, i.e., grids are used in academic fields and clouds are rather used in business even though the cloud providers target academic users and programs exist to obtain computing resources at low prices or free of charge. The details of management or implementation of service in clouds are hidden from users and users cannot control the data administration in clouds. Specifically, users cannot confirm security measured by cloud. This is why SLA (Service Level Agreement), which defines services between the service provider and the user formally, is regarded important in the business field (Walker, 2012). On the other hand, techniques of authentication and authorisation are already well established in the grid. GSI (Grid Security Infrastructure) is the security infrastructure provided in Globus Toolkit and the de facto standard for grid middlewares. GSI provides many useful services for grid, such as secure authentication, communication protection and authorisation by using public key encryption, X.509 proxy certificates and so on.

As an example of a medical cloud, in the area of human genomics, the EU Biobankcloud project (Kuhn et al., 2014) develops a cloud-computing Platform as a Service (PaaS) for the storage, analysis and inter-connection of biobank data.

4. Summary and conclusions

Grid systems have many roles in medical informatics and medical data analysis. Although the processing and storage capacity of individual computers has grown considerably in recent decades, the capacity of instruments and medical systems to produce data has grown at the same rate or even faster. Here, we discussed several grid technologies (distributed computing, data access), frameworks (BAXSQL, middlewares) and applications to help with the processing of large quantities of medical data.

Grids seem to be established in multi-organization medical collaborations. It is less evident whether they are widely used in individual hospitals. As an example, Pitkanen et al. (2008) interviewed health care and public sector professionals. The professionals were well aware of distributed computing technologies and virtualization, and saw the benefit of harnessing the power of office PC’s as a “desktop Grid”. Some desktop grid systems have
been used in pilot projects in hospitals, too, but to our knowledge they have not been adopted for long-term use. Some authors (Philbin, Prior, & Nagy, 2011) see utility in cloud computing offering professional service quality and a sustainable solution with clear interfaces. On the other hand, grid and cloud technologies are regarded as complementary (Rings et al., 2009). Volunteer computing systems have likewise been used in medical computing (Abdennadher, Evequoz, & Billat, 2008; Ben Belcagem, Abdennadher, & Niinimaki, 2012).

The comparison of the grid evolution between Switzerland and Japan revealed that not only grid middlewares but also the social demand for information services are different. As a future direction, soft inter-connection technologies using Clouds in connection with biomedical GRIDs on-demand may be more optimized for many medical institutions.

5. References


