Paper

Health applications and grid technologies

Fadi-Sotiris Salloum

Abstract— Biology and genomics in the near future will play a major role in day to day activities related with health. Today biologists and doctors work together in order to decode and analyze large amount of data extracted from the DNA analysis of proteins. Advanced health applications will be needed in order to store, retrieve, process the large amount of data – being produced today by genomics and bioinformatics analysis – in order to extract useful results in a reasonable time frame. In this paper we present the results from our research regarding the use of grid technologies with health applications. We present the current status of standardization activities and working groups, which are currently involved with the specification of health applications and the standardization of needed components such us security, functionality, etc., which are being introduced by the use of grid technologies.

Keywords—grid technologies, health applications, health grid, genomic medicine, grid enable pharmaceutical, open grid services, distributed systems.

1. Introduction

In this paper we investigate the current status regarding the use of grid technologies in health applications. We begin by giving a definition of grid and describing the basic types of grid and some of current practical implementations. Afterwards we present the main reasons for using grid technologies in health applications and we present also the most important issues for health applications. The second part of this paper, present three examples of health application incorporating the use of grid technology, more specifically we present the use of grid for medical imaging and analysis, grid enabled pharmaceutical R&D and grid and genomic medicine. The third part of this paper present the overall conclusions related with health applications and grid technologies.

2. What is grid?

Technically speaking we all understand grid as an computing infrastructure which enables the virtualization of distributed computing and data resources such as processing, network bandwidth and storage capacity to create a single system image, granting users and applications seamless access to vast IT capabilities [1–3]. Just as an Internet user which views a unified instance of content via the web, a grid user essentially will see a single, large virtual computer. At its core, grid computing is based on an open set of standards and protocols, e.g., open grid services architecture (OGSA), that enable communication across heterogeneous, geographically dispersed environments. With grid computing, organizations can optimize computing and data resources, pool them for large capacity workloads, share them across networks and enable collaboration.

Today grid is not a revolution (like the Internet) it is a kind of a latest and most complete evolution of familiar developments such us distributed computing, the web, peer-to-peer computing and virtualization technologies. This evolution was a natural next step due to the increase use of information technology on every aspect of our life. The grid paradigm offers CPU and data handling capabilities to the user, the same way that today we consume electric energy. Grids use a middleware to interconnect computers so they are able to exchange information without intervention of the user. This differs from the Internet where the user has to choose on which machine he wants to connect and which information he wants to retrieve among the ones available.

On a grid, the exchange of information between computers is hidden from the user. High level services (resource broker, distributed file system…) hide the underlying infrastructure required to respond to the user requests. This transparency from the user point of view requires an extra layer of software, called middleware. Beside research projects dedicated to developing new middlewares or enhancing the performances of the existing ones, grids should be deployed to address the needs of the biomedical community using the state of the art of the middleware technology.

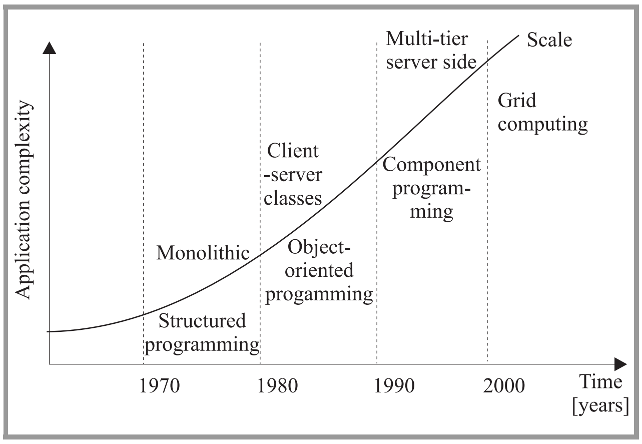

Fig. 1. Grid evolution.

In the above diagram (Fig. 1) the relation between the scale of application complexity and the evolution of programming is given. During the 70's structured programming was appropriate for the small application complexity, with the introduction of client server applications during the 80's and the need for more managed applications. Object oriented programming was introduced with the first remote communication between applications processes. During the 90's the application complexity was increased further and multi-tier server side applications where introduced, the component programming and the standardization for components developments has taken place. During the last years the application complexity has increased resulting on increased need for processing power and flexible data/content management, the grid computing with the parallel help from evolutions in data communications and integrated circuits technologies seem to be the main player.

2.1. Grid types

Currently grids can be classified into three categories, computational, content and collaborative grids [4]. The goal of a computational grid is to create a virtual supercomputer, which dynamically aggregates the power of a large number of individual computers in order to provide a platform for advanced high-performance and/or high-throughput applications that could not be tackled by a single system. Content grids include data and information or knowledge grids. Data grids focus on the sharing of vast quantities of data while information and knowledge grids extend the capabilities of data grids by providing support for data categorization, information discovery, and knowledge sharing and reuse. It must be noted here that the grid-enabled application service providers approach combines aspects from both of computational and content grids to provide appropriate services to users or a grid application environment. The last type collaborative grids establish a virtual environment, which enables geographically dispersed individuals or groups of people to cooperate, as they pursue common goals. Collaborative grid technologies also enable the realization of virtual laboratories or the remote control and management of equipment, sensors, and instruments. Currently example of collaborative grids can be found mainly in car and airplane industries but in the future it is estimated that can be found for health sector, an example could be the setting of a virtual laboratory for fighting viruses (i.e., SARS virus last year).

2.2. Are grid's practical enough?

This is a question made by major businessmen which like to invest in grids to play roles such us grid application service providers, today the opportunity exists to begin providing practical grids, everyday we witness dramatic increases in a wide range of areas such VLSI technology, data communications (increased bandwidth), storage technology and recently in health care devices. The continuous cost decrease, has been the main factor for increased interest today in the implementation of grids. It is not surprisingly that grid farms can be build for less than 100 000 € immediately accessible by everyone having an xDSL connection or even from a WLAN network.

But the above advantages are not enough, grid vision (namely the vision of large scale resource sharing) has not yet become a reality, from our research the following main reasons were identified:

- the lack of commonly accepted standards (for management, for service development, for power accessing);
- the diversity and fragmentation of available grid middleware, tools and services.

Even if the solution is not available today, we estimate based on our research that the vision will be achieved in the next two to five years, not only for government grids. The reason for this conclusion is simple, major companies and institutes have identified grid as the way forward in order to handle the extremely complicated processing needs of the 21st century. This has resulted in two main middleware for grid implementations, the globus toolkit [5, 6] (used by more than 20 existing project currently and has been voted as the most preferable) and unicore. Additionally the need for standards, resulted major companies to support actively the global grid forum (GGF) and it is reasonable to say that soon commonly accepted standards will be available. An example is the open grid services architecture approach with support from the largest server vendors like IBM, Oracle Sun and Microsoft. The open grid services architecture approach combines perfectly web services with grid solutions and it is estimated to increase in the next years the involvement of 3rd party software providers to produce grid targeted software.

2.3. Health applications

After having described grid, in this section we explain the reasons why health applications are stongly related with grid technologies. Before continuing it is usefull to present some facts regarding modern health applications. First it is well known that health applications are characterized by the availability of large amount of data (clinical, genomic, proteomic, etc.) which are additionally found in heterogeneous sources and formats. The last four years, we are witnessing a rapid progress in fields such as computer based drug design, medical imaging and medical simulations, this progress has lead to a growing demand for large computational power and easy accessibility to heterogeneous data sources in the health domains.

Modern health applications include also research from large biology and bioinformatics teams spread all over the world currently located on large research centers with the target to be located in hospitals. The true success for this new approach (biomedicine) will be if every biologist has immediate access to processing power in order to treat immediately their patients. With the introduction of medical imaging in modern hospitals there a growing need for large computations analysis of 2D, 3D and 4D images, pattern matching and other complex software applications [7].

Health grids are grid infrastructures comprising applications, services or middleware components that deal with the specific problems arising in the processing of medical data. Resources in health grids are databases, computing power, medical expertise and even medical devices. Health grids are thus closely related to e-health. Although the ultimate goal for e-health in Europe would be the creation of a single health grid, i.e., a grid comprising all e-health resources, naturally including security and authorisation features to handle subsidiarity of independent nodes of the health grid, the development path will mostly likely include a set of specific health grids with progressive inter-grid interaction/interoperation capabilities.

From these facts it is obvious that good potential exist for grid technologies in health applications, In the next years information technologies will play more and more important role, the introduction of bioinformatics applications for diagnosis and production of personalized therapies is a vision which is steadily becoming a reality and most probably will be here to witness it. Based on this facts modern and future health applications produce the following requirements, first there is a strong demand of computing power, a categorized storage power (including not only data but also information and knowledge which required some additional specialized indexing software) and geographically distributed information (a recent example is the SARS virus and it is obvious to understand the importance of having such infrastructure in such cases). But for all of these to be successful the most important requirement is to have transparent and secure access from everywhere, in order to enable the use by the simple physician and biologist.

In order for grid technologies to be used in order to satisfy these requirements, existing identified problems in current grid technologies must be resolved. Firstly the lack of adequate standardized infrastructure, a problem which most probably will be resolved within the next years with the global grid forum and the open grid services architecture which is backed up by industry leaders. Secondly is the lack of user's confidence which is strongly related with the shortage of applications and the lack of investments from software vendors (a trend which is stopped with the recent examples from Oracle and IBM).

2.3.1. Health applications needs

From the above thoughts it is obvious that are needs, for applications dealing with large-sized information processing, such as medical imaging. Breast cancer imaging has been the focus of several successful grid projects; and the efforts have been concentrated on federating and sharing the data and the implementation of semi-automatic processing tools that could improve the specificity and sensibility of breast cancer screening programs [8–12]. Additionally many efforts have been invested to reduce the information needed to be exchanged and to protect privacy of the information, such us:

- Applications dealing with person-centric grid for health approaches, making the information available to the whole health community (patient, relatives, physicians, nursery...) considering always access rights and language limitations.
- Applications dealing wit bioinformatics where the study of the contents and information flow in biological process and systems requiring large processing powers and with main challenge the development and maintenance of an infrastructure for the storage, access, transfer and simulation of biomedical information and processes.
- Applications collecting and automatic processing genomics and post-genomics data around the world, with immediate availability of the information on all levels from molecular to population.

All this needs have been discussed thoroughly for the last two years by major centers all over Europe, the result of all this discussions is the vision which is presented as a mission statement on the health grid organizations. The vision of the health grid [8, 9, 13] is to create, based on the grid technologies an environment where information at the 5 levels (molecule, cell, tissue, individual, population) can be associated to provide individualized healthcare.

2.3.2. Health applications features

In order for this vision to be achieved six key features of health grids have been identified:

- Security, since health grids are more closely related to data it is important to leverage security up to a trustworthy level of confidence that could release a generalized access to data from the outside (see also below). Due to the fact also that data storage is a responsibility for hospitals, many business opportunities can arise from data sharing and the increased security will also protect investments made.
- Security in grid infrastructures even if it is currently sufficient for research, in the future it must be improved to ensure privacy of data. Encrypted transmission and storage is not sufficient, integrity of data and automatic pseudo-anonymisation must be ensured to guarantee that data is complete and reliable and privacy leakages can not appear due to unattended use of the resources.
- Robustness and fault tolerance of grids fits very well to the needs for "always on" medical applications. Grid technologies can ease the access to replicated resources and information, just requiring the user to have a permanent Internet connection.
- Flexibility is needed for the controlling and managing the grid infrastructures, the management of resources should be more precise and dynamic, depending on

many criteria such as urgency, users' authorization or other administrative policies.

- The ability to access health grid resources from everywhere is very important feature of health grids, the combination of WLAN's availability and the evolution of handheld PC's make this feature increasingly important.
- The standardization of protocols for plug and play health devices with grid infrastructures is very important feature since it will minimize the needed time and make more useful the use of such devices.

2.3.3. EU activities

Europe invests heavily at the moment in health grids, several European projects from the 6th framework deal with health grid approach with several implementations [8]. The importance of the use of grid technologies in health applications has been recognized and the most active consortium in this field is the health grid organization. The first conference was held in Lyon in 2003 and from this conference the main mission statement was the development of an intelligent environment that enables ubiquitous management of citizen's health status and to assist health professionals in coping with some major challenges, risk management and the integration into clinical practice of advances in health knowledge.
The conference lead to the creation of the health grid association a non-profit research association legally located in France but formed from a wide community of European researchers and institutions sharing expertise in health grids. The second conference was held in Clermont-Ferrand this year with increased participants from all over Europe.
The most recent activity was to publish the *HealthGrid White Paper* which can be downloaded from the health grid website.

2.4. Examples of health applications with grid

After we have described the main needs of health applications and the solutions which can be provided by grids, in his section we will present some examples of health applications with grid technologies. The selected applications are:

- medical imaging and analysis;
- pharmaceutical R&D grid enabled applications;
- genomic medicine applications.

2.4.1. Medical imaging and analysis

Medical imaging has been one of science's greatest contributions to medicine. It has evolved over a century, from the earliest X-rays to the latest developments in functional MRI, to become one of the principal tools of the physician in diagnosis, therapy, and research. This field has been evolved due to IT technologies which resulted, in a growing range of available data available to the clinician, from X-ray (currently increasingly digital to ultrasounds, scans, etc.). Additionally due to the evolution in data communication with the introduction of high bandwidth systems for picture archiving and communications there is a trend of increasing use in digital format, as a result several picture archiving and communication systems (PACS) have been deployed in hospitals, but today systems suffer many limitations. As an example most of the time these systems are not connected with the hospital administrative information systems, and they are often proprietary solutions of medical imaging companies due to the absence of open standards to ease communication between different PACS. Additionally most of that system doesn't ensure security and confidentiality of personal data, the solution for the above is the use of a standardized datagrid supporting distributed and mass storage of data, and a standardized middleware could provide access via web services.
Beside the medical imaging (storing, archiving the image data) another important aspect is the medical image processing, computerized medical image analysis algorithms have been developed for two decades or so grid technologies will offer an access to large processing power suited to processing full datasets in reasonable time, compatible with the needs for experiencing new algorithms. Grid technologies will also ease the sharing of algorithms developed by different research groups thus encouraging comparative studies. Based on the above grid technologies are expected to boost the production of medical image analysis algorithms and to facilitate their quality improvement.
Medical imaging has been one of science's greatest contributions to medicine. It has evolved over a century, from the earliest X-rays to the latest developments in functional MRI, to become one of the principal tools of the physician in diagnosis, therapy, and research. As with imaging in any other field, doctors have to apply very subtle knowledge and skill in interpreting the images they make. This can be improved by providing them with greater access to images and associated histories, with standards, with quality processes and with better and broader-based training.

2.4.2. Grid enabled pharmaceutical R&D

Our second example includes grid enabled pharmaceutical R&D, in general pharmaceutical R&D combines the utilization of diverse and complex information including:

- chemical identities;
- biological sequences, structures;
- genomic, toxicological and clinical trials patient data.

Most advanced current modern pharmaceutical developments include billion character long encoded sequences: full text scientific reports and millions of records of prescriptions and physician encounter re-imbursements.
It is obvious that all the above yield technical challenges in several fields such us knowledge representation and integration, distributed systems search and access control, data mining and knowledge management, real-time modeling and simulations, algorithm development and of course computational complexity.
Grid seems to be an optimum solution for pharmas, providing extremely large CPU power to perform computing intense tasks in a transparent way by means of an automated job submission and distribution facility. Grids could provide transparent and secure access to storage and archiving of large amount of data in an automated and self organized mode. Additionally grids could be used as a framework for connecting and analyzing structures data and information in a transparent mode according to pre-defined rules. Combining all these it is realistic to say that pharmaceutical grids could open the perspective of cheaper and faster drug development and in the near future of personalized medicine.

2.4.3. Grid and genomic medicine

Our third and last example describes a scenario with grid and genomic medicine. Genomic medicine as all current innovative health technologies, include diverse and complex information, such information are genomic and proteomics information:

- information describing relationships between genotype, phenotype and environmental parameters;
- use and integration of data for populations, diseases, patients along with genomics and post genomics data.

Of course the availability of such information is only the one part; the other is that this care delivery must be empowered by knowledge intensive tools, assisting the professional with integrated view of a patient's conditions. Again grids can provide solution in several ways, firstly grid enabled infrastructures for life sciences and care provision can catalyze the use of common practices in a domain that is highly fragmented. Secondly biomedical grids can deliver secure distributed platforms for health applications. As a result the contribution of grid technology could help genomic medicine realization with:

- the development of models and digital simulations of cells and diseases;
- improved biomedical knowledge management;
- introduction of advanced techniques such us molecular imaging;
- genetic epidemiology.

Combined all these will increase the development of tools for supporting clinical decision making.

3. Conclusions

In this paper we presented the status of the use of grid technologies in health care applications. Our research was aimed first to identify the current status of grid technologies, and secondly to investigate the possible use of grid technologies in health applications. From our research, the first finding was the fact that currently a strong need exist of immediate access to high processing power and transparent data storage in health related applications. The advances in the genomic and bioinformatics industry soon will result on even more increased need for processing and data retrieval.
From our research we concluded that current grid technologies are not yet matured for immediate use in health applications, mainly due to the lack of commonly accepted standards for several aspects such us security, management, application development, health devices integration, and the diversity and fragmentation of available grid middleware, tools and services. Today in the EU several activities and EU funded projects are working towards to overpass those obstacles. Based on the current activities from our research we estimate that in the next two to five years grid technologies will be mature enough, mainly due to the activities taken from global grid forum, which is one of the most active standardization bodies which are driving global standardization efforts for grid services, protocols, and interfaces.
Another important fact which was also identified from our research is the open source community which helps the faster adaptation of standards between the communities. Currently several grid solutions exist (with the most known the open globus toolkit) for the establishment of open source based grid services.
Additionally another fact which also increases our approach is the the combination of grids with web services and Java/XML based middleware utilizing technologies such us WSDL. We believe that the above mentioned activities and facts about current grid standardization activities will increase dramatically the development of grid applications in general and more specifically for health applications.
As a final conclusion, from our research and based on the above mentioned facts, it is clear that Europe invest heavily at the moment in health grids with several european projects from the 6th framework dealing with health grid approach with several implementations.
The importance of the use of grid technologies in health applications has been recognized and in Europe the most active consortium in this field is the health grid organization, which attempts to define health applications involving grid technologies and proceed to test trials in cooperation with European Union research projects. The major

result from these activities is the *HealthGrid White Paper* available from the health grid website, which present several initial health applications involving grid technologies.

References

[1] I. Foster and C. Kesselman, *The Grid Bluebrint for a New Computing Infrastructure*. San Francisco: Morgan Kaufmann Publ., 1999.

[2] I. Foster, C. Kesselman, J. Nick, and S. Tuecke, "The physiology of the grid, an open services architecture for distributed systems integration", in *Proc. 5th Global Grid Forum Worksh. GGF5*, Edinburgh, Scotland, 2002.

[3] I. Foster, C. Kesselman, and S. Tuecke, "The anatomy of the grid, enabling scalable virtual organization", *Int. J. Supercomput. Appl*., vol. 2001, http://www.globus.org/research/papers/anatomy.pdf.

[4] Globus toolkit papers, www.globus.org/research/papers.html

[5] Introduction to globus project, joint work of USC Information Sciences Institute and Argonne National Laboratory, http://www.globus.org

[6] H. Tian, "Grid computing as an integrating force in virtual enterprises". Master thesis of engineering in civil and environmental engineering, Massachusetts, MIT, June 2003.

[7] I. Bilykh, Y. Bychkov, D. Dahlem, J. H. Jahnke, G. McCallum, C. Obry, A. Onabajo, and C. Kuziemsky, "Can GRID services provide answers to the challenges of national health information sharing?", in *Proc. Conf. Cent. Adv. Stud. Collabor. Res*., Toronto, Canada, 2003.

[8] HealthGrid.org, "HealthGrid-White_Paper-Draft_v.1.1-5", 22/09/2004, http://www.healthgrid.org/main/library/images/doc.gif

[9] Sofie NØRAGER, European Commission DG INFSO, eHealth unit, "The vision of a HealthGrid", 2003.

[10] T. Hey, "The UK e-science, next generation grid applications", *Int. J. High Perform. Comput. Appl*., vol. 18, no. 3, pp. 285–291, 2004.

[11] "Grid research integration development & support center", http://www.grids-center.org

[12] H. Müller, A. Garcia, J. P. Vallée, and A. Geissbuhler, "Case study: grid computing at the University Hospitals of Geneva", University Hospitals of Geneva, Division of Medical Informatics, 2002.

[13] V. Breton, "Health grid presentation", ISGC, 2003.

Fadi-Sotiris Salloum is a member of IEEE Computer and Communication Society with more than 6 year's experience in the development of data communication protocols and advanced collaboration applications utilizing VoIP, P2P and distributed technologies. He holds a M.Sc. in data communications from Brunel University and a B.Sc. in electronics from TEI of Athens. He was involved in several European projects in the field of data communications in several positions as a software engineer (IST SECOMS, PANORAMA), technical coordinator (IST GEMINI) and as project coordinator (IST EVOLUTE). He currently investigate bioinformatics and health applications with the use of open source grid solutions such us the globus toolkit and monitors the activities of HealthGrid organization.
e-mail: ssal@intracom.gr
INTRACOM SA
19.5 km Markopoulo Avenue
GR-19002 Peania, Greece