Bioprofiling over Grid for Early Detection of Dementia
(Invited Paper)

L. Sun[1], P. Hu[1], C. Goh[1], B. Hamadicharef[1], M. Hess[1], E. Ifeachor[1]
I. Barbounakis[2,3], M. Zervakis[2], N. Nurminen[4], A. Varri[4]

[1]School of Computing, Communications and Electronics
University of Plymouth, Plymouth PL4 8AA, U.K.
Email:{l.sun; pin.hu; c.goh; b.hamadicharef; m.hess; e.ifeachor}@plymouth.ac.uk

[2]Telecommunication System Institute, Technical University of Crete
University Campus, Kounoupidiana, Chania 73100, Greece
Email: ioannisb@ieee.org; michalis@danai.systems.tuc.gr

[3]Technological & Educational Institute of Crete, Chania, 73133, Greece

[4]Institute of Signal Processing, Tampere University of Technology
P.O. Box 553, FIN-33101, Finland
Email: reunamo@cs.tut.fi; alpo.varri@tut.fi

Abstract

The primary aim of this paper is to present a new concept, bioprofiling over Grid, and to illustrate how Grid computing may be used to support individualisation of healthcare in future, with the aid of a new test bed, the BIOPATTERN Grid. The BIOPATTERN Grid is designed to facilitate seamless sharing of geographically distributed bioprofile databases and to support the analysis of bioprofiles to combat major diseases such as brain diseases and cancer within a major EU project, BIOPATTERN (www.biopattern.org). The main objectives in this paper are 1) to report the development of the BIOPATTERN Grid for biopattern analysis and bioprofiling in support of individualisation of healthcare; 2) to illustrate how the BIOPATTERN Grid could be used for biopattern analysis and bioprofiling for early detection of dementia. We present the architecture and general functionalities of BIOPATTERN Grid, and the development of a prototype test bed (including a Grid Portal and Grid services for early detection of dementia). We illustrate the concept of bioprofiling over Grid and discuss issues such as scalability in high throughput computing for biodata analysis associated with bioprofiling for dementia. Results show benefits in both high throughput computing in biodata analysis and for individualisation of healthcare using Grid computing which makes it possible to access geographically distributed patient's information for subject-specific data analysis for early detection of dementia.

1. Introduction

Grid computing is aimed at the provision of a global Information Communication Technology (ICT) infrastructure that will enable flexible, seamless and secure sharing of geographically distributed resources (e.g. data, storage, computers, software, tools, applications, instruments and networks) at anytime and anywhere. In recent years, there is a growing interest in the application of grid computing to healthcare to support data-, computation- and/or knowledge-intensive tasks in areas such as diagnosis, prognosis, disease prediction and drug discovery. Often, this involves the acquisition, analysis and visualisation of biomedical data (medical informatics + bioinformatics). Examples of healthcare applications include distributed mammography data retrieval and processing (e.g. the MammoGrid [1] and the eDiaMoND [16] projects), and multi-centre neuroimaging analysis for brain diseases (e.g. the BIRN [11] and the MEGrid [13] within the BioGrid [10] projects). There is a trend in modern medicine towards individualisation of healthcare and, potentially, grid computing can also play a role in this by allowing sharing of resources and expertise to improve the quality of care.

INFOSCALE '06. Proceedings of the First International Conference on Scalable Information Systems, May 29-June 1 2006, Hong Kong

In this paper, we report efforts to exploit grid computing to support individualisation of healthcare to combat major diseases such as brain diseases within the BIOPATTERN project (www.biopattern.org).

BIOPATTERN is a major EU-funded, Network of Excellence (NoE) project. The Grand Vision of the project is to develop a pan-European, coherent and intelligent analysis of a citizen's bioprofile; to make the analysis of this bioprofile remotely accessible to patients and clinicians; and to exploit bioprofiles to combat major diseases such as cancer and brain diseases. A biopattern is the basic information (pattern) that provides clues about underlying clinical evidence for diagnosis and treatment of diseases. Typically, it is derived from specific data types, e.g. genomics information and vital biosignals, such as Electroencephalogram (EEG) and Magnetic Resonance Imaging (MRI). A bioprofile is a personal 'fingerprint' that fuses together a person's current and past medical history, biopatterns and prognosis. It combines data, analysis, and predictions of possible susceptibility to diseases. The idea of the Grand Vision of BIOPATTERN is to move away from 'local solutions to local problems' and towards 'European wide solutions to European problems'. The joint activities of the project include making information from distributed databases available in a secure way over the Internet, and providing on-line algorithms, libraries and processing facilities, e.g. for intelligent remote diagnosis and consultation. It is obvious that a grid-enabled network can facilitate the seamless sharing and pervasive access to distributed databases and for online bioprofiling analysis and diagnosis.

The main objectives in this paper are 1) to report the development of a new Grid test bed, the BIOPATTERN Grid, for biopattern analysis and bioprofiling in support of individualised healthcare. We introduce the general architecture, main functionalities and design targets of BIOPATTERN Grid; 2) to illustrate how the BIOPATTERN Grid could be used for biopattern analysis and bioprofiling for early detection of dementia in subjects on an individual basis. A goal of the BIOPATTERN Grid is to build a Grid-enabled test bed within the BIOPATTERN Consortium to demonstrate the concept of biopattern analysis and bioprofiling from ' birth to death', to combat major diseases such as brain diseases and cancer. Currently, the BIOPATTERN Grid connects five sites - the University of Plymouth (UoP), UK; the Telecommunication System Institute (TSI), Greece; Tampere University of Technology (TUT), Finland; the University of Pisa (UNIPI), Italy and Synapsis S.r.l (Synapsis), Italy. UoP, TSI and TUT are involved in the case study for bioprofiling over Grid for dementia which is presented in the paper.

The remainder of the paper is organized as follows: In Section 2, the basic infrastructure, main functionalities and design targets of BIOPATTERN Grid are described. In Section 3, a case study for Bioprofiling over Grid for early detection of dementia is presented, and the implementation of BIOPATTERN Grid prototype, Portal and Grid services and the demonstration of two applications – scalability in high-throughput computing and bioprofiling of Grid for detection of dementia are discussed. Section 4 concludes the paper.

2 Architecture of BIOPATTERN Grid

2.1 BIOPATTERN Grid Architecture

The architecture of the BIOPATTERN Grid is organized into five layers from top to bottom as illustrated in Figure 1.

The top layer is the client application layer which provides a friendly user interface to allow an end user (e.g. a clinician or a researcher) to access the Grid portal via a web browser. After user authentication (login/password), the end user can then make use of grid services for data/information query (e.g. query a patient's clinical information), update clinical information database (e.g. when a patient's new biorecord, such as the EEG, is ready, this can be uploaded to the patient's database which may be located locally or remotely), data analysis using selected algorithms (e.g. using Fractal Dimension (FD) [12] for analysis of the biorecord to support diagnosis of dementia (see later for more details)), viewing results (e.g. showing the FD index in canonograms) and/or visualization (e.g. rendering MRI image). The connection and data transfer between a client and a server is based on HTTPS protocol.

The next layer which is at the server side is the Web server layer. It contains web servers (e.g. main web server and backup server) and related components (e.g. databases and Java classes), which are used to authenticate end-users information and to establish connections to the lower layer (i.e. grid application layer). This layer is responsible for handling HTTP(S) requests from end-users, forwarding such requests to the lower layers, and retrieving as well as sending results back to end-users. Apache Tomcat [3] is used as the Web server in the BIOPATTERN Grid.

The Grid application layer is the third layer in the architecture. This layer plays two major roles. One role is to translate requests from the higher layers into detailed action requisitions based on pre-determined rules for different grid services (e.g. EEG analysis services) and then send those action requisitions to grid servers for execution. The action requisitions may include creating a job or transferring a file. The custom APIs (Application Program Interfaces), such as data analysis and visualization, are responsible mainly for providing the above mentioned functions. Another role of the Grid application layer is to provide efficient and effective management of grid resources and jobs based on the latest resource and job status which can be obtained from grid monitoring components in a secure manner. At this level, the authentication service is responsible for checking the grid users' credentials. The Information Service provides the function of translating the resource information obtained from the grid resource monitoring tools and passing this on for a variety of purposes (such as updating of resource information and recording of information for analysis). A broker offers the function of finding appropriate resources based on the job requirements and the resource information provided by the Information Services and of generating jobs together with job management services. File and data services are responsible for managing file/data transfers, etc.

The Grid Middleware Layer is responsible for dealing with those action requisitions that come from the grid application layer and for enabling such requisitions to become real actions. For the BIOPATTERN Grid, some advanced grid middleware technologies have been implemented in this layer, including Globus Toolkit 4 (GT4) [7], Condor [6] [18], Grid Monitoring Architecture [19] based distributed resource monitoring tools.

GT4 is the core component in this layer, and consists of a set of software components that implement Web Services (WS) mechanisms for building distributed systems. It has four main mechanisms: security, resource management, data management and information services. The Grid Security Infrastructure (GSI) enables secure authentication and communication over an open network. GridFTP offers high performance, secure, reliable data transfers, optimised for high-bandwidth wide-area networks. The WS-Grid Resource Allocation Management (WS-GRAM) provides functions to submit, monitor, and cancel jobs on grid resources. Other functions of job and resource management are left for Condor to provide. These functions include job queuing mechanism, scheduling policy, local priority scheme, local resource monitoring, and local resource management. WS-GRAM and Condor have different roles in job and resource management, the former focuses more on global scale management and the latter is mainly responsible for the management of local resources. Open Grid Service Architecture Data Access and Integration (OGSADAI) provides generic grid data services for access to and integration of data held in relational database management systems, as well as semi-structured data held in XML repositories. WS- Monitoring and Discovery System (WSMDS) provides services to monitor and discover resources and services on Grids.

The bottom layer is the grid resources layer. It consists of three main types of grid resources: computational resources (e.g. CPUs), data resources (e.g. relational databases) and knowledge resources (e.g. software implementations of scientific algorithms and/or visualization algorithms).

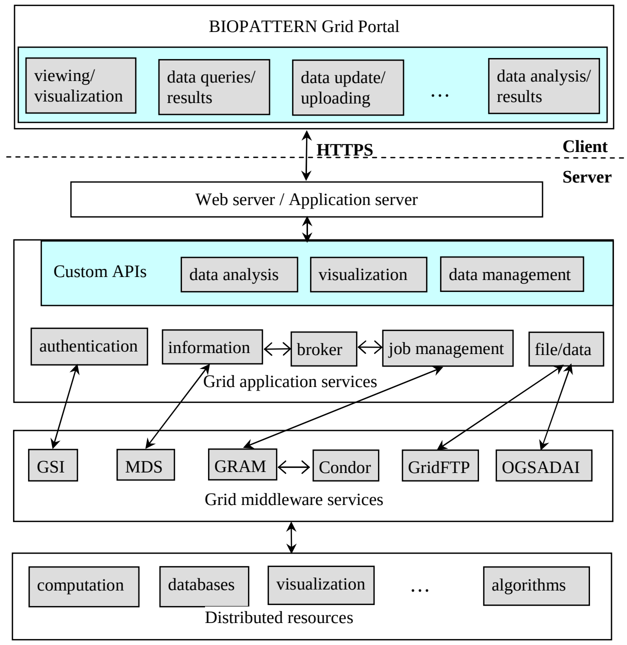

Figure 1. BIOPATTERN Grid Architecture

2.2 Functionalities and design targets of BIOPATTERN Grid

In the design of the BIOPATTERN Grid, we have considered the following objectives and functionalities:

1). To provide support for future 'Virtual Organizations' (VOs) or Groups for brain diseases and cancer.

The BIOPATTERN project integrates the research effort of 30 institutions across Europe to tackle and reduce fragmentation in the new field of biopattern and bioprofile analysis, which will underpin eHealthcare in the post genome era. It has several special interest areas (e.g. cancer, brain diseases and bioinformatics). In line with the ongoing work within the BIOPATTERN project, the BIOPATTERN Grid aims to support several virtual organizations (VOs) [8] in the areas of brain diseases and cancer within the Consortium. The clinical data and analysis algorithms are different for each such VO. For example, for brain diseases, the patient data may include physiological data, such as EEG and neuroimaging data such as MRI. For cancer, the patient data may include clinical information, ultrasound images, micro array analysis, etc. For many years, partners in BIOPATTERN Consortium have developed analysis algorithms for brain diseases and cancer. Such algorithms form part of distributed resources which can be shared via the BIOPATTERN Grid. Patient's data (e.g. EEG and MRI) are also located in different sites which can be shared across the BIOPATTERN Grid. The provision of different VOs or Groups (e.g. BIOPATTERN Cancer Grid, BIOPAT-

TERN Dementia Grid) should facilitate collaboration between partners with common interests by enabling data sharing, analysis and comparison of analysis methods and evaluation.

2). To create different user-tailored Grid portals.

Even with the same medical speciality, different users (e.g. researchers, clinicians and/or consultants) may have different requirements from the BIOPATTERN Grid. Several user-tailored Grid portals will be designed to meet the different requirements and requests. For example, if a researcher logs in, he/she can access a portal tailored to researchers in their interested domain. In this case, the focus will be on algorithms, analysis/comparison and cross-centre evaluation. If a clinician logs in, a clinician-oriented portal will be accessed. The focus in this case will be on disease detection, diagnosis and treatment. The authorised clinician can access/review patients clinical data and access analysis algorithms for supporting clinical decision-making.

3). To develop a variety of grid services to meet end users' need.

The BIOPATTERN Grid aims to provide a variety of grid services to meet the needs of end users. Typical grid services can be divided into the following three categories: data query/update; data analysis; data visualization. The data analysis services may include typical analysis algorithms/tools developed by BIOPATTERN Consortium to tackle specific problems (e.g. combined EEG and genomics analysis for dementia diagnosis, combined medical informatics and bioinformatics for prognosis and survival analysis for breast cancer). The algorithms may include non-linear dynamical analysis techniques; neural network and neural-fuzzy based algorithms. Due to the existence of different data formats and data structures, a data conversion service will also be provided to support the automatic conversion of data formats. These grid services will be provided via a user-friendly Grid portal.

4). To provide high performance computing (HPC) and high throughput computing (HTC) resources.

The BIOPATTERN Grid aims to provide both high performance computing and high throughput computing resources to meet the need for computation-intensive applications from the BIOPATTERN Consortium. The HPC will be based on distributed Grid computing resources and HPC clusters provided by the BIOPATTERN partners. This will facilitate some near realtime medical analysis services and applications for supporting near-real-time clinical decision making. The HTC will be based on distributed clusters of computers (e.g. based on Condor pools) to facilitate non-real-time computational requirements for computation-intensive applications. In the next section, we will illustrate one application based on HTC.

3 Case Study – Bioprofiling over Grid for Early Detection of Dementia

In this section, we present a case study on 'Bioprofiling over Grid' for early detection of dementia. We begin by providing some background on dementia and then move on to a discussion of the development of the BIOPATTERN Grid prototype, Portal and grid services. Using the prototype and the Portal, we will then highlight two applications of Grid computing to early detection of dementia: (i) the issue of scalability and high throughput computing and (ii) the concept of bioprofiling over grid.

3.1 Background on Dementia

Dementia is a degenerative cognitive disorder that affects mainly elderly people. On the whole, approximately 10% of those over 65 years of age and 50% of those over 85 years old will develop dementia [15]. There is, at present, no cure for this disorder, although several acetylcholinesterase inhibitors have been registered for the symptomatic relief of dementia of the Alzheimer's type (DAT). Nonetheless, unless sufferers are diagnosed in the early stages, they cannot reap the maximum benefit of the treatments [2].

There are currently several clinical modalities that can be used for the early detection of dementia. Amongst others, the Electroencephalogram (EEG), which is a non-invasive, real time depiction of electrical activities in the brain, offers the potential for providing an acceptable and affordable method in the routine screening of dementia in the early stages [12].

Using the current clinical criteria, the delay between the actual onset and clinical diagnosis of dementia is typically 3 to 5 years. It is therefore possible that poor treatment effects reported in some studies could be due to this delay in initiating treatment. This impeded advancement is mainly due to the fact that diagnosis has been concentrated on group comparisons, that is, attempting to separate individuals into groups (Normal, Alzheimer's, Parkinsons, etc.) based on their individual conditions. An alternative to this is individualized care through subject-specific analysis where a subject's "bioprofile" is obtained and used in the diagnosis. For each subject, we can compare his/her current and previous conditions to look for trends that arise over time (bioprofile) rather than comparing it to what is generally normal within the population. This is because the conditions of every individual are subjective and may change dramatically from one subject to another, or for one subject over time. Figure 2 shows changes in a hypothetical index from a subject's EEG. It can be seen that the index is initially normal and only becomes "abnormal" some time after the onset of the disease when it falls outside the normal spread. This illustrates why a subject-specific method that compares an

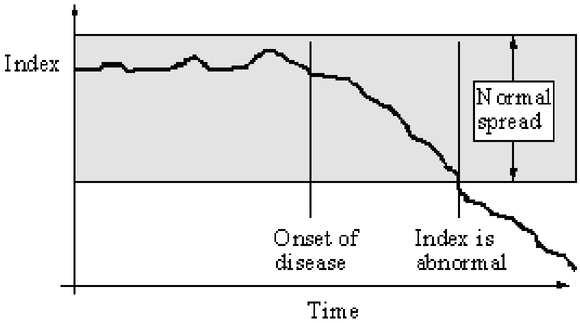

Figure 2. Illustration of Subject Specific Analysis

EEG to those taken previously from the same subject has the potential to provide earlier detection of a disease than a method that compares an EEG to what is generally normal within the population.

3.2 BIOPATTERN Grid Prototype and Portal for Detection of Dementia

In order to prove the concept of Bioprofiling over Grid for early detection of dementia, we have built up a BIOPATTERN Grid prototype as shown in Figure 3. At present, the prototype uses three partner sites - TUT, TSI and UoP. The bioprofile databases are distributed between the three sites. The pool of algorithms, which includes algorithms for computing, for example, the Fractal Dimension, zero-crossing intervals, mutual information, correlation dimension and lyapunov exponent from the EEG, is located at UoP. In addition, UoP provides four Grid nodes installed with Globus (GT4), a condor pool with 50 nodes. A web server is hosted in UoP to provide a Grid/Web Portal serving as an interface between an end user and grid services/resources. Besides the bioprofile database, the TSI site also provides four Grid nodes installed with Globus (GT4), a condor pool with 4 nodes and it has setup a grid-portal in collaboration with Myproxy grid service in an attempt to support the basic grid services through a user-friendly interface (browser).

In the prototype, all clinical data are stored in different types of databases which include patient information databases, algorithm databases, EEG databases and related meta-databases. The meta-databases store information that describes the EEG data (including sizes of EEG data file, locations, access permissions, owners, modification details) and algorithms (including such as algorithm names and programming languages). The OGSADAI-WSRF 2.0 is employed as an API in order to provide data services for access to and integration of data held in those databases distributed into three major sites (i.e. UoP, TSI and TUT).

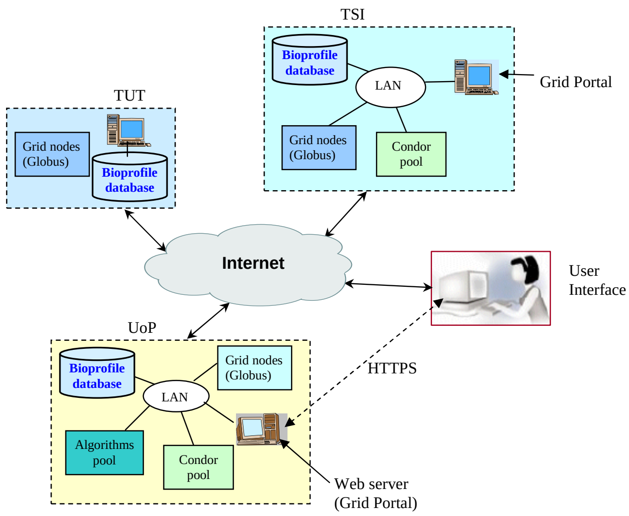

Figure 3. BIOPATTERN Grid Prototype for Early Detection of Dementia

The entry point for users to start using the prototype is through a Grid Portal. Currently, the prototype offers the following main services:

1). **Data Query** – for querying patient records, i.e. clinical information and EEG data files. Patient records are geographically distributed.

2). **Data Update/Upload** – for updating/uploading patient clinical information and EEG data files

3). **Data Analysis** – Perform some analysis on patients data using selected analysis algorithms, e.g. using Fractal Dimension (FD) to analyse EEG data. The analysis can be carried out based on High Performance Computing (HPC) or High Throughput Computing (HTC) resources.

Figure 4 shows a snapshot of choosing EEG analysis from the Portal. By keying in a specified patient's name (e.g. Mark Thompson) and/or his other personal information for query, the EEG analysis service can locate the patient's distributed EEG data files. Further by selecting the patient's one or all EEG files, the analysis service will start to analyse the EEG data and to show the relevant results. The detailed implementation of EEG analysis service will be given in the next section.

3.3 Implementation of BIOPATTERN Grid Services

In BIOPATTERN Grid protopyte, we have implemented the data query, data update/upload and data analysis services. To give an insight view of these grid services, we use

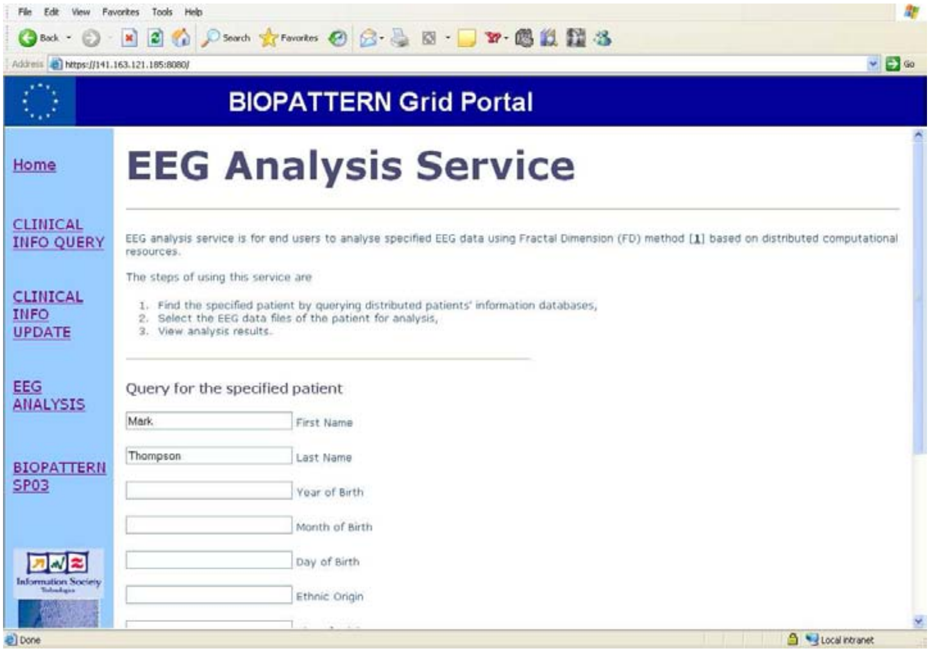

Figure 4. Biopattern Grid Portal – EEG Analysis

analysis service as an example to illustrate how it is implemented.

The analysis service (e.g. EEG analysis) is designed based on a scenario that an end-user (e.g. a clinician) can select EEG analysis algorithms from the BIOPATTERN Grid Portal to analyse a patient's EEG data for detection and diagnosis of dementia. To achieve this, the clinician has to go through several steps via the Portal, e.g. 1) query for a specified patient; 2) select EEG data files for analysis for the patient and submit analysis jobs; 3)choose to view analysis results in canonograms and/or bargraphs (see Figure 4).

The analysis service is implemented in four levels from the top customer GUI level to the bottom grid resource level as shown in Figure 5.

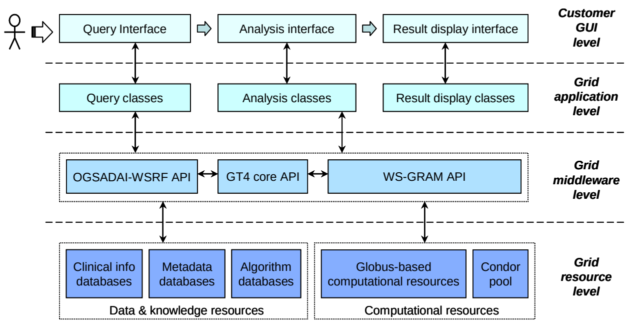

Figure 5. Implementation of Analysis Service

The customer GUI level provides user friendly interfaces to let end users seamlessly access grid-enabled services. The query, analysis and result display interfaces are designed to enable users to access the EEG analysis service without any grid knowledge or aware there is a grid network behind. The functionalities provided in this level mainly include handling HTTPS requests (e.g. obtainment of requests from end-users by filling e-forms) and forwarding such requests to the lower level. JSP and HTML are the languages used in the implementation of these interfaces. The codes are held by Tomcat containers.

In grid application level, three types of Java classes for query, analysis and result display are implemented. They are responsible for translating specified requests from the above level into detailed action requisitions and for further sending those requisitions to grid servers for execution. The query classes can provide concurrent access of distributed databases via OGSADAI-WSRF in order to obtain information as descriptions of specified patients, EEG data locations, analysis algorithm locations, etc. The analysis classes mainly offer generation of grid jobs (e.g. generation of Resource Specification Language files for job description), finding appropriate computational grid resources (e.g. looking for light-loaded resources based on job requirements and resource information), submission of jobs, retrieving analysis results and forwarding the results to the higher level for display. The result display classes enable end-users to view such analysis results in canonograms and/or bargraphs etc. The direct connection between result display classes and grid middleware level will be established in the future in order to support more complex functionalities such as 3D visualisation.

The grid middleware level contains certain APIs which mainly include OGSADAI-WSRF API, GT4 core API and WS-GRAM API. All these APIs are responsible for dealing with action requisitions from the higher level and for implementing such requisitions by accessing grid resources (e.g. algorithm databases and condor pool) in the lower level.

The grid resource level holds both data and computational resources, which are described as services held by Globus containers. Data and knowledge resources that support the EEG analysis service includes clinical information databases (e.g. databases for patient's personal information and EEG records), metadata databases (e.g. database for descriptions of EEG data and analysis algorithms) and algorithm databases (algorithms's executable code implemented in C/C++ or matlab). Computational resources cover both HTC and HPC resources. The Condor pool, currently, plays a key role in HTC-based EEG analysis. The Globus-based computational resources support the HPC-based EEG analysis at present.

3.4 Scalability Issues in High Throughput Computing

The Grid provides the resources required for high performance/throughput computing and analysis. It has been reported in numerous publications that in some cases algorithms which take several days to run on standalone PCs were completed in a matter of a few hours when jobs were distributed on the Grid [1] [4] [9].

To investigate scalability issues and to demonstrate how the computational resources of the Grid can be employed in the early detection of dementia, we created a hypothetical patient pool consisting of 400 subjects, each with 3 EEG recordings. For consistency reasons, the synthetic EEG data were created by adding random noise to 16 cases of original EEG recordings - 8 ADs and 8 normals. These data are hypothetical representation of recordings taken at 3 time instances akin to longitudinal studies carried out in reality. Each dataset consists of 21 channels of recording and is 1.3Mbytes. The recording duration is 4 mins and the sampling rate is 128 Hz. The data format is .EDF (European Data Format) [14]. The FD analysis algorithm is used to compute the FD index for each synthetic EEG file.

The scalability test was carried out on the UoP-based Condor pool, which has one job submit node/PC (Pentium 4 3GHz CPU, 1GB RAM) with Linux OS (kernel 2.6.8), and 50 execute nodes/PCs with a majority setup as Pentium 4 2.8GHz CPU, 512MB RAM and Windows XP SP2 OS. All nodes/PCs are connected to the UoP LAN, which provides 100 Mbps Ethernet connections.

We created a total number of 1200 tasks based on the 400 subjects (each has three EEG records), and generated a further four jobs (each with 100, 400, 800 and 1200 tasks, respectively) in order to investigate the performance of the Condor pool. We scaled the number of Condor nodes from 6, 9, 12 up to 50 PCs. Table 1 shows the observed execution time for running each specified number of tasks on one machine and at different scales of the Condor pool (nodes from 6 to 50). Figure 6 illustrates the speed-up comparison of running different jobs at different scales of the Condor pool.

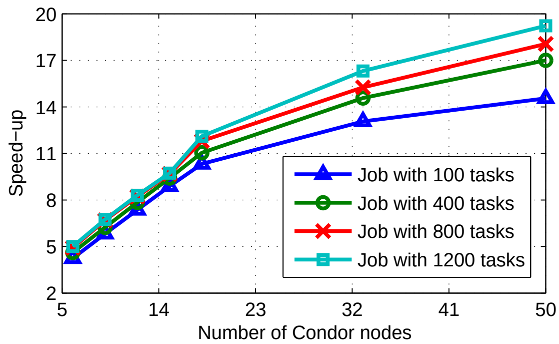

Figure 6. Speed-up comparison of running different jobs at different scales of Condor Pool

Preliminary results show that the Condor pool with 50 nodes can speed the execution time more than 19 times when compared with the time on a stand-alone PC. We also observed that running a job with more tasks is relatively more efficient than running a job with fewer tasks. This may be due to the time saving for the Condor to create and start a job and the time to transfer executables. This work demonstrats the benefit of EEG analysis using high throughput computing (e.g. by Condor Pool over Grid). However, we notice that the payoff is low when executing the EEG analysis on more than 18 PCs. This may be due to the following reasons:

- Waste of CPU time. The CPU time may be wasted due to the time required to create and start a job, to transfer input files, output files and executables to each execute node/PC, or due to some technical problems of Condor (e.g. preemption).

- Lack of checkpoint services. Condor does not provide checkpoint services on Windows machines to date, so that when a running task is interrupted by an unknown incident or other programs on a machine, it requires a restart of the task.

- Processing and networking load of the Condor pool. The running of a Condor job may be interfered by

Table 1. Performance evaluation results for scalability (with execution time in minutes and speed up factor compared with 1 PC)

No. of Tasks	100		400		800		1200	
No. of PCs in condor pool	exe. time	speedup	exe. time	speedup	exe. time	speedup	exe. time	speedup
1	64	-	255	-	511	-	765	-
6	15.1	4.2	55.3	4.6	103.8	4.9	152.9	5.0
9	11	5.8	40.9	6.2	76.3	6.7	113.5	6.7
12	8.7	7.4	32.5	7.9	62.2	8.2	92.1	8.3
15	7.2	8.9	27.1	9.4	52.8	9.7	78.6	9.7
18	6.2	10.3	23.1	11.0	43.3	11.8	63.2	12.1
33	4.9	13.1	17.5	14.6	33.5	15.3	46.9	16.3
50	4.4	14.5	15.0	17.0	28.3	18.1	39.8	19.2

other applications on a processor or under the local network.

3.5 Bioprofiling for early detection of dementia

An emerging issue in individualized care and early detection of dementia is the mobility of the individual. This is illustrated using the scenario shown in Figure 7, which illustrates the life of a fictitious individual called Mark in the hypothetical patient pool created earlier. Mark was born in France and lived there until he was 20 years old. He then went to Germany to study and work. When Mark was 40, he relocated to Italy to take up another job. He retired at 60 years old and went to the UK to live with his children. At the age of 62, Mark is showing early signs of Alzheimer's Disease (AD) (e.g. memory loss). To support medical diagnosis, his GP in the UK needs to look at his past and present medical information (Mark's bioprofile). However, as he has lived in several countries throughout his life, it is difficult and tedious to obtain all these geographically scattered information.

The Grid, with its secured and flexible infrastructure thus offers a way forward by providing a platform for seamless information retrieval, storage, exchange and analysis. For example, Mark goes to see the GP for a follow-up check at a memory clinic in Plymouth UK in December 2005. Using the information query services, the GP loads Mark's information (bioprofile and personal information) by simply keying in some of his particulars into the form provided. Upon submission, all of Mark's information (such as his previous medical records) will be returned to the GP in real-time without him having to worry about where the data is located or how it is being accessed. After a review of Mark's records, the GP performs another EEG recording. While this is taking place, the GP uses the visualization tools to see the changes in Mark's bioprofile - other clinical findings and FD indexes computed from his earlier recordings. When the recordings are completed, they are uploaded and analysed using the selected algorithm(s). Following this, a textual report detailing the results of the analysis can be generated.

From the current BIOPATTERN Grid Portal, the EEG analysis results can be displayed using canonograms or bargraphs for Mark's EEG records (distributed in TSI, TUT and UK, respectively, representing Mark's EEG records taken in his 40s, 50s, and 60s) as shown in Figures 8. Using the canonograms, clinicians can see changes in the brain electrical activities and evaluate how Mark's illness is progressing. In Figure 8, the canonograms (from left to right) show the FD value (or index) of the Mark's EEG taken at time instances of 1 (data located at TSI), 2 (data located at TUT) and 3 (data located at UoP) respectively. The FD value for the left canonogram indicates Mark in a normal condition with high brain activity, whereas the FD value for the right canonogram indicates Mark in a probable Alzheimer Disease with low brain activity. The middle one shows the stage in between. The changes (or trends) in the FD values provide some indication on the disease progression. This can help clinicians to detect dementia at an early stage, to monitor the progression of the disease and to monitor the response of the patient for a treatment [5].

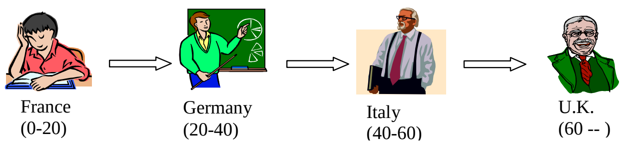

Figure 7. Mark's life itinerary

It is apparent from the results that the advantages that Grid computing brings to individualized healthcare goes beyond the benefits of pure high throughput and high performance computing. In addition to a flexible and practical

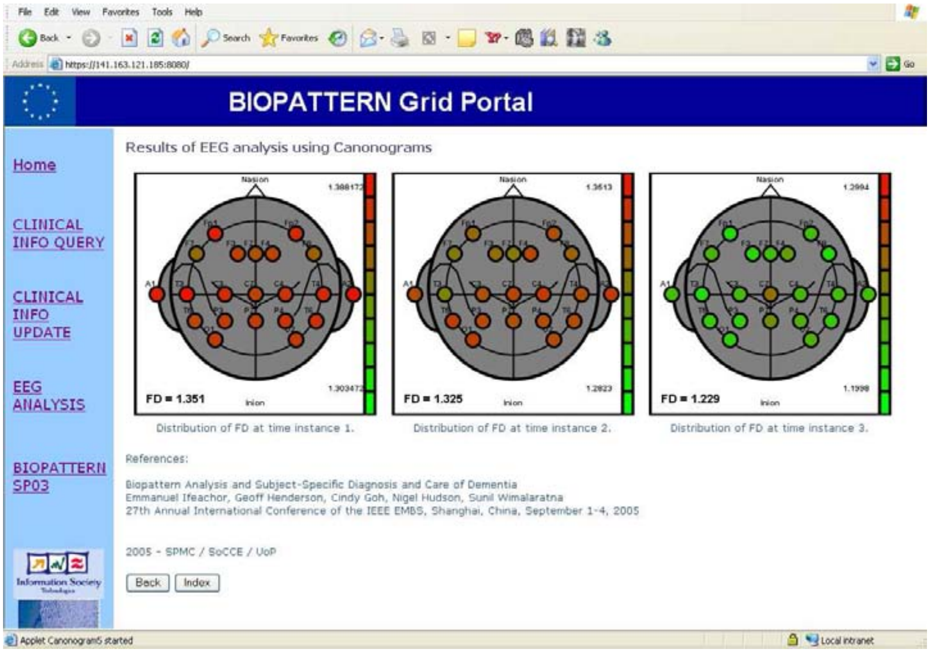

Figure 8. EEG Analysis Results for Patient 'Mark' in Canonograms

platform to alleviate the problem of access to geographically distributed data due to patient mobility, it also allows clinicians access to different algorithms for data analysis and comparison. In some cases the superior calculation resources also enable the use of more accurate analysis methods such as the application of the matching pursuit method [17] with a large wave dictionary for the EEG frequency analysis.

4 Conclusions and Future Work

In this paper, we have presented our work on the BIOPATTERN Grid which aims to build a Grid enabled network within the BIOPATTERN Consortium to facilitate the sharing of distributed bioprofile databases and analysis of bioprofiles. We discussed the architecture of the BIOPATTERN Grid, its general functionalities, main objectives and the development of the BIOPATTERN Grid prototype, Portal and grid services. We illustrated the use of the BIOPATTERN Grid in a case study on Bioprofiling over Grid for early detection of dementia. We demonstrated two key concepts in this application - scalability issues in high throughput computing and the concepts of bioprofiling. From the results presented, it is clear that grid computing not only offers the benefits of providing high throughput computing, potentially it can also offer an elegant solution to some of the emerging key problems involved in individualisation of healthcare, e.g. by alleviating the problem of geographically distributed medical data due to mobility of individuals.

The BIOPATTERN Grid is an ongoing project and the results presented here are limited in scale. In the near future, the BIOPATTERN Grid prototype will be extended to include more partners and more grid nodes. More clinical data, algorithms and computing resources are being added to the testbed, and more grid/web services to the Portal. Computational resources will also be expanded to a larger scale. For example, the current UoP Condor pool will soon be expanded to over 2000 PCs within the Campus Grid Programme. The test bed will also be connected to high performance computing clusters located in UNIPI. This will provide more powerful HTC and/or HPC resources for the BIOPATTERN Grid. Furthermore, the incorporation of powerful visualization tools will provide clinicians with a wider choice to interpret results for a more comprehensive and robust diagnosis of diseases for the benefits of individual patient.

Due to the nature of eHealthcare, the BIOPATTERN Grid will need to address several issues before it can move from academic/research prototype to clinical use, e.g.: (i) ethical and regulatory issues - to what extent can we move/share the data distributed in several sites?), (ii) privacy and security issues - how best should we address these, given that some users may not trust an on-line system?, (iii) QoS issues - can the system guarantee/provide adequate QoS to meet user expectations? and (iv) scalability issues - can the network/system cope when the data/sites scale up?. These are important issues that will need to be addressed before the BIOPATTERN Grid will become accepted in practice. However, they should not prevent us from looking into the future possibilities in ehealthcare with Grid computing.

Acknowledgment

The authors would like to thank Dr. C. Bigan from EUB, Romania for providing the original EEG data sets and Dr. G. Henderson for providing the algorithms for computing the FD index. We acknowledge the financial support of the European Commission (The BIOPATTERN Project, Contract No. 508803) for part of this work.

References

[1] S. R. Amendolia, F. Estrella, C. D. Frate, J. Galvez, W. Hassan, T. Hauer, D. Manset, R. McClatchey, M. Odeh, D. Rogulin, T. Solomonides, and R. Warren. Development of a Grid-based Medical Imaging Application. In *Proceedings of Healthgrid 2005, from Grid to Healthgrid*, pages 59–69, April 2005.

[2] R. Anand. Rivastigmine - Clinical efficacy and tolerability. *Clinician*, 16(5):14–22, 1998.

[3] Apache Tomcat Project. `http://tomcat.apache.org/`.

[4] J. Binns, M. Mccrory, M. E. Papka, J. C. Silverstein, and R. Stevens. Developing a Distributed Collaborative Radiological Visualization Application. In *Proceedings of Healthgrid 2005, from Grid to Healthgrid*, pages 70–79, April 2005.

[5] R. Bullock. New drugs for Alzheimer's disease and other dementias. *The British Journal of Psychiatry*, 180:135–139, 2002.

[6] Condor Project. `http://www.cs.wisc.edu/condor/`.

[7] I. Foster. Globus toolkit version 4: Software for service-oriented systems. In *IFIP International Conference on Network and Parallel Computing*, pages 2–13, l 2005. Springer-Verlag LNCS 3779.

[8] I. Foster, C. Kesselman, and S. Tuecke. The Anatomy of the Grid: Enabling Scalable Virtual Organizations. *International Journal of Supercomputer Applications*, 15(3):200–222, 2001. `http://www.globus.org/research/papers/anatomy.pdf`.

[9] J. Fritschy, L. Horesh, D. Holder, and R. Bayford. Applications of GRID in Clinical Neurophysiology and Electrical Impedance Tomography of Brain Function. In *Proceedings of Healthgrid 2005, from Grid to Healthgrid*, pages 138–145, April 2005.

[10] K. Fujikawa, W. Jin, S.-J. Park, T. Furuta, S. Takada, H. Arikawa, S. Date, and S. Shimojo. Applying a Grid Technology to Protein Structure Predictor "ROKKY". In *Proceedings of Healthgrid 2005, from Grid to Healthgrid*, pages 27–36, April 2005.

[11] J. S. Grethe, C. Baru, A. Gupta, M. James, B. Ludaescher, M. E. Martone, P. M. Papadopoulos, S. T. Peltier, A. Rajasekar, and S. Santini. Biomedical Informatics Research Network: Building a National Collaboratory to Hasten the Derivation of New Understanding and Treatment of Disease. In *Proceedings of Healthgrid 2005, from Grid to Healthgrid*, pages 100–109, April 2005.

[12] G. T. Henderson, E. C. Ifeachor, H. S. K. Wimalartna, E. Allen, and N. R. Hudson. Prospects for routine detection of dementia using the fractal dimension of the human electroencephalogram. *MEDSIP00*, pages 284–289, 2000.

[13] K. Ichikawa, S. Date, Y. Mizuno-Mastumoto, and S. Shimojo. A Grid-enabled System for Analysis of Brain Function. In *Proceedings of CCGrid 2003 (3rd IEEE/ACM International Symposium on Cluster Computing and the Grid)*, May 2003.

[14] B. Kemp, A. Värri, A. C.Rosa, K. D. Nielsen, and J. Gade. A simple format for exchange of digitized polygraphic recordings. *Electroencephalography and Clinical Neurophysiology*, 82(5):391–393, May 1992.

[15] D. Knopman, S. DeKosky, J. Cummings, H. Chui, J. Corey-Bloom, N. Relkin, G. Small, B. Miller, and J. Stevens. Practice parameter: diagnosis of dementia (an evidence-based review): report of the quality standards subcommittee of the american academy of neurology. *Neurology*, 56(9):1143–1153, 2001. `http://www.neurology.org/cgi/reprint/56/9/1143`.

[16] S. Lloyd, M. Jirotka, A. C. Simpson, R. P. Highnam, D. J. Gavaghan, D. Watson, and J. M. Brady. Digital mammography: a world without film? *Methods of Information in Medicine*, 44(2):168–169, 2005.

[17] S. Mallat and Z. Zhang. Matching pursuits with time-frequency dictionaries. *IEEE Transactions on Signal Processing*, 41(12):3397–3415, December 1993.

[18] D. Thain and M. Livny. Building reliable clients and servers. In I. Foster and C. Kesselman, editors, *The Grid: Blueprint for a New Computing Infrastructure*. Morgan Kaufmann, 2003.

[19] B. Tierney, R. Aydt, D. Gunter, W. Smith, V. Taylor, R. Wolski, and M. Swany. A grid monitoring architecture. Grid forum working group document, Grid Forum, February 2001. http://www.gridforum.org.