A Survey On Medical Image Retrieval Implementations In Data Grid

R. Kingsy Grace*  B. Priyavarshini**  J. Rajasri**  H. Rajeshwari**

*Assistant Professor, Sri Ramakrishna Engineering College, Coimbatore-22
**Final Year B.E. Students, Sri Ramakrishna Engineering College, Coimbatore-22

Abstract

In the last few years there has been a tremendous increase in the storage and processing of data but at the cost of speed and storage space. But still some problems in various fields that cannot be solved even by the usage of supercomputers exist. One such problem is medical image retrieval. Here, various existing techniques for medical image retrieval in data grids are surveyed and analysed. Discussions on relational databases which are used to store images of DICOM standards are included. The object definition methods using Extensible Markup Language Schema documents (XML Schema) and methods like Grid Resource Allocation Manager (GRAM) and Grid Security Infrastructure (GSI) which are used to achieve various non-functional requirements like performance, reliability, accuracy, interoperability, confidentiality and security are also discussed. Discussions on web portal designed using Php: Hypertext Pre-processor (PHP) and Java Script, through which user interaction with the grid architecture is established, are also added in this survey. The results of grid performance tests with 2 current operative nodes are also discussed.

KEY WORDS: Health grid, Digital Imaging and Communications in Medicine (DICOM), grid computing, health level 7 (HL7).

1. Introduction

Grid Computing is the interconnection of computers to form a network in order to achieve a common goal from different federations or sources. Grids are loosely coupled, geographically distributed and heterogeneous collection of systems. Grid computing uses middleware to communicate among processes and devices. Grid computing is a form of distributed computing where different devices across different regions can be integrated to form a network irrespective of the distance between them.

Here we use data grid architecture for medical image storage and retrieval. The main advantage of using data grid is that we can add features to support computation of data whereas in computational grid, implementing features to store data is hard. Data grid pulls data and other resources together from multiple administrative domains which are hospitals here that are spread geographically and then presents it to users upon requests and allows users to access, modify and transfer large volume of data (here images). This acts as a tool that enables researchers to investigate diseases and rare conditions, as well as provide doctors with the new ways to diagnose and treat patients. The data in data grid can be located at a single site or at multiple sites where each site is governed by its own administrators for security purposes. Specifically developed data grid middleware is what handles the integration between users and the data they request by controlling access while making it available as efficiently as possible.

Data grid under grid computing is one of the most increasingly used technologies in all fields including medical image retrieval. The process of storing and retrieving medical images for reference, integration and distribution across multiple sources is very tedious. One promising possibility is grid computing, a networking technology that provides shared access to the processing capabilities of distributed computing resources around the world. Several grid networking projects are already underway. Medical images such as X-Rays, results of MRI scans, CT scans etc need to be shared among the physicians who diagnose diseases and technicians who process these images. If there is no proper access mechanism to the system which stores such images, then there may be a bottleneck in the system caused due to uncoordinated access. Existing systems such as PACS (Picture Archiving and
Communications System) suffers from this problem. When people from different parts of the world try to access images from a common storage media, due to the different speeds with which they access, problems may be created. In order to avoid this, many new systems implementing DICOM (Digitul Imaging and Communications in Medicine) are being developed based on the Globus toolkit. Looking closer, it seems inevitable that high-performance computing will find its way into all areas of medicine. Today e-health and grid computing together are shaping the future of medical field and this proves that our project has good scope in mere future.

2. Medical Image Retrieval Systems in Grid

Grid computing is playing a key role in healthcare by making it more integrated into our lives. Using this technology, doctors can gain access to relevant medical images of various patients and patients can have a personalized form of information (images) with high security.

2.1 mantisGRID

mantisGRID [1] is an open source grid architecture which provides facilities to exchange medical images along with its information. It uses Open Grid Service Architecture – Data Access and Integration (OGSA-DAI) developed using Globus Toolkit and it complies with DICOM (Digital Imaging and Communications in Medicine) and HL7 standards for image accessing. Extensible Mark-up Language Schema (XML) is used to represent objects that are stored in Relational Databases. mySQL databases are used to store the images to maintain interoperability for the images while transfer. It is currently implemented in Colombian High Technology Academic Network. mantisGRID uses Tomcat for Web Services and OGSA Data Query Processing (OGSA-DQP) to support queries over resources stored in databases. For transfer of data from another node to local machine, File Thrower Application is used. This is available in all the nodes. Database can be updated with less effort by using uploading GRID portal. DQP-coordinator is used to evaluate the queries provided at various nodes and this acts as a main interaction point to all clients. This uses Query Evaluation Service and develops an XML document to structure the queries based on source ID. Priorities are set based on the node which submitted the query first. Once this is done, the plan is distributed to various evaluators (nodes) which handle a part of query or one whole query. This set up is established by making one node as server and all other nodes as clients. The data which being stored in database has a primary field: “patient UID (globally unique id). Patients records when accessed for research purposes, measures are taken for anonymization of the data. If data is accessed for clinical purposes, patient’s identity will be available. Data Transmission is made secure by using Secure Socket Layers and only authorized people can access the data from this architecture. Various tests were conducted to ensure the performance of this architecture. Checks to ensure interoperability and anonymization of data were made. Checks like Database engine test was also conducted (only information and not images). This proved that the mantisGRID performance is better than that achieved by the Distributed architecture. Here PHP is used to develop web service. So extra overhead is introduced to convert all the query results from Java into PHP. But this increases the performance if many records are accessed. Future works include extending this architecture to medical digital photography for dermatology and dentistry. Apart from HL7 and DICOM, various others standards are yet to be established.

2.2 GRIPLAB 1.0

GRIPLAB 1.0 [2] is a software layer which is used to provide image processing solution by using GLite middleware in the grid architecture. This layer has several image processing algorithms which are run on the middleware. This middleware handles system authentication, resource authorization and system security. GRIPLAB 1.0 gets user inputs through UI (User Interface) and has separate logical part called grid gateway to handle sending and receiving of images/data between nodes. User inputs are stored as text files and further used while image processing. The main advantage in using GRIPLAB 1.0 is that the user can specify the required filters and algorithms which are later applied to the images. Here once the user enters the input parameters like file name and algorithms, it is stored in extensible form and is dynamically executed in the GLite middleware. After executing the user input file, the image to be processed is split into different parts and is submitted for processing by the local machine (sender) to the grid architecture. This is received at the receiver side either same machine or another node and is processed/ reconstructed by considering the users’ inputs and also takes overlap parameter into account. User at the sender side is updated with the status of the job. Once image reconstruction/ processing is over, it is again joined into single image and is sent back from grid architecture to the GLite middleware which then directs the processed image...
Advantages in using GRIPLAB 1.0 include:

- Parallel processing of different parts of single image can be achieved thus resulting in decrease in execution time of processing algorithms.
- Flexible to the users.
- User needn’t have thorough knowledge about the architecture.

### 2.3 MedIGrid

MedIGrid [3] is an application tool to implement fast and easy medical image diagnosis especially used by nuclear doctors. It was developed to overcome the disadvantages like limited storage space and processing speed, usage of proprietary softwares resulting in problems to deal with images generated using various equipments which are experiences while using PACS (Picture Archiving and Communication System) which includes image storage and transfer process through web. MedIGrid is a grid architecture which is set among a group of hospitals where they can store, retrieve, visualize and reconstruct the digital images. GASS (Globus Access to Secondary Storage) is used to provide access to remote files whereas GSI (Grid Security Infrastructure) and GRAM (Globus Resource Allocation Manager) are used to establish security features and resource allocation respectively. Users can retrieve images from the storage and can view it using GUI (Graphical User Interface). If needed the image can be reconstructed using the high performance computing systems in the grid architecture. To interact with other systems in MedIGrid, user has to log-in. Once authentication through log-in is done, user can interact or access any other system in the MedIGrid. All the transfers from one system to another is through SSL (Secure Socket Layer). For reconstruction, the parameters denoting the reconstruction level that are specified in metadata files are used. These metadata files are sent with the raw image to the computing server. ImageJ which is an open source program with algorithms to read, display, analyse, process and print images in the user required format. Future works which can be expected include the replacement of GUI with the web portal through which image reconstruction requests can be provided, visualization of the reconstructed image or the status of the job being carried out.

Advantages of using MedIGrid include:

- Secured access to various digital medical images.
- Quick reconstruction of digital medical images.
- Open source and the user needn’t worry about interoperability and quality.
- Easy extensibility on adding new nodes.

### 2.4 Extended MedIGrid

Grid Enabled PSE is the extension of MedIGrid [4] which was designed for nuclear doctors to store, retrieve, visualize and reconstruct the images. This adds extra features like workflow, runtime discovery and allocation of resources and fault tolerance mechanisms to improve the experiences of the common users. Discovery mechanism is based on the parameters like expected OS and minimum memory requirements. In MedIGrid we used metadata files to set reconstruction parameters but here the doctor enters the parameters through GUI (Graphical User Interface). To access data or applications that are in various nodes, user has to log-in to the grid. For reconstruction, after logging in, user provides request through GUI. The request provided by the user is taken by the Application Manager (AM) which by using discovery and selection algorithms or through previous executions selects the target node to perform the user’s expected task. User is provided with the facility to know the status of the job and in case of any failure in the process of accessing or reconstruction, AM selects another resource to perform the job. Apart from mere image reconstruction, user can also upload and transfer images to/from other nodes with the help of GUI. Result of reconstruction is viewed at the user’s side through MRlcro (external application) or ImageJ (as in MedIGrid) and can also be stored in the local database. This also has the feature of performing local database search. Apart from recovery and monitoring, this provides migration feature in which the half run application in case of failure (machine), will be transferred to another machine which is free at that point of time resulting in high reliability and performance. The advantages in using this include user friendly GUI to select resource either to transfer, access or reconstruct. Also provides high performance due to the proper workflow management which in turn avoids load overheads. Future works are focused in implementing advanced visualization techniques and tools, more useful services and also to implement migration strategies in different cases like convenience and need rather than just failure.

### 2.5 MIFAS

There are two major problems in using the existing PACS (Picture Archiving and communication Systems) – one caused due to limited bandwidth and locations and the other due to the cost of maintenance. Limited bandwidth causes bottleneck
in the system when people from different locations with different speeds access the system. Hence a PACS system based on data grids which utilizes MIFAS (Medical Image File Accessing System) [5] has been proposed. Here the images are stored and retrieved from a co-allocation data grid.

A data grid is a system composed of multiple servers that work together to manage data and communications. MIFAS supports replicas of images for failover recovery. The proposed co-allocation model consists of 3 main components: An information service, a broker/co-allocator and local storage systems.

Applications specify the characteristics of desired data and pass attribute descriptions to a broker, which in turn searches for the available resources, gets replica locations from information services and replica management service to retrieve a list of physical locations. Out of the implemented co-allocation schemes, ARAM (Anticipative Recursively Adjusting Mechanism) is proved to be the most efficient. This scheme adjusts the workload on selected replica servers and handles unpredictable variations in network performances. It uses the finish rates of previously assigned transfers to anticipate the bandwidth status for the next selection adjusts workloads and reduces file transfer time in grid environments.

When users request medical images or DICOM Image Storage nodes, it is processed. If any storage node is available, it is obtained from the remote medical image storage via the GridFTP protocol. The cyber agent uses MIFAS co-allocation to download images in parallel. The cyber agent transformer accelerates data transfer rates and manages replicas over various sites too. It can be invoked with either the logical name of a data file or a list of replica source host names. When files are searched by logical file names, it searches the Replica Location Services to find all corresponding replicas and each source starts parallel transfer. The system ranks all replica servers according to the replica selection model and users can choose the best servers for parallel downloading.

The performance of ARAM is better than WebPACS and the average transfer time is also better. Co-allocation downloading is better than DICOM protocol. ARAM overcomes network faults too. MIFAS provides fast recovery from failure. Thus MIFAS has four major advantages:

- It reduces co-allocation transfer time
- It is a fast, secure and reliable system for obtaining medical images.
- It enables parallel download from co-allocation data grid.
- Management is easy, expenses are reduced and stability is increased.

2.6 Grid-DICOM

Grid-DICOM [6] is a protocol which enables transfer of DICOM images in grid. Existing systems have to convert DICOM into the GridFTP protocol, which makes us lose many of the advantages of DICOM and transfer files. Since many imaging devices have to be integrated, we go in for a standard format called DICOM to exchange medical files and related data. It provides interoperability and asynchronous communication. The major requirements for an imaging application are dynamic anonymization for patient IDs, transparent communication between existing systems in different locations and data encryption, integrity and other security measures.

In order to enable DICOM standard in grids, we need:

- **GSI (Grid Security Infrastructure)** that provides security features for the grid.
- **Grid-FTP protocol** which adapts GSI enables parallel data transfer and compresses data before sending it.
- **DICOM based on OGSA (Open Grid Service Architecture)** that uses web services to exchange objects via the grid.
- **DGIS (DICOM Grid Interface Service)** in which each device connects to a gateway to send and receive images from the grid. The gateway exports the images in DICOM format and compresses the files in a single lossless form. Its address is stored in the grid-wide database along with its metadata. Other gateways can also connect to this one and retrieve the files as and when needed.
- **Medical Data Manager (MDM)** which implements a gateway between DICOM and gLite Storage Resource Manager (SRM) using any protocol such as Grid-FTP.
- **Storage Resource Broker (SRB)** which stores retrieves or replicates data with its own protocol and also contains metadata.

In the proposed system, DICOM protocol has been modified for use in the grid as Grid-DICOM which supports GSI and can directly transfer images while retaining DICOM’s integrity and asynchronous communication. A new routing software which routes data between the existing and new protocol has also been developed for transfer to and from the grid. The Grid DICOM protocol uses TCP/IP as the base and GSI for security. The Grid-DICOM router is available at every entry point to the grid. It implements the DICOM protocol at the outbound interface and GSI on the inbound interface. The
router preserves the asynchronous communication ability of DICOM by listening to messages continuously on both interfaces. This implementation has been tested on MediGRID. It has been concluded that the throughput has increased and transfer rate has decreased. The Grid-DICOM router has almost the same transfer rate as the existing systems. It provides uninterrupted communication among the devices in the grid and preserves integrity using GSI. It also provides direct access to DICOM data, preserves asynchronous communication of DICOM and is firewall friendly.

2.7 MiddleWare

MiddleWare [7] is a new system that provides a virtual storage for medical images using middleware based on grid technologies. This integrates different systems and platforms and provides global searching, progressive image transmission and automatic encryption services. The MiddleWare developed is based on TCP/IP protocol and is used purely for transferring images. The middleware developed consists of the following layers:

- **Core MiddleWare Layer** which is the lowest layer where grid services are integrated. This communicates with the upper layers using XML thus providing interoperability, portability and integration of new devices.
- **Grid Services Server layer** which performs server tasks and communicates with the Core MiddleWare layer through interfaces.
- **Communication Layer** which enables communication between Core MW layer and Grid Services Server layer or between Grid Services Server layer and Components MW layer.
- **MiddleWare Components Layer** that implements services and components and provides a uniform interface to the different storage devices.
- **Application Layer** which interfaces with the applications that use MW.

The grid services deployed are:

- **Core MiddleWare Grid Services** which uses DICOM_storage to interact directly with devices and search for the files and help transfer it. This interacts with the Grid Services Server Layer.
- **Grid Services Server** which manages all services from Core Mw and Grid Services Server layer. It also supports all the services in the grid.
- **Storage Broker** service to manage all DICOM_storage information in the grid. It also optimizes creation and searching index of the components used.

One of the MiddleWare components used is DICOM image package which manages all images and interacts with all the other components too. The security system used in Grid MiddleWare is GSI (Grid Security Infrastructure) that provides security for information, secure communication and single sign on. Users entering the Grid are authenticated using certificates containing Distinguished Names. It is compared with the available Access Control Lists before granting access. The MiddleWare developed is very efficient for transferring images and increases the productivity of developers. It has an open architecture that can be easily understood by all and the components developed can be migrated to any language.

The comparisons of various medical image retrieval systems are presented in Table. 1

3. Conclusion

In this literature survey, medical image retrieval problem in data grid is discussed. Several algorithms available for this problem are studied in detail. The major focus of all these algorithms is efficient and fast retrieval from data grids. From this study, it can be concluded that a few changes in mantisGrid can go a long way towards improving the existing systems for medical image retrieval. In addition to advancements, new medical image standards and algorithms are also implemented.

4. References


Table 1. Comparison on Medical Image Retrieval Systems

<table>
<thead>
<tr>
<th>S. No.</th>
<th>Medical Image Retrieval Systems using Grid and Year</th>
<th>Middleware Used</th>
<th>Descriptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>mantaGRID 2011</td>
<td>Globus Toolkit</td>
<td>Open-source grid architecture to transfer medical images among various Universities and hospitals in Colombia. Uses XML and PHP based web portals. Efficiency is high.</td>
</tr>
<tr>
<td>2.</td>
<td>GRiLAB 1.0 2008</td>
<td>GLite</td>
<td>GRiLAB 1.0 is a software layer integrated to grid architecture through GLite middleware and is used for image processing. User friendly.</td>
</tr>
<tr>
<td>3.</td>
<td>MedIGrid 2002</td>
<td>Picture Archiving and Communication System (PACS), Globus Toolkit, ImageJ</td>
<td>This is used to store, retrieve, visualize, reconstruct and share medical images among various hospitals and uses metafiles. Highly secure but not easy to use.</td>
</tr>
<tr>
<td>5.</td>
<td>MIFAS (Medical Image File Accessing System) 2010</td>
<td>Globus Toolkit, Picture Archiving and Communication System (PACS), Cyber Agent Service</td>
<td>Implements PACS on co-allocation data grid and used to store and retrieve medical images from the database. Uses Anticipative Recursively Adjusting Mechanism (ARAM) to retrieve images.</td>
</tr>
<tr>
<td>6.</td>
<td>Middleware 2008</td>
<td>Globus Toolkit, Extensible Markup Language (XML)</td>
<td>Provides a virtual storage for medical images based on grid and used to transfer medical images across networks using the TCP/IP protocol. It has an open architecture and can be migrated to any language.</td>
</tr>
<tr>
<td>7.</td>
<td>GRID-DICOM 2005</td>
<td>Globus Toolkit, DICOM based on OGSA (Open Grid Service Architecture), DGIS (DICOM Grid Interface Service)</td>
<td>Implements DICOM in Globus based medical grids as GRID-DICOM instead of Grid-FTP so that none of the advantages of DICOM is lost. It integrates the existing DICOM imaging system with OSI based grids.</td>
</tr>
</tbody>
</table>