Data Grid for large-scale medical image archive and analysis

Conference Paper · January 2005

DOI: 10.1145/1101149.1101357 · Source: DBLP

Aifeng Zhang
University of Illinois at Chicago

Zheng Zhou
Baylor College of Medicine

Nelson King
No affiliation

Lawrence Wing-Chi Chan
The Hong Kong Polytechnic University

Data Grid for Large-Scale Medical Image Archive and Analysis

H.K. Huang, Aifeng Zhang, Brent Liu, Zheng Zhou, Jorge Documet, Nelson King, L.W.C. Chan

Image Processing & Informatics Laboratory, Departments of Radiology and Biomedical Engineering,
University of Southern California,
Suite 601, 4676 Admiralty Way, Marina del Rey, California 90292, USA

hkhuang@aol.com

ABSTRACT

Storage and retrieval technology for large-scale medical image systems has matured significantly during the past ten years but many implementations still lack cost-effective backup and recovery solutions. As an example, a PACS (Picture Archiving and Communication system) in a general medical center requires about 40 Terabytes of storage capacity for seven years. Despite many healthcare centers are relying on PACS for 24/7 clinical operation, current PACS lacks affordable fault-tolerance storage strategies for archive, backup, and disaster recovery. Existing solutions are difficult to administer, and often time consuming for effective recovery after a disaster. For this reason, PACS still encounters unexpected downtime for hours or days, which could cripple daily clinical service and research operations.

Grid Computing represents the latest and most exciting technology to evolve from the familiar realm of parallel, peer-to-peer, and client-server models that can address the problem of fault-tolerant storage for backup and recovery of medical images. We have researched and developed a novel Data Grid testbed involving several federated PAC systems based on grid computing architecture. By integrating grid architecture to the PACS DICOM (Digital Imaging and Communication in Medicine) environment, in addition to use its own storage device, a PACS also uses a federated Data Grid composing of several PAC systems for off-site backup archive. In case its own storage fails, the PACS can retrieve its image data from the Data Grid timely and seamlessly. The design reflects the Globus Toolkit 3.0 five-layer architecture of the grid computing: Fabric, Resource, Connectivity, Collective, and Application Layers. The testbed consists of three federated PAC systems, the Fault-Tolerant PACS archive server at the Image Processing and Informatics Laboratory, the clinical PACS at Saint John's Health Center, and the clinical PACS at the Healthcare Consultation Center II, USC Health Science Campus.

In the testbed, we also implement computational services in the Data Grid for image analysis and data mining. The federated PAC systems can use this resource by sharing image data and computational services available in the Data Grid for image analysis and data mining application.

In the paper, we first review PACS and its clinical operation, followed by the description of the Data Grid architecture in the testbed. Different scenarios of using the DICOM store and query/retrieve functions of the laboratory model to demonstrate the fault-tolerance features of the Data Grid are illustrated. The status of current clinical implementation of the Data Grid is reported. An example of using the digital hand atlas for bone age assessment of children is presented to describe the concept of computational services in the Data Grid.

MM'05, November 6-11, 2005, Singapore.

Categories and Subject Descriptors

[Multimedia Medical Image Retrieval Design]: multimedia archive and retrieval system, medical image, picture archiving and communication system, data grid, and computing grid.

General Terms: Design

Keywords: PACS, Grid Computing, data grid, fault-tolerance archive, image analysis, image data mining, bone age assessment of children, computational services

1. INTRODUCTION

We start our discussion with a clinical large-scale imaging system PACS (Picture Archiving and Communication system), its requirements for fault-tolerance archive, and clinical image recovery after disaster. These characteristics are the precursors of the development of the Data Grid.

1.1 Picture Archiving and Communication systems (PACS)

A PACS is a system integration of computers, servers, workstations, communication networks, and software to form a system for radiological image information archive, distribution, and display. It consists of the following components:

- A data acquisition gateway connected to the Radiology Information System (RIS) and the Hospital Information system (HIS) for patient and examination related data,
- Several image acquisitions gateways connected to various radiology modalities including film digitizer, CR (computed radiography), DR (digital radiology), DM (digital mammography), US (ultrasonic), CT (computed tomography), and MRI (magnetic resonance image),

- A PACS Controller and Archive Server including various storage devices, and
- Image display workstations (WS).

These components are integrated together by digital networks, communication protocols, and software shown in Figure 1. PACS is an integrated component in a total healthcare delivery system for daily 24/7 clinical diagnosis. PACS uses Digital Imaging and Communications in Medicine (DICOM) standard for data communication protocol and image data format.

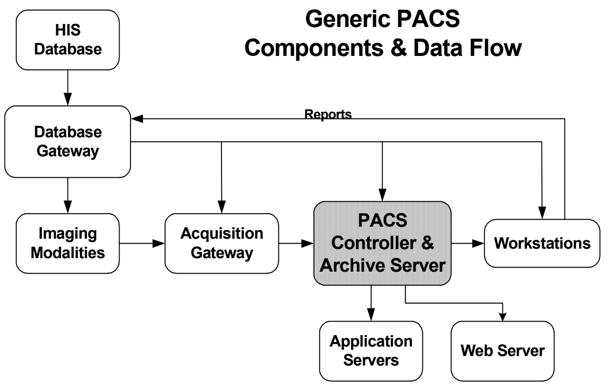

Figure 1. PACS and its basic components. This paper discusses the PACS Controller and Archive Server (shaded box) related to the Data Grid.

1.2 Clinical Image Recovery after Disaster

PACS, developed originally to smooth the operation of hospital radiology departments, has evolved into the most commonly found computing tool of a mission critical nature in the modern hospital. The component technologies of PACS have matured over the past several years, from the image acquisition devices to display workstations, archive servers, networks and, perhaps most significantly, the DICOM standard and IHE (Integrating Healthcare Enterprise), dataflow profiles. Among the components shown in Figure 1, the PACS controller and the archive server serve as the image data archive system consisting of arrays of hard disks, RAID, and DLT (Digital Linear Tape). This component archives all clinical images for seven years, as mandated by the recently adopted Health Insurance Portability and Accountability Act (HIPAA). [1] For an average 300–500 bed hospital in USA, this translates to about 40 terabytes of storage without image compression; but simply possessing the storage capability is not enough to assure compliance as no single image in this archive can be lost under any circumstances.

Fault Tolerance, long a necessity in most other applications of mission critical computing, is now expected in medical applications; however, practical field experience demonstrates that PACS archive servers can and do go down without warning. [21, 22] Once down, the malfunction cripples the clinical operation and dramatically affects the quality of healthcare services. Currently, most PAC systems use a backup archive solution for image data recovery. This not-so-fool-proof solution is entirely inadequate given its expense, in dollars and man-hours, as well as its slowness in recovering the image data, which has a negative impact on healthcare delivery. [2-4] New approaches to this problem are greatly desired.

2. GRID TECHNOLOGY

2.1 Grid Technology and the Globus Toolkit 3.0

Grid computing is the integrated use of geographically distributed computers, networks, and storage systems to create a virtual computing system environment for solving large-scale, data-intensive problems in science, engineering, and commerce. [5-11] A grid is a high-performance hardware and software infrastructure providing scalable, dependable and secure access to the distributed resources. Unlike distributed computing and cluster computing, the individual resources in grid computing maintain administrative autonomy and are allowed system heterogeneity; this aspect of grid computing guarantees scalability and vigor. Therefore, the grid's resources must adhere to agreed-upon standards to remain open and scalable. A formal taxonomy, composed of five layers (as shown in Fig. 2) has been defined by the Globus Toolkit 3.0 of grid computing to assure this standardization. [7]

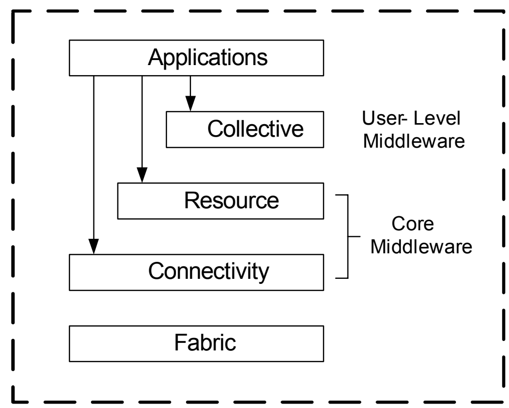

Figure 2. The five layers of the grid computing technology based on Globus Toolkit 3.0

At its core, grid computing is based on an open set of standards and protocols - e.g., the Open Grid Services Architecture (OGSA) [9, 10].

In this paper, we address the computational services and the data services of the Globus Toolkit 3.0. [6-8]

a. *Computational Services* support specific applications on distributed computational resources, such as supercomputers. A grid for this purpose is often called a Computational Grid (see Section 4).

b. *Data Services* allow the sharing and management of distributed datasets. A grid for this purpose is often called a Data Grid. (See Sections 3 and 4)

We use the Globus 3.0 toolkit co-developed by ANL and ISI, USC as the guide for implementing the Data Grid architecture [18]. (Globus 4.0 is now available for Alpha test)

2.2 Use of SAN Technology in PACS and Data Grid

A current data storage trend in large-scale archiving is Storage Area Network (SAN) technology, and PACS is no exception in this trend [12]. With this new configuration, the PACS server will still have a short-term storage solution in local disks of the display workstations containing unread patient studies. However, for long-term storage, the PACS data is stored in a SAN. This SAN is a stand-alone data storage repository with a single Internet Protocol (IP) address. File management and data backup can be achieved with a combination of digital media (e.g., RAID and DLT.) smoothly and with total transparency to the user. In addition, the SAN can be partitioned into several different repositories each storing different data file types. The storage manager within the SAN is configured to recognize and distribute the different clients' data files and store them to distinct and separate parts of the SAN. Figure 3 shows an example of a SAN configuration where three different clients, PACS server, IT server and Data Grid, store their data to the SAN. They each access the same IP address and make a request for data storage or retrieval. The SAN is pre-configured such that the storage manager knows which client is requesting storage service and where the data should be stored within the SAN. In the case where a client requests data to be retrieved from the SAN, the storage manager again knows which client is requesting data and where that data needs to be retrieved from within the SAN. In addition, the SAN can be partitioned in a heterogeneous fashion where the different volume sizes can be scaled to either the data file sizes of the different clients or the total volume of data needed for storage by a particular client as compared to another.

In the Data Grid system to be described in Section 3, both clinical PAC systems use one partition in its respective SAN for its own long-term archive. The Data Grid uses the second partition in each respective PACS' SAN for the back up image data of other federated PACS. The resource allocation method based on the Globus 3.0 toolkit is discussed in Section 3.

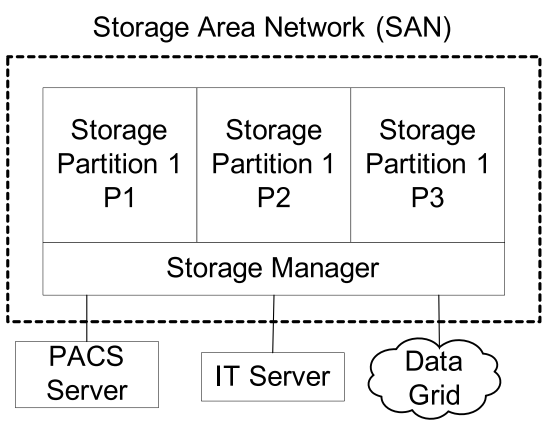

Figure 3. SAN Technology integrated with a clinical PACS. The SAN can be partitioned for orthogonal types of data storage such as an Information Technology server (e.g., email) or used as a resource for the proposed Data Grid. In this example, the data from the PACS server is sent to the storage manager within the SAN and then configured to be stored in partition 1 (P1). Likewise, data from an IT server is stored in partition 2 (P2). Finally data from the proposed Grid would be stored in partition 3 (P3). Each partition is separate and distinct.

3. DATA GRID ARCHITECTURE

3.1 Data Grid for PACS Archive Application

The Data Grid testbed consists of three sites. The first site is the IPI (Image Processing and Informatics Laboratory) where the major resources are the PACS Simulator and the DICOM Fault-Tolerant backup archive. Both components are the resources in the Data Grid. The second and the third sites are the Saint John's Health Center (SJHC) and the Healthcare Consultation Center II (HCC II) at USC/HSC. Both sites have a clinical PACS with a SAN archive system. A partition of each SAN, which does not handle the site's clinical PACS image data, is used as backup archive resources in the Data Grid. It is important to note that the SAN partitions belonging to each of the two sites are completely independent and the data stored in these partitions are orthogonal and separate from the clinical data partitions that are integrated with each of the respective clinical PACS. From these available resources in this testbed, we have developed the five layers based on the Globus Toolkit 3.0 (GT) and some PACS DICOM resources shown in Figure 4.

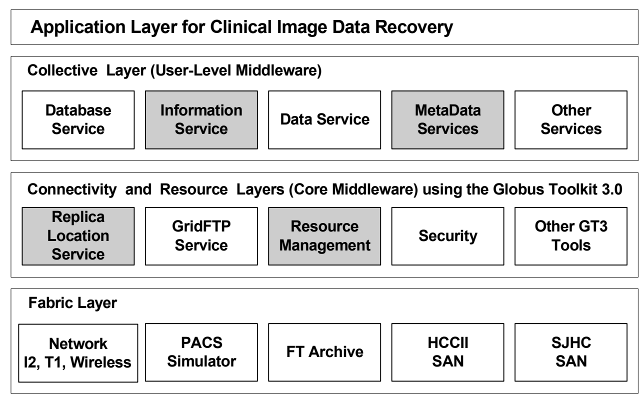

Figure 4. Five-layer architecture and the contents of the Data Grid in the testbed. For shaded boxes, see Section 4.

1. **Fabric Layer**: This layer consists of a DICOM compliant fault-tolerant (FT) backup Archive Server and a PACS simulator at IPI; two SANs (Storage Area Network) at two PACS clinical sites – HCCII, USC (Healthcare Consultation Center 2, University of Southern California Health Science Campus), and SJHC (Saint John's Health Center, Santa Monica, CA); and communication network systems including LAN (local area network), Internet 2 and broadband WAN (wide area network - T1).

2. & 3. **Connectivity and Resource Layers (Core Middleware) using the Globus Toolkit 3.0** [13-17]. This layer consists of a set of services in the Globus Toolkit 3.0 (GT3) developed at the Argonne National Laboratory, and ISI (Information Science Institute), USC. GT3 is an open and free toolkit based on Open Grid Service Architecture (OGSA) [11, 18, 19] mechanisms, which has the same five layer grid architecture. GT3 provides a set of services, such as Grid Security (GSI), remote job submission and control using the Grid Resource Allocation and Management (GRAM), high-performance secure data transfer (GridFTP), Replica Location Service (RLS)

and other core tools for building the Core middleware layer. GT3 is common amongst a large user community in Grid research and applications.

4. **Collective Layer (User-Level Middleware)** This layer consists of services to interact between the User Applications and the services in the Core Middleware, such as database service (to find the best available database in the Data Grid), information service (to monitor the current active services in the Data Grid), and data service (to find the physical address of the logical data) as well as other services. In this layer, GT3 has only the Information Service (Fig. 4 shaded), all others, such as Database, Data, Metadata, and Other Services are currently not available (Fig. 4). We are developing these services in conjunction with our image data recovery application.

5. **Data Grid Application Layer:** This layer consists of several applications, such as the recovery of clinical PACS image data.

3.2 A Three-site Testbed and Its Major Image Data Storage Resources

Figure 5 shows the configuration of the testbed Data Grid including the three sites mentioned above. A clinical PACS workstation outside of the Data Grid is able to access the grid for services.

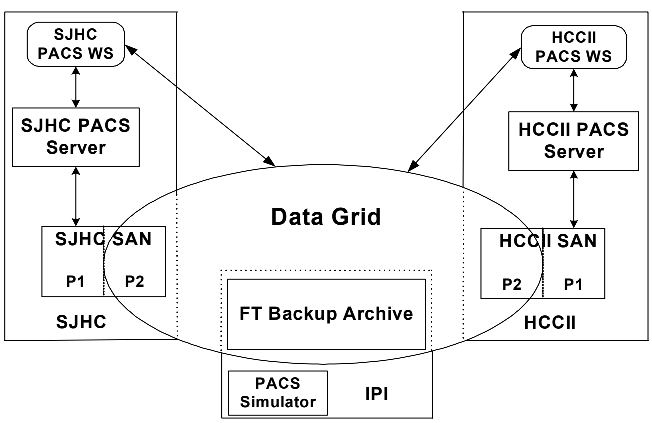

Figure 5. Configuration of three sites data storage systems that comprises of image data storage resources in the Data Grid. Workstations outside of the Data Grid can still access the grid for particular services. In this configuration, P1 at each site is used for clinical data pertinent to it PACS, whereas P2 becomes a shared resource of the Data Grid.
SJHC: Saint Johns Healthcare Center
HCCII: Healthcare Consultation Center II, University of Southern California (USC)
IPI: Image Processing and Informatics Laboratory, USC

The SAN architecture of the two sites is heterogeneous which provides a good basis for a more robust testbed development and evaluation. In both cases, the SAN is utilized as long-term storage of clinical PACS data. Each of the PACS servers has short-term storage, or, it can be considered as an unread buffer of clinical PACS studies. However, with each of the long-term storage solutions, there are major differences within the SAN architecture. Figure 6 shows the major differences between the two SAN architectures that are used as part of the Data Grid resources. In the case of the SJHC SAN (Architecture A), the RAID is only a temporary holding place for the data before it is transferred to the main storage area, which is the digital tape library holding 13.5 TB. Therefore, the RAID is only a small amount (270 GB) of data storage volume. In the case of HCCII's SAN the RAID is used as the primary storage area and therefore data volumes are larger (7.2 TB). In addition, the RAID capacity is scalable up to 52 TB. The digital tape library (14.4 TB), in this case, is only used as disaster recovery/backup to the RAID. Therefore, the tape read/write throughput is much slower as compared to the read/write throughput of SJHC's digital tape library within its SAN. Table 1 lists the backup storage resources, which are allocated from each PACS and Archive Server in the Data Grid.

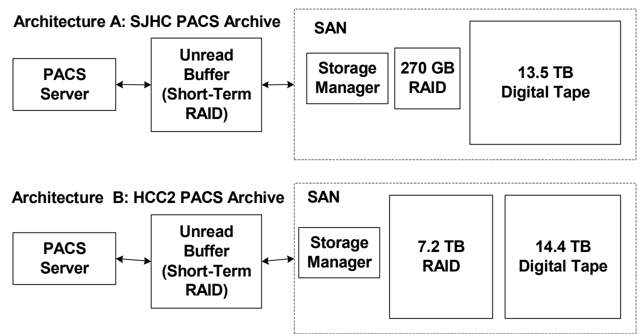

Figure 6. Two different SAN architectures that are utilized within the proposed Data Grid.

Table 1. Backup Storage Resources in the Data Grid Testbed

Resource	Location	Types (Capacity)
IPI FT Backup Archive	IPI	RAID (300 GB), DLT (1.6 TB)
SAN (P2)	SJHC	RAID & DLT (1.3 TB)
SAN (P2)	HCCII	RAID & DLT (2 TB)

3.3 The Backup Image Data Storage in the Data Grid [20-23]

Clinical images generated in the three sites, IPI, SJHC, and HCCII can be backed up by the Data Grid using the following protocol shown in Table 2. There are always two backup copies of the image data acquired from any site in the Data Grid. For example, image data acquired from SJHC PACS will have two backup copies in the Data Grid. One is stored in HCCII SAN (P2), and the second in the IPI FT backup archive. Similarly, image data acquired from HCCII PACS will be backed up in the SJHC SAN (P2) and IPI FT backup archive [20], and data from

IPI PACS Simulator will be backed up in the SJHC SAN (P2) and HCCII SAN (P2). The IPI PACS Simulator has a connection to the clinical PACS at USC for clinical image data. A database in the Data Grid based on the PACS data model is used to track every patient comes in contact with the Data Grid. Figure 7 illustrates an example of the backup procedure.

Table 2. Backup Policy in the Data Grid

Site	Clinical Image Data	Backup Copy 1	Backup Copy 2
SJHC	SJHC SAN (P1)	HCCII SAN (P2)	IPI FT Backup Archive
HCCII	HCCII SAN (P1)	SJHC SAN (P2)	IPI FT Backup Archive
IPI PACS Simulator	PACS Simulator Archive	SJHC SAN (P2)	HCCII SAN (P2)

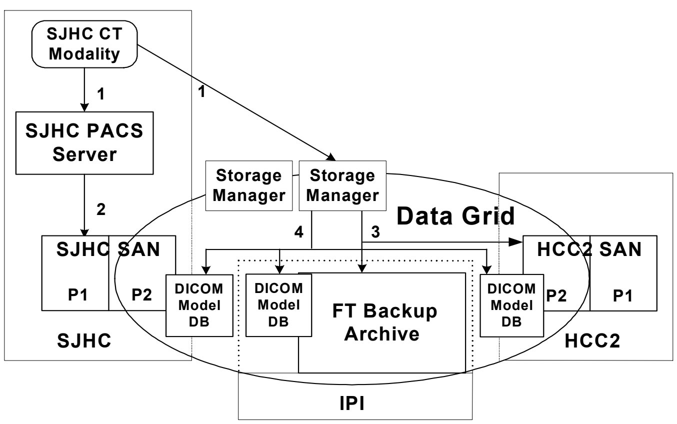

Figure 7. The backup procedure of image data from SJHC CT Modality to two storage sites (IPI FT Backup Archive and HCCII SAN (P2) in the Data Grid).

Follow the numerals described in Figure 7:

1. After an examination is completed at the SJHC CT scanner, it sends out two copies of the image data, one to the SJHC PACS Server, and the second to the Data Grid for backup.
2. The SJHC PACS Server receives the first copy of the image data and stores it in the SJHC SAN (P1) as its own archive.
3. The Data Grid activates one service Storage Manager to receive the second copy of the image. The Data Grid monitors the status of Storage Manager service and can activate a second one if the first Storage Manager fails. After receiving the image data, the Storage Manager automatically sends two copies, one to the IPI FT Backup Archive and the second to the HCCII SAN (P2). Since two copies are distributed to two different storage sites by the Data Grid, a single-point-of failure can be avoided. After the image data has been successfully archived in the two storage sites, the physical location will be added to the Replica Location Table of the Data Grid. The Replica Location Table keeps the physical storage information of each copy of image data.
4. The Storage Manager adds the patient information of the image data (not the image data itself) to a DICOM Data Model database in the Data Grid which makes three copies distributed to the FT backup Server at IPI, the SJHC SAN (P2), and the HCC SAN (P2). Three copies are used to avoid the single-point-of-failure of the database.

3.4 Interfacing the PACS and Data Grid

The utilization of Data Grid for PACS applications are mainly in DICOM Image Store and Query/Retrieve (Q/R), and Computational Services. In this section we use two examples of DICOM image Store, and DICOM image Q/R to demonstrate the fault tolerance features of the Data Grid based on three sites (see Figure 5 and Table 1). Section 4 describes the computational service. Figure 8 shows a DICOM WS using the Data Grid for image store. Under normal operation condition (solid lines), the image is sent from the DICOM client to the Data Grid through the Grid Access point 1. Two copies of the image are stored in two Grid Storage Resource 1, and 2 respectively. Suppose the Grid Access Point 1 fails (cross lines), as shown in Figure 9, the Data Grid has the intelligence to find Grid Access Point 2 (dotted line) from there it finds (dotted lines) Grid Storage Resource 1, and 2 for archive.

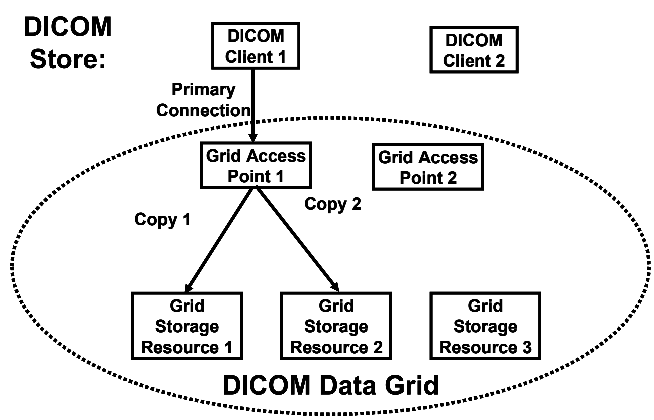

Figure 8. Normal DICOM Store (solid Lines).
The three Storage Resources are those shown in Figure 5 and Table 2.

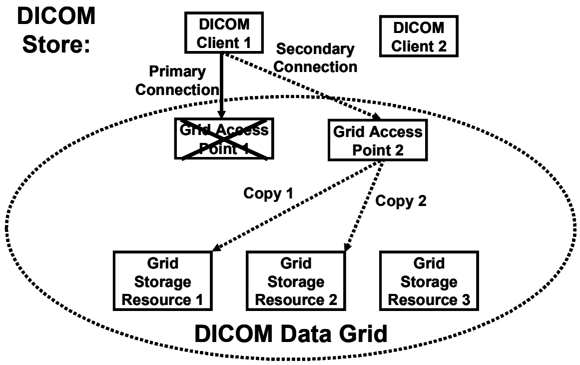

Figure 9. If Grid Access Point 1 fails, the Data Grid automatically switches to Grid Access Point 2 to complete the DICOM Store process (dotted lines)

Figures 10 and 11 show the second example which is for DICOM Q/R. Under normal operation condition (solid lines), the Q/R processes go through Grid Access Point 1 to find the required information at Grid Storage Resource 1. The information is retrieved and sent to the DICOM Client. Suppose the Grid Storage Resource 1 fails during the Q/R (cross lines), the Grid Access Point 1 finds the Grid Storage Resource 2 which contains the required information, Resource 2 then fetches the information and sends to the DICOM Client 1 (dotted lines).

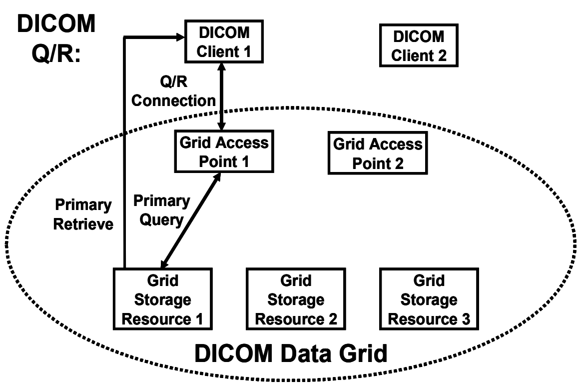

Figure 10. Normal DICOM Query/Retrieve (Q/R) (solid lines)

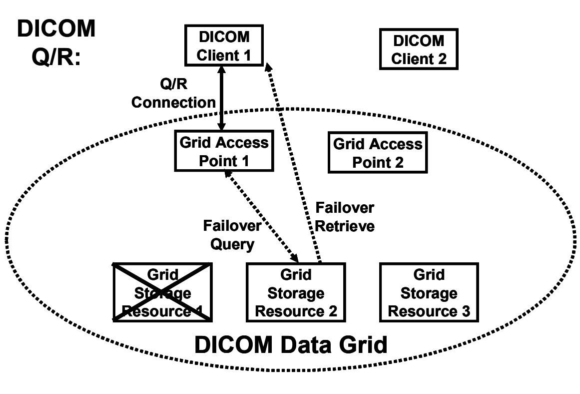

Figure 11. If Grid Storage Resource 1 fails, Data Grid automatically switches to Resource 2 and completes the Q/R process. (dotted lines)

3.5 EXTENSION OF THE DATA GRID SERVICE TO MORE PAC SYSTEMS

This paper discusses the data grid model which incorporates three PACS sites. When the data grid grows with more than three heterogeneous PAC systems, the Data Grid concept can be extended without much difficulty because only storage resource (see Fig. 5, 7 and 8) which is not used by each of the PAC systems is involved.

4. COMPUTATIONAL SERVICES IN THE DATA GRID FOR IMAGE ANALYSIS

A computational service infrastructure in the Data Grid provides dependable, consistent, pervasive, and inexpensive access to computational capabilities. Globus toolkit [13] described earlier includes core and user-level middleware services (refer to Fig. 4 shaded boxes) that enable users to obtain information about the services available, component software, data files, and the execution environment. The Grid execution environment includes computing and storage services with diverse capabilities. [24] We develop four computational services using the Globus toolkit in the Data Grid infrastructure. We first describe the basic infrastructure and then in Section 4.2, we apply this infrastructure to bone age assessment. [25]

4.1 Computational Services Architecture in the Data Grid

In Grid environment, an application component can be implemented in different source files; each complied to run in a different type of target architecture. Exact replicas of the executable file can be stored in many locations, which helps reduce execution time. Data files can also be replicated in various locations. Each file has a description of its contents in terms of application-specific metadata. The Metadata Service including Catalog Service, see Figure 4, responds to queries based on application-specific metadata and returns the logical names of files containing the required data, if they already exist. Given a logical file name that uniquely identifies a file without specifying a location, the Replica Location Service (RLS), see Figure 4, can be used to find physical location for the file on the Grid.

A specific application may require a certain type of resource for execution. Figure 12 shows the operation architecture of the computational services. First, the client requests resources from MDS (Monitoring and Discovery System) server (Fig. 12, Numeral 1), which manages the resources and distributes the jobs to the computational services. The index service finds resources appropriate to the requirements of application components and notifies the client to send the application to the Grid Resource Allocation and Management (GRAM) service (Fig. 12, Numeral 2). The GRAM service acknowledges MDS after it receives the application. (Fig. 12, Numeral 3). After the GRAM service receives the application, jobs that completely specified for execution are sent to schedulers that manage the resources and monitor execution progress. Execute acknowledges MDS server the completion of the job (Fig. 12, Numeral 3) which in turn notifies the client (Fig. 12, Numeral 4).

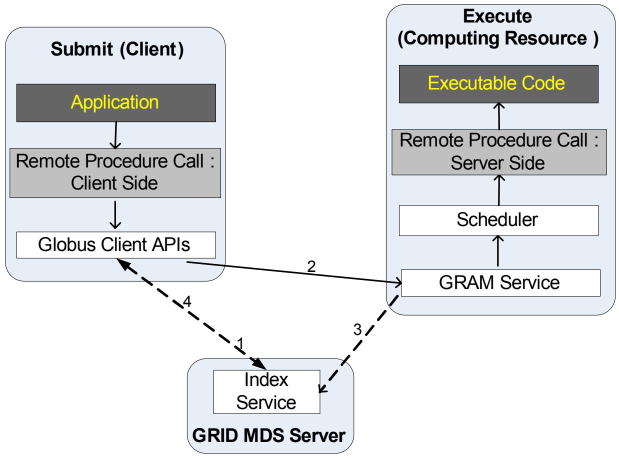

Figure 12. Operation architecture of a computational grid. Numerals represent the workflow.

4.2 Computerized Bone Age Assessment

We use the bone age assessment (BAA) of children as an example to illustrate how the computational services in the Data Grid can be used in BAA application. Bone age assessment is a procedure performed in pediatric patients to evaluate parameters of maturation and growth of the child from a left hand and wrist radiograph (Figure 13).

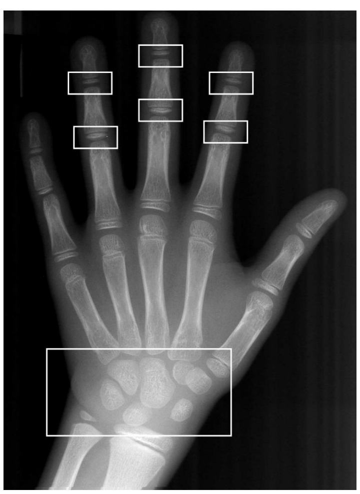

Figure.13. A left hand and wrist radiograph with seven regions of interest (ROI). Top: Six epi-metaphyseal ROIs; Bottom: Carpal bones ROI

Recent works in the field of computerized approach to BAA have already been completed from our previous work. [26] We have acquired a total of 1,080 digitized hand images of normally developed children, 5 images for each group of children (0 - 9 years old) and 10 images for children (10 -18 years old) from the Childrens Hospital Los Angeles, evenly distributed from boys and girls of European, African, Hispanic and Asian descent. The images in our collection have been subjected to fully automatic procedure of image processing yielding a vector of features for each successfully processed region of interest (ROI). There are seven ROIs (Fig. 13) which provide evidences of bone growth.

The overall image processing procedure is broken down into 3 software layers shown in Figure 14: pre-processing, region of interest analysis, and high-end application. The hand image processing and analysis starts with background removal based on a histogram analysis. Then the hand object is identified with the phalangeal axes along digits II, III and IV. The six epi-metaphyseal joints and carpal bones ROIs are segmented out as separate subimages, shown in Figure 13, for further analysis.

An overview of medically accepted diagnostic method indicates that epi-metaphyseal ROIs appear to be the most sensitive area reflecting the skeletal development stage. At the early stage of skeletal development, the epiphysis is separated from the metaphysis. 11 distance-related features, which form the feature vector of a ROI of a phalange, are extracted using the Gibbs random fields. In later stage (starting from 9 to 12 years old), the gap between epiphysis and metaphysis becomes radiographically inapparent. The ROI is subject to wavelet decomposition in order to assess the stage of fusion. Carpal bones part is another important ROI in bone age assessment. The inclusion of number, size, and shape features of carpal bones can be used to augment epi-metaphyseal ROIs to obtain higher accuracy of bone age assessment. The separation of carpal bones is most advantage to be used for bone age assessment for children of ages 0-5 (female) or 0-7(male), and the amount of bone overlapping can be used as an indication of ages 5-12 (female) or 7-12 (male).

The features from the six epi-metaphysis ROIs and the carpal ROI are integrated to estimate the bone age of a child from the hand radiograph by using a fuzzy logic system.

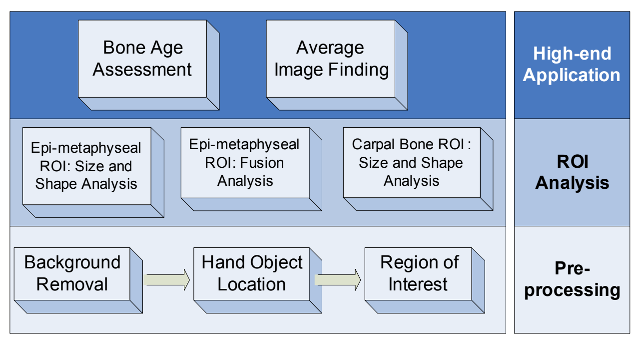

Figure 14. Three-layer procedure of computerized BAA

As data including the image, features extracted from each ROI, along with growth factors in textual form, are augmented and organized for each normal subject for all 1,080 subjects in a large-scale database, a digital atlas can be formed. Two high end applications can be implemented. One is to assess the bone age of a child from a hand radiograph based on image analysis using the digital atlas discussed earlier. The other is the “average” reference image in the digital atlas which can be selected for each of the groups of normal developed children with the best representative skeletal maturity based on specific bony features using image mining technique. Figure 15 shows an example of the average image of African American girls’ category from 1 to 12 years old in the digital atlas. [27]

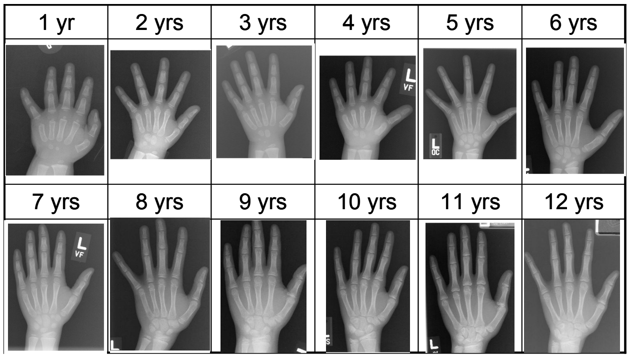

Figure 15. The average images for age groups (1-12 years) of African American girls (size not in scale). Each average image is obtained by image mining from all normal images based on their feature vectors in the same category (for example, ethnic origin and gender) in the digital hand atlas.

4.3 Use of Data Grid for Bone Age Assessment

The already developed BAA methodology has one disadvantage which is the computational requirements in identifying the ROIs, feature extraction from each ROI, and the use of fuzzy logic for bone age assessment. The Data Grid with computational services can be used to speed up the computational requirements. In order to accomplish that, we modified the current methodology in BAA shown in Figure 14 to facilitate the computational services in the Data Grid shown in Figure 4 and Figure 12 by subdividing the entire hand image into multiple regions of interest, including six epi-metaphyseal ROIs of three phalanges and the carpal bones ROI. These ROIs are processed by remote grid resource nodes; later, all results are merged together through the Data Grid MDS (Monitoring and Discovering System) server.

Let us assume the user submit a patient's hand image to the Data Grid with BAA computational services, the overall operation workflow of bone age assessment is shown in series (1, 2, 3) in Figure 16.

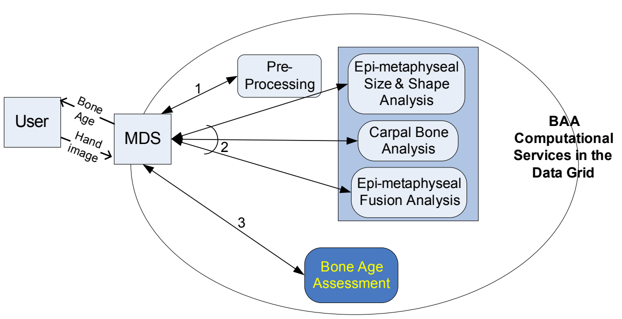

Figure 16. Operation workflow of BAA computational services. The MDS allocates the computation requirements according to the available resources in the Data Grid. Numerals represent the workflow.

4.4 Advantages of Computational Services in the Data Grid for Bone Age Assessment

The advantages of developing the computational services in the Data Grid for BAA versus using the conventional method are:

1. It utilizes the existing Data Grid technology which saves the job distribution and computation time.
2. It does not require to significant rewrite the image processing codes, this approach has resulted in substantial acceleration of the data analysis speed.
3. It facilitates the upgrading of the segmentation and feature extraction methodology easily.
4. With an ever-expanding digital hand atlas, such an object-oriented designed system can assure a continuous accrual of the best representations of the "average maturity" of a normal population by using a systematic image analysis method.

5. CONCLUSION

Grid computing is a powerful tool for large-scale computation and storage requirements. In this paper we present a novel concept of Data Grid for medical image application, in particular, for daily clinical PACS on-site archive, off-site backup, and disaster recovery. The Data Grid utilizes the SAN technology, and imbeds the DICOM standard as an integrated system. A testbed with two clinical sites and one research site is being implemented to evaluate the performance of the Data Grid. The allocation and distribution of resources at each site is described. The backup procedure during disaster is presented with dataflow using the DICOM Store and DICOM Query/retrieval functions.

Taking advantage of the Data Grid infrastructure, we also implemented computational services in the Data Grid for bone age assessment of children. A large-scale digital atlas containing over 1,000 digital hand radiographs of normal children from age 0 to 18, male and female, and four ethnic origins has been developed in an organized database. Each radiograph has seven computer extracted regions of interest (ROI), and each ROI has a feature vector containing pertinent information of bone age indicators. These feature vectors in the atlas are used to assess the bone age of a child when his/her radiograph is submitted. A feature vector is extracted from the radiograph and compared with those in the digital atlas by using Fuzzy logic. This conventional computation procedure has been modified and implemented in the computational services of the Data Grid. Evaluation of the Data Grid method versus the conventional computation methods in terms of computational effectiveness and efficiency is being conducted.

6. ACKNOWLEDGEMENTS

This research has been partially supported by NIH R01 EB 00298 and NIH R01 LM07606 grants.

7. REFERENCES

[1] Cao F, Huang, HK, Zhou XQ. 2003 Medical Image Security in a HIPAA Mandated PACS Environment. *Comp Med Imag & Graphics* V27, Issues 2-3, 185-196.

[2] Huang HK. 2003. Enterprise PACS and Image Distribution, *Comp Med Imaging & Graphics* V27, Issues 2-3, 241-253.

[3] Liu BJ, Cao F, Zhou MZ, Mogel G, 2003 Trends in PACS image Storage and Archive. *Comp Med Imaging & Graphics* V27, Issues 2-3, 165-174.

[4] Huang HK, Cao F, Zhang JG, Liu BJ, Tsai ML 2000. Fault tolerant Picture Archiving and Communication System and Teleradiology Design. In Reiner B, Siegel EL, Dwyer SJ*: Security Issues in the Digital Medical Enterprise, SCAR, Chapter 8,* 57-64.

[5] What is grid computing, http://www-1.ibm.com/grid/about_grid/what_is.shtml

[6] Grids and Grid technologies for wide-area distributed computing, Mark Baker, etc. *SPIE,* 2002.

[7] The Grid: A New Infrastructure for 21st Century Science, http://www.aip.org/pt/vol-55/iss-2/p42.html,

[8] Computational Grids, *The Grid: Blueprint for a New Computing Infrastructure*, Chap 2, Morgan-Kaufmann, 1999.

[9] The Physiology of the Grid: An Open Grid Services Architecture for Distributed Systems Integration. I. Foster, C. Kesselman, J. Nick, S. Tuecke, Open Grid Service Infrastructure WG, *Global Grid Forum*, June 22, 2002.

[10] The Anatomy of the Grid: Enabling Scalable Virtual Organizations. I. Foster, C. Kesselman, S. Tuecke. *International J. Supercomp Applications*, 15(3), 2001.

[11] Grid Services for Distributed System Integration. I. Foster, C. Kesselman, J. Nick, S. Tuecke. *Computer*, 35(6), 2002.

[12] SAN Technology: http://www.storage.ibm.com/ibmsan/whitepaper.html

[13] Globus Toolkit 3, http://www.globus.org/toolkit/gt3-factsheet.html

[14] The Globus Striped GridFTP Framework and Server. W. Allcock, J. Bresnahan, R. Kettimuthu, M. Link, C. Dumitrescu, I. Raicu, I. Foster. *Proceedings of Super Computing* 2005 (SC05), November 2005.

[15] Security for Grid Services. V. Welch, F. Siebenlist, I. Foster, J. Bresnahan, K. Czajkowski, J. Gawor, C. Kesselman, S. Meder, L. Pearlman, S. Tuecke. *Twelfth International Symposium on High Performance Distributed Computing (HPDC-12)*, IEEE Press, to appear June 2003.

[16] Grid Information Services for Distributed Resource Sharing. K. Czajkowski, S. Fitzgerald, I. Foster, C. Kesselman. *Proceedings of the Tenth IEEE International Symposium on High-Performance Distributed Computing (HPDC-10)*, IEEE Press, August 2001.

[17] Grid Resource Management. J. Nabrzyski, J.M. Schopf, J. Weglarz (Eds). Kluwer Publishing, Fall 2003.

[18] Globus Toolkit 3 Core White Paper, http://www-unix.globus.org/toolkit/documentation.html

[19] Open Grid Services Infrastructure (OGSI) Version 1.0. S. Tuecke, K. Czajkowski, I. Foster, J. Frey, S. Graham, C. Kesselman, T. Maguire, T. Sandholm, P. Vanderbilt, D. Snelling; *Global Grid Forum Draft Recommendation*, 6/27/2003.

[20] Liu BJ, Huang HK, Cao F, Zhou MZ, Zhang J, Mogel GT. 2004, A Complete Continuous -Availability PACS Archive Server Solution, *Radiographics*, 1203-1209.

[21] Huang HK, Liu BJ, Zhou Z. 2004, A CA Server for Medical Imaging Application, *Academic Radiology*, V.11, No.7 767-778.

[22] Huang HK, 2005, Medical Imaging Informatics Research and Development Trends – An Editorial. *Comp Med Imaging & Graphics* V.29, Issues 2-3, 91-93.

[23] Huang, HK, 2004. *PACS and Image Informatics: Basic Principles and Applications*. 703 pages, Cloth. John Wiley & Sons, Hoboken, NJ

[24] Jim Blythe, Ewa Deelman, Transparent Grid Computing: a Knowledge-Based Approach. *Fifteenth Innovative Applications of Artificial Intelligence Conference* (IAAI-03),Acapulco, August 12-14 2003.

[25] Performance Analysis of the Globus Toolkit Monitoring and Discovery Service, MDS2. X. Zhang and J. Schopf. *Proceedings of the International Workshop on Middleware Performance (MP 2004), part of the 23rd International Performance Computing and Communications Workshop (IPCCC)*, April 2004.

[26] E. Pietka, S. Pospiech, A. Gertych, F. Cao, Integration of Computer Assisted Bone Age Assessment with Clinical PACS, *Computerized Medical Imaging and Graphics*, 1-12, 2002.

[27] A. Zhang, et al. Data mining and visualization of average images in a digital hand atlas. *Proceedings of SPIE Medical Imaging*, Vol. 5748, pp65-72, February 2005